Journal of Enzyme Inhibition and Medicinal Chemistry

ISSN: 1475-6366 (Print) 1475-6374 (Online) Journal homepage: informahealthcare.com/journals/ienz20

# QSAR and docking studies of anthraquinone derivatives by similarity cluster prediction

Alexandra M. Harsa, Teodora E. Harsa & Mircea V. Diudea

**To cite this article:** Alexandra M. Harsa, Teodora E. Harsa & Mircea V. Diudea (2016) QSAR and docking studies of anthraquinone derivatives by similarity cluster prediction, Journal of Enzyme Inhibition and Medicinal Chemistry, 31:3, 508-515, DOI: 10.3109/14756366.2015.1046061

**To link to this article:** https://doi.org/10.3109/14756366.2015.1046061

Journal of Enzyme Inhibition and Medicinal Chemistry

www.tandfonline.com/ienz
ISSN: 1475-6366 (print), 1475-6374 (electronic)

J Enzyme Inhib Med Chem, 2016; 31(3): 508–515
 DOI: 10.3109/14756366.2015.1046061

informa healthcare

RESEARCH ARTICLE

# QSAR and docking studies of anthraquinone derivatives by similarity cluster prediction

Alexandra M. Harsa, Teodora E. Harsa, and Mircea V. Diudea

*Faculty of Chemistry and Chemical Engineering, Babeş-Bolyai University, Cluj, Romania*

**Abstract**

Forty anthraquinone derivatives have been downloaded from PubChem database and investigated in a quantitative structure-activity relationships (QSAR) study. The models describing log *P* and LD50 of this set were built up on the hypermolecule scheme that mimics the investigated receptor space; the models were validated by the leave-one-out procedure, in the external test set and in a new version of prediction by using similarity clusters. Molecular docking approach using Lamarckian Genetic Algorithm was made on this class of anthraquinones with respect to 3Q3B receptor. The best scored molecules in the docking assay were used as leaders in the similarity clustering procedure. It is demonstrated that the LD50 data of this set of anthraquinones are related to the binding energies of anthraquinone ligands to the 3Q3B receptor.

**Keywords**

3Q3B Receptor, anthraquinone, AUTODOCK Vina, docking, hypermolecule, leave-one-out, log *P*, QSAR, similarity

**History**

Received 13 March 2015
Revised 20 April 2015
Accepted 24 April 2015
Published online 28 May 2015

## Introduction

Anthraquinones are aromatic compounds usually present as one specific isomer, 9,10-anthraquinone (IUPAC: 9,10-dioxoanthracene). Anthraquinones are found in various organisms, including bacteria, fungi, plants, as well as in some marine animals and terrestrial insects[1–3]. In higher plants, anthraquinones serve as secondary metabolites and display numerous biological activities[4].

The notion of similarity is strongly dependent on the current use to which similarity is addressed. Molecules can be described in various ways: by molecular graphs, by atoms position, by molecular fields, etc. Quantitative similarity measures can be developed for each of the above descriptions[5].

Quantitative Structure-Activity Relationship (QSAR) is a powerful method for the design of bioactive compounds and prediction of their activity or physical-chemical properties. The aim of this work was to determine predictive QSAR models[6] for log *P* and LD50 of 40 anthraquinone derivatives downloaded from PubChem Database.

The octanol–water partition coefficient (log *P*) is related to the hydrophobicity of molecules and their transport to biological receptors[7]. LD50 refers to the toxicity of molecules, being the concentration needed to kill 50% of the tested animals[8].

## Structural molecular data

A set of 40 anthraquinones were taken from PubChem Database[9] (Table 1); the set was divided into a training set (25 molecules) and a test set (15 molecules), taken randomly. The property chosen for modeling was log *P* (calculated, Table 1) and LD50 (on rat, oral route administrated, Table 2).

