

Antiprotozoal Nitazoxanide Derivatives: Synthesis, Bioassays and QSAR Study Combined with Docking for Mechanistic Insight

Thomas Scior^{*1}, Jorge Lozano-Aponte¹, Subhash Ajmani², Eduardo Hernández-Montero³, Fabiola Chávez-Silva⁴, Emanuel Hernández-Núñez⁴, Rosa Moo-Puc⁵, Andres Fraguera-Collar³ and Gabriel Navarrete-Vázquez⁴

¹Laboratorio de Simulaciones Moleculares Computacionales, Facultad de Ciencias Químicas, BUAP, Puebla, Puebla 72000, México

²Department of Computational Chemistry, Jubilant Biosys Limited, #96, Industrial Suburb, 2nd Stage, Yeshwanthpur, Bangalore - 560 022, India

³Posgrado en Matemáticas, Facultad de Ciencias Físico Matemáticas, BUAP, Puebla, Puebla 72000, México

⁴Laboratorio de Química Farmacéutica, Facultad de Farmacia, Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos 62209, México

⁵Unidad de Investigación Médica Yucatán, Unidad Médica de Alta Especialidad del Centro Médico Nacional Ignacio García Téllez, IMSS Mérida, Yucatán 97000, México



Thomas Scior

Abstract: In view of the serious health problems concerning infectious diseases in heavily populated areas, we followed the strategy of lead compound diversification to evaluate the near-by chemical space for new organic compounds. To this end, twenty derivatives of nitazoxanide (NTZ) were synthesized and tested for activity against *Entamoeba histolytica* parasites. To ensure drug-likeness and activity relatedness of the new compounds, the synthetic work was assisted by a quantitative structure-activity relationships study (QSAR). Many of the inherent downsides – well-known to QSAR practitioners – we circumvented thanks to workarounds which we proposed in prior QSAR publication. To gain further mechanistic insight on a molecular level, ligand-enzyme docking simulations were carried out since NTZ is known to inhibit the protozoal pyruvate ferredoxin oxidoreductase (PFOR) enzyme as its biomolecular target.

Keywords: 3D-QSAR, 4D-QSAR, 5D-QSAR, mathematical regularization, nitrothiazole, PFOR, QSAR pitfalls, tizoxanide.

I. INTRODUCTION

In the 1930s, chemists were beginning to explore the effects of structural modifications on the kinetic processes of chemical reactions, resulting in the birth of physical organic chemistry [1]. Decades before, other scientists observed structure-activity relationships, e.g. ethers or alcohols (*Cros*, “Action de l’alcool amylique sur l’organisme”, University of Strasbourg, France, 1863) showed a correlation between lipo-solubility and toxicity. Around 1900, Meyer and Overton, independently, established the linear dependency between the narcotic action and water / oil partitioning (*Meyer, 1899; Overton, 1901* in [2]). One of the seminal works at an early stage of Quantitative Structure-Activity Relationships (QSAR) was the study on electronic effects of benzoic acids substituents and the descriptors were named after its inventor Hammett σ constants (*Hammett, 1937; Hammett, 1940* in [2]). After the 1960s, QSAR modeling was carried out using multiple linear regression (MLR) on

numeric independent variables describing structural features (descriptors) by Hansch and Fujita (*Hansch, Fujita, 1964* in [2]) under the assumption that an “inherent association between chemical structure and biological activity” exists [3]. Not more than twenty physicochemical descriptors are commonly used in QSAR studies descriptors [4], like octanol/water partition coefficient (log P), the Hammett σ constant, and the fragmental lipophilicity parameter π . Others were derived from quantum chemical calculations, namely: dipole moment, partial charges, HOMO/LUMO energies (*Kubinyi, 1993* in [2]) or were based on molecular graph theory and topology concepts by Wiener (*Wiener, 1947* in [2]). Kier and Hall (*Kier, Hall, Murray, 1975* in [2]) and Randić (*Randić, 1975; Kier, Hall, 1976; Kier, Hall, 1986* in [2]).

QSAR studies were applied on many occasions and recently reviewed [5-7]. QSAR studies also predicted antiprotozoal potency of pentamidines. The authors generated their own descriptors and used them in the final multiple linear regression equations. The comparison to our descriptors is delayed, awaiting the free *Desmol* software [8]. Drug screens from natural and synthetic sources have focused on antiprotozoal activities as part of the Drugs for Neglected Diseases Initiative [9-13]. A kind of knowledge-based approach is the combination of activity data base and

*Address correspondence to this author at the Department of Pharmacy, Facultad de Ciencias Químicas, Benemérita Universidad Autónoma de Puebla, Ciudad Universitaria, Edificio 105 C/106, C.P. 72570 Puebla, PUE., Mexico; Tel: xx52-222-2 295500, Ext. 7529; Fax: xx52 -222-2 295584; E-mail: tscior@gmail.com; Deprecated Institutional E-mail: thomas.scior@correo.buap.mx

QSAR prediction, called multi-species complex networks for antiparasite drugs [14]. As early as the 1980s nitro-aromatic antiparasitic drugs were already well-known, *e.g.* metronidazole [15]. Structurally related compounds to our nitrothiazole scaffold are also reported, *e.g.* benznidazole, nifurtimox and megalol 5-nitroimidazole derivative highly active *in vitro* and *in vivo* against parasites like *Trypanosoma cruzi*, or thiadiazines [16, 17].

With growing population worldwide, infectious diseases are considered a serious health problem. In particular, amoebiasis is a protozoal infection caused by a living organism called *Entamoeba histolytica*. Upon contaminated food uptake or other oral expositions, the clinical symptoms of patients include severe dysenteries, mucosal ulcers and even peritonitis, amoebic granulomas and/or fulminant colitis [18-20].

In 1974, Rossignol and Cavier conducted synthesis of nitazoxanide (NTZ) which was patented two years later [21]. Nowadays the drug is employed for a plethora of treatments, such as: anaerobic bacteria (*Helicobacter pylori*, *Clostridium difficile*); protozoa (*Entamoeba histolytica*, *Balantidium coli*, *Giardia intestinalis*, *Trichomonas vaginalis*); helminthes (*E. vermicularis*, *A. lumbricoides*, *Strongyloides stercoralis*, *T. trichuria*, *Taenia spp.*, *H.nana*, *Fasciola hepática*) in addition to certain viruses (rotavirus, hepatitis B and C and influenza) [22-26].

The common sense indicates that the biological activity of NTZ relies on the presence of the 5-nitrothiazole ring as the core substructure, *i.e.* the essential scaffold (pharmacophore) which is kept constant during chemical derivatization efforts (Figs. 1, 2) [27]. Deacetylation of NTZ produces tizoxanide (TIZ). It is a known active metabolite and included in the study (Tables 1a and 1b) [27-30].

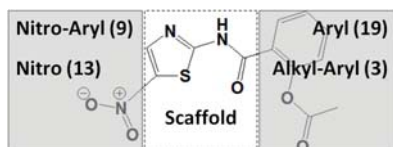


Fig. (1). The structure of nitazoxanide and its pharmacologically active substructure or scaffold (dotted area). The shaded boxes indicate the chemical variations introduced into the lead compound nitazoxanide chemically named as 2-[(5-nitro-1,3-thiazol-2-yl)carbamoyl]phenyl acetate (IUPAC: 2-acetyloxy-N-(5-nitro-2-thiazolyl)benzamide).

In 2011, Navarrete-Vazquez *et al.* synthesized and characterized two benzologues of nitazoxanide and tizoxanide [27], with proven antiprotozoal activity. After that, new compounds were designed and synthesized (Fig. 3), and clustered into two series containing 11 thiazole analogues and 9 benzothiazoles as benzologues. They all (n

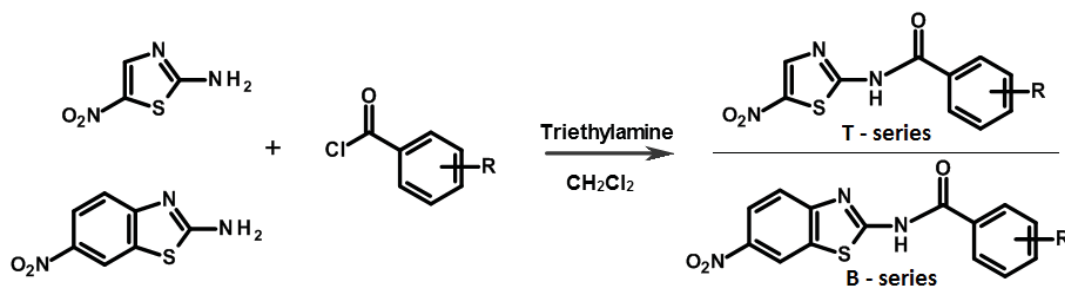


Fig. (3). Synthesis of the thiazole (T) and benzothiazole (B) series.

