

Receptor-Dependent 4D-QSAR Analysis of Peptidemimetic Inhibitors of *Trypanosoma cruzi* Trypanothione Reductase with Receptor-Based Alignment

Samuel Silva da Rocha Pita^{1,***}, Magaly Girão Albuquerque^{1,*}, Carlos Rangel Rodrigues², Helena Carla Castro³ and Anton J. Hopfinger⁴

¹Universidade Federal do Rio de Janeiro, PPGQu, IQ, CT, LabMMol, 21949-900, Rio de Janeiro, RJ, Brazil

²Universidade Federal do Rio de Janeiro, CCS, Faculdade de Farmácia, ModMolQSAR, 21941-590, Rio de Janeiro, RJ, Brazil

³Universidade Federal Fluminense, CEG, Instituto de Biologia, LaBioMol, 24210-130, Niteroi, RJ, Brazil

⁴University of New Mexico, College of Pharmacy, 87131-0001, Albuquerque, New Mexico, USA

*Corresponding author: MGA, magaly@iq.ufrj.br; SSRP, samuel.pita@ufba.br

** Permanent address: Universidade Federal da Bahia, Faculdade de Farmácia, LABIMM, 40170-115, Salvador, BA, Brazil

Receptor-dependent four-dimensional quantitative structure–activity relationship (RD-4D-QSAR) studies were applied on a series of 21 peptides reversible inhibitors of *Trypanosoma cruzi* trypanothione reductase (TR) (Amino Acids, 20, 2001, 145). The RD-4D-QSAR (J Chem Inform Comp Sci, 43, 2003, 1591) approach can evaluate multiple conformations from molecular dynamics simulation and several superposition structure alignments inside a box composed by unitary cubic cells. The descriptors are the occupancy frequency of the atoms types inside the grid cells. We could develop 3D-QSAR models that were highly predictive (q^2 above 0.71). The 3D-QSAR models can be visualized as a spatial map of atom types that are important on the comprehension of the ligand–enzyme interaction mechanism, pointing main pharmacophoric groups and TR subsites described in the literature. We were able also to identify some TR subsites for further development in the drug discovery process against tropical diseases not yet studied.

Key words: Chagas' disease, molecular modeling, receptor-dependent four-dimensional quantitative structure–activity relationship, three-dimensional quantitative structure–activity relationship, trypanothione reductase

Received 28 September 2011, revised 25 November 2011 and accepted for publication 6 January 2012

Since 1960s and earlier works of Hansch and Fujita (1,2), quantitative relationships between structures of many classes of compounds and their biological response have been applied in several examples of drug–receptor problems.

The development of computer programs and understanding of the mechanism of drug–receptor interaction allowed to correlate not only the chemical structure–derived two-dimensional (2D) descriptors with the variation of biological response of a series of compounds (e.g., pharmacological activity or toxicity), but also the chemical structure–derived three-dimensional (3D) descriptors. The most popular applied QSAR approach using 3D descriptors, namely 3D-QSAR, was developed by Cramer *et al.* (3) and called Comparative Molecular Field Analysis (CoMFA). Since the mechanism of drug–receptor interaction is dependent of both, ligand and receptor structures, the knowledge of their 3D structures and conformational behavior in biological media is pivotal. Therefore, although the knowledge of the 3D structure of the receptor and the conformation profile of the ligand was increasing, most of the 3D-QSAR studies did not consider them.

Hopfinger *et al.* (4) in last decades introduced the 'multiple conformation' concept in a new 3D-QSAR approach named 4D-QSAR method. The 4D-QSAR method is able to construct 3D-QSAR models as 3D pharmacophoric maps (4), precisely the type of models one would like to have to complement CoMFA models (3). 4D-QSAR analysis (4) generates their models as a function of alignment for a set of analogues upon both receptor-independent (RI) (4) and receptor-dependent (RD) geometry (5–7). The merits of 4D-QSAR method are in its ability to (8) (i) incorporate ligand-conformational flexibility, (ii) explore multiple alignments, (iii) evaluate ligand-embedded pharmacophore groups in so as part of the QSAR models building and optimization process and (iv) propose an 'active' ligand conformation. Because the capability of the 4D-QSAR in exploring large degrees of freedom of both conformational and alignment freedoms in the search for active conformation binding mode for each compound investigated, the 4D-QSAR analysis has been successfully applied to a variety of structure–activity training sets (4–9) and has proven useful and reliable in both RI-4D-QSAR (4,8,9) and RD-4D-QSAR (5–7). Although the RD approach takes into account the receptor topology, the recent RD-4D-QSAR applications perform the alignment

step using the ligand structure. In this report, we have applied the RD-4D-QSAR formalism to construct 3D-QSAR models for a series of inhibitors of *Trypanosoma cruzi* trypanothione reductase (TR) (E.C. 1.6.4.8) using the receptor (TR enzyme) structure in the alignment step.

Since its discovery in 1909, Chagas' disease (American trypanosomiasis) has been affecting people in Latin America. Nowadays, there are 200 000 people infected by protozoa *T. cruzi* and 18–20 million at risk of acquiring this disease (10). Many efforts have been developed to overcome trypanosomiasis, and selective inhibition of vital enzymes of parasites is the most successful (10). Trypanothione reductase (TR) is a homodimeric enzyme responsible for maintaining milieu cellular in an oxidative state, reducing its substrate, trypanothione (N^1, N^3 -bis-glutathionyl spermidine, TS_2), to the active form, dihydro-trypanothione ($T(SH)_2$), a parasite antioxidant agent. The TR is homologous to the human glutathione reductase (GR), but their structural differences in the active site to recognize their specific substrates made the development of TR selective inhibitors a possible issue (11–14). The 3D structure of TR co-crystallized to its natural substrate (trypanothione) was solved by Bond *et al.* (11), and it is available at the Protein Data Bank (PDB) under the entry code 1BZL.

Some classes of compounds have been tested against TR (13,14), and the most promising one was a series of 21 peptide mimetics, developed by McKie *et al.*, (12) because they can act as selective TR inhibitors through a substrate competitive mechanism. Many classes of compounds were tested against TR and presented broad results, that is, values for inhibition concentration varied from millito nanomolar range (11,15–24). As there are no crystallographic data available about peptides bounded to TR, we have applied the 4D-QSAR method to the McKie *et al.* (12) series of peptide mimetics to get same insight into their binding mode. Moreover, this work represents the first RD-4D-QSAR application, where the alignment step was performed based on the receptor structure. Our results were compared with structural information derived from experimental and theoretical results (12,25).

