SUPPLEMENTARY MATERIAL

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

Scior Thomas¹, Lozano-Aponte Jorge¹, Ajmani Subhash², Hernández-Montero Eduardo³, Chávez-Silva Fabiola⁴, Hernández-Núñez Emanuel⁴, Moo-Puc Rosa⁵, Fraguela-Collar Andres³ and Navarrete-Vázquez Gabriel⁴

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

Supplementary Data #1

**BIOLOGICAL ASSAYS: IN VITRO ANTIPROTOZOAL ASSAY**

*E. histolytica* strain HM1-IMSS was cultured in TYI-S-33 medium, supplemented with 10% bovine serum. In vitro susceptibility assays were performed using a method previously described [SD1, SD2, SD3]. Precisely, 4 x 10⁴ trophozoites of *E. histolytica* were incubated for 48 h at 37 °C with increasing concentrations of synthesized compounds, including NTZ and tizoxanide (TIZ). As the negative control, trophozoites were incubated in culture medium with DMSO used in the experiments. After the incubation, trophozoites were washed and subcultured for another 48 h in fresh medium alone. At the end of this period, trophozoites were counted and the 50% inhibitory concentration (IC₅₀) was calculated by Probit analysis. Experiments were carried out in triplicate and repeated at least twice.

Supplementary Data #2

**HOMOLOGY MODELING OF BIOMOLECULAR TARGET PFOR**

Prior to protein homology modeling, the target and template sequences were retrieved from the UniProt databases (id: Q24818, and P94692, respectively) and aligned [SD4, SD5, SD6]. Thanks to a 53% identity score all active site amino acids were found to be conserved in the target (Suppl. Table SD1) [39, 50]. The resulting *E. histolytica* PFOR model was automatically deposited into the ModBase database (Primary Database Link Q24818, Q24818 tr Original Database ID [51, SD7]) Model quality was checked with Swiss PDB Viewer software [52, SD8].

**Table SD1.** Multiple sequence alignment (MSA) of different PFOR species (*No reported structure. **Crystal structure available in PDB, code 2C42).**

| Name                  | Length | Name                | Length | Score |
|-----------------------|--------|---------------------|--------|-------|
| *E. histolytica_PFOR  | 1162   | **D. africanus_PFOR| 1232   | 53.0  |

Supplementary Data #3

**DOCKING STUDY SETTINGS**

When the prosthetic group does not interact directly, its atom types do not matter. Two complementary studies were designed: (i) back docking of observed cofactor TPP for validation, and (ii) competitive replacement docking of NTZ and TIZ versus TPP to verify the capacity of the former to compete with the latter in its binding site. The grid dimensions were 55 x 55 x 55 Å³. The Lamarckian genetic algorithm was applied using default parameters except for: runs 256, elitism 3, root mean square deviation
(RMSD) 2.5 Å, dstep 15, qsteps 15, tstep 0.5. A cluster analysis based on RMSD values, with reference to the starting geometry, was subsequently performed and both, the lowest binding energy conformations as well as the most populated clusters were considered as solutions for subsequent interpretation [44, 53, 54].

Supplementary Data #4

HIGHER DIMENSIONAL QSAR MODELS GENERATION

In their seminal work to evaluate descriptors calculated on three-dimensional structures Marquardt et al. observed how systematical errors are loaded on the 3D-descriminator values, leading to wrong structure-property relationships [82]. A. Hopfinger et al. also identified error-prone steps in QSR and introduced a computational fourth dimension to circumvent such pitfalls by an automated search step for target binding conformers (4D-QSR) [41, SD9]. We followed a protocol of conventional 2D-QSAR and 3D-QSAR methods, as well as the protein based alignment approach to QSAR setup which is reported in the literature [SD10, SD11, SD12, SD13]. This way, we inspected some occupational features in the hydrophobic areas of the binding pocket [51].

In total, over five hundred descriptors were gathered as potential independent variables for the forward stepwise MLR equation generation. At the later stage (4D-QSR Symposar module of BiografX®), some of them were recalculated due to their conformation-dependencies (so-called 3D descriptors). Statistica 8 and Bioestat 5 while the plethora of 2D- and 3D descriptors came from Dragon 3.0, Vega ZZ as well as HyperChem 7 [44, 74, 75].

Concerning the higher dimensional QSAR [45, 46, 61, 62], and the resulting conformational arrangements of the compounds (4D-QSR output), a 5D-QSAR study was conducted (Raptor module of BiografX®) [61, 62]. Raptor maps physicochemical property projections on the molecule and projects them into the space around the aligned molecules to reflect the common parts of a virtual active site. In particular, it generates a Dual Shell representation where an inner shell reflects the solvation shell while the outer shell reflects the active site properties of the hypothetical binding site (surrogate). The molecular fields consider hydrophobicity, topology, entropy and hydrogen bonding information, all of which are used by Raptor to generate QSAR equations, too (5D-QSAR equations). To this end, the MLR Training (n=18) and Test sets (n=4) were applied here again.

Supplementary Data #5

LIGAND STRUCTURE GENERATION DETAILS

The crystal structure of NTZ was retrieved from Cambridge Crystallographic Data Centre (refcode QUZWOY, CCDC 775321[45, 46]) as a 3D template for later conformational rearrangements of 4D-QSAR (conformational search) under Symposar within the BiografX® package [45, 46]. The other molecules were generated by BioX® and underwent a conformational search under the Yeti force field (AMBER derivative) [43, SD14, SD15, SD16]. Then Gasteiger partial atomic charges were assigned within the Vega ZZ package [44].

Supplementary Data #6

DOCKING RESULTS

It is assumed that the common substructure to all 22 compounds reflects the scaffold bearing the pharmacophore (structure-activity relationships). After homology modeling of PFOR as target [50] the compounds were docked as ligands into the active site of TPP in PFOR (Autodock 4.2 [53, 54]) [SD17].

In general, ligands with a common target may be superposed using their final docked poses from docking simulations (protein based-alignment procedure). The protein-based alignment (PBA) approach for structural superposition is the starting point for 3D- and higher dimensional QSR modeling. Intriguingly, the ligand binding energies did not reflect the measured activities in all cases. A highly active compound (T03) failed and raises the question whether it can be assumes that the docked poses reflect true binding mode(s) when discrepancies between predicted and measured activities exist (Fig. SD1)? Inversely, ligand-based alignments, where all ligands are fitted onto their scaffold, also have some implications: is the scaffold rigid, and how extended is the net of side chains to the scaffold? [SD18].
Fig. (SD1). Resulting ligand conformations of docking with the PFOR model of *E. histolytica*. The conformations with highest $\Delta G_{\text{binding}}$ are shown at each case (TPP as reference). Blue: NTZ, Green: TPP, Brown: T03, Yellow: T22 (TIZ is not included). The aminoacid side chains in color indicate the principal interaction with the corresponding colored ligand. Pir: Pyruvate, Fed: Ferredoxin.

Fig. (SD2). A: Manual docking of NTZ (blue) in relation with TPP (green). B: Top view of A. C: Superposition of two NTZ conformations (final pose of *Autodock* 4 in red, *versus* manually docked pose of crystal structure from CCDC [47] in blue).