= 20) were tested for their antiprotozoal activities (IC_{50}) against *Entamoeba histolytica* [personal communication, 2011]. The molecular structures are displayed in Figs. 1, 2), Tables 1a and 1b. Moreover, Hoffman suggested NTZ acts as an inhibitor of the pyruvate ferredoxin oxidoreductase (PFOR) enzyme as a possible mechanism of action [31, 32, personal communication 2012].

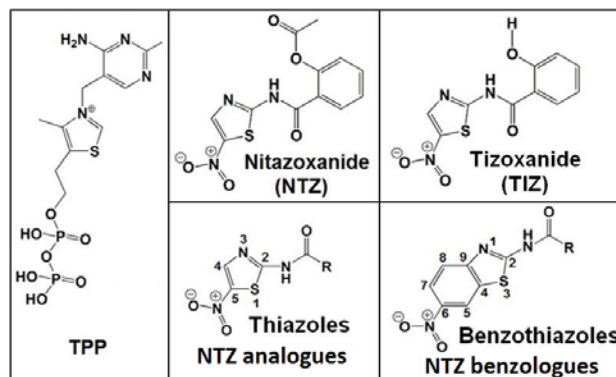


Fig. (2). The structures of lead compound (nitazoxanide), active metabolite (tizoxanide), physiological coenzyme bound to active site of PFOR (thiamine pyrophosphate, TPP) and the new, active derivatives of the present study (benzo / thiazoles).

Pharmacological reports relate the antiprotozoal activity of NTZ to an anaerobic microorganism-specific enzyme as its drug target, namely PFOR [31, 32, 35-37]. The physiological role of PFOR is known as an oxidative decarboxylation of pyruvate to form acetyl-CoA and carbon dioxide, paralleled by the reduction of ferredoxin or flavodoxin in the presence of the coenzyme Thiamine Pyrophosphate (TPP, Vitamin B1) [32, 37-39]. One possible explanation for the essential step in the anaerobic energy metabolism was proposed in that the antiprotozoal drug (candidates) inhibits the electron transfer reaction in a PFOR-dependent fashion [24, 30, 32]. Another possible pathway was suggested to explain the antiprotozoal activity [40]. P. S. Hoffman *et al.* proposed that NTZ directly interacts with TPP, what is a unique case of drug targeting a coenzyme. They observed NTZ was not replacing TPP from the active site at all in their *Helicobacter pylori* and *Campylobacter jejuni* assays [32].

On occasion of practicing a full-size antiprotozoal drug development cycle including *in silico* design, syntheses and *in vitro* bioassays, we also inspected the quantitative structure-activity relationships (QSAR) models and analyzed their equations in order to identify the pitfalls encountered

Table 1a. Listing of molecular structures of NTZ and its eleven thiazole derivatives (T series). The inhibition concentration pIC50 (against *E. histolytica*) is given and used as input data for QSAR. Note: tizoxanide (TIZ) is the hydrolysis product of NTZ: deacetyl-nitazoxanide. Note: recently, T17 without *E. histolytica* activity data was used in a published study (cf. Table 1 on page 1627 in [33]). And also recently T18 appeared without *E. histolytica* activity data (cf. Fig. 1 on page 3159 in another article by GNV [34]). Molecules of test set are denoted by asterisk (*).

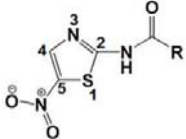
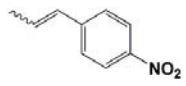
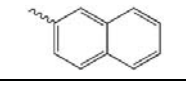
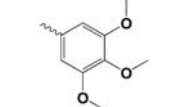
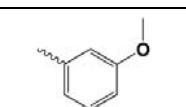
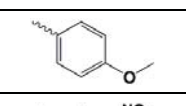
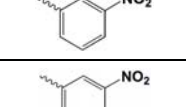
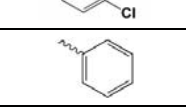
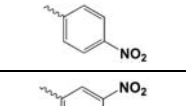
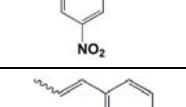
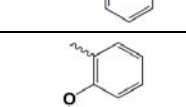
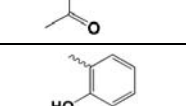


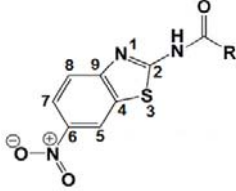
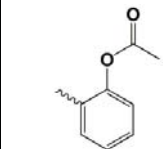
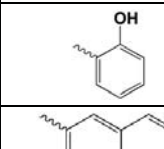
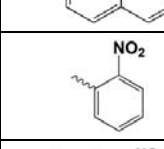
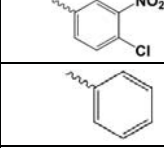
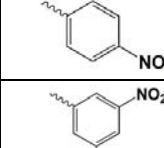
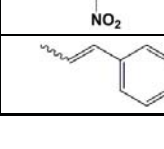
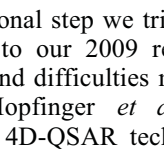
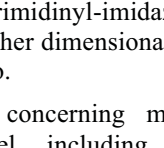
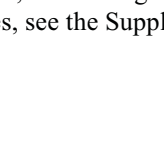
THIAZOLES (T - Series) - Analogues Scaffold		
		
id	Radical	pIC50 (-log IC ₅₀)
T03		8.30
T04*		7.92
T05		8.15
T06		7.46
T07		7.92
T08		7.82
T09		7.92
T17		6.07
T18*		6.18
T19		5.83
T22		6.34
NTZ		6.30
TIZ		5.91

Table 1b. Listing of molecular structures of NTZ and its nine benzothiazole derivatives (B series). The inhibition concentration pIC50 (against *E. histolytica*) are given and used as input data for QSAR. Note: B01 and B02 combined with *E. histolytica* activity data were already documented by the group leader GNV (cf. Table 1 on page 3169 in [27]). Molecules of test set are denoted by asterisk (*).

BENZOTHAZOLES (B - Series) - Benzologues Scaffold		
		
id	Radical	pIC50 (-log IC ₅₀)
B01		5.45
B02		5.10
B10*		5.19
B11		4.98
B12		5.38
B25		5.04
B26		4.81
B27		5.07
B30*		4.72

therein. In an additional step we tried to circumvent them if possible according to our 2009 report on a collection of frequent problems and difficulties met in QSAR studies [3]. Intriguingly, A. Hopfinger *et al.* successfully applied predictive 3D- and 4D-QSAR techniques to two series of pyridinyl- and pyrimidinyl-imidazoles [41, 42], so we decided to apply higher dimensional QSAR protocols for our NTZ derivatives, too.

For more details concerning modeling, bioassays and mathematical model, including leave one out cross-validation procedures, see the Supplementary Material.

2. MATERIALS AND METHODS

2.1. Organic Synthesis and Characterization of New Compounds

Triethylamine (1.2 equiv.) was added to a solution of 2-amino-5-nitrothiazole or 2-amino-6-nitrobenzo[d]thiazole (0.0015 mol) in dichloromethane. Fifteen minutes of stirring was necessary at 5°C. To this mixture a solution of acylchloride (adequately substituted) in dichloromethane was added. Under continuous stirring at room temperature during several hours the reaction process was monitored by thin layer chromatography between 4 to 24 hours until the transformation was complete. After distillation of dichloromethane was achieved, the resulting residues were neutralized with a saturated NaHCO₃ solution. Prior to washing with brine (3×20 mL) the aqueous layer was extracted with ethyl acetate (3×15 mL). Ethyl acetate was removed in vacuum. Then the precipitated solids were either recrystallized from a mixture of solvents or purified by column chromatography.

2.2. QSAR Modeling and Molecular Docking

Upon three-dimensional (3D) model generations [43-46], their molecular geometries adopted a conformation taken from the crystal structure of NTZ [47, 48]. Prior to retrieving the crystal structure of NTZ, we correctly predicted (15°) the out of coplanarity tilt (18°) between the 5- and 6-membered rings of NTZ.