Materials and Methods

The RD-4D-QSAR analysis was performed on the 4D-QSAR software package, version 3.0 (26), installed on a Silicon Graphics workstation (processor IP32, 150 MHz CPU). The general steps of the RD-4D-QSAR method was described elsewhere (5), and it will be exposed a short explanation pointing out the major differences in the present study.

A series of 21 peptide inhibitors of *T. cruzi* TR (Table 1) were retrieved from McKie *et al.* (12). The assay values of I_{50} (12) were converted in their negative logarithmic units, pI_{50} , which comprises the set of dependent variables in the 4D-QSAR analysis. The range in activity for the peptide mimetics is about 3.5 pI_{50} units.

Step 1

All 3D structures of each 21 peptide mimetic sets (Table 1) were built using the HYPERCHEM 7.0 software (27). The peptide structures

Table 1: Numbering, structures, and pI_{50} (M) values of the 21 compounds tested against TR of *Trypanosoma cruzi* (12)

No.	Peptide mimetics structures	pI_{50}
1	benzoyl-Arg- <i>p</i> -nitroanilide	3.26
2	benzyloxycarbonyl-Arg- <i>p</i> -nitroanilide	3.40
3	H-Arg- β -naphthylamide	2.82
4	H-Trp- β -naphthylamide	2.57
5	benzyloxycarbonyl-Lys-4-methoxy- β -naphthylamide	3.09
6	H-Trp-OH (tryptophan amino acid)	2.57
7	benzyloxycarbonyl-Arg-Arg-4-methoxy- β -naphthylamide	4.41
8	benzyloxycarbonyl-Arg-Arg- <i>p</i> -nitroanilide	4.22
9	H-His-Trp-His-OH	2.76
10	H-His-Trp-Lys-OH	2.41
11	H-Phe-Arg-Trp-OH	3.40
12	H-Phe-Pro-Arg-4-methoxy- β -naphthylamide	3.26
13	benzoyl-Leu-Arg-Arg- β -naphthylamide	4.93
14	<i>tert</i> -butyl-Leu-Arg-Arg-7-amide-methylcoumarin	3.45
15	<i>tert</i> -butyl-Leu-Lys-Arg-7-amide-methylcoumarin	3.38
16	benzoyl-Gly-Arg-Arg-Leu- β -naphthylamide	3.29
17	benzyloxycarbonyl-Gly-Gly-Arg-7-amide-methylcoumarin	3.53
18	benzyloxycarbonyl-Ala-Arg-Arg-4-methoxy- β -naphthylamide	4.98
19	benzoyl-Lys-Phe-Arg- <i>p</i> -nitroanilide	3.62
20	benzoyl-Phe-Val-Arg-7-amide-4-methylcoumarin	3.61
21	H-Phe-Met-Arg-Phe-NH ₂	2.56

were assembled using the HyperChem amino acids' database, and the most extended beta-sheet conformation was adopted. Each structure was energy-minimized until the gradient reached 0.001 kcal/mol using the conjugated-gradient algorithm under the AMBER force field (28) on HyperChem. Partial atomic charges were computed using AM1 semi-empirical method (29) also implemented in this program. These 21 modeled structures were used as the initial structures, which will be docked at the TR Receptor Model (see Step 2).

Step 2

The co-ordinates of TR complexed with their natural substrate was extracted from the Protein Data Bank (30) (PDB ID: 1BZL). As it is computationally demanding to sample this entire structure using molecular dynamics (MD) simulation, the receptor pruning technique was applied (5), also intending to achieve reasonable ensemble profile and performing practical RD-4D-QSAR analysis in terms of time and computational resources. Therefore, receptor pruning was performed using the WEBLAB VIEWER software (31). All residues in a radius of 11.5 Å of all atoms of trypanothione (TS_2) were included in the pruned receptor, named *Receptor Model*. To retain the integrity of the local geometric environment of the receptor, hydrogen atoms were added to complete the terminal fragments of the residues in the pruned model. AMBER partial atomic charges were assigned (28) to all atoms of the TR Receptor Model. As the protein was not neutral at the assay (12) and many charged residues are important for electrostatic interaction at the active site (18,25,32), we maintained the AMBER partial atomic charges at the enzyme. The largest inhibitor of our data set, peptide **16** (Table 1), is posed into the *Receptor Model* (Figure 1), revealing that all 21 peptides fit into the enzyme pruned at 11.5 Å cutoff.

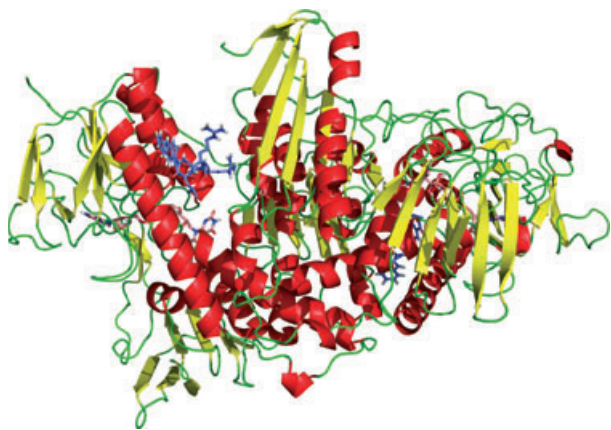


Figure 1: Schematic representation of *Trypanosoma cruzi* trypanothione reductase (TR) *Receptor Model* bounded with peptide 16. The peptide mimetics are represented in thick purple licorice, FAD (cofactor) is shown in salmon stick, and the enzyme secondary structure presented the alpha-helices colored in red, beta-sheets in yellow, and loop regions in green.

Step 3

All the inhibitors have significant structural differences from the bounded substrate (TS_2), and the alignment used to generate the initial binding geometry for each peptide–TR complex followed the orientation reported by McKie *et al.* (12) and our previous docking studies (33).

After the alignment, the TS_2 was removed from the *Receptor Model* and the corresponding set of peptide–TR complexes was generated. To acquire reliable structures for MD simulation, these 21 systems were energy-minimized until a gradient reached 0.001 kcal/mol on HYPERCHEM 7.0 software (27), using the AMBER force field (28), applying this sequential procedure: the peptide docked into the pruned protein was minimized, followed by the hydrogen atoms, the backbone atoms, and finally, with a restrained potential at the C- and N-termini of the protein, the entire complex.