**Supplementary Data #7**

**HIGHER DIMENSIONAL QSAR RESULTS DETAILS**

The BiografX® module Symposar is able to take into account just one conformer per molecule or, instead, each molecule is considered by an ensemble of conformers comprising different orientations and protonation states. This treatment is called the fourth dimension in QSAR techniques [45]), thereby reducing the bias associated with the choice of the bioactive conformation [80].
In BiografX®, the 4D-QSAR concept (Symposar) was developed not only to consider local induced fit and hydrogen-bond (HB) flip-flop, i.e. prototropic changes between two HB donors, when one serves as acceptor. This allows the representation of some ligands by a set (or ensemble) of conformation and orientations. The contribution of a single entity within this set to the total ligand-receptor interaction energy is determined by Boltzmann criterion. The three-dimensional surrogate is represented by a family of receptor-surface models, populated with atomistic properties, hydrogen bonds, salt bridges, hydrophobic interactions in addition to solvent-mapping [45].

5D-QSAR

In the so-called fifth dimension of QSAR modeling an ensemble of different induced-fit ligand-receptor scenarios is included [80]. Even with a set of active conformers at hand (4D-QSAR step), there is still need to improve the ligand-receptor interactions due to the possible adaptations (induced fit) at the binding pocket combined with the individual ligand topologies (4D-QSAR). While Biograf’s Symposar work implies the outcome dependency on a single induced fit scenario, the 5D-QSAR module Raptor extends the 4D-QSAR concept by an additional degree of freedoms the fifth dimension which is the data collection of induced fit scenarios all of which merge into a new consensus solution. The authors of the software demonstrated “that the simulated evolution converges to a single model and that 5D-QSARs due to the fact that model selection may vary throughout the entire simulations yields less biased results than 4D-QSAR where only a single induced fit model can be evaluated at a time” [61].

5D-QSAR EQUATIONS AND PROJECTIONS (SURROGATE, THE HYPOTHETICAL BINDING SITE)

Different views of surrogate active site generated by 5D-QSAR approach (Suppl. Fig. SD3). A1, A2: side views, B: top view, C: bottom view, Fig. (6), which represents the consensus of the three 5D-QSAR equations (equation JL-2. Color code in surrogate active site: Blue: H-bond donor; Red: H-bond acceptor; Brown: Strongly hydrophobic; Yellow: weakly hydrophobic).
Fig. (SD3). A.2. Counter side view of Fig. (6).

Fig. (SD3). B. Top view of Fig. (6).

Fig. (SD3). C. Bottom view of Fig. (6).
Supplementary Data #8

PITFALLS DETAILS

Overview Concerning Pitfalls

Among the most dangerous pitfalls of QSAR studies can be found: (1) descriptor number too large, overfitting; (2) descriptor range too wide, over-weighting; (3) closely related data used for incoherent extrapolation to generate new data, over-interpretation; (4) apparent linear relationships only for a small data range, true parabolic behavior; (5) descriptor redundancy by similarity or reversion, co-linearity; (6) coincidental versus representative test sets; (7) only synthetically feasible series, synthetic bias; (8) unique scaffold superimposition versus multiple binding modes, called structural noise; (9) molecular derivatives with similarity noise; and finally (10) chance correlation in MLR [3]. A more general objective of any QSAR study is to find a predictive rule that performs better than this "no rule at all" guess of using the average. The number of components describing the degree of complexity of the model should never exceed one third of the ligand number in the training set. At some point adding more descriptors corresponds to fitting the data to noise, and the predictive ability begins to diminish.

In order to quantify true ligand-receptor interactions and to avoid that corresponding atoms in homologous series correlate only coincidentally and are mistakenly connected to the biological activity A. Hopfinger and colleagues generated higher dimensional QSAR strategies. Applying the ergodic hypothesis to molecular dynamics in which case one single small molecule can be treated during pico-seconds to simulate realistic conformational behavior of its ensemble, they exclude meaningless conformations by generation of volume occupancy descriptors at the binding site [SD9].

Pitfall: Common Action Mechanism and Multiple Binding Modes

In absence of crystallographic data, many authors avoid docking and QSAR and switch to virtual screens of compound data bases [4, SD17, SD19, SD20].

Pitfall: Prodrug Function and Instable Products of Synthesis

NTZ is reported as an active agent [25, 32], whereas in other reports its deacetylation product as the main active metabolite, i.e. TIZ is considered the true active compound. [26, 31, SD21, SD22].

Most probably, the discrepancy arises from the indistinctly used in vitro and in vivo data and the confusion thereof, since in vitro experiments lack the metabolic fate of blood and liver biotransformation in living animals. The specific deacetylation is carried out by deacetylases in all mammalians and nonspecific esterases for drug metabolism are generally found in blood, liver, kidneys and intestine. Nonetheless, the deacetylization through ester cleavage may also occur in a nonenzymatic way. A text book rule of thumb states “electron-withdrawing substituents on an aryl system of an ester enhance nonenzymatic hydrolysis by nucleophilic attack of bases (H₂O or OH⁻)” [SD23, SD24]. The hydrolysis rate for the cleavage of esters in aqueous solution depends on the electronic influences of substitution patterns (mechanistic reactivities). Besides NTZ, the study encompasses one benzologue (B01) with an acetylated hydroxyl group connected to a nitro group through resonance. The bond conjugation can be described as a mesomeric “ortho” effect which in turn parallels a known case study of para-nitrophenyl benzoate esters. They are more reactive than the corresponding benzoate esters themselves because of the strongly electron-withdrawing nitro group (-I, -M effects). In addition, the NTZ structure can be regarded as an acetylsalicylic acid (aspirine) derivative which is known for its acetylation capacity of COX target enzymes [SD25]. The common amide group bridging the two aromatic systems seen on all congeners is by far more stable for four reasons: (i) it is part of an extended push - pull system between an electron-rich and a phenyl ring system; (ii) the decreasing hydrolysis rate for acidic halids > acidic anhydrides > esters > amides (R-CO-Cl > RCO-O-CO-R > R-CO-OR > R-CO-NH-R); (iii) the higher stability – measured as lower pKₐ values of its conjugate acidic form – of the leaving group leads to a greater hydrolysis rate (HCl > R-COOH (here: R = -CH₃) > R-OH > R-NH₂); and finally (iv) the amide group lies in between two substituted (bulky) rings reducing the hydrolysis rate by sterically hindering the approaching nucleophilic/base (OH⁻). The leaving group acetate (CH₃-COO) is also favored because nonenzymatic base-catalyzed hydrolysis reacts more swiftly if the carboxyl group is not conjugated with a pi-electron or aryl systems.

To date, it remains unknown whether TIZ alone is the active moiety or not. NTZ may constitute an inactive prodrug, or on the contrary NTZ can be considered as active agent, too. Intriguingly, besides B01 and NTZ itself, no further member of the two series bears an acetyl group on the phenyl ring, which infers that it may be not prerequisite for activity. In vivo, enzymatic acetylation may occur throughout the living body with an inverted consequence of having all compounds decorated with the acetyl group like NTZ and B01 (Table 4).