Prior to docking of NTZ into the protozoal enzyme PFOR, the missing structure of the target enzyme PFOR from *Entamoeba histolytica* (RCSB Protein Data Bank) [49] was built from a similar PFOR crystal structure of *Desulfovibrio africanus* as three-dimensional template (PDB code: 2C42 [50]) [51, 52]. Automated flexible ligand docking was applied to the PFOR model of *E. histolytica* as rigid target receptor under *Auto-Dock 4.2* & *AutoDockTools* [53, 54]. Adjacent to the binding cleft we kept the prosthetic group ferredoxin and a magnesium cation to give the site its natural form.

Parts of our multicenter molecular modeling contributions were carried out on a Linux workstation using Schrodinger/CANVAS software [55]. To obtain QSAR equations by stepwise forward MLR method, an in-house program was developed using C language on a Windows platform.

Various 2D topological, structural, constitutional and physicochemical descriptors were calculated using CANVAS yielding a set of 757 descriptors [55]. In addition, a binary variable which has values 0 or 1 for indicating absence or presence of benzothiazole scaffold respectively was included in the study. The data preprocessing of independent variables was performed by removing the descriptors common to more than 90 % of the compounds and with a homogeneity index < 20 (*i.e.* deviating more than 80% from the ideal homogenous distribution). The homogeneity index quantitatively indicates the degree of uniformity of a variable (descriptor) distribution within its range [56]. Data preprocessing led to a significant reduction of descriptors *viz.* 217 descriptors for further study.

Furthermore for building multiple linear regression based QSAR models auto-scaled descriptors were utilized.

For the statistical procedure we followed the reasoning about internal and external validation processes by M.N. Noolvi *et al.* and applied their equations appropriately, besides alternative solutions [57, 58]. On the other side, the drawbacks and shortcomings concerning cross-validation procedures were published in a seminal work by D. Krstajic and coworkers [59]. I. Tetko (see below for *Ochem* online QSAR server [60]) recommended that the final step should always be the validation of hitherto unseen compounds and done only once (private communications, Helmholtz-Zentrum Munich, Germany, 2013, 2014). We followed this tenet wherever the software allowed us to maintain all 22 molecules together for QSAR modeling input, while Raptor was the sole tool which requires an *a priori* manual division into a test and training set [61, 62]. Hence our data set was only divided into two groups for Raptor application. To this end, the Raptor training set contained 18 compounds to generate the 5D-QSAR model complemented by a test set with 4 hand-selected compounds for the external (final) model validation. The molecules labeled B10, B30, T04 and T18 were manually selected because they embrace almost all the chemical features of the training set. The underlying trade-off meant to reflect a minimum number of molecules in the test set, while maximizing the chemical representativeness in that test set. The leave one out (LOO, Q²) method was carried out for internal/cross-validation of the models. For validation of either internal/cross or external models, values of either Q² or pred R², and other statistical parameters (Q²_{F1}, Q²_{F2}, Q²_{F3}, r²_m) were calculated using formulae reported in the literature and references cited therein [57, 63-70].

While Q² values describe the internal stability, the pred R² values reflects the predictive power of a model [57]. Moreover, to the influence of chance correlations on the model building of the training set was evaluated by Y-randomization (randomization of the biological activity) test and 'Z score' as reported in literature [3, 71]. The Y-randomization test was performed to ascertain that the obtained statistical significance of the QSAR model was not due to chance correlation or overfitting.

With another tool box, we created additional QSAR linear equations using widely accepted concepts [44, 72-75]. The acidic dissociation constants (pK_a) were estimated in the first place based on the known functional fragments and later determined by two online resources (*Marvin*, *SPARC*) [76, 77]. A higher-dimensional QSAR approach included a 4D-QSAR [45, 46] as well as 5D-QSAR study [61, 62]. For external validation of the QSAR equations, newly, all molecules were hand-selected into a training set (n=18) and test set (n=4; B10, B10, T04, T18) [3]. The composition of the test set was changed, once the descriptors were selected for the final equations in order to get rid of any bias and compensate the small amount of compounds. The selection criterion for MLR equations was established with an R² ≥ 0.4 [3].

Note, an additional fully automated 2D-QSAR approach (*Ochem*) is available online at <https://ochem.eu/home/show.do> [60]. In addition, the Supplementary Material presents another approach based on more physical grounds:

the general regularization theory for ill-posed inverse problems was employed as an alternative to the MLR statistic technique in order to obtain a more general and robust linear QSAR equation. To this end, a smooth dependence from the activity data on the molecular descriptors was the only a priori assumption needed.

3. RESULTS

3.1. Synthesis

The organic synthesis resulted in yields in the range of milligrams and hereafter each compound could be crystallized and characterized. The melting points of all twenty synthesized compounds were each measured with a fully automated device known as EZ-Melt MPA120 and left uncorrected. Reaction monitoring was performed by thin layer chromatography (0.2 mm pre-coated silica gel 60 F254 plates from E. Merck). ¹H-NMR spectra were collected on a Varian Oxford (400 MHz) and ¹³C NMR (100 MHz) instruments. Chemical shifts were recorded in part per million (tetramethylsilane, $\delta = 0$) in deuterated dimethylsulfoxide; J values are represented in Hertz unities. Mass spectrometry analyses were carried out using a JEOL JMS-700 spectrometer under fast atom bombardment [FAB(+)]. Starting reagents were purchased from Sigma-Aldrich and the reagents did not undergo any purification before their use in the corresponding reaction procedures.

For more details concerning the spectrometric characterizations see the Supplementary Material.

3.2. Docked Ligand-PFOR Binding Analysis and 2D QSAR Modeling

Upon inspection of the docked ligand-enzyme poses, the interacting amino acids were identified (Table 2) to answer questions about the possible activity enhancements related to chemical derivatization. This way we gained mechanistic insight and observed whether or not all ligands have one common binding mode. The proposed poses were compared with closely related crystallographic data to avoid a common pitfall, namely multiple binding modes, see Supplementary Material, Figs. SD1, SD2 [3]. In theory, the NTZ derivatives could replace TPP in a competitive way. Actually, none showed higher binding energies than TPP itself. On the contrary, their affinities at best are found to be 1000-fold lower than TPP, so they cannot replace TPP as stronger binders to PFOR at all. Our computed finding is in

excellent keeping with the postulated binding mechanism proposed by Hoffman *et al.* [32].

The 2D-QSAR model for biological response was build using stepwise forward multiple regression using all the calculated descriptors (as independent variables) and pIC50 values of biological response (as a dependent variable). Stepwise forward multiple regression adds descriptor one by one in the regression model until there is no significant improvement in training set R² value. This analysis led to a multiple regression QSAR model with reasonable statistical parameters using one subset of four descriptors. Table 3 reports the descriptors, regression coefficients and statistical parameters associated with the developed 2D-QSAR model (equation SA-1). The observed and predicted pIC50 values of training and test sets by using the 2D-QSAR model are shown in Fig. (4).

The model revealed major role of BTZ_indicator descriptor *i.e.* 43.9% in determining *Entamoeba histolytica* activity, and it is inversely proportional to activity data. This suggests that compounds with thiazole scaffold are preferred over benzothiazole to attain higher potency against *E. histolytica*.

$$pIC50 = -20.72 + 0.96 aasC_Cnt + 0.50 aOm_Cnt + 6.62 ssNH_Sum - 3.45 BTZ_indicator \quad (\text{equation SA-1})$$

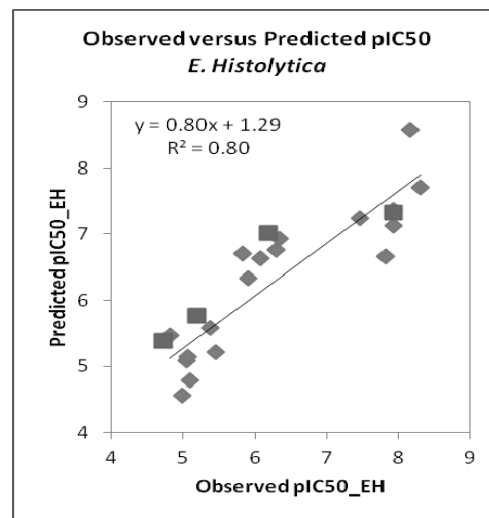


Fig. (4). The final plot graphically displays the prediction power (the distance from predicted value to observed values). Diamonds and rectangles represent compounds in training and test sets, respectively.

Table 2. Matrix of interactions representing Docking results (H-bonds, donors or acceptors). The gray boxes represent the interaction with the corresponding amino acid of PFOR model of *E. histolytica*. The experimental pIC50 and the computed $\Delta G_{\text{binding}}$ energies are included in the right side (T03 doesn't show qualitative relationship between both values).