A hypermolecule (Figure 1) that mimics the investigated receptor space was bult up from the common features of the molecules in the dataset. Superposition of actual molecular structures over the hypermolecule was performed by HyperChem 8.0 program (http://www.hyper.com/) in order to minimize the sum of square distances between equivalent atoms[10,11]. The result of this superposition/mapping was a binary vector that collects the mapping information. Later, values 1 will be changed with the corresponding mass fragments and partial charges, respectively (Section ''Results and discussion''). The protein glycogen synthase kinase-3 beta receptor (Figure 2) was downloaded from RCSB protein data bank and bears the PDB code-3Q3B[12].

## Docking setup

Anthraquinone derivatives (optimized at Hartree-Fock HF (3-21 g(p)) level of theory) were docked to the target 3Q3B receptor with the protein molecule considered as a rigid body and the ligands being flexible. The Lamarckian genetic algorithm was used to search for the best conformers; it searches for an empirical binding free energy that allows the prediction of binding affinity for docked ligands[13]. Grid menu was toggled, after loading protein.pdbqt and the map files were selected directly with setting up the grid points with $40 \times 40 \times 40$ Å3 dimensions, at 0.375 Å cell, centered on (x,y,z) 24.569, −0.448, 21.386; (3Q3B), with 41 non-bonded atoms. The investigated anthraquinone derivatives were loaded and their torsions along the rotatable bonds (Table 2) were assigned, next their files were saved as ligand.pdbqt[14].

## Docking results

The ligands docked at Glycogen synthase kinase-3 beta (3Q3B) protein have shown the best fit (Root Mean Square Difference

Address for correspondence: Mircea V. Diudea, Faculty of Chemistry and Chemical Engineering, Babeş-Bolyai University, 400028 Cluj, Romania. E-mail: diudea@gmail.com

Table 1. Anthraquinone molecular structures and their log *P* (taken from PubChem).