Pitfall: Experimental Errors and Inappropriate Bioassays

In our case we obtained a mean value out of three replicates, but could not seize the individual values from the students’ laboratory journal, and thereof inspect the data spread and determine statistical means, like the standard deviations etc.
Pitfall: Data Size and Variety

The two compound groups including the two reference drugs comprise 22 tested molecules and the chemical variability is fairly limited which also affects the even distribution of chemical representatives in the training and test sets (Table 4). If the representativeness is at stake due to a small amount of chemical group members, random distribution may deliver uneven or inhomogeneous sets for training and testing. For instance, if there are only 2 ester members among 30 otherwise bonded molecules, the chance to see both esters in one set is higher than to see one in each set. So, hand-selecting would be a work around to put one ester in the training set and the other into the test set.

Pitfall: Meaningless Descriptor Selection

Concerning the molecular descriptor Dipole moment (DM), it depends on the geometry or conformation, for this reason is not a commonly considered parameter in QSAR studies (Table 5).

Pitfall: Collinearity and overfitting

If seen as (property) vectors in (chemical) space, two differently named parameters could load the same basic features, thusly expressing redundancy (Collinearity problem). If the two (or more) descriptors are determined with differing concepts (algorithms) they could lead to a larger “spread” of allowed points in (vector) space. Figuratively, their vectors partially load desired features but also unwanted aspects from the differing numerical approaches, all of which leads to an apparent improvement. But the gain reflects only “impurities” (unwanted aspects), that tolerate value deviations in the experimental and computed variables. Inversely, in the present case, lipole and dipole are based on the same algorithm, which is fed with values only differing by the physical measurement units: dipole computes partial (atomic) charges, while lipole calculates with partial (fragmental) lipophilicities. The hidden overlap here is given when the charges are mapped as electron densities which reflect polar and nonpolar zones, which in turn stand for hydrophilic and hydrophobic molecular segments, for instance, projected onto the molecular surface as PSA (non/polar surface area). To reduce this unnatural plasticity (and for other reasons like overfitting) it is mandatory to reduce the number of descriptors in the final equations.

Supplementary Data #9

NEW ANTIPROTOZOAL MOLECULAR DESIGN BASED ON GENERATED QSAR EQUATIONS

The 3D QSAR equation with the highest predictive power (equation JL-1) was employed to design a new antiprotozoal molecule. Based on the most active molecule treated in this study (here: pIC50=8.3), we examine the effect in the selected descriptors of different substituents in the molecular scaffold of large thiazole derivatives (most active molecules), which show the best biological activity values (threshold=8.3). The designed molecule is shown in the Suppl. Fig. SD4. A chemical inspection of molecular structure suggest a little deviation of the phenyl ring in relation of the thiazole plane ring due the formation of an Hydrogen bond between the ether Oxygen and the near proton, very similar with X-ray structure of NTZ. Since we are employing a QSAR equation considering crystal structure of NTZ, we have to imitate this 3D conformation as possible. Curiously, when we evaluate the new designed compounds with both equations (generated by MLR from crystal conformation and mathematical regularization), we can observe the highest pIC50 predicted value in both cases (different values).

Fig. (SD4). Left: New antiprotozoal molecule which was designed using the most predictive QSAR equations. Center and right: Rotational perspectives of the new designed molecule, where a small phenolic ring deviation related to thiazole plane ring can be observed.
Supplementary Data #10

SUMMARY OF QSAR EQUATIONS

Table SD2.a. QSAR Equations summary generated by MLR in this study.

| Eq. | Equation or model | Training set (n=18) | Test set (n=4) |
|-----|------------------|---------------------|--------------|
|     |                  | R² | Q² | r²m | R² | Q² | r²m | Comments/Platform                           |
| SA-1 | pIC50 = -20.72 + 0.96 aasC_Cnt + 0.50 aOm_Cnt + 6.62 ssNH_Sum - 3.45 BTZ_indicator | 0.80 | 0.55 | 0.48 | 0.72 | 0.72 | 0.70 | 0.69 | 0.69 | Difficult interpretation/ 2D-QSAR |
| JL-1 | pIC50 = -3.85 + 2.17pKa + 0.41MD + 0.9Lipole | 0.85 | 0.82 | 0.53 | 0.96 | 0.94 | 0.94 | 0.92 | 0.92 | Geometry-dependent descriptors (MD, lipole)/ 2D-QSAR, 3D-QSAR |
| JL-3 | pIC50 = -3.82 + 2.48pKa + 0.006AA | 0.71 | 0.60 | 0.33 | 0.71 | 0.42 | 0.42 | 0.50 | 0.50 | NO Geometry dependent descriptors/ 3D-QSAR |
| JL-4 | pIC50 = -27.7 + 3.54pKa + 26.4Mv + 0.0125AA | 0.77 | 0.70 | 0.41 | 0.90 | 0.44 | 0.34 | 0.36 | 0.62 | NO Geometry dependent descriptors/ 3D-QSAR |

MD: Dipole Moment (MD_Z: Z axis), Mv: Molecular van der Waals Volume, AA: Apolar area.
IF: Internal Factor (Topology), TdS: Entropy, HO: Hydrophobicity, HB: H-Bonds formation.
R²: Multiple coefficient of determination.
r²m: Square root function [69].
Q²: Cross-validated squared correlation coefficient.

Table SD2.b. QSAR equations summary generated with all molecules by MLR in this study under cross-validation.

| Eq. | QSAR Equation (n=22) | R² | R²α | Q² | r²m | Comments |
|-----|----------------------|----|-----|----|-----|----------|
| JL-5 | pIC50 = -1.58 + 0.585Lipole + 0.35Dipole + 2.02pKa | 0.74 | 0.70 | 0.65 | 0.37 | Based on all 22 molecules in ionized form (Table 4) |
| JL-6 | pIC50 = -3.825 + 0.46Lipole + 2.11pKa + 0.88Lipole | 0.87 | 0.84 | 0.85 | 0.56 | Input data from equation JL-1 (22 neutral molecules, geometry-dependent descriptors and activity values, equation JL-1) |
| JL-7 | pIC50 = -4.33 + 2.54pKa + 0.0068AA | 0.71 | 0.68 | 0.59 | 0.33 | Input data from JL-3 (22 neutral molecules, geometry-independent descriptors and activity values, equation JL-3) |
| JL-8 | pIC50 = -20.3 + 3.22pKa + 18.18Mv + 0.01AA | 0.74 | 0.70 | 0.66 | 0.37 | Input data from JL-4 (22 neutral molecules, geometry-independent descriptors and activity values, equation JL-4) |

Supplementary Data #11

BUILDING QSAR EQUATIONS BY MATHEMATICAL REGULARIZATION

A way to build a QSAR linear equation begins accepting the smooth local dependence of the biological activity to molecular descriptors surrounding the scaffold. By first order Taylor’s formula it is possible to obtain a linear approximation of the activity in a neighborhood of \( \mathbf{x}_0 \), as function of a vector of molecular descriptors.