ID	Glu60	Glu808	Cys831	Asp957	Gly958	Trp959	Thr985	Ser989	Asn990	pIC50	$\Delta G_{\text{binding}}$
TPP										---	-12
T03										8.30	-3
T22										6.34	-6.5
NTZ										6.30	-6.9
TIZ										5.91	-6.4

Table 3. Statistical parameters and descriptor definitions of the regression model (equation SA-1). Cross-validation standard error: cvSE; external-validation standard error: predSE; cross-validated Zscore: Zscore_cv; cross-validated alpha: alpha_cv.

Statistical Parameter	Value	Statistical Parameter	Value	Statistical Parameter	Value
Train/Test (n)	18/4	Q^2_{F1}	0.72	cvSE	0.97
Descriptors (k)	4	Q^2_{F2}	0.70	predSE	0.68
R^2	0.80	Q^2_{F3}	0.69	Zscore	3.99
Q^2 (LOO*)	0.55	r^2_m	0.39	Zscore_cv	2.42
pred R^2	0.72			alpha	0.00
SEE	0.65	r^2_m (LOO*)	0.48	alpha_cv	0.01
Descriptor	Definition				
aasC_Cnt	Count of atom-type E-State: :C:-				
aOm_Cnt	Count of atom-type E-State: :O-				
ssNH_Sum	Sum of atom-type E-State: -NH-				
BTZ indicator	A binary variable either 0 or 1 to indicate the absence or presence of a benzothiazole scaffold in a given molecule				

Note: (LOO*) is a particular parameter only applied to the initial training set (18 compounds).

3.3. 3D-QSAR Equations for Activity Prediction

In principle, the QSAR equations predict new activities for hitherto unseen structures and therein support the decisions to take for a new cycle of synthesis, tests and improved design, although in this paper we focus more on the diagnostic strength of QSAR and possible shortcomings [3, 41, 42].

$pIC50 = -3.85 + 2.17 pK_a + 0.41 MD + 0.9 Lipole$ (equation JL-1)

Analyzing the 3D-QSAR equations for the tilted compounds (equation JL-1, Fig. 5) yields the following design message: the higher the descriptor values for a given compound – the more active it will become, since all three linear descriptors are positive with similar contribution weight. The first variable represents molecular acidity (pK_a) which is formally located on the common exocyclic “>N-H” group and originated on the acidifying electron-withdrawing decoration of the mesomeric systems. The third one (lipole)

is calculated in analogy to the dipole moment (DM) when molecular lipophilicity is broken down to all atoms of the molecule and assigned as fragmental atomic values to feed the dipole formula and thereby becoming what is called the lipole measure. The statistical analyses yield the following linear determination coefficients of $R^2_{Train}=0.85$, $R^2_{Test}=0.96$ with $p<0.0001$ (equation JL-1, Fig. 5). In each case there is no risk of overfitting thanks to the 6:1 ratio between numbers of molecules versus descriptors [3]. The test set was hand-selected in the first place, but then during 5D-QSAR its composition was unattended, and the R^2 did neither drop nor rise significantly (R^2 ranged between 0.8 and 0.9, Fig. 6).

In addition we present numerical results for the entire compound set, adding the four preselected compounds (B10, B30, T04, T18) of the test set. As can be noted for equation JL-1, the overall performance of our final QSAR model does by no means change, yet the values remain constant applying the LOO to the entire set of 22 compounds (see S).

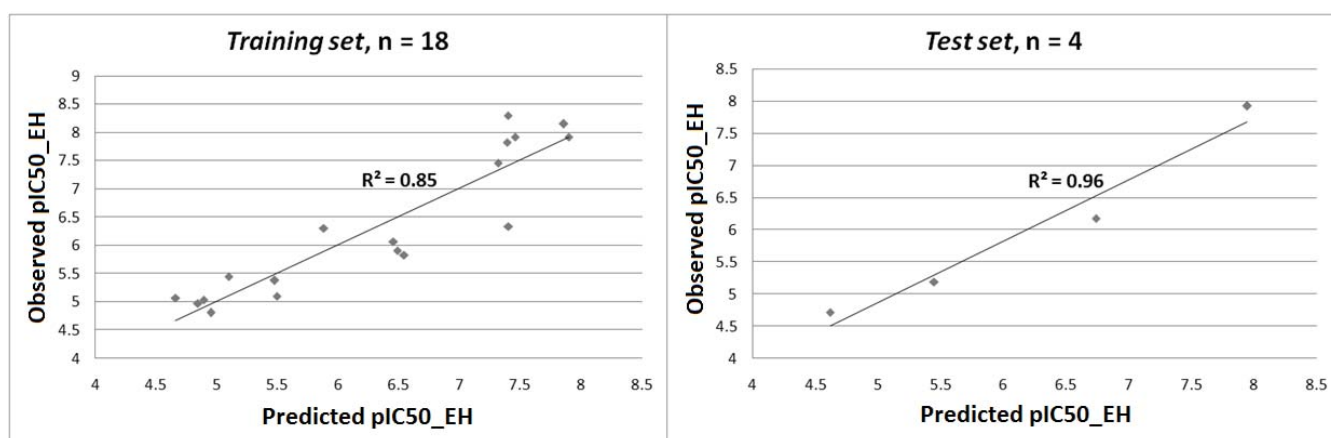


Fig. (5). 3D-QSAR equation generated by MLR for *E. histolytica* (equation JL-1). The three descriptors are: pK_a : Acidic dissociation constant; MD: Dipole Moment; Lipole: 3D-lipophilic distribution.

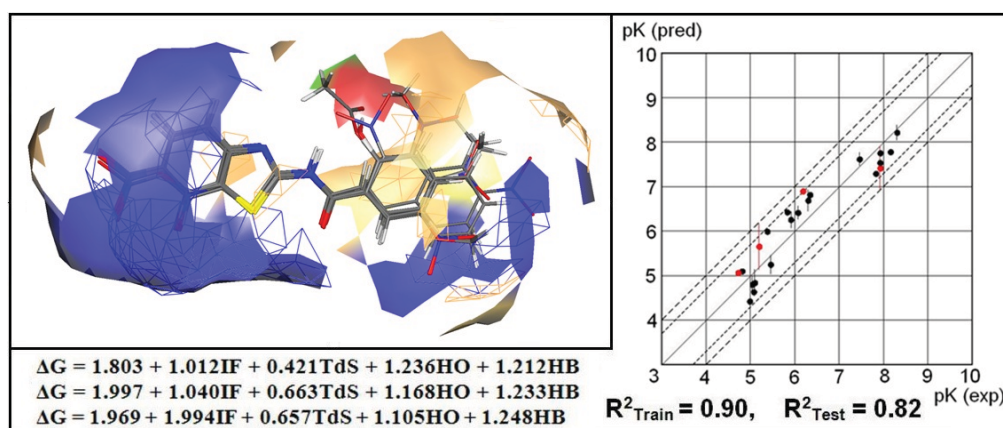


Fig. (6). Display of the active site projections (surrogate) generated by 5D-QSAR approach which represents the consensus of the three 5D-QSAR equations (equation **JL-2**; Bottom; IF: Internal Factor of Topology; TdS: Entropy; HO: Hydrophobicity; HB: H-Bonds formation.), dispersion plot of which is shown to the right. Color code in surrogate active site: blue: H-bond donor; red: H-bond acceptor; brown: strongly hydrophobic; yellow: weakly hydrophobic.

3.4. Higher Dimensional QSAR

Concerning the higher dimensional QSAR (5D-QSAR), the model quality (3 equations **JL-2**, Fig. (6), Bottom) is not significantly improved over the canonical QSAR linear equations (equation **SA-1**; $R^2_{Train}=0.80$, $n=18$, $R^2_{Test}=0.72$, $n=4$) or the 3D-QSAR equation (equation **JL-1**; $R^2_{Train}=0.85$, $n=18$, $R^2_{Test}=0.91$, $n=4$).

3.5. Inspecting the QSAR Models for Possible Pitfalls

We performed all QSAR studies on a training set ($n=18$) leaving four hand-selected molecules for the test set (B10, B30, T04, T18, Table 1).

After inspection of two online resources (Marvin, SPARC [76, 77].) mayor inconsistencies were detected for the extended mesomeric systems concerning nitro groups [76] (pitfall: non constant “constants” [3]).