Step 4

All atoms of the peptide–TR complexes were assigned according to the atom types, named Interaction Pharmacophore Elements (IPE), available in the 4D-QSAR software (26), which permits the classification of enzyme–inhibitor interactions and were defined as follows (5): any type of atom (*any*); nonpolar atom (*np*); polar atom of positive partial charge (*p+*); polar atom of negative partial charge (*p-*); hydrogen bond acceptor (*hba*); hydrogen bond donor (*hbd*); and aromatic atoms (*ar*).

Step 5

In the 4D-QSAR method, the MD simulation is used to generate the conformational ensemble profile (CEP) of each peptide–TR complex (5). Each low-energy complex resulting from Step 3 was used as the initial structure for the MD simulation, which was performed in the MOLSIM software (34) available in the 4D-QSAR package (26). Three hundred conformations of each complex were obtained for sampling,

with a time step of 0.001 ps, at a temperature of 298 K, resembling the biological assay (12), and a molecular dielectric of $3.0 \epsilon_0$.

Because the potential energy of the system was stabilized at the last 100 steps (data not shown), these final conformations were evaluated for constructing the CEP and the corresponding grid cell occupancy descriptors (GCOD) of each peptide–TR complex model. The RD-4D-QSAR method does not use a single conformation when constructing a 4D-QSAR model, but the intrinsic conformational flexibility of each complex is taken into account through its CEP (9).

Step 6

Alignments in the RD-4D-QSAR study can introduce sterically forbidden overlaps of parts of a ligand with parts of the binding model; overall, flexible and/or non-equivalent atoms among ligands are not good alignment atoms because of the high probability of introducing structural damage to the pruned receptor–ligand complex (5). As a result of immense differences in the structures of the database (Table 1), pointing out that the peptide–TR complex model contained the same residues originated from the TR crystallographic structure and the 21 peptides bind to the pruned enzyme in a diverse mode, we choose the alpha-carbon atoms ($C\alpha$) from the backbone for the alignment and tested various $C\alpha$ sets. The current 4D-QSAR algorithm (26) considers 'three-ordered atom match' alignment rule, and seven alignments were chosen to cover the entire topology of peptide–TR complex model, including residues previously related to interacting with the substrate and other inhibitors (11,16,18,25,32). The corresponding alignments and residue numbers are listed in Table 2.

Step 7

The CEPs of each peptide–TR complex model (consisting of 100 conformations recorded from each MD sampling) were placed in a reference cell lattice, according to the seven alignments (see Step 6). The resolution of the grid cell lattice was set to 2.0 Å, and the overall size of the box is automatically fitted to enclose all peptides of the database set. The atom occupancy of each grid cell, GCOD, is a descriptor in the 4D-QSAR analysis (see Step 4), and it was computed considering the seven IPE atom types. The normalized absolute occupancy of each grid cell, classified through its IPE for a given alignment, provides the trial pool of RD-4D-QSAR independent variables referred to as GCOD (5).

Step 8

The 4D-QSAR method may retain a very large number of spurious variables, which can be deleterious for the models (4,9,35,36). Moreover, the possibility of highly correlated GCOD owing to factors such as ligand–receptor interactions, induced-fit ligand and/or receptor conformational changes, and ligand modulated receptor allosteric effects can lead to a large number of QSAR models with similar measures of significance, which further complicates both models validation and interpretation (5).

Partial least-squares (PLS) regression analysis (37) was employed as a data reduction tool between the observed dependent variable

Table 2: Statistical values for the different alignments tested in the RD-4D-QSAR study of the 21 peptide mimetic inhibitors of TR

Alignments ^a	C _α atoms	Variables	r _{adj} ²	q _{adj} ²	Lack-of-fit	Least-squares error
1	W ²² /F ^{396'} /H ^{461'}	3	0.707	0.515	0.295	0.150
		4	0.720	0.458	0.374	0.144
		5	0.780	0.406	0.410	0.113
2	F ^{396'} /P ^{398'} /L ^{399'}	3	0.682	0.527	0.320	0.163
		4	0.748	0.568	0.336	0.129
		5	0.792	0.621	0.257	0.106
3	H^{461'}/E^{467'}/E^{466'}	3	0.763	0.657	0.371	0.121
		4	0.882	0.778	0.157	0.060
		5	0.904	-0.029	0.046	0.802
4	W ²² /Y ¹¹¹ /M ¹¹⁴	3	0.730	0.190	0.138	0.643
		4	0.734	0.539	0.355	0.136
		5	0.802	0.210	0.101	0.560
5	W ²² /E ¹⁹ /L ¹⁸	3	0.727	0.496	0.427	0.135
		4	0.617	0.336	1.082	0.199
		5	0.906	-0.034	0.046	0.834
6	E ¹⁹ /R ^{472'} /E ^{467'}	3	0.811	0.700	0.096	0.188
		4	0.850	0.057	0.076	0.766
		5	0.948	-0.131	0.026	0.710
7	V⁵⁶/T^{410'}/L²⁷	3	0.806	0.702	0.194	0.099
		4	0.820	0.675	0.239	0.092
		5	0.867	0.625	0.246	0.067

^aThe alignments **3** and **7** were selected for building two independent RD-4D-QSAR models. The best results from alignments **3** and **7** are highlighted in boldface.

(pl₅₀, Table 1) and the corresponding GCOD values for each trial alignment (9). Additionally, PLS identify the most highly weighted GCOD from enormous data set of a local grid cell (5). In our study, independent variable columns with variances over the set of peptide complexes <2.0 were eliminated. The automated reduction data by the PLS analysis provide the selection of the descriptors having the highest individual weightings to the observed biological activity measures (9). Then, PLS reduction was applied for each alignment of each peptide-TR complex model (Table 2).

Step 9

The number of descriptors (GCOD) that could be part of a QSAR equation (model) depends on the ligand-receptor system; generally, larger and/or more flexible ligands will generate a high number of GCOD. However, the statistics impose a limit of four to five observation compounds per descriptor (9). If the number of GCOD in the grid cell occupancy profile for a grid cell is <500, all those GCOD are retained for the genetic function approximation (GFA) analysis. Otherwise, the top 500 GCOD from the PLS (see Step 8) are used. In the current work, the 500 most highly ranked PLS descriptors, determined in the previous step for a grid cell of 2.0 Å, were chosen to form the trial descriptor basis sets for model optimization by GFA analysis (38).