| Mol. | Canonical SMILES | CID | log *P* |
|---|---|---|---|
| 1 | C1 = CC = C2C(=C1)C(=O)C3 = CC = CC = C3C2 = O | 6780 | 3.4 |
| 2 | C1 = CC = C2C(=C1)C(=O)C3 = C(C2 = O)C = C(C = C3)O | 11796 | 3 |
| 3 | CC1 = CC(=C2C(= C1)C(= O)C3 = C(C2 = O)C(=CC = C3)O)O | 10208 | 3.5 |
| 4 | C1C2 = C(C(= CC = C2)O)C(= O)C3 = C1C = CC = C3O | 2202 | 3.2 |
| 5 | C1 = CC2 = CC3 = C(C(=CC = C3)O)C(= C2C(=C1)O)O | 10187 | 3.9 |
| 6 | C1 = CC2 = C(C(= C1)O)C(= O)C3 = C(C2 = O)C = CC = C3O | 2950 | 3.2 |
| 7 | CC1 = C(C2 = C(C = C1)C(=O)C3 = C(C2 = O)C = CC(= C3O)O)O | 442756 | 3.3 |
| 8 | CC1 = C(C = C2C(=C1)C(= O)C3 = CC = CC = C3C2 = O)O | 10889963 | 2.9 |
| 9 | C1 = C(C = C2C(=C1O)C(= O)C3 = C(C = C(C = C3C2 = O)O)O)O | 3016789 | 2 |
| 10 | CC1 = CC(=C2C(=C1)C(=O)C3 = CC(=C(C(=C3C2 = O)O)O)O)O | 12548 | 2.4 |
| 11 | CC1 = CC(=C2C(=C1)CC3 = CC(=CC(=C3C2 = O)O)O)O | 122635 | 3.2 |
| 12 | CC1 = CC2 = C(C=C1)C(=O)C3 = C(C2 = O)C(=CC = C3)O | 155237 | 3.9 |
| 13 | CC1 = C(C2 = C(C = C1)C(=O)C3 = C(C2 = O)C = CC(=C3)O)O | 124063 | 3.1 |
| 14 | CC1 = C(C2 = C(C = C1)C(=O)C3 = CC(=C(C = C3C2 = O)O)O)O | 25202820 | 2.7 |
| 15 | CC1 = C(C(=C2C(=C1)C(=O)C3 = C(C2 = O)C = CC = C3O)O)O | 12322346 | 3.3 |
| 16 | CC1 = CC2 = C(C = C1)C(=O)C3 = C(C2 = O)C = CC(=C3O)O | 5319503 | 3.1 |
| 17 | CC1 = CC = CC2 = C1C(=O)C3 = C(C2 = O)C(=C(C = C3)O)O | 57536669 | 3.1 |
| 18 | CC1 = C(C(=C2C(=C1)C(=O)C3 = CC = CC = C3C2 = O)O)O | 429241 | 3.1 |
| 19 | CC1 = CC(=C2C(=C1)C(=O)C3 = C(C2 = O)C(=C(C = C3)O)O)O | 12313148 | 2.7 |
| 20 | CC1 = C(C(=C2C(=C1)C(=O)C3 = C(C2 = O)C(=CC = C3)O)O)O | 442759 | 2.7 |
| 21 | C1 = CC2 = C(C(=C1)O)C(=O)C3 = CC(=C(C = C3C2 = O)O)O | 11196140 | 2.8 |
| 22 | C1 = CC2 = C(C(=C1)O)C(=O)C3 = C(C2 = O)C(=C(C = C3)O)O | 436367 | 3.4 |
| 23 | C1 = CC2 = C(C = C1O)C(=O)C3 = C(C2 = O)C(=C(C = C3)O)O | 65739 | 2.4 |
| 24 | C1 = CC = C2C(=C1)C(=O)C3 = C(C2 = O)C(=C(C = C3)O)O | 6293 | 3.2 |
| 25 | C1 = CC2 = C(C = C1O)C(=O)C3 = C(C2 = O)C = CC(=C3O)O | 1320 | 2.4 |
| 26 | CC1 = C2C(=CC(=C1O)O)C(=O)C3 = CC = CC = C3C2 = O | 11391150 | 2.5 |
| 27 | C1 = CC(=C(C2 = C1C(=O)C3 = C(C2 = O)C = CC(=C3O)O)O)O | 69440 | 2.5 |
| 28 | CC1 = C(C = C2C(=C1)C(=O)C3 = C(C2 = O)C(=CC = C3)O)O | 71368906 | 3.1 |
| 29 | C1 = CC(=C(C2 = C1C(=O)C3 = C(C2 = O)C(=C(C = C3)O)O)O)O | 22643725 | 2 |
| 30 | CC1 = C(C = C2C(=C1C)C(=O)C3 = C(C2 = O)C = CC = C3O)O | 57745748 | 3.4 |
| 31 | CC1 = C(C2 = C(C = C1)C(=O)C3 = CC(=C(C(=C3C2 = O)O)O)O)O | 25203424 | 2.4 |
| 32 | CC1 = C(C(=C2C(=C1)C(=O)C3 = CC(=CC(=C3C2 = O)O)O)O)O | 11818503 | 2.4 |
| 33 | CC1 = C2C(=CC(=C1)O)C(=O)C3 = C(C2 = O)C(=CC = C3)O | 3085033 | 3.1 |
| 34 | C1 = CC2 = C(C(=C1)O)C(=O)C3 = C(C2 = O)C = CC(=C3)O | 14886011 | 3.2 |
| 35 | C1 = CC2 = C(C(=C1)O)C(=O)C3 = C(C2 = O)C = C(C = C3)O | 12628831 | 3.2 |
| 36 | C1 = CC = C2C(=C1)C(=O)C3 = CC(=C(C = C3C2 = O)O)O | 11031986 | 2.7 |
| 37 | CC1 = C(C = C2C(=C1)C(=O)C3 = C(C2 = O)C = C(C = C3)O)O | 10060853 | 2.5 |
| 38 | CC1 = CC(=C(C2 = C1C(=O)C3 = C(C = C(C = C3C2 = O)O)O)O)O | 9817337 | 2.9 |
| 39 | C1 = CC(=C(C2 = C1C(=O)C3 = C(C = CC(=C3C2 = O)O)O)O)O | 5004 | 2.5 |
| 40 | C1 = C2C(=CC(=C1O)O)C(=O)C3 = CC(=C(C = C3C2 = O)O)O | 44300874 | 1.4 |

(rmsd) value are calculated relative to the best mode and use only movable heavy atoms)[15]; docking data refer to the best nine ligand conformers[16]. The compound #1, 4, 5 and 6 had shown the lowest affinity (−8 kcal/mol) while molecules #7, 10, 16, 19, 37 and 38 the highest affinity (−8.8 kcal/mol), see Table 3; among the ligands with the highest affinity to 3Q3B protein, will be employed in the similarity clustering procedure (Section ''Similarity cluster validation''). Figure 3 shows the binding energies of the ligand docking[17].