According to Table 1, seven out of twenty compounds possess higher activity than NTZ or TIZ in the *in vitro* susceptibility assays (T03, T04, T05, T06, T07, T08, and T09). The activity spreads over three orders of magnitude (pIC50 values) which complies with the rule of thumb concerning the data range pitfall [3]. However, the data set falls short of expectation concerning the covered chemical space, which is a serious setback when only a restricted variation of synthetic paths limit the functional enrichment of the produced compounds (cf. pitfall: data size and variety [3]). Despite obvious structural differences (cf. scaffold types of either thiazoles or benzothiazoles) we fused both clusters following the more pivotal constraint of dealing with a small data size and the limited chemical space therein (here: $n_{Total}=22$) [3].

In general, QSAR studies must be very carefully performed, *i.e.* under a balanced protocol (bias of *training set*, size and chemical diversity and activity range, representative internal and external test set, outlier handling) to ensure a meaningful interpretation of the chosen descriptors and require understanding of the mathematical operations that takes time and experience [78-84]. While

practicing 2D- or 3D-QSAR modeling certain problems arose and were collected (Table 4).

4. DISCUSSION

The application of computer aided drug design tools demonstrates the (more) rational approach to drug candidate development using medicinal chemistry concepts to support drug research. To guide new efforts in the near future into promising new chemical space, we predicted the theoretical activity of a new promising compound, see Supplementary Material (Fig. **SD4**).

Mechanistic insight comes from the docking studies indicating that NTZ and its derivatives do not replace cofactor TPP as a PFOR ligand at the binding site as proposed by Hoffman *et al.* [32].

As an asset, the present work sheds also light on the implications and limitations which the medicinal chemists and QSAR practitioners should be aware of as a *sine qua non* condition. If the physicochemical data space is huge, the linearity may be at stake (outliers, activity cliffs) whereas, if it becomes smaller and smaller or even ill-explored – like in the present case with only ether, nitro and halogen substituents – then QSAR prediction falls short of expectation without remedy. Apparently, very few data points are addressed and new structure predictions tend to become either intrapolations with no fresh ideas or extrapolations into uncovered chemical space.

Another challenge becomes the sheer number of descriptors available for QSAR model generation which makes it more probable that solutions are found “by chance”, which is obviously more the case in automated, unattended QSAR approaches with hundreds to thousands of descriptors [85]. In contrast, the meaningful descriptors can be hand-selected from the very beginning to build stepwise forward QSAR models, *i.e.* reiteratively increase the number of descriptors bit by bit (stepwise forward procedure). QSAR studies carried out without considering the active conformation compounds (tautomers, ionic forms, prodrugs) are omitting critical information reducing the quality of the

Table 4. Listing of detected QSAR shortcomings and pitfalls which were reported in the literature [3].

Pitfalls	Comments
Small simple and limited chemical variability	It does not exist an ideal sample number size for QSAR, but is clear with larger sample size, the results become more representative. Here, the compound number is small (n=22), and the chemical variety is so limited. Basically it consist in ester (-O-CO-), ether (-O-), nitro (-NO ₂) and chloride, all of which occupy different positions of the phenyl ring of the molecules.
Composition of training and test sets	With a small sized series, the distribution of chemical items may not be representative in terms of the activity and chemical variability.
Meaningless descriptor selections	A common QSAR descriptor is pK _a with high correlation to biological activity, which is not the case of dipole moment (DM, dipole). DM takes on different values with changing conformations. When DM is calculated for artificially held planar molecules, it is loaded onto the Z-axis only, while even small torsional changes are reflected by huge changes in DM values.
not constant coefficients and constants	The calculated pK _a values are not equal to literature reports. It takes different values in different programs; moreover, each software consider different ionized forms, making difficult its selection/consideration for QSAR equations. Depending on the software, DM takes on different values due to normalization of input data. Certain descriptors like molar refractivity (MR) are very similar in magnitude in most programs. In contrast, polar surface area (PSA) should change according to the conformation of the molecule which implies that PSA is a 3D descriptors [44]. Other programs, however, calculate a “flat” PSA based on 2D data (atom connectivities and radii).
Starting geometries for 3D-QSAR	Albeit, the active conformation is not necessarily identical to the observed crystal structure, and since no NTZ-PFOR complex has been solved, two pieces of information were taken into account to assess the active conformation of the NTZ scaffold: (1) its crystallographic record deposited in CCDC [47] and (2) the final pose of NTZ docked into the ligand binding site of the cofactor TPP-PFOR complex. As a direct result, both geometries are practically the same, see Supplementary Material (Fig. SD2-C).
Errors of descriptor calculations (acidity, dissociation)	The experimental acidity value of NTZ is reported as pK _a ≈6 [32] for the conjugated acid / neutral thiazole system ([B-H ⁺] / [B]) which corresponds to approx. 90% neutral species under physiological conditions. The calculated value, pK _a ≈8 [76], however, inverts the cationic/nonionic portions (10% neutral species). With no experimental value at hand, the (wrong) cationic forms would have been taken as input for the QSAR and docking studies.
Lipole-dipole collinearity	The algorithm of Lipole calculation is derived from the dipole moment equation (DM = q*r, q = atomic partial charge, r = VDW atomic radius) and atomic lipophilic values replace atomic partial charges. Despite different scale and units (charges and lipophilic fragments, same VDW radii), the equal calculation protocol generates collinearity.
Linearity hypothesis	The a priori assumption of linearity might be the main drawback in QSAR studies. Since data sampling is not complete, because no scientist would seek to explore the weaker, less active or more toxic data segments, it is often not clear if linearity is a first principle of nature or just appears due to insufficient data spread. Outliers and activity cliffs are first signs of nonlinear relationships between independent variables and response (biological activity, dependent variable).
Ligand based alignment (LBA)	The X-ray (crystal) conformation of NTZ may not constitute the biological active conformation. The hitherto unsolved structure of the NTZ-PFOR complex constitutes a disadvantage in case of higher dimensional QSAR where reliable conformational data is required. Results based on 2D descriptors (connectivities, drawings, SMILES, etc.) do not need special information while ligands can be superposed on their more rigid substructure or common scaffold (LBA).
Multiple solutions	We generated different equations based on different conformations and methodologies. It is not clear whether modeling based on NTZ X-ray conformation reflects realistic molecule geometries for binding site interaction, because the NTZ liganded binding site complex has not been elucidated. According to the pK _a value of NTZ (here: 6.2, located on exothiazolic N amide), it can be inferred that all molecules treated here, present their activity at anionic form. Then, a new QSAR equation generation step based on descriptors calculated considering anionic compounds (without H at exothiazolic N amide, same training set), give us smaller R ² values that those obtained with X-ray data. The ideal case is considering the anionic form directly related with biological activity, because a small structural change can be reflected in huge descriptor magnitudes differences. This last QSAR equation generated with ionic compounds ($pIC50 = -2.36 + 2.28 pK_a + 0.17DM + 0.62 Lipole$; R ² _{Train} =0.74, Q ² =0.65, r ² _m =0.37, n=18; R ² _{Test} =0.75, Q ² _{F1} =0.75, Q ² _{F2} =0.75, Q ² _{F3} =0.75, r ² _m =0.68, n=4) could be seen as a poor predictive equation, but it becomes a better reflection of the biological behavior of our molecules.
Prodrugs and active metabolites	Some publications describe NTZ as a prodrug, albeit the biological activity of NTZ itself has been reported, too. Nevertheless, upon hydrolysis of the acetyl group, the metabolite TIZ shows comparable antiprotozoal potency.
Incompatible concepts and contradictions (chance correlation)	Sometimes, linear equations in 2D QSAR include conformation-dependent descriptors in a way where spatial information about structural requirements for the ligands and the binding site remains unknown. Hence, conformation-dependent descriptors contribute to establish the “rules” governing the relations between structures and activities, without any reason to be present in the equation except for chance correlations: “... because the relevant features only appear in molecules that also contain the wrong features” [3].

input data needed for equations which in turn lose predictive power, ending up in an incorrect molecular design [3, 81, 82].

As an internal validation we used the final QSAR equations to predict the pIC₅₀ values of the designed compound applying our linear equations, see Supplementary Material (Figs. SD4). Equation SA-1 yields 8.3 while equation JL-1 yields 8.4 which are practically the same.