Our goal is to find unique vector \( \mathbf{a} \) with a correct dimension, such that

\[
pIC50_\mathbf{x} - pIC50_\mathbf{x}_0 = \mathbf{a} \cdot (\mathbf{x} - \mathbf{x}_0) + \mathbf{\epsilon}(\mathbf{x}, \mathbf{x}_0);
\]

where \( \mathbf{x} \) denotes the vector of previously selected molecular descriptors as variable, \( \mathbf{x}_0 \) the value of \( \mathbf{x} \) at scaffold, \( \mathbf{x}_i \) the respective value at the \( i-th \) molecule in a \( m \) size sample of the molecular family, and \( |\mathbf{\epsilon}(\mathbf{x} - \mathbf{x}_0)| / \|\mathbf{x} - \mathbf{x}_0\| \) converges to 0 when \( \mathbf{x} \) tends to \( \mathbf{x}_0 \).
Given an activity's measurements set from the sample, we will assume that the error \( \epsilon \) is an unbiased and bounded variance statistic. The error is like in a MLR problem, but in this case we proceed from viewpoint of regularization theory [SD26, SD27].

To the case exposed in this article we worked with a matrix from a previously selected sample of descriptors with QSAR chemistry sense. We opted to use regularization techniques because we had few information about the error order, further, the smallest singular value was less than \( 10^{-1} \).

The parameters identification problem was solved by proposing \( a = a_\lambda \), where \( a_\lambda \) is Tikhonov's solution with regularization parameter \( \lambda \), that is, the solution of the optimization problem

\[
\min_{a \in \mathbb{R}^n} \| X'a - pIC50' \|_2^2 + \lambda^2 \| a \|_2^2 ;
\]

the natural \( n \) denotes how many descriptors were previously selected \((n < m)\), while \( X' \) is the matrix with \( x_i - x_0 \) as \( i \)-th row, and the coordinates of \( pIC50' \) are of the form \( pIC50_{xi} - pIC50_{x0} \).

Tikhonov's solution is here an alternative to the usual MLR problem with coordinates' origin at \((x_0; pIC50_{x0})\) and null the independent coefficient.

The L-curve technique was employed to choose the regularization parameter \( \lambda \). Our results about it are summarized in Suppl. Fig. SD5.

Without the activity data for the scaffold, we had to get an alternative for this development. A molecule in the family with the smallest squared residuals sum observed was chosen instead of scaffold. Results of the linear fit by Tikhonov's solution can be appreciated in Suppl. Fig. SD6.

![L-curve](image)

**Fig. (SD5).** L-curve for the regularization problem, alternative solution to usual MLR problem.
Fig. (SD6). Graph of residuals respect to linear model obtained by Tikhonov solution.

Equation of Tikhonov, for vector $a_\lambda$ optimization problem for QSAR models:

$$\min_{a}\|X'a - pIC50'\|^2 + \lambda^2 \|a\|^2;$$

$X'$ and $pIC50'$ are matrices for the measures of descriptors and regression.

Regarding the LOO cross-validation procedure, we present a mathematical regularization:

The QSAR equation as regularized solution of an ill posed inverse problem is an alternative way to get a linear model of activity, depending on a pre-selected set of molecular descriptors (a priori information).

Our main idea in problem formulation from this point of view is that any molecule in the family, seen as vector of molecular descriptor values, can be understood as a perturbed state of another molecule in the family which works as the geometric center of the family.

In more general terms, procedures like our in association with the given selected descriptors, the so-called regularization techniques are an appropriate alternative to the statistics.
Supplementary Data #12

LIST OF COMMON DESCRIPTORS IN QSAR

Table SD3. Comparison of descriptors employed in this work with reported ones from the literature.

| Eq. or [Ref.] | Equations and Descriptors (QSAR models) | $R^2_{\text{train}}$ (n=18) | $R^2_{\text{test}}$ (n=4) | Comments |
|---------------|----------------------------------------|-----------------------------|---------------------------|----------|
| SA-1          | pIC50 = - 20.72 + 0.96 aasC_Cnt + 0.50 aOm_Cnt + 6.62 ssNH_Sum - 3.45 BTZ_indicator  
|               | 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 which has values 0 or 1 for indicating absence or presence of benzothiazole scaffold in a given molecule respectively. | 0.80 | 0.72 | overfitting risk; Number of data points : num. of descriptors ratio = 4 : 1 |
| JL-1          | pIC50 = - 3.85 + 2.17 pK_a + 0.41 MD + 0.9 Lipole  
|               | pK_a: Acidic dissociation constant;  
|               | *MD: Dipole Moment;  
|               | *Lipole: 3D-lipophilic distribution. | 0.85 | 0.96 | *geometry-dependent descriptors, values change with torsions |
| JL-3          | pIC50 = -3.82 + 2.48pKa + 0.006 AA  
|               | pK_a: Acidic dissociation constant;  
|               | AA: Apolar Area (AA = Area_{total} - PSA). | 0.71 | 0.71 | high predictability w/o overfitting, and no geometry-dependent descriptors used |
| [3]           | number of hydrogen-bond donors and acceptors  
|               | logS (solubility in water)  
|               | log P or log D (octanol/water partition coefficient) pK_a (acidity constant)  
|               | PSA (polar surface area)  
|               | Lipinski’s Rule of Five  
|               | ADME: volume of distribution, bioavailability, permeation tests  
|               | logBB (Blood/brain partitioning)  
|               | pIC50, DL50, K_i (biol. activities) | | | chemically well-understood and henceforth interpretable (classic) descriptors in QSAR |

Supplementary Data #13

SPECTROMETRIC CHARACTERIZATION

Thiazoles

(2E)-3-(4-nitrophenyl)-N-(5-nitro-1,3-thiazol-2-yl)prop-2-enamide (T03):

Yield: 61.9 %, mp: 218.6–220.6 °C.
1H NMR (200 MHz, DMSO-d6) δ 7.02 (d, J=16.2 Hz, 1H, H-a), 7.87 (t, J=8.8 Hz, 3H, H-2', H-6', H-b), 8.27 (d, J=8.6 Hz, 2H, H-3’, H-5’), 8.63 (s, 1H, H-4) ppm.
13C NMR (50 MHz, CDCl3, DMSO-d6) δ 132 (C-5), 143 (C4), 162.6 (C-2), 1164.6 (C=O), 123.1, 124.1, 129.2, 140.3, 141.2 ppm.
MS (FAB⁺): m/z 321 (M+H)⁺.
Elemental analysis calculated for $C_{12}H_8N_2O_5S$: C, 45.00; H, 2.52; N, 17.49. Found: C, 41.8; H, 2.85; N, 17.09.

$N$-(5-Nitro-1,3-thiazol-2-yl)naphthalene-2-carboxamide (T04):

![Structure of T04]

Yield: 87.18 %, mp: 266.8–268.6 °C.

$^1H$ NMR (400 MHz, DMSO-d$_6$) $\delta$ 7.62 (dd, $J=6.2$, $J=1.1$ Hz, 1H, H-6'), 7.65 (dd, $J=4.4$, $J=1.4$ Hz, 1H, H-8'), 8.0 (d, $J=8$, 1H, H-3'), 8.05 (t, $J=6.6$, 1H, H-5'), 8.06 (t, $J=7.8$ Hz, 1H, H-4'), 8.10 (dd, $J=8.8$, $J=1.6$ Hz, 1H, H-7'), 8.69 (s, 1H, H-4), 8.80 (s, 1H, H-2') ppm.