To obtain the pharmacophore for the interaction of anthraquinones with the 3Q3B protein, which could be inferred in their toxicity, the conformers with the highest affinity, as resulted from the docking procedure, have been selected; these are ligands 7, 10, 16, 19, 37 and 38 (binding energy −8.8 kcal/mol). The resulting pharmacophore is shown in Figure 4(a) and (b).

## Computational details

Molecular structures have been optimized at HF (3-21 g(p)) level of theory, in gas phase, by Gaussian 09[18]. Topological indices have been computed by TOPOCLUJ software; some of them (Sum-descriptor $SD_k$, sum of distances (i.e. the Wiener index[19]) Di, sum of genuine distances D3D, HOMO energy, total adjacency Adj and Cluj indices (on detour CfDe and on distance CFDi, respectively)[20] are listed in Tables 4 and 5.

The QSAR models fit abilities were assessed by the leave one out analysis[21] using a dedicated software[22,23].

## Results and discussion

Two cases are discussed in the Hypermolecule description: (1) mass fragments (log *P*) and (2) partial charges (as computed by Gaussian at HF level of theory) (for LD50).

### Mass fragments description (for log *P*)

According to the binary vector of ligand superposition over the hypermolecule, the 1-values were changed with the mass number of each vertex, thus resulted in a more specific description of physico-chemical properties of ligands[24].

#### *Data reduction*

The descriptors with variance <10% (i.e., the variance of non-zero values) and intercorrelation larger than 0.80 (it means two highly correlated descriptors bring quite the same information on the topology of molecule, one of the two being sufficient) were discarded. Correlation weighing was performed on all the positions of hypermolecule: the correlating coefficients of the statistically significant positions of the hypermolecule were used to multiply the local descriptors, thus resulting new weighted vectors $CD_{ij}$. Next, the new correlating descriptors are summed to

Table 2. List of ligands showing their molecular weight and formula, hydrogen bond acceptors, hydrogen bond donors and torsions.