CONCLUSION

The synthetic afford yielded seven new derivatives with a ten-fold or even stronger antiprotozoal activity than nitazoxanide. Concerning the relationships between structures and activities only mechanistically interpretable and fast to calculate descriptors were used in the final linear equations to predict even better candidates. Finally, we

devised a hitherto unseen molecule with a 100-fold higher potency than NTZ, see Supplementary Material. The external validation was achieved upon manual and Random separation of molecules into a training and test set to guarantee representativeness despite the small input data size. This downside and other pitfalls were presented and solutions thereof proposed if possible. In addition, mechanistic insight came from the docking studies indicating that NTZ and its derivatives do not replace cofactor thiamine pyrophosphate as a pyruvate ferredoxin oxidoreductase binder at the active site as proposed in the scientific literature.

LIST OF ABBREVIATIONS

CCDC	= The Cambridge Crystallographic Data Centre
DM	= Dipole Moment.
IC ₅₀	= 50% Inhibitory Concentration.
MLR	= Multiple Linear Regression.
NTZ	= Nitazoxanide.
pIC ₅₀	= -log ₁₀ (IC ₅₀).
PFOR	= Pyruvate Ferredoxin Oxidoreductase.
QSAR	= Quantitative Structure-Activity Relationships.
R ²	= Linear coefficient of determination.
RMSD	= Root mean square distance.
TIZ	= Tizoxanide.
TPP	= Thiamine Pyrophosphate, Vitamin B1

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This work was financially supported by Dr. Ygnacio Martínez L., VIEP, Benemérita Universidad Autónoma de Puebla (BUAP), Mexico. Many thanks to Prof. Dr. Paul S. Hoffman at University of Virginia Health Systems, Charlottesville, USA; as well as to Dr. Igor Tetko, Helmholtz Zentrum München, D-85764 Neuherberg, Germany for invitation, OChem support and discussions in 2013. Not to forget Prof. Dr. Anton Hopfinger, College of Pharmacy, University of New Mexico, USA, for discussing trial descriptor pools. This work was also supported in part by internal fund from Facultad de Farmacia, UAEM. Lozano-Aponte acknowledge the Mexican CONACyT for funding, CONACyT - Mexico City (2010-2012/302746) for project support. We are grateful for financial support by Mexican PROMEP funds for BUAP CA-120 in 2013.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's web site along with the published article.

REFERENCES

- King, F. D. Medicinal chemistry: principles and practice, 2nd ed.; Royal Society of Chemistry, **2002**.
- Carbó-Dorca R.; Robert D.; Amat Ll.; Gironés X.; Besalú E. Molecular quantum similarity in QSAR and drug design. 1st ed.; Springer-Verlag Berlin Heidelberg, **2000**.
- Scior, T.; Medina-Franco, J.L.; Do, Q-T.; Martínez-Mayorga, K.; Yunes-Rojas, J.A.; Bernard, P. How to Recognize and Workaround Pitfalls in QSAR Studies: A Critical Review. *Curr. Med. Chem.*, **2009**, *16*(32), 4297-4313.
- Scior, T.; Bernard, P.; Medina-Franco, J.L.; Maggiora, G.M. Large compound databases for structure-activity relationships studies in drug discovery. *Mini Rev. Med. Chem.*, **2007**, *7*(8), 851-860.
- Castillo-Garit, J.A.; Abad, C.; Rodríguez-Borges, J.E.; Marrero-Ponce, Y.; Torrens, F. A review of QSAR studies to discover new drug-like compounds actives against leishmaniasis and trypanosomiasis. *Curr. Top. Med. Chem.*, **2012**, *12*(8), 852-865.
- Cramer, R.D. Rethinking 3D-QSAR. *J. Comput. Aided Mol. Des.*, **2011**, *25*(3), 197-201.
- Cramer, R.D. The inevitable QSAR renaissance. *J. Comput. Aided Mol. Des.*, **2012**, *26*(1), 35-38.
- García-Doménech, R.; Manhenje, I.D.; Monje, Y.; López-Gonzalez, A.; Marco, R.; Tacho, C.; Gálvez, J. Search of QSAR models for predicting the antiprotozoal activity and cytotoxicity *in vitro* of a group of pentamidine analogous compounds. *Lett. Drug Des. Discov.*, **2011**, *8*(2), 172-180.
- Pelosi, G. Thiosemicarbazone Metal Complexes: From Structure to Activity. *Open Crystallogr. J.*, **2010**, *3*, 16-28.
- Graebin, C.S.; Uchoa, F.D.; Bernardes, L.S.C.; Campo, V.L.; Carvalho, I.; Eifler-Lima, V.L. Antiprotozoal agents: an overview. *Curr. Med. Chem.*, **2009**, *8*(4), 345-366.
- Monzote, L.; Siddiq, A. Drug development to protozoan diseases. *Open Med. Chem. J.*, **2011**, *5*, 1-3.
- Zucca, M; Savoia, D. Current developments in the therapy of protozoan infections. *Open Med. Chem. J.*, **2011**, *5*, 4-10.
- Schmidt, T.J.; Khalid, S.A.; Romanha, A.J.; Alves, T.M.; Biavatti, M.W.; Brun, R.; Da Costa, F.B.; de Castro, S.L.; Ferreira, V.F.; de Lacerda, M.V.; Lago, J.H.; Leon, L.L.; Lopes, N.P.; das Neves Amorim, R.C.; Niehues, M.; Ogungbe, I.V.; Pohlit, A.M.; Scotti, M.T.; Setzer, W.N.; de N. C. Soeiro, M.; Steindel, M.; Tempone, A.G. The potential of secondary metabolites from plants as drugs or leads against protozoan neglected diseases - part I. *Curr. Med. Chem.*, **2012**, *19*(14), 2128-2175.
- Prado-Prado, F.J.; González-Díaz, H.; de la Vega, O.M.; Ubeira, F.M.; Chou, K.C. Unified QSAR approach to antimicrobials. Part 3: first multi-tasking QSAR model for input-coded prediction, structural back-projection, and complex networks clustering of antiprotozoal compounds. *Bioorg. Med. Chem.*, **2008**, *16*(11), 5871-5880.
- Jefford, C.W.; Cadby, P.A.; Smith, L.C.; Pipe, D.F. Highly active nitro-aromatic antiparasitic drugs. *Pharmazie*, **1982**, *37*(6), 395-402.
- Soeiro, M.N.; de Castro, S.L. Screening of potential anti-trypanosoma cruzi candidates: *in vitro* and *in vivo* studies. *Open Med. Chem. J.*, **2011**, *5*, 21-30.
- Bermello, J.C.; Piñeiro, R.P.; Fidalgo, L.M.; Cabrera, H.R.; Navarro, M.S. Thiadiazine derivatives as antiprotozoal new drugs. *Open Med. Chem. J.*, **2011**, *5*, 51-60.
- Sargeant, P. G.; Williams, J. E.; Kumate, J.; Jimenez, E. The epidemiology of Entamoeba histolytica in Mexico City. A pilot survey I. *Trans. R. Soc. Trop. Med. Hyg.*, **1980**, *74*(5), 653-656.
- Harries, J. Amoebiasis: a review. *J. R. Soc. Med.*, **1982**, *75*(3), 190-197.
- Bansal, D.; Malla, N.; Mahajan, R. C. Drug resistance in amoebiasis. *Indian J. Med. Res.*, **2006**, *123*(2), 115-118.
- US patent: (a) Rossignol y Cavier, US3950351, 1976; (b) Rossignol y Cavier, US3957812, 1976; (c) Rossignol, US4315018, 1982; (d) Rossignol, US 5578621, 1996; (e) Rossignol, US5856348, 1999; (f) Rossignol, US5859038, 1999; (g) Rossignol, US5965590, 1999; (h) Rossignol, US6020353, 2000; (i) Rossignol, US5387598, 1995.