$^{13}$C NMR (100 MHz, CDCl$_3$, DMSO-d$_6$) $\delta$ 142.1 (C-5), 142.6 (C-4), 162.7 (C-2), 166.4 (C=O) 124.2, 127.2, 127.8, 127.9, 128.5, 128.9, 129.4, 130.0, 131.9, 135.2 ppm.

MS (CI): m/z 298 (M$^+$).

Elemental analysis calculated for $C_{16}H_{16}N_2O_5S$: C, 56.18; H, 3.03; N, 14.04. Found: C, 56.22; H, 2.95; N, 14.05.

3,4,5-Trimethoxy-$N$-(5-nitro-1,3-thiazol-2-yl)benzamide (T05):

Yield: 61.9 %, mp: 215.1–217.2 °C.

![Structure of T05]

$^1H$ NMR (200 MHz, CDCl$_3$, DMSO-d$_6$) $\delta$ 3.76 (s, 3H, 4'-OCH$_3$), 3.80 (s, 6H, 3'-OCH$_3$, 5'-OCH$_3$), 7.36 (s, 2H, H-2', H-6'), 8.24 (s, 1H, H-4) ppm.

$^{13}$C NMR (50 MHz, CDCl$_3$, DMSO-d$_6$) $\delta$ 125.7 (C-5), 141.4 (C-4), 163.4 (C-2), 165.8 (C=O), 56.5, 61.0, 61.1, 106.2, 142.4, 142.9, 153.2 ppm.

MS (FAB$^+$): m/z 340 (M+H$^+$). HRMS (FAB$^+$) calculated for $C_{13}H_{14}N_2O_6S$ [M+H$^+$] 339.2958, found $C_{13}H_{14}N_2O_6S$: 340.0626.

Elemental analysis calculated for $C_{13}H_{13}N_2O_6S$: C, 46.01; H, 3.86; N, 12.38. Found: C, 46.12; H, 3.85; N, 12.35.

3-Methoxy-$N$-(5-nitro-1,3-thiazol-2-yl)benzamide (T06):

![Structure of T06]

Yield: 92.5 %, mp: 218.9–221.1 °C.

$^1H$ NMR (200 MHz, CDCl$_3$, DMSO-d$_6$) $\delta$ 3.79 (s, 3H, 3'-OCH$_3$), 7.05 (dd, $J=2.6$, $J=2.2$ Hz, 1H, H-4'), 7.32 (t, $J=8.8$ Hz, 1H, H-5'), 7.6 (s, 1H, H-2'), 7.64 (d, $J=8$, 1H, H-6'), 8.30 (s, 1H, H-4) ppm.

$^{13}$C NMR (50 MHz, CDCl$_3$, DMSO-d$_6$) $\delta$ 55.1 (OCH$_3$), 127.0 (C-5), 140.9 (C-4), 159.2 (C-2), 162 (C=O), 74.0, 112.6, 119.4, 120.4, 129.2, 131.7, 106.2, 125.7, 141.4, 142.4 ppm.

MS (FAB$^+$): m/z 280 (M+H$^+$).
Elemental analysis calculated for C11H9N3O4S: C, 47.31; H, 3.25; N, 15.05. Found: C, 47.29; H, 3.22; N, 15.01.

4-Methoxy-N-(5-nitro-1,3-thiazol-2-yl) benzamide (T07):

![Chemical structure of 4-Methoxy-N-(5-nitro-1,3-thiazol-2-yl) benzamide (T07)]

Yield: 89.6 %, mp: 233.1–235.3 °C.

$^1$H NMR (200 MHz, CDCl$_3$, DMSO-d6) δ 3.79 (s, 3H, OCH$_3$-4’), 6.89 (t, $J$=7.4, $J$=1.2 Hz, 2H, H-3’, H-5’), 8.03 (t, $J$=7.6, $J$=0.8 Hz, 2H, H-2’, H-6’), 8.29 (s, 1H, H-4) ppm.

$^{13}$C NMR (50 MHz, CDCl$_3$, DMSO-d6) δ 55.7 (OCH$_3$), 123.15 (C-5), 141.5 (C-4), 163.7 (C-2), 165.8 (C=O), 114.1, 116.0, 130.8, 142.7, 149.8 ppm.

MS (FAB+): m/z 280 (M+H)$^+$

Elemental analysis calculated for C11H9N3O4S: C, 47.31; H, 3.25; N, 15.05. Found: C, 47.35; H, 3.25; N, 15.10.

3-Nitro-N-(5-nitro-1,3-thiazol-2-yl) benzamide (T08):

![Chemical structure of 3-Nitro-N-(5-nitro-1,3-thiazol-2-yl) benzamide (T08)]

Yield: 17.9 %, mp: 142.1–143.1 °C.

$^1$H NMR (200 MHz, CDCl$_3$, DMSO-d6) δ 7.59 (s, $J$=8 Hz, 1H, H-5’), 8.2 (s, 1H, H-6’), 8.39 (s, 1H, H-4), 8.52 (d, $J$=7.6 Hz, 1H, H-4’), 9.02 (s, 1H, H-2’) ppm.

$^{13}$C NMR (50 MHz, CDCl$_3$, DMSO-d6) δ 126.2 (C-5), 148.8 (C-4), 152.9 (C-2), 175.9 (C=O), 128.6, 130.7, 134.4, 139.7, 142.7, 144.3 ppm.

MS (FAB+): m/z 295 (M+H)$^+$

Elemental analysis calculated for C10H6N4O5S: C, 40.82; H, 2.06; N, 19.04. Found: C, 40.88 H, 2.10; N, 19.10.

4-Chloro-3-nitro-N-(5-nitro-1,3-thiazol-2-yl) benzamide (T09):

![Chemical structure of 4-Chloro-3-nitro-N-(5-nitro-1,3-thiazol-2-yl) benzamide (T09)]

Yield: 95.7 %, mp: 167.8–170.0 °C.

$^1$H NMR (200 MHz, CDCl$_3$, DMSO-d6) δ 7.59 (s, $J$=8 Hz, 1H, H-2’), 8.22 (dd, $J$= 8.2, $J$= 2 Hz, 2H, H-5’, H-6’), 8.72 (d, $J$=1.8 Hz, 1H, H-4) ppm.

$^{13}$C NMR (50 MHz, CDCl$_3$, DMSO-d6) δ 127 (C-5), 141.4 (C-4), 163.5 (C-2), 164.1 (C=O), 126.0, 131.6, 132.5, 133.3, 147.8 ppm.

MS (FAB+): m/z 379 (M+H)$^+$.
Elemental analysis calculated for C₁₀H₇N₃O₃S: C, 48.19; H, 2.83; N, 16.86. Found: C, 48.22; H, 2.85; N, 16.92.

N-(5-Nitro-1,3-thiazol-2-yl)benzamide (T17):

Yield: 87.2 %, mp: 258.6–261.0 °C [31].