| Ligand | Molecular weight [g/mol] | Molecular formula | H-bond donor | H-bond acceptor | Torsions |
|---|---|---|---|---|---|
| 1 | 208.212 | $C_{14}H_8O_2$ | 0 | 2 | 0 |
| 2 | 224.211 | $C_{14}H_8O_3$ | 1 | 3 | 1 |
| 3 | 254.237 | $C_{15}H_{10}O_4$ | 2 | 4 | 2 |
| 4 | 226.227 | $C_{14}H_{10}O_3$ | 2 | 3 | 2 |
| 5 | 226.227 | $C_{14}H_{10}O_3$ | 3 | 3 | 3 |
| 6 | 240.211 | $C_{14}H_8O_4$ | 2 | 4 | 2 |
| 7 | 270.237 | $C_{15}H_{10}O_5$ | 3 | 5 | 3 |
| 8 | 238.238 | $C_{15}H_{10}O_3$ | 1 | 3 | 1 |
| 9 | 272.210 | $C_{14}H_8O_6$ | 4 | 6 | 4 |
| 10 | 286.236 | $C_{15}H_{10}O_6$ | 4 | 6 | 4 |
| 11 | 256.253 | $C_{15}H_{12}O_4$ | 3 | 4 | 3 |
| 12 | 238.238 | $C_{15}H_{10}O_3$ | 1 | 3 | 1 |
| 13 | 254.237 | $C_{15}H_{10}O_4$ | 2 | 4 | 2 |
| 14 | 270.237 | $C_{15}H_{10}O_5$ | 3 | 5 | 3 |
| 15 | 270.237 | $C_{15}H_{10}O_5$ | 3 | 5 | 3 |
| 16 | 254.237 | $C_{15}H_{10}O_4$ | 2 | 4 | 2 |
| 17 | 254.237 | $C_{15}H_{10}O_4$ | 2 | 4 | 2 |
| 18 | 254.237 | $C_{15}H_{10}O_4$ | 2 | 4 | 2 |
| 19 | 270.237 | $C_{15}H_{10}O_5$ | 3 | 5 | 3 |
| 20 | 270.237 | $C_{15}H_{10}O_5$ | 3 | 5 | 3 |
| 21 | 256.210 | $C_{14}H_8O_5$ | 3 | 5 | 3 |
| 22 | 256.210 | $C_{14}H_8O_5$ | 3 | 5 | 3 |
| 23 | 256.210 | $C_{14}H_8O_5$ | 3 | 5 | 3 |
| 24 | 240.211 | $C_{14}H_8O_4$ | 2 | 4 | 2 |
| 25 | 256.210 | $C_{14}H_8O_5$ | 3 | 5 | 3 |
| 26 | 254.237 | $C_{15}H_{10}O_4$ | 2 | 4 | 2 |
| 27 | 272.210 | $C_{14}H_8O_6$ | 4 | 6 | 4 |
| 28 | 254.237 | $C_{15}H_{10}O_4$ | 2 | 4 | 2 |
| 29 | 272.210 | $C_{14}H_8O_6$ | 4 | 6 | 4 |
| 30 | 268.264 | $C_{16}H_{12}O_4$ | 2 | 4 | 2 |
| 31 | 286.236 | $C_{15}H_{10}O_6$ | 4 | 6 | 4 |
| 32 | 286.236 | $C_{15}H_{10}O_6$ | 4 | 6 | 4 |
| 33 | 254.237 | $C_{15}H_{10}O_4$ | 2 | 4 | 2 |
| 34 | 240.211 | $C_{14}H_8O_4$ | 2 | 4 | 2 |
| 35 | 240.211 | $C_{14}H_8O_4$ | 2 | 4 | 2 |
| 36 | 240.211 | $C_{14}H_8O_4$ | 2 | 4 | 2 |
| 37 | 240.211 | $C_{14}H_8O_4$ | 2 | 4 | 2 |
| 38 | 286.236 | $C_{15}H_{10}O_6$ | 4 | 6 | 4 |
| 39 | 272.210 | $C_{14}H_8O_6$ | 4 | 6 | 4 |
| 40 | 272.210 | $C_{14}H_8O_6$ | 4 | 6 | 4 |

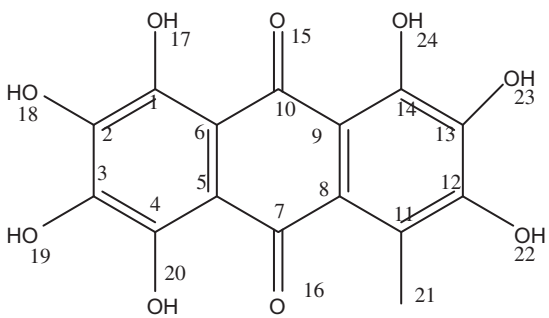


Figure 1. The hypermolecule comprising common features of the dataset.

give a global descriptor, $SD_i = \sum_j CD_{ij}$. This new descriptor is a linear combination of the local correlating descriptors for the significant positions in the hypermolecule (e.g.). It correlates with log $P$ as below:

$$\log P = 1.001 \times \mathrm{SD} + 21.040$$

$$N = 40;\ R^2 = 0.901;\ s = 0.162;\ F = 349.283$$

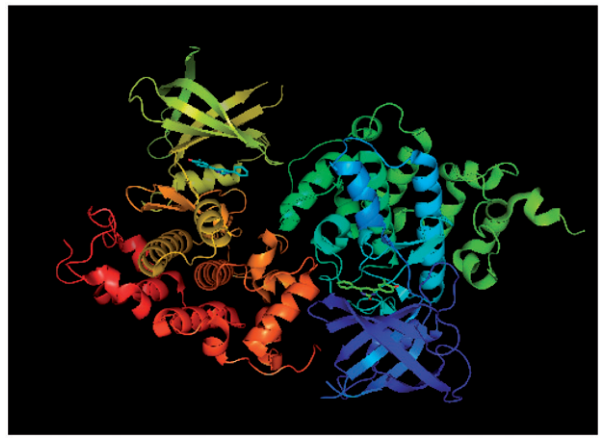

Figure 2. Glycogen synthase kinase-3 beta receptor, PDB Entry ID: 3Q3B, obtained from RCBS Protein data bank.