The GFA analysis searches which descriptor (GCOD) combination scores well, rather than identifying good individual high-scoring descriptors. Thus, it is possible that one (or more) descriptor in a good RD-4D-QSAR model does not individually contribute significantly to the overall quality of a model, and these low-scoring descriptors could cause confusion when trying to interpret a model (5). To compare models with different numbers of selected variables, the

cross-validated correlation coefficient (q^2) and the linear correlation coefficient (r^2) were normalized (39), resulting in the adjusted q^2 (q_{adj}^2) and adjusted r^2 (r_{adj}^2) values, respectively.

The GFA optimizations were initiated using 500 randomly generated 4D-QSAR models. Twenty thousand to 5000 crossover operations were applied, and the number of mutation operations over the crossover cycle was set at 50–70%. The smoothing factor is used to alter the number of independent variables (GCOD) into the models, to reduce the least-squares error (LSE) and impair overfitting (9). Smoothing factor values of 0.5–2.0 were tested to fit to the optimal number of descriptors in the 4D-QSAR models, resulting in lower values of Friedman's lack-of-fit (LOF) factor, which is a penalized LSE measure (39,40)

Aiming to build a RD-4D-QSAR model with high predictive power, the top 10 best models (i.e., equations that have highest q_{adj}^2 , without data overfitting), obtained from GFA optimization, are determined and investigated. The number of top models investigated can be varied, and the numbers of times each unique GCOD is used on the top 10 best models could be recorded. This selection seems to capture the GCOD with higher contributions to activity profile. Any GCOD used more than one for each grid cells among the top 10 best models is retained for the next model building GFA analysis.

Diagnostics measures to analyze the resultant QSAR models are determined as a part of GFA optimization. The diagnostic measurement includes descriptor usage as a function of crossover operation, linear cross correlation among descriptors and/or biological activity measures, and measures of model significance, including the q^2 , r^2 , q_{adj}^2 , r_{adj}^2 , and Friedman's LOF value.

Step 10

Steps 7–9 were repeated until all seven alignments were included in the 4D-QSAR analysis.

Step 11

The final step in RD-4D-QSAR analysis is to hypothesize an 'active conformation' of the composite ligand–receptor complex (5). The 21 single complexes were evaluated. They were extracted from 100 conformations obtained from MD and presented the best energy criterion, i.e., lower potential energy of the conformation sampled.

Results

The seven trial alignments studied are shown in Table 2. Only alignments **3** and **7** provides the best RD-4D-QSAR models, defined by the highest q_{adj}^2 value. Moreover, these two alignments provide 4D-QSAR models with smallest values of LSE and Friedman's LOF. The complete statistical measures, including the values of q_{adj}^2 , r_{adj}^2 , LSE, LOF, and the number of variables of the top 10 best models for each alignment, are presented in Table 2. The remarkable differences between these two alignments compared to the others cannot only be analyzed based on statistical parameters, for example higher q_{adj}^2 values, but it should take into account the 'three-ordered atoms' used in alignments **3** and **7**. They were clearly diverse from the others, including the catalytic residues (alignment **3**) or including only residues close to the border of the pruned model (alignment **7**). Thus, only these two alignments will be analyzed.

The best 10 top 4D-QSAR models have many variables, and some of them are not to be significant for analysis; hence, only models with 3, 4, or 5 descriptors were considered. To determine whether these variables, incorporated into the models, are providing common or distinct structure–activity information, the correlation coefficient matrix among the residuals of models with 3, 4, and 5 variables for both alignments was computed (Table S1). The idea of determining the residual pair correlations is that equivalent models will have near-identical residues, while distinct models should have noncorrelated residuals (4,34,38).

Table S1 reveals that all the models are highly correlated with each other. Thus, the RD-4D-QSAR model with the highest q_{adj}^2 value was selected as the representative model for each alignment.

The best 4D-QSAR model for alignment **3** (model 1) is defined by eqn 1:

$$\begin{aligned} \text{pl}_{50} = & +3.362 + 0.962(-10, -4, 0, \text{any}) \\ & +1.080(-5, -1, 6, \text{any}) - 1.797(-7, 0, 4, \text{any}) \\ & -1.414(-2, 8, 10, \text{any}) \end{aligned} \quad (1)$$

$$[n = 21, r^2 = 0.8811, q^2 = 0.8112, q_{\text{adj}}^2 = 0.791]$$

The best 4D-QSAR model for alignment **7** (model 2) is defined by eqn 2:

$$\begin{aligned} \text{pl}_{50} = & +3.366 + 0.977(5, 3, 2, \text{any}) - 1.239(1, -2, -3, \text{any}) \\ & -0.211(-1, 3, 6, \text{any}) - 8.672(-3, 13, 3, \text{any}) \end{aligned} \quad (2)$$

$$[n = 21, r^2 = 0.8189, q^2 = 0.7239, q_{\text{adj}}^2 = 0.786]$$

In eqns 1 (model 1) and 2 (model 2), the three sequential numbers in the parenthesis correspond to the Cartesian coordinates of each descriptor (GCOD), and *any* code corresponds to the IPE of the corresponding GCOD (see Materials and Methods).

It is noteworthy that the two best models contain four variables, and they are composed, in both cases, only by one type of atom type (IPE): *any* atoms. Moreover, model 1 shows two cells (–7, 0, 4 and –2, 8, 10) with negative coefficients, which decrease the activity, and two cells (–10, –4, 0 and –5, –1, 6) with positive coefficients, which increase the activity, while model 2 shows simply one cell (2,3,5) with positive coefficient.

The 'active conformation' of each peptide–TR complex was hypothesized considering the models 1 and 2 from alignments **3** and **7**, respectively. The GCOD of each resulting set of low-energy conformations was employed to predict the activities for each ligand using eqns 1 and 2, and the conformer with the highest activity was selected as the 'active conformation' of each complex. Figure 2 shows the most (**18**) and least (**10**) active compounds of this series for each model.