- [22] Dubreuil, L.; Houcke, I.; Mouton, Y.; Rossignol, J.F. *In vitro* evaluation of activities of nitazoxanide and tizoxanide against anaerobes and aerobic organisms. *Antimicrob. Agents Chemother.*, **1996**, *40*(10), 2266-2270.
- [23] Adagu, I.S.; Nolder, D.; Warhurst, D.C.; Rossignol, J.F. *In vitro* activity of nitazoxanide and related compounds against isolates of *Giardia intestinalis*, *Entamoeba histolytica* and *Trichomonas vaginalis*. *J. Antimicrob. Chemother.*, **2002**, *49*(1), 103-111.
- [24] Anderson, V. R.; Curran M. P. Nitazoxanide: A review of its use in the treatment of gastrointestinal infections. *Drugs*, **2007**, *67*(13), 1947-1967.
- [25] Korba, B.E.; Montero, A.B.; Farrar, K.; Gaye, K.; Mukerjee, S.; Ayers, M.S.; Rossignol, J.F. Nitazoxanide, tizoxanide and other thiazolides are potent inhibitors of hepatitis B virus and hepatitis C virus replication. *Antiviral Res.*, **2008**, *77*(1), 56-63.
- [26] Stachulski, A.V.; Pidathala, C.; Row, E.C.; Sharma, R.; Berry, N.G.; Iqbal, M.; Bentley, J.; Allman, S. A.; Edwards, G.; Helm, A.; Hellier, J.; Korba, B. E.; Semple, J.E.; Rossignol J.F. Thiazolides as novel antiviral agents. I. Inhibition of hepatitis B virus replication. *J. Med. Chem.*, **2011**, *54*(12), 4119-4132.
- [27] Navarrete-Vazquez, G.; Chavez-Silva, F.; Argotte-Ramos, R.; Rodríguez-Gutiérrez, M.C.; Chan-Bacab, M.J.; Cedillo-Rivera, R.; Moo-Puc, R.; Hernandez-Nunez, E. Synthesis of benzologues of Nitazoxanide and Tizoxanide: A comparative study of their *in vitro* broad-spectrum antiprotozoal activity. *Bioorg. Med. Chem. Lett.*, **2011**, *21*(10), 3168-3171.
- [28] Broekhuysen, J.; Stockis, A.; Lins, R.L.; De Graeve, J.; Rossignol, J.F. Nitazoxanide: pharmacokinetics and metabolism in man. *Int. J. Clin. Pharmacol. Ther.*, **2000**, *38*(8), 387-394.
- [29] Esposito, M.; Stettler, R.; Moores, S.L.; Pidathala, C.; Müller, N.; Stachulski, A.; Berry, N.G.; Rossignol, J.F.; Hemphill A. *In vitro* efficacies of nitazoxanide and other thiazolides against neospora caninum tachyzoites reveal antiparasitic activity independent of the nitro group. *Antimicrob. Agents Chemother.*, **2005**, *49*(9), 3715-3723.
- [30] Gargala, G. Drug treatment and novel drug target against cryptosporidium. *Parasite*, **2008**, *15*, 275-281.
- [31] Sisson, G.; Goodwin, A.; Raudonikiene, A.; Hughes, N.J.; Mukhopadhyay, A.K.; Berg, D.E.; Hoffman, P.S. Enzymes Associated with Reductive Activation and Action of Nitazoxanide, Nitrofurans, and Metronidazole in *Helicobacter pylori*. *Antimicrob. Agents Chemother.*, **2002**, *46*(7), 2116-2123.
- [32] Hoffman, P.S.; Sisson, G.; Croxen, M.A.; Welch, K.; Herman, D.W.; Cremades, N.; Morash, M.G. Antiparasitic drug nitazoxanide inhibits the pyruvate oxidoreductases of *Helicobacter pylori*, selected anaerobic bacteria and parasites, and campylobacter jejuni. *antimicrob. Agents Chemother.*, **2007**, *51*(3), 868-876.
- [33] Nava-Zuazo, C.; Chávez-Silva, F.; Moo-Puc, R.; Chan-Bacab M.J.; Ortega-Morales B.O.; Moreno-Díaz, H.; Díaz-Coutiño, D.; Hernández-Núñez, E.; Navarrete-Vázquez, G. 2-acylamino-5-nitro-1,3-thiazoles: preparation and *in vitro* bioevaluation against four neglected protozoan parasites. *Bioorg. Med. Chem.*, **2014**, *22*(5), 1626-1633.
- [34] Valladares-Méndez, A.; Hernández-Núñez, E.; Cedillo-Rivera, R.; Moo-Puc, R.; Barbosa-Cabrera, E.; Orozco-Castellanos, L.M.; Rivera-Leyva, J.C.; Navarrete-Vázquez, G. Synthesis, *in vitro* and *in vivo* giardicidal activity, and pharmacokinetic profile of a new nitazoxanide analog. *Med. Chem. Res.*, **2014**, *23*(6), 3157-3164.
- [35] Micromedex 2.0 Home Page. <http://www.thomsonhc.com/micromedex2/librarian>
- [36] Upcroft, P.; Upcroft, J.A. Drug targets and mechanisms of resistance in the anaerobic protozoa. *Clin. Microbiol. Rev.*, **2001**, *14*(1), 150-164.
- [37] Hemphill, A.; Muller, N.; Muller, J. Structure-function relationship of thiazolides a novel class of antiparasitic, investigated in intracellular and extracellular and extracellular protozoan parasites and larval-stage cestodes. *Curr. Med. Chem.*, **2007**, *6*(4), 273-282.
- [38] Chabrière, E.; Charon, M.; Volbeda, A.; Pieulle, L.; Hatchikian, E.C.; Fontecilla-Camps, J.C. Crystal structures of the key anaerobic enzyme pyruvate:ferredoxin oxidoreductase, free and in complex with pyruvate. *Nat. Struct. Biol.*, **1999**, *6*(2), 182-190.
- [39] Ragsdale, S.W. Pyruvate ferredoxin oxidoreductase and its radical intermediate. *Chem. Rev.*, **2003**, *103*(6), 2333-2346.
- [40] Fox, L.M.; Saravolatz, L.D. Nitazoxanide: a new thiazolide antiparasitic agent. *Clin. Infect. Dis.*, **2005**, *40*(8), 1173-1180.
- [41] Romeiro, N.C.; Albuquerque, M.G.; de Alencastro, R.B.; Ravi, M.; Hopfinger, A.J. Construction of 4D-QSAR models for use in the design of novel p38-MAPK inhibitors. *J. Comput. Aided Mol. Des.*, **2005**, *19*, 385-400.
- [42] Romeiro, N.C.; Albuquerque, M.G.; de Alencastro, R.B.; Ravi, M.; Hopfinger, A.J. Free-energy force-field three-dimensional quantitative structure-activity relationship analysis of a set of p38-mitogen activated protein kinase inhibitors. *J. Mol. Model.*, **2006**, *12*, 855-868.
- [43] Dobler, M. BioX 3.0 a versatile molecular-modeling program. Biographics Laboratory 3R, Basel/Switzerland. 2005-2011.
- [44] Pedretti, A.; Villa, L.; Vistoli, G. VEGA: a versatile program to convert, handle and visualize molecular structure on Windows-based PCs. *J. Mol. Graph. Model.*, **2002**, *21*(1), 47-49.
- [45] Vedani, A.; McMasters, D.R.; Dobler, M. Multi-conformational ligand representation in 4D-QSAR: reducing the bias associated with ligand alignment. *Quant. Struct-Act. Relat.*, **2000**, *19*, 149-161.
- [46] Lill, M.A.; Vedani, A. Combining 4D pharmacophore generation and multidimensional QSAR: modeling ligand binding to the bradykinin B2 receptor. *J. Chem. Inf. Model.*, **2006**, *46*(5), 2135-2145.
- [47] "CCDC 775321 contains the supplementary crystallographic data for this paper. These data can be obtained free of charge from The Cambridge Crystallographic Data Centre via www.ccdc.cam.ac.uk/data_request/cif".
- [48] Bruno, F.P.; Caira, M.R.; Monti, G.A.; Kassuha, D.E.; Sperandio, N.R. Spectroscopic, thermal and X-ray structural study of the antiparasitic and antiviral drug nitazoxanide. *J. Mol. Struct.*, **2010**, *984*(1-3), 51-57.
- [49] Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The protein data bank. *Nucleic Acids Res.*, **2000**, *28*(1), 235-242.
- [50] Cavazza, C.; Contreras-Martel, C.; Pieulle, L.; Chabrière, E.; Hatchikian, E.C.; Fontecilla-Camps, J.C. Flexibility of thiamine diphosphate revealed by kinetic crystallographic studies of the reaction of pyruvate-ferredoxin oxidoreductase with pyruvate. *Structure*, **2006**, *14*(2), 217-224.
- [51] Pieper, U.; Eswar, N.; Braberg, H.; Madhusudhan, M.S.; Davis, F.P.; Stuart, A.C.; Mirkovic, N.; Rossi, A.; Marti-Renom, M.A.; Fiser, A.; Webb, B.; Greenblatt, D.; Huang, C.C.; Ferrin, T.E.; Sali, A. MODBASE, a database of annotated comparative protein structure models, and associated resources. *Nucleic Acids Res.*, **2004**, *32*(Database issue), D217-222.
- [52] Sciort, T.; Wahab, A. Structure prediction of proteins with very low homology: A comprehensive introduction and a case study on aminopeptidase. In: Stanley P. Kaplan, editor. *Drug Des. Res. Perspect.*, **2007**, 675-708.
- [53] Morris, G.M.; Goodsell, D.S.; Halliday, R.S.; Huey, R.; Hart, W.E.; Belew, R.K.; Olson, A.J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.*, **1998**, *19*, 1639-1662.
- [54] Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.*, **2009**, *30*(16), 2785-2791.
- [55] Maestro, version 8.0, Schrödinger, LLC, New York, NY, 2007. LigPrep. 2012 version 2.5. Schrödinger. LLC: New York, NY. Maestro. 2012. version 9.3, Schrödinger. LLC: New York, NY.
- [56] Viswanadhan, V.N.; Ghose, A.K.; Weinstein, J.N. Mapping the binding site of the nucleoside transporter protein: a 3D-QSAR study. *Biochim. Biophys. Acta*, **1990**, *1039*, 356-366.
- [57] Noolvi, M.N.; Patel, H.M.; Bhardwaj, V. 2D QSAR studies on a series of 4-anilino quinazoline derivatives as tyrosine kinase (EGFR) inhibitor: An approach to design anticancer agents. *Dig. J. Nanomater. Bios.*, **2010**, *5*(2), 387-401.
- [58] Patterson, D.E.; Cramer, R.D.; Ferguson, A.M.; Clark, R.D.; Weinberger, L.E. Neighborhood behavior: a useful concept for validation of "molecular diversity" descriptors". *J. Med. Chem.*, **1996**, *39*(16), 3049-3059.
- [59] Krstajic, D.; Buturovic, L.J.; Leahy, D.E.; Thomas, S. Cross-validation pitfalls when selecting and assessing regression and classification models. *J. Cheminform.*, **2014**, *6*(1), 10.
- [60] Sushko, I.; Novotarskyi, S.; Körner, R.; Pandey, A.K.; Rupp, M.; Teetz, W.; Brandmaier, S.; Abdelaziz, A.; Prokopenko, V.V.; Tanchuk, V.Y.; Todeschini, R.; Varnek, A.; Marcou, G.; Ertl, P.;