### *QSAR models*

The models were performed on the training set (the first 25 structures in Table 1) and the best results (in decreasing order of $R^2$) are listed below and in Table 6.

(i) Monovariate regression
$\log P = 22.350 + 1.071 \times \mathrm{SD}$

(ii) Bivariate regression
$\log P = 22.791 + 1.110 \times \mathrm{SD} + 0.001 \times \mathrm{D3D}$

(iii) Three-variate regression
$\log P = 27.550 + 1.147 \times \mathrm{SD} - 0.293 \times \mathrm{Adj} + 0.004 \times \mathrm{Di}$

(iv) Five-variate regression
$\log P = 41.197 + 1.087 \times \mathrm{SD} - 1.087 \times \mathrm{Adj} + 0.004 \times \mathrm{D3D} + 0.1015 \times \mathrm{CfDe}$

### *Model validation*

*Leave-one-out.* The performances in leave-one-out analysis related to the models listed as best in Table 6 are shown in Table 7. The values of $R^2$–$Q^2$ show a good predictability of models.

*External validation.* The values log $P$ for the test set of anthraquinones were calculated by using equation in Table 6, entry 11. Data are listed in Table 8 and the monovariate correlation: $\log P = 0.934 \times \log P_{\mathrm{calc.}} + 0.298$; $n = 15$; $R^2 = 0.754$; $s = 0.201$; $F = 39.749$ is plotted in Figure 5.

*Similarity cluster validation.* Clusters of similarity were performed by using as leaders the 15 molecules in the external set; each leader will have its own cluster, selected by 2D similarity among the 25 structures of the initial learning set. The values log $P_{\mathrm{calc.}}$ were computed by 15 new equations (the leader being left out) with the same descriptors as in Table 6, entry 11. Data are listed in Table 9 and the monovariate correlation: $\log P = 1.039 \times \log P_{\mathrm{calc.}} - 0.042$; $n = 15$; $R^2 = 0.961$; $s = 0.080$; $F = 317.747$ is plotted in Figure 6.

The prediction of log $P$ is much better done by using the clusters of similarity (Table 9) that by the classical external validation of the model (Table 8).

## Partial charges description; LD50

In this section, the weighted vector was completed by weighting the binary vector of ligand superposition over the hypermolecule by partial charges (computed at HF (3−21 g(p)) level of theory) for every molecule.

Table 3. Final lamarckian genetic algorithm docked state – binding energy for nine ligand conformations.