Outlier compounds were defined as the compounds in which predicted activities calculated by the model were higher than twice the standard-error deviations among the residuals, that is, the residual differences between experimental (pl_{50}) and predicted activities ($\text{pl}_{50\text{calc}}$). Table 3 shows the residual values considering the RD-4D-QSAR models 1 and 2 from alignments **3** and **7**, respectively. In the peptide–TR complex model 1 (alignment **3**), compounds **3** and **11** were considered *outliers*, and in the peptide–TR complex model 2 (alignment **7**), the *outliers* were compounds **3** and **4**.

Golbraikh and Tropsha (41) stated that higher q^2 (or q_{adj}^2) values are not a sufficient condition for model validation. Therefore, to have a qualitative validation and access more details about the model utility and limitations (42), the correlation between the GCODs of models **1** and **2** was analyzed by the calculation of the percentage of occupancy of the grid cells (Table 4) of the most (**18**) and the least (**10**) active compounds and the *outliers* of models **1** (**3** and **11**) and **2** (**3** and **4**), which was compared with some crystallographic structures deposited in PDB (11,43).

Discussion**Model 1**

The GCODs (–5, –1, 6) and (–7, 0, 4) of model 1 (eqn 1, Figure 2A,B) revealed an opposite effect on the predicted activity despite their geometric proximity (~3.0 Å), where the occupancy of the grid cell (–5, –1, 6) increases the activity, while the occupancy

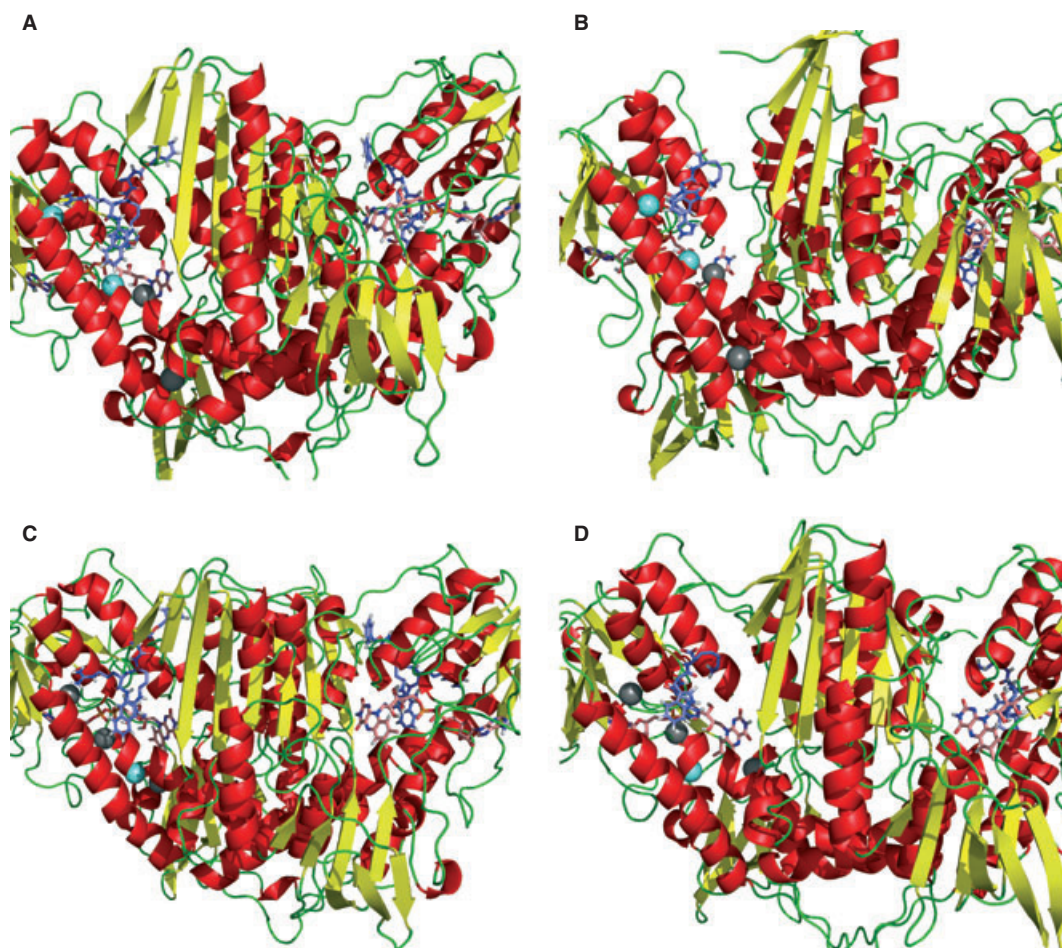


Figure 2: RD-4D-QSAR with most active peptide (**18**, A and C) and less active (**10**, B and D) for model 1 (A and B) and model 3 (C and D). The representations are the same as Figure 1; the gray and cyan spheres represent the grids that decrease and increase the biological activities, respectively.

of the grid cell $(-7, 0, 4)$ decreases the activity, because the former grid cell has a positive coefficient and the latter a negative one. Analyzing the crystallographic structure of TR complexed with their natural substrate (**11**), it was noted that the spatial position of these grid cells corresponds to the catalytic cysteines C^{53} and C^{58} , respectively. To understand the reasons of this behavior, we attended for the role of each cysteine at the mechanism of TR catalysis proposed by Leichus *et al.* (44). The side chain of C^{53} residue acts as a nucleophile attacking the substrate during the catalysis, and the thiol group of C^{58} makes a *charge transfer complex* with the FAD aromatic ring (44). It is remarkable that our model could distinguish through these nearby residues and assigned the contributions for each one descriptor similar to their biological role in catalysis, that is, the substrate interaction with the enzyme residues increases (positive GCOD) the activity, while its interaction with the cofactor (FAD) decreases (negative GCOD) the activity.

The positions of other two remaining GCODs $(-10, -4, 0)$ and $(-2, 8, 10)$ of model 1 were not helpful in our study, because the first was not occupied in model 1 (Table 4) and the second is located ~ 2.45 Å away from the backbone oxygen atom of S^{15} , which interacts through H-bond with glycine-I of TS_2 (**11**).

Interestingly, during the MD simulation, the complex of the most activity compound (**18**) occupies only one positive GCOD $(-5, -1, 6)$ and does not occupy any negative GCOD, while the complex of compound **10** (the least-potent inhibitor) occupies the positive GCOD, as complex **18**, but both negative GCODs $(-7, 0, 4)$ and $(-2, 8, 10)$ (Table 4). Table 3 shows that the activities of these compounds were well predicted, resulting in residual values of -0.034 and 0.211 , respectively, for **18** and **10**.