¹H NMR (200 MHz, DMSO-d₆) δ 7.34-7.60 (m, 3H, H-3’, H-4’, H-5’), 8.09 (dd, Jmeta= 1.6 Hz, Jorto= 8.4 Hz, 2H, H-2’, H-6’), 8.33 (s, 1H, H-4) ppm.

¹³C NMR (50 MHz, DMSO-d₆) δ 128.5 (2C, C-3’, C-5’), 128.7 (2C, C-2’, C-6’), 130.7 (C-5), 142.0 (C-1’), 142.5 (C-4), 162.4 (C-2), 166.3 (C=O) ppm.

MS (FAB+): m/z 250 (M+H)+. HRMS (FAB+) calculated for C₁₀H₇N₃O₃S [M+H]+ 250.247, found: 250.0271.

Elemental analysis calculated for C₁₀H₇N₃O₃S: C, 48.19; H, 2.83; N, 16.86. Found: C, 48.22; H, 2.85; N, 16.92.

4-Nitro-N-(5-nitro-1,3-thiazol-2-yl)benzamide (T18):

Yield: 95.1 %, mp: 172.0–174.6 °C [32].

¹H NMR (400 MHz, DMSO-d₆) δ 8.33 (ddd, Jmeta= 2.0 Hz, Jmeta= 2.0 Hz, Jorto= 9.0 Hz, 2H, H-2’, H-6’); 8.39 (ddd, Jmeta= 2.0 Hz, Jmeta= 2.0 Hz, Jorto= 9.0 Hz, 2H, H-3’, H-5’) ppm.

¹³C NMR (100 MHz, DMSO-d₆) δ 121.7 (C-4´), 128.7 (2C, C-2’, C-6’); 130.0 (2C, C-2’, C-6’); 136.5 (C-5); 142.2 (C-1’); 142.2 (C-4); 149.8 (C-4’); 162.4 (C-2); 165.2 (C=O) ppm.

MS (FAB+): m/z 295 (M+H)+. HRMS (FAB+) calculated for C₁₀H₆N₄O₅S [M+H]+ 295.245, found: 295.0134.

Elemental analysis calculated for C₁₀H₆N₄O₅S: C, 48.19; H, 2.83; N, 16.86. Found: C, 48.22; H, 2.85; N, 16.92.

3,5-Dinitro-N-(5-nitro-1,3-thiazol-2-yl)benzamide (T19):

Yield: 12.4 %, mp: 131.4–134.0 °C.

¹H NMR (200 MHz, DMSO-d₆) δ 8.32 (s, 1H, H-4); 9.12 (t, Jmeta= 2.0 Hz, 1H, H-4’); 9.43 (d, Jmeta= 2.0 Hz, 2H, H-2’, H-6’) ppm.

¹³C NMR (50 MHz, DMSO-d₆) δ 121.7 (C-4’), 128.7 (2C, C-2’, C-6’), 134.4 (C-5), 140.3 (2C, C-3’, C-5’), 142.9 (C-1’), 148.3 (C-4), 162.2 (C-2), 166.0 (C=O) ppm.

MS (FAB+): m/z 340 (M+H)+.
Supplementary Material

Elemental analysis calculated for C_{10}H_{5}N_{5}O_{7}S: C, 35.40; H, 1.49; N, 20.64. Found: C, 35.34; H, 1.53; N, 20.62.

(2E)-N-(5-Nitro-1,3-thiazol-2-yl)-3-phenylprop-2-enamide (T22):

Yield: 87.9 %, mp: 243.6–246.6 °C.

\(^1\)H NMR (400 MHz, DMSO-d6) \(\delta\) 6.92 (d, \(J_{a-b} = 16.0\) Hz, 1H, COCH=CH); 7.48 (dd, \(J_{meta} = 2.4\) Hz, \(J_{orto} = 6.4\) Hz, 3H, H-3’, H-4’, H-5’); 7.68 (dd, \(J_{meta} = 3.0\) Hz, \(J_{meta} = 7.0\) Hz, 2H, H-2’, H-6’); 7.85 (d, \(J_{a-b} = 16.0\) Hz, 1H, COCH=CH), 8.67 (s, 1H, H-4) ppm.

\(^{13}\)C NMR (100 MHz, DMSO-d6) \(\delta\) 118.1 (OCCH), 125.6 (C-5´), 128.2 (2C, C-3’, C-5’), 129.1 (2C, C-2’, C-6’), 130.8 (C-4´), 133.8 (C-5), 141.9 (C-1´), 142.8 (CHPh), 144.5 (C-4), 161.8 (C-2), 164.7 (C=O) ppm.

MS (FAB+): m/z 276 (M+H)+. HRMS (FAB+): calculated for C_{12}H_{9}N_{3}O_{3}S [M+H]+ 276.284, found: 276.0477.

Elemental analysis calculated for C_{12}H_{9}N_{3}O_{3}S: C, 52.36; H, 3.30; N, 15.26. Found: C, 52.34; H, 3.35; N, 15.22.

Benzothiazoles

2-[(6-Nitro-1,3-benzothiazol-2-yl)carbamoyl]phenyl acetate (B01):

Yield: 73.2 %, mp: 300.0–303.5 °C [25].

\(^1\)H NMR (400 MHz, DMSO-d6) \(\delta\) 2.25 (s, 3H, CH_{3}), 7.03 (t, \(J_{orto} = 7.52\) Hz, 1H, H-5’), 7.08 (d, \(J_{orto} = 8.20\) Hz, 1H, H-3’), 7.52 (td, \(J_{meta} = 1.74\) Hz, \(J_{orto} = 8.60\) Hz, 1H, H-4’), 7.88 (d, \(J_{orto} = 8.90\) Hz, 1H, H-4), 7.98 (dd, \(J_{meta} = 1.75\) Hz, \(J_{orto} = 7.90\) Hz, 1H, H-6’), 8.31 (dd, \(J_{meta} = 2.40\) Hz, \(J_{orto} = 8.90\) Hz, 1H, H-5), 9.10 (d, \(J_{meta} = 2.0\) Hz, 1H, H-7), 12.76 (br, 1H, NH) ppm.

\(^{13}\)C NMR (100 MHz, DMSO-d6) \(\delta\) 22.9 (CH_{3}), 117.3 (C-3’), 119.1 (C-7), 120.6 (C-4), 121.8 (C-5), 122.0 (C-1´), 130.6 (C-6), 132.2 (C-4´), 135.0 (C-6´), 135.0 (C-7a), 142.9 (C-5´), 143.1 (C-2´), 143.1 (C-4a), 153.5 (C-2), 163.5 (NHC=O), 170.2 (OC=O) ppm.

MS (FAB+): m/z 358 (M+H)+. HRMS (FAB+): m/z 358.0497 [M+H]+ (calcd for C_{16}H_{11}N_{5}O_{5}S [M+H]+ 358.0498).

Elemental analysis calculated for C_{16}H_{11}N_{5}O_{5}S: C, 53.78; H, 3.10; N, 11.76. Found: C, 53.81; H, 3.15; N, 11.78.