- Potemkin, V.; Grishina, M.; Gasteiger, J.; Schwab, C.; Baskin, I.I.; Palyulin, V.A.; Radchenko, E.V.; Welsh, W.J.; Kholodovych, V.; Chekmarev, D.; Cherkasov, A.; Aires-de-Sousa, J.; Zhang, Q.Y.; Bender, A.; Nigsch, F.; Patiny, L.; Williams, A.; Tkachenko, V.; Tetko, I.V. Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. *J. Comput. Aided Mol. Des.*, **2011**, *25*(6), 533-535.
- [61] Vedani A.; Dobler, M. 5D-QSAR: the key for simulating induced fit? *J. Med. Chem.*, **2002**, *45*, 2139-2149.
- [62] Lill, M.A.; Vedani, A.; Dobler, M. Raptor: combining dual-shell representation, induced-fit simulation, and hydrophobicity scoring in receptor modeling: application toward the simulation of structurally diverse ligand sets. *J. Med. Chem.*, **2004**, *47*(25), 6174-6186.
- [63] Ajmani, S.; Kulkarni, S.A. Application of QSAR for scaffold hopping and lead optimization in multitarget inhibitors. *Mol. Inf.*, **2012**, *31*(6-7), 473-490.
- [64] Zheng, W.; Tropsha, A. Novel variable selection quantitative structure-property relationship approach based on the k-nearest-neighbor principle. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*(1), 185-194.
- [65] Hawkins, D.M.; Basak, S.C.; Mills, D. Assessing model fit by cross-validation. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*(2), 579-586.
- [66] Kraker, J.J.; Hawkins, D.M.; Basak, S.C.; Natarajan, R.; Mills, D. Quantitative Structure-Activity Relationship (QSAR) modeling of juvenile hormone activity: Comparison of validation procedures. *Chemometr. Intell. Lab.*, **2007**, *87*(1), 33-42.
- [67] Schüürmann, G.; Ebert, R.; Chen, J.; Wang, B.; Kühne, R. External validation and prediction employing the predictive squared correlation coefficient s test set activity mean vs training set activity mean. *J. Chem. Inf. Model.* **2008**, *48*(11), 2140-2145.
- [68] Consonni, V.; Ballabio, D.; Todeschini, R. Evaluation of model predictive ability by external validation techniques. *J. Chemometrics*, **2010**, *24*(3-4), 194-201.
- [69] Roy, P.P.; Roy, K. On some aspects of variable selection for partial least squares regression models. *QSAR Comb. Sci.*, **2008**, *27*(3), 302-313.
- [70] Roy, K.; Mitra, I.; Kar, S.; Ojha, P.K.; Das, R.N.; Kabir, H. Comparative studies on some metrics for external validation of QSPR models. *J. Chem. Inf. Model.*, **2012**, *52*(2), 396-408.
- [71] Shen, M.; Xiao, Y.; Golbraikh, A.; Gombar, V.K.; Tropsha, A. Development and validation of k-nearest-neighbor qspr models of metabolic stability of drug candidates. *J. Med. Chem.*, **2003**, *46*, 3013-3020.
- [72] StatSoft, Inc. (2004). STATISTICA (data analysis software system), version 7. www.statsoft.com.
- [73] BioEstat 5.0. Aplicaciones estadísticas para ciencias Bio-Médicas. Manuel Ayres. **2007**; Univerdidade Federal do Pará.
- [74] Dragon 3.0 Web Version – 2003. Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M. Milano chemometrics and QSAR Research Group. Department of Environmental Sciences. University of Milano-Bicocca. http://www.taletе.mi.it/dragon_exp.htm
- [75] HyperChem: HyperChem(TM) Professional 7.51, Hypercube, Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA.
- [76] "Marvin was used for drawing, displaying and characterizing chemical structures, substructures and reactions, Marvin n.n.n (version number), 201n (insert year of version release), ChemAxon (<http://www.chemaxon.com>)" "Calculator Plugins were used for structure property prediction and calculation, Marvin n.n.n (insert version number), 201n (insert year of version release), ChemAxon (<http://www.chemaxon.com>)"
- [77] Hilal, S.H.; Karickhoff, S.W.; Carreira, L.A. A rigorous test for SPARC's chemical reactivity models: estimation of more than 4300 ionization pKa's. *Quant. Struct.-Act. Relat.*, **1995**, *14*(4), 348-355.
- [78] Thibaut, U.; Folkem, G.; Klebe G.; Kubinyi H.; Men, A.; Rognan, D. Recommendations for CoMFA studies and 3D QSAR publications. *Quant. Struct. Act. Relat.*, **1994**, *13*, 1-3.
- [79] Rognan, D. Molecular dynamics simulations: a tool for drug design. In: *3D QSAR in drug design*; Kubinyi, H.; Folkers, G.; Martin, Y.C., Eds.; Springer, **1998**, Vol. 9-11, pp. 181-209.
- [80] Langer, T.; Hoffmann R.D. Application of Structure-Based Alignment Methods for 3D QSAR Analyses. In: *Pharmacophores and Pharmacophore Searches*; Mannhold, R.; Kubinyi, H.; Folkers, G., Eds.; Wolfgang Sippl. Published, **2006**; Vol. 32, pp. 223-249.
- [81] Huang, J.; Fan, X. Why QSAR fails: an empirical evaluation using conventional computational approach. *Mol. Pharm.*, **2011**, *8*(2), 600-608.
- [82] Hechinger, M.; Leonhard, K.; Marquardt, W. What is wrong with quantitative structure-property relations models based on three-dimensional descriptors? *J. Chem. Inf. Model.*, **2012**, *52*(8), 1984-1993.
- [83] Li J.; Li S.; Bai C.; Liu H.; Gramatica P. Structural requirements of 3-carboxyl-4(1H)-quinolones as potential antimalarials from 2D and 3D QSAR analysis. *J. Mol. Graph. Model.*, **2013**, *44*, 266-277.
- [84] Del Rio A.; Piras P.; Roussel C. Enantiophore modeling in 3D-QSAR. A data mining application on Whelk-O1 chiral stationary phase. *Chirality*, **2006**, *18*(7), 498-508.
- [85] Tseng Y.J.; Hopfinger, A.J.; Esposito E.X. The great descriptor melting pot: mixing descriptors for the common good of QSAR models. *J. Comput. Aided Mol. Des.*, **2012**, *26*(1), 39-43.