Model 1 has two *outliers* (**3** and **11**). During the MD simulation, the complex of *outlier 3* does not occupy any GCOD (Table 4), and because of this fact, the predicted activity of this compound is higher than the experimental one (residual = 0.544 , Table 3). It could be explained from eqn 1, where the high constant term (3.362) will result in an 'overpredicted' activity by no occupancy of any GCOD.

The complex of *outlier 11* occupies the same GCODs as complex **10**, but while complex **10** occupies more negative GCODs, complex **11** occupies more the positive GCOD, and then the predicted activity should be high (residual = 0.565 , Table 3). However, it should be noted that the high constant term (3.362) could not justify alone

Table 3: Residual differences between experimental (p_{150}) and predicted ($p_{150\text{calc}}$) activities of the RD-4D-QSAR models **1** and **2**

No. ^a	p_{150}	Model 1		Model 2	
		$p_{150\text{calc}}$	Residual	$p_{150\text{calc}}$	Residual
1	3.26	3.409	0.147	3.210	-0.052
2	3.40	3.037	-0.368	3.595	0.190
3	2.82	3.362	0.544	3.408	0.590
4	2.57	2.518	-0.056	3.128	0.554
5	3.09	3.283	0.195	3.010	-0.079
6	2.57	2.540	-0.034	2.551	-0.023
7	4.41	4.269	-0.146	4.457	0.043
8	4.22	4.313	0.090	4.093	-0.130
9	2.76	2.694	-0.065	2.734	-0.025
10	2.41	2.618	0.211	2.280	-0.128
11	3.40	3.967	0.565	3.109	-0.294
12	3.26	3.362	0.098	3.339	0.074
13	4.93	4.636	-0.292	4.957	0.029
14	3.45	3.326	-0.123	3.285	-0.164
15	3.38	3.372	-0.007	3.454	0.075
16	3.29	3.370	0.077	3.312	0.019
17	3.53	3.254	-0.275	3.267	-0.262
18	4.98	4.949	-0.034	4.835	-0.148
19	3.62	3.362	-0.254	3.324	-0.292
20	3.61	3.362	-0.245	3.149	-0.458
21	2.56	2.533	-0.029	3.041	0.479

^aThe outliers from models **1** (**3** and **11**) and **2** (**3** and **4**) are underlined.

the calculated activity, because the 'positive' occupancy in complex **11** resulted in activity values higher than the experimental results from McKie *et al.* (12). Taken these results together, it could be concluded that the high constant term summed with the fourfold occupancy in the positive grid cells culminate in 'overpredicted' activity for peptide **11**.

Model 2

The unique positive GCOD (2,3,5) of model 2 (eqn 2, Figure 2C,D) is located near the residues N⁵⁵ and V⁵⁴ (~1.74 Å away from V⁵⁴-C α). This region is a hydrophobic site near the TR active site and contiguous to Z-site (11,16), a promising location that interacts with diverse classes of known TR inhibitors (12,13,16,18). Moreover, Bond *et al.* (11) reported that residue V⁵⁴ has two main hydrophobic interactions with TS₂: one interaction through its backbone atoms with cysteine and the other with the glutamate-I side chain. The behavior of this descriptor increasing the predicted activity corroborates with these previous crystallographic data.

Table 4: Percentage of occupancy of grid cell occupancy descriptors (GCODs) selected in the RD-4D-QSAR models **1** and **2**, respectively, during the MD simulation of the 21 peptide-TR complexes. The results are shown for the most (**18**) and the least (**10**) active compounds and outliers of models 1 (**3** and **11**) and 2 (**3** and **4**)

GCOD	Model 1				GCOD	Model 2			
	3	10	11	18		3	4	10	18
+(-10, -4, 0)	0	0	0	0	+(5, 3, 2)	0.330	0	0	1.560
+(-5, -1, 6)	0	0.290	0.840	1.470	-(-1, 3, 6)	1.330	1.130	0.092	0.260
-(-7, 0, 4)	0	0.100	0.160	0	-(-3, 13, 3)	0	0.470	0	0
-(-2, 8, 10)	0	0.620	0.100	0	-(1, -2, -3)	0	1.240	0.720	0

The negative GCOD (1, -2, -3) (eqn 2, Figure 2C,D) is ~3.8 Å away from the gamma-carbon atom of P⁶⁰ side chain. This result reinforces the accuracy of this model when penalizes the grid cell occupancy of GCODs that does not interact close to the TR active site, because the P⁶⁰ is far from the main interactions observed in crystal structures of TR-inhibitor complexes (11,43).

The occupancy of the other negative GCOD (-3, 13, 3) decreases the predicted activity, and the main reason was attributable to the spatial location of this grid cell ~4.3 Å away from the oxygen atom of G¹² backbone. Both at the crystal structure (11,43) and in our peptide-TR complex model, there is not any amino acid interacting with this residue. The inspection of the TR-NADP complex crystal structure [PDB ID: 1TYT (43)] shows that G¹² is near the FAD adenine ring. Therefore, the occupancy of this site decreases the predicted activity, because flavin is the TR cofactor, which donates a hydride during the enzymatic catalysis (44).

The last negative GCOD (-1, 3, 6) also decreases the calculated activity according to model **2**, and it is located between the nitrogen backbone of A⁴⁸ and oxygen backbone of G⁵¹, with distances of ~1.37 and ~2.54 Å, respectively. We noted that both residues interacts with the polyphosphate chain of FAD, a spatial obstructed region of the enzyme in the crystal complex (43), which corroborates with the negative signal of this descriptor decreasing the activity.

The complex of compound **18**, as revealed in Table 4, occupies the unique positive GCOD (2,3,5) almost five times more than the negative one (-1, 3, 6), during the MD simulation, resulting in activity values near to the experimental ones. The complex of peptide **10** occupies only two negative GCODs (-1, 3, 6 and 1, -2, -3), resulting in decreased activity. In model 2, as in model 1, compounds **18** and **10** were well predicted, since their residual values were -0.148 and -0.128, respectively (Table 3).

Model 2 has two outliers (**3** and **4**). Again, the outlier from complex **3** is also an outlier in model 2. Although distinct from model 1, where its behavior was attributable to an 'empty occupancy' of all GCODs, in model 2, the complex **3** occupies a negative GCOD (-1, 3, 6) four times more than it occupies the positive one (2,3,5) (Table 4), resulting in a decreased activity (residual = 0.590, Table 3).