2-Hydroxy-N-(6-nitro-1,3-benzothiazol-2-yl)benzamide (B02):
Yield: 81.8 %, mp: 317.0–320.9 °C [25].

\[ ^1H \text{NMR} \\text{(400 MHz, DMSO-}d_6\text{)} \delta \ 6.95 \text{ (t, } J_{orto} = 7.90 \text{ Hz, 1H, H-5' }, 7.08 \text{ (d, } J_{orto} = 7.96 \text{ Hz, 1H, H-3' }, 7.51 \text{ (td, } J_{meta} = 1.76 \text{ Hz, } J_{orto} = 8.48 \text{ Hz, 1H, H-4' }, 7.87 \text{ (d, } J_{orto} = 9.16 \text{ Hz, 1H, H-4}, 7.97 \text{ (dd, } J_{meta} = 1.68 \text{ Hz, } J_{orto} = 7.92 \text{ Hz, 1H, H-6' }, 8.28 \text{ (dd, } J_{meta} = 2.46 \text{ Hz, } J_{orto} = 8.94 \text{ Hz, 1H, H-5}), 9.07 \text{ (d, } J_{meta} = 2.40 \text{ Hz, 1H, H-7), 12.76 (br, 1H, NH) ppm.}
\]

\[ ^{13}C \text{NMR (100 MHz, DMSO-}d_6\text{)} \delta 117.3 \text{ (C-3'), 119.0 (C-7), 120.6 (C-4), 121.8 (C-5), 122.0 (C-1'), 130.6 (C-6), 132.2 (C-4'), 135.0 (C-6'), 135.0 (C-7a), 142.9 (C-5'), 143.1 (C-4a), 153.5 (C-2), 157.5 (C-2'), 163.5 (\text{NHC=O) ppm.}
\]

MS (FAB+): m/z 316 (M+H)+. HRMS (FAB+): m/z 316.0356 [M+H]+ (calcld for C_{14}H_{12}N_{3}O_{4}S+ 316.0392).

Elemental analysis calculated for C_{14}H_{9}N_{3}O_{4}S: C, 53.33; H, 2.88; N, 13.33. Found: C, 53.39; H, 2.90; N, 13.38.

\[ N-(6-Nitro-1,3-benzothiazol-2-yl)naphthalene-2-carboxamide (B10): \]

Yield: 98.8 %, mp: 264.7–266.6 °C.

\[ ^1H \text{NMR} \text{(400 MHz, DMSO-}d_6\text{)} \delta 7.72 - 7.64 \text{ (m, 2H, H-4', H-6'), 7.94 (d, } J = 8.8 \text{ Hz, 1H, H-4), 8.04 (d, } J = 8 \text{ Hz, 1H, H-8'), 8.1 (t, } J = 8.4, J = 4.4 \text{ Hz, 2H, H-3', H-5'), 8.16 (d, } J = 8.4 \text{ Hz, 1H, H-7'), 8.3 (dd, } J = 8.8, J = 2 \text{ Hz, 1H, H-5), 8.86 (s, 1H, H-2'), 9.09 (d, } J = 2.2, 1H, H-7 \text{) ppm.}
\]

\[ ^{13}C \text{NMR (100 MHz, DMSO-}d_6\text{)} \delta 118.9 \text{ (C-7), 124.3 (C-5), 134.8 (C-8), 142.9 (C-9) 152.2 (C-6), 164.6 (C-4), 166.6 (C=O), 127.3, 127.7, 128.6, 128.8, 129.3, 129.7, 131.9, 132.2 \text{ ppm.}
\]

MS (FAB+): m/z 350 (M+H)+. 

Elemental Analysis calculated for C_{18}H_{11}N_{3}O_{3}S: C, 61.88; H, 3.17; N, 12.03. found : C, 61.35 ; H, 3.19 ; N, 12.04.

\[ 2-Nitro-N-(6-nitro-1,3-benzothiazol-2-yl)benzamide (B11): \]

Yield: 32.5 %, mp: 241.0–244.1 °C.

\[ ^1H \text{NMR} \text{(200 MHz, DMSO-}d_6\text{)} \delta: 7.69 - 7.73 \text{ (m, H4, H-3',H-4', H-5', H-6'), 8.19 (d, } J = 8 \text{Hz, 1H, H-4), 8.29 (dd, } J = 8.8, J = 2.4 \text{ Hz, 1H, H-5), 9.07 (d, } J = 2.6 \text{ Hz, 1H, H-7) ppm.}
\]

MS (Cl+): m/z 343 (M-H)+.
**Elemental analysis** calculated for C\textsubscript{14}H\textsubscript{9}N\textsubscript{3}O\textsubscript{4}S: C, 53.33; H, 2.88; N, 13.33. Found: C, 53.39; H, 2.90; N, 13.38.

4-Chloro-3-nitro-N-(6-nitro-1,3-benzothiazol-2-yl)benzamide (B12):

```
O^+  
N
S
H
O
+  
O^-  
N
O
```

Yield: 95.1 %, mp: 322.0 °C (dec).

\textsuperscript{1}H NMR (200 MHz, DMSO-d\textsubscript{6}) \(\delta\): 7.46 (d, \textit{J}= 8.6 Hz, 1H, H-4), 7.5 (d, \textit{J}= 8 Hz, 1H, H-2'), 8.0 (dd, \textit{J}= 8.8, 1H, H-6', H-5'), 8.2 (dd, \textit{J}= 8.4, \textit{J}= 2 Hz, 1H, H-5), 8.68 (d, \textit{J}= 2 Hz, 1H, H-7) ppm.

MS (FAB+): m/z 379 (M+H)+. HRMS (FAB+) Calculated for C\textsubscript{14}H\textsubscript{7}ClN\textsubscript{4}O\textsubscript{5}S [M+H] 378.7458, found: 378.9987.

**Elemental analysis** calculated for C\textsubscript{14}H\textsubscript{7}ClN\textsubscript{4}O\textsubscript{5}S: C, 44.40; H, 1.86; N, 14.79. Found: C, 44.35; H, 1.90; N, 14.77.

N-(6-Nitro-1,3-benzothiazol-2-yl)benzamide (B25):

```
O^+  
N
S
H
O
```

Yield: 89.3 %, mp: 275.0–278.8 °C.

\textsuperscript{1}H NMR (400 MHz, DMSO-d\textsubscript{6}) \(\delta\): 7.59 (t, \textit{J}_{\text{orto}}= 8.0 Hz, 2H, H-3', H-5'), 7.70 (t, \textit{J}_{\text{orto}}= 8.0 Hz, 1H, H-4'), 8.16 (d, \textit{J}_{\text{orto}}= 7.30 Hz, 2H, H-2', H-6'), 8.32 (dd, \textit{J}_{\text{meta}}= 2.40 Hz, \textit{J}_{\text{orto}}= 8.90 Hz, 1H, H-5), 9.11 (s, 1H, H-7), 13.30 (br, 1H, NH) ppm.