The outlier from complex **4** occupies only the negative grid cells (-1, 3, 6; -3, 13, 3; and 1, -2, -3) of model 2, during the MD

simulation, which decreases the calculated activity, where the high constant term (3.366) 'overpredicted' its activity (residual = 0.554, Table 3).

Conclusions

In the present work, we have constructed two 3D-QSAR models, applying the receptor-dependent 4D-QSAR (RD-4D-QSAR) approach, for a series of peptide mimetic inhibitors of *T. cruzi* trypanothione reductase (TR, PDB ID: 1BZL) synthesized and tested by McKie *et al.* (12). The models were derived from 21 peptide-TR complexes, constructed using a TR *Receptor Model*, a sphere of 11.5 Å radii around all atoms of the substrate containing some residues of the enzyme.

The best equations generated by the RD-4D-QSAR method resulted in good models, that is, statistically meaningful: $q^2 > 0.72$, $q_{adj}^2 > 0.78$ and $r^2 > 0.81$ (Table 2). Besides its statistical significance, both models revealed some regions at the TR receptor of particular interest for access in future steps of the drug discovery process against tropical diseases, for example the disulfide bond region between the TR catalytic cysteines (C⁵³ and C⁵⁸) closest to the C⁵³ that may act preventing the nucleophile attack to the substrate; the *Z-site* and its adjacent regions showing hydrophobic characteristic, which has been reported before (16,18,43,44); and the FAD region whose occupancy will decrease the activity of the compounds. It is noteworthy that the *Receptor Model* does *not* include *any* atoms of the FAD ring, and despite it, in the two best models (equations), at least one descriptor (GCOD) related to this region was selected.

These results show that new insights can be obtained with the RD-4D-QSAR approach, as it includes atoms from the receptor and not only atoms of the ligand, as it was more usual, that is, the receptor-independent (RI) approach (5,8,9,36,42). Possibly, including the complete TR enzyme will result in more useful models that could reveal other affordable regions to explore more diversely than that reported here.

Acknowledgments

We gratefully acknowledge the funding support by the following Brazilian governmental agencies: CAPES, CNPq, and FAPERJ. SSRP is also grateful to CAPES for scholarship support.

References

- Hansch C. (1969) A quantitative approach to biochemical structure-activity relationships. *Acc Chem Res*;2:232–239.
- Hansch C., Fujita T. (1964) Rho-Sigma-Pi analysis. Method for correlation of biological activity + chemical structure. *J Am Chem Soc*;86:1616–1626.
- Cramer R.D., Patterson D.E., Bunce J.D. (1988) Comparative molecular-field analysis (CoMFA) .1. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc*;110:5959–5967.
- Hopfinger A.J., Wang S., Tokarski J.S., Jin B.Q., Albuquerque M., Madhav P.J., Duraiswami C. (1997) Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *J Am Chem Soc*;119:10509–10524.
- Pan D., Tseng Y., Hopfinger A.J. (2003) Quantitative structure-based design: formalism and application of receptor-dependent RD-4D-QSAR analysis to a set of glucose analogue inhibitors of glycogen phosphorylase. *J Chem Inf Comput Sci*;43:1591–1607.
- Pan D., Liu J., Senese C., Hopfinger A.J., Tseng Y. (2004) Characterization of a ligand-receptor binding event using receptor-dependent four-dimensional quantitative structure-activity relationship analysis. *J Med Chem*;47:3075–3088.
- Santos O.A., Hopfinger A.J. (2006) Structure-based QSAR analysis of a set of 4-hydroxy-5,6-dihydropyrones as inhibitors of HIV-1 protease: an application of the receptor-dependent (RD) 4D-QSAR formalism. *J Chem Inf Model*;46:345–354.
- Hong X., Hopfinger A.J. (2003) 3D-pharmacophores of flavonoid binding at the benzodiazepine GABA(A) receptor site using 4D-QSAR analysis. *J Chem Inf Comput Sci*;43:324–336.
- Pasqualoto K.F., Ferreira E.I., Santos-Filho O.A., Hopfinger A.J. (2004) Rational design of new antituberculosis agents: receptor-independent four-dimensional quantitative structure-activity relationship analysis of a set of isoniazid derivatives. *J Med Chem*;47:3755–3764.
- Urbina J.A., Docampo R. (2003) Specific chemotherapy of Chagas disease: controversies and advances. *Trends Parasitol*;19:495–501.
- Bond C.S., Zhang Y., Berriman M., Cunningham M.L., Fairlamb A.H., Hunter W.N. (1999) Crystal structure of *Trypanosoma cruzi* trypanothione reductase in complex with trypanothione, and the structure-based discovery of new natural product inhibitors. *Structure*;7:81–89.
- McKie J.H., Garforth J., Jaouhari R., Chan C., Yin H., Besheya T., Fairlamb A.H., Douglas K.T. (2001) Specific peptide inhibitors of trypanothione reductase with backbone structures unrelated to that of substrate: potential rational drug design lead frameworks. *Amino Acids*;20:145–153.
- Schirmer R.H., Müller J.G., Krauth-Siegel R.L. (1995) Disulfide-reductase inhibitors as chemotherapeutic agents: the design of drugs for trypanosomiasis and malaria. *Angew Chem Int Ed Engl*;34:141–154.
- Schmidt A., Krauth-Siegel R.L. (2002) Enzymes of the trypanothione metabolism as targets for antitrypanosomal drug development. *Curr Top Med Chem*;2:1239–1259.
- Benson T.J., McKie J.H., Garforth J., Borges A., Fairlamb A.H., Douglas K.T. (1992) Rationally designed selective inhibitors of trypanothione reductase. Phenothiazines and related tricyclics as lead structures. *Biochem J*;286(Pt 1):9–11.
- Chan C., Yin H., Garforth J., McKie J.H., Jaouhari R., Speers P., Douglas K.T., Rock P.J., Yardley V., Croft S.L., Fairlamb A.H. (1998) Phenothiazine inhibitors of trypanothione reductase as potential antitrypanosomal and antileishmanial drugs. *J Med Chem*;41:148–156.
- Chan C., Yin H., McKie J.H., Fairlamb A.H., Douglas K.T. (2002) Peptoid inhibition of trypanothione reductase as a potential antitrypanosomal and antileishmanial drug lead. *Amino Acids*;22:297–308.

18. Faerman C.H., Savvides S.N., Strickland C., Breidenbach M.A., Ponasik J.A., Ganem B., Ripoll D., Krauth-Siegel R.L., Karplus P.A. (1996) Charge is the major discriminating factor for glutathione reductase versus trypanothione reductase inhibitors. *Bioorg Med Chem*;4:1247–1253.
19. Garforth J., Mckie J.H., Jaouhari R., Benson T.J., Fairlamb A.H., Douglas K.T. (1994) Rational design of peptide-based inhibitors of trypanothione reductase as potential antitrypanosomal drugs. *Amino Acids*;6:295–299.
20. Garforth J., Yin H., McKie J.H., Douglas K.T., Fairlamb A.H. (1997) Rational design of selective ligands for trypanothione reductase from *Trypanosoma cruzi*. Structural effects on the inhibition by dibenzazepines based on imipramine. *J Enzyme Inhib*;12:161–173.
21. Khan M.O., Austin S.E., Chan C., Yin H., Marks D., Vaghjiani S.N., Kendrick H., Yardley V., Croft S.L., Douglas K.T. (2000) Use of an additional hydrophobic binding site, the Z site, in the rational drug design of a new class of stronger trypanothione reductase inhibitor, quaternary alkylammonium phenothiazines. *J Med Chem*;43:3148–3156.
22. Linares G.E., Ravaschino E.L., Rodriguez J.B. (2006) Progresses in the field of drug design to combat tropical protozoan parasitic diseases. *Curr Med Chem*;13:335–360.
23. Marsh I.R., Bradley M. (1997) Substrate specificity of trypanothione reductase. *Eur J Biochem*;243:690–694.
24. Moutiez M., Aumercier M., Parmentier B., Tartar A., Sergheraert C. (1995) Compared recognition of di- and trisulfide substrates by glutathione and trypanothione reductases. *Biochim Biophys Acta*;1245:161–166.
25. Bailey S., Smith K., Fairlamb A.H., Hunter W.N. (1993) Substrate interactions between trypanothione reductase and N1-glutathionylspermidine disulphide at 0.28-nm resolution. *Eur J Biochem*;213:67–75.
26. Hopfinger A.J. (2001) 4D-QSAR Package User's Manual, v.3.0. The Chem21 Group Inc, Lake Forest, IL, USA.
27. Hypercube (2001) HyperChem Professional v.7.0. Hypercube, Inc, 1115 NW, 4th street, Gainesville, FL, USA.
28. Cornell W.D., Cieplak P., Bayly C.I., Gould I.R., Merz K.M., Ferguson D.M., Spellmeyer D.C., Fox T., Caldwell J.W., Kollman P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc*;117:5179–5197.
29. Dewar M.J.S., Zoebisch E.G., Healy E.F., Stewart J.J.P. (1985) The development and use of quantum-mechanical molecular-models .76. AM1 – a new general-purpose quantum-mechanical molecular-model. *J Am Chem Soc*;107:3902–3909.
30. Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. (2000) The protein data bank. *Nucleic Acids Res*;28:235–242.
31. Molecular Simulation (2000) WebLab ViewerLite. v.4.0: Molecular Simulation, Inc., San Diego, California, USA.
32. Hunter W.N., Bailey S., Habash J., Harrop S.J., Helliwell J.R., Aboagye-Kwarteng T. *et al.* (1992) Active site of trypanothione reductase. A target for rational drug design. *J Mol Biol*;227:322–333.
33. da Rocha Pita S.S., Cirino J.J., de Alencastro R.B., Castro H.C., Rodrigues C.R., Albuquerque M.G. (2009) Molecular docking of a series of peptidomimetics in the trypanothione binding site of *T. cruzi* trypanothione reductase. *J Mol Graph Model*;28:330–335.
34. Doherty D.C. (1994) Molsim User's Guide. Molsim v.3.0. The Chem21 Group, Inc., 1780 Wilson Drive, Lake Forest, IL 60045, USA.
35. Ooms F. (2000) Molecular modeling and computer aided drug design. Examples of their applications in medicinal chemistry. *Curr Med Chem*;7:141–158.
36. Liu J., Pan D., Tseng Y., Hopfinger A.J. (2003) 4D-QSAR analysis of a series of antifungal p450 inhibitors and 3D-pharmacophore comparisons as a function of alignment. *J Chem Inf Comput Sci*;43:2170–2179.
37. Dunn W.J. III, Scott D.R., Glen W.G. (1989) Principal components analysis and partial least squares regression. *Tetrahedron Comput Methodol*;2:349–376.
38. Rogers D., Hopfinger A.J. (1994) Application of genetic function approximation to quantitative structure-activity-relationships and quantitative structure-property relationships. *J Chem Inf Comput Sci*;34:854–866.
39. Friedman J. (1991) Multivariate Adaptive Regression Splines. *Ann Statist*;19:1–67.
40. Rogers D. (1991) G/SPLINES: a hybrid of Friedman's multivariate adaptive regression splines (MARS) algorithm with Holland's genetic algorithm. Eds. Belew R.K., Booker L.B., Proceedings of the Fourth International Conference on Genetic Algorithms. San Francisco: Morgan Kaufman Publishers; 91.10:384–391.
41. Golbraikh A., Tropsha A. (2002) Beware of q²! *J Mol Graph Model*;20:269–276.
42. Romeiro N.C., Albuquerque M.G., Alencastro R.B., Ravi M., Hopfinger A.J. (2005) Construction of 4D-QSAR models for use in the design of novel p38-MAPK inhibitors. *J Comput Aided Mol Des*;19:385–400.
43. Bailey S., Fairlamb A.H., Hunter W.N. (1994) Structure of trypanothione reductase from *Crithidia fasciculata* at 2.6 Å resolution; enzyme-NADP interactions at 2.8 Å resolution. *Acta Crystallogr D Biol Crystallogr*;50:139–154.
44. Leichus B.N., Bradley M., Nadeau K., Walsh C.T., Blanchard J.S. (1992) Kinetic isotope effect analysis of the reaction catalyzed by *Trypanosoma congolense* trypanothione reductase. *Biochemistry*;31:6414–6420.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Correlation coefficient matrix between the residual values of the RD-4D-QSAR models from alignments **3** and **7** with 3, 4 or 5 variables included at each equation.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.