\textsuperscript{13}C NMR (100 MHz, DMSO-d\textsubscript{6}) \(\delta\): 119.2 (C-7), 120.7 (C-4), 120.7 (C-1'), 121.9 (C-5), 128.5 (2C, C-2', C-6'), 128.8 (2C, C-3', C-5'), 131.4 (C-6), 132.3 (C-7a), 133.3 (C-4'), 143.1 (C-4a), 153.5 (C-2), 166.4 (NH=O) ppm.

MS (FAB+): m/z 300 (M+H)+.

**Elemental analysis** calculated for C\textsubscript{14}H\textsubscript{7}ClN\textsubscript{4}O\textsubscript{5}S: C, 56.18; H, 3.03; N, 14.04. Found: C, 56.01; H, 3.10; N, 14.08.

4-Nitro-N-(6-nitro-1,3-benzothiazol-2-yl)benzamide (B26):

```
O^+  
N
S
H
O
```

Yield: 95.5 %, mp: 282.0–285.5 °C.

\textsuperscript{1}H NMR (400 MHz, DMSO-d\textsubscript{6}) \(\delta\): 7.97 (d, \textit{J}_{\text{orto}}= 8.72 Hz, 1H, H-4); 8.33 (dd, \textit{J}_{\text{meta}}= 2.54 Hz, \textit{J}_{\text{orto}}= 9.0 Hz, 1H, H-5); 8.26-8.45 (m, 4H, H-2', H-3', H-5', H-6'); 9.14 (d, \textit{J}_{\text{meta}}= 2.32 Hz, 1H, H-7); 13.57 (br, 1H, NH) ppm.

\textsuperscript{13}C NMR (100 MHz, DMSO-d\textsubscript{6}) \(\delta\): 119.3 (C-7); 122.0 (C-4); 122.0 (C-5); 123.8 (2C, C-2', C-6'); 129.0 (2C, C-2', C-6'); 130.2 (C-7a); 128.9 (C-1'); 130.2 (C-6); 143.3 (C-4a); 148.28 (C-2'), 150.0 (C-4'); 150.0 (C-2); 165.0 (C=O) ppm.
Supplementary Material

MS (FAB+): m/z 345 (M+H)+.

Elemental analysis calculated for C_{14}H_{18}N_{6}O_{3}S: C, 48.84; H, 2.34; N, 16.27. Found: C, 48.80; H, 2.38; N, 16.27.

3,5-Dinitro-N-(6-nitro-1,3-benzothiazol-2-yl)benzamide (B27):

Yield: 54.9 %, mp: 220.0–224.6 °C.

H NMR (400 MHz, DMSO-d6) δ: 7.41 (d, Jorto= 8.92 Hz, 1H, H-4); 8.25 (s, 3H, H-2’, H-4’, H-6’); 8.32 (dd, Jmeta= 2.48 Hz, Jorto= 8.92 Hz, 1H, H-5); 8.69 (d, Jmeta= 2.48 Hz, 1H, H-7) ppm.

C NMR (100 MHz, DMSO-d6) δ: 131.6 (C-7); 134.2 (C-7a); 143.4 (C-4a); 156.8 (C-2); 156.6 (C=O) ppm.

Elemental analysis calculated for C_{14}H_{18}N_{6}O_{3}S: C, 43.19; H, 1.81; N, 17.99. Found: C, 43.22; H, 1.78; N, 17.95.

(2E)-N-(6-Nitro-1,3-benzothiazol-2-yl)-3-phenylprop-2-enamide (B30):

Yield: 82.1 %, mp: 270.0–272.3 °C.

H NMR (400 MHz, DMSO-d6) δ: 6.53 (d, Ja-b= 16.04 Hz, 1H, Ha); 7.33 (d, Ja-b= 16.04 Hz, 1H, Hb); 7.42 (dd, Jmeta= 2.51 Hz, Jorto= 4.60 Hz, Jorto= 6.46 Hz, 2H, H-2’, H-6’); 7.49 (dd, Jmeta= 1.82 Hz, Jorto= 5.02 Hz, 1H, H-4’); 7.68 (dd, Jmeta= 1.68 Hz, Jorto= 3.60 Hz, Jorto= 5.04 Hz, 2H, H-2’, H-6’); 7.92 (d, Jorto= 8.96 Hz, 1H, H-4); 8.29 (dd, Jmeta= 2.44 Hz, Jorto= 8.92 Hz, 1H, H-5); 9.09 (d, Jmeta= 2.44 Hz, 1H, H-7); 13.01 (br, 1H, NH) ppm.

C NMR (100 MHz, DMSO-d6) δ: 117.9 (C-a); 119.1 (C-7); 119.1 (C-4’); 120.6 (C-4); 121.9 (C-5); 128.3 (2C, C-2’, C-6’), 129.0 (2C, C-3’, C-5’); 130.6 (C-6); 134.0 (C-1’); 134.2 (C-7a); 143.0 (C-4a); 144.7 (C-b); 153.6 (C-2), 167.6 (C=O) ppm.

Elemental analysis calculated for C_{16}H_{11}N_{3}O_{3}S: C, 59.07; H, 3.41; N, 12.92. Found: C, 59.05; H, 3.39; N, 12.94.

SUPPLEMENTARY MATERIAL REFERENCES

[SD1] Cedillo-Rivera, R.; Chávez, B.; González-Robbles, A.; Tapia, A.; Yépez-Mulía, L. In vitro effect of nitazoxanide against Entamoeba histolytica, Giardia intestinalis and Trichomonas vaginalis trophozoites. J. Eukaryot. Microbiol., 2002, 49(3), 201-208.

[SD2] Hernández-Núñez, E.; Tlahuext, H.; Moo-Puc, R.; Torres-Gómez, H.; Reyes-Martínez, R.; Cedillo-Rivera, R.; Nava-Zuazo, C.; Navarrete-Vazquez, G. Synthesis and in vitro trichonemicidal, giardicidal and amebicidal activity of N-acetamide(sulfonamide)-2-methyl-4-nitro-1H-imidazoles. Eur. J. Med. Chem., 2009, 44(7), 2975-2984.

[SD3] Nava-Zuazo, C.; Estrada-Soto, S.; Guerrero-Alvarez, J.; León-Rivera, I.; Molina-Salinas, G.M.; Saíd-Fernández, S.; Chan-Bacab, M.J.; Cedillo-Rivera, R.; Moo-Puc, R.; Mirón-López, G.; Navarrete-Vazquez, G. Design, synthesis, and in vitro antiprotozoal, antimycobacterial activities of N-[2-(2-[7-chloroquinolin-4-yl]amino)ethyl]ureas. Bioorg. Med. Chem., 2010, 18(17), 6398-6403.

[SD4] UniProt Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Res., 2012, 40(Database issue), D71-75.

[SD5] Cates, S. NCBI: National Center for Biotechnology Information, Connexions Web site. http://cnx.org/content/m11789/1.3/, Feb 21, 2006. National Center for Biotechnology Information, U.S. National Library of Medicine 8600 Rockville Pike, Bethesda MD, 20894 USA.

[SD6] Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; Thompson, J.D.; Gibson, T.J.; Higgins, D.G. ClustalW and ClustalX version 2.0. Bioinformatics, 2007, 23(21), 2947-2948.