| Ligand | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Docked energy (kcal/mol) |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | −8.0 | −7.9 | −7.9 | −7.9 | −7.9 | −7.9 | −7.9 | −7.4 | −7.4 | −8.0 |
| 2 | −8.1 | −8.1 | −8.0 | −7.9 | −7.9 | −7.8 | −7.7 | −7.6 | −7.6 | −8.1 |
| 3 | −8.5 | −8.3 | −8.2 | −8.2 | −8.2 | −8.2 | −8.1 | −8.1 | −8.1 | −8.5 |
| 4 | −8.0 | −8.0 | −7.9 | −7.9 | −7.5 | −7.5 | −7.4 | −7.3 | −7.3 | −8.0 |
| 5 | −8.0 | −8.0 | −8.0 | −7.9 | −7.9 | −7.9 | −7.8 | −7.7 | −7.7 | −8.0 |
| 6 | −8.0 | −8.0 | −7.9 | −7.9 | −7.9 | −7.9 | −7.9 | −7.7 | −7.6 | −8.0 |
| 7 | −8.8 | −8.6 | −8.6 | −8.4 | −8.4 | −8.3 | −8.2 | −8.1 | −8.0 | −8.8 |
| 8 | −8.4 | −8.2 | −8.2 | −8.1 | −8.0 | −8.0 | −7.8 | −7.7 | −7.6 | −8.4 |
| 9 | −8.2 | −8.2 | −7.9 | −7.9 | −7.9 | −7.9 | −7.9 | −7.9 | −7.9 | −8.2 |
| 10 | −8.8 | −8.7 | −8.5 | −8.3 | −8.3 | −8.2 | −8.1 | −8.1 | −8.0 | −8.8 |
| 11 | −8.4 | −8.3 | −8.2 | −8.2 | −8.1 | −8.0 | −8.0 | −7.8 | −7.8 | −8.4 |
| 12 | −8.3 | −8.3 | −8.3 | −8.3 | −8.2 | −8.1 | −8.1 | −8.0 | −7.9 | −8.3 |
| 13 | −8.7 | −8.5 | −8.3 | −8.2 | −8.0 | −8.0 | −8.0 | −7.9 | −7.8 | −8.7 |
| 14 | −8.6 | −8.2 | −8.1 | −8.1 | −8.0 | −7.7 | −7.7 | −7.7 | −7.7 | −8.6 |
| 15 | −8.5 | −8.5 | −8.5 | −8.5 | −8.4 | −8.3 | −8.0 | −7.9 | −7.7 | −8.5 |
| 16 | −8.8 | −8.6 | −8.5 | −8.4 | −8.4 | −8.2 | −8.2 | −7.9 | −7.9 | −8.8 |
| 17 | −8.6 | −8.5 | −8.4 | −8.4 | −8.3 | −8.2 | −8.0 | −7.7 | −7.7 | −8.6 |
| 18 | −8.6 | −8.5 | −8.5 | −8.5 | −8.5 | −8.3 | −8.3 | −8.1 | −7.9 | −8.6 |
| 19 | −8.8 | −8.7 | −8.6 | −8.4 | −8.3 | −8.2 | −8.2 | −8.1 | −8.1 | −8.8 |
| 20 | −8.7 | −8.7 | −8.6 | −8.5 | −8.4 | −8.2 | −8.1 | −8.0 | −7.9 | −8.7 |
| 21 | −8.4 | −8.2 | −8.1 | −8.1 | −8.1 | −8.0 | −8.0 | −7.9 | −7.8 | −8.4 |
| 22 | −8.3 | −8.3 | −8.2 | −8.1 | −8.1 | −8.0 | −8.0 | −7.9 | −7.7 | −8.3 |
| 23 | −8.4 | −8.3 | −8.3 | −8.2 | −8.0 | −8.0 | −8.0 | −7.9 | −7.5 | −8.4 |
| 24 | −8.3 | −8.2 | −8.2 | −8.2 | −8.1 | −8.0 | −7.8 | −7.7 | −7.6 | −8.3 |
| 25 | −8.6 | −8.5 | −8.2 | −8.2 | −8.0 | −8.0 | −7.9 | −7.9 | −7.6 | −8.6 |
| 26 | −8.3 | −8.3 | −8.3 | −8.3 | −8.1 | −8.0 | −8.0 | −7.9 | −7.9 | −8.3 |
| 27 | −8.5 | −8.3 | −8.3 | −8.3 | −8.1 | −8.1 | −8.0 | −8.0 | −7.9 | −8.5 |
| 28 | −8.7 | −8.4 | −8.3 | −8.3 | −8.2 | −8.1 | −8.1 | −8.0 | −8.0 | −8.7 |
| 29 | −8.6 | −8.6 | −8.3 | −8.3 | −8.3 | −8.2 | −8.2 | −8.1 | −8.1 | −8.6 |
| 30 | −8.5 | −8.5 | −8.4 | −8.3 | −8.2 | −8.2 | −8.2 | −8.0 | −8.0 | −8.5 |
| 31 | −8.6 | −8.5 | −8.4 | −8.4 | −8.3 | −8.3 | −8.3 | −8.1 | −8.0 | −8.6 |
| 32 | −8.6 | −8.6 | −8.6 | −8.5 | −8.4 | −8.3 | −8.2 | −8.1 | −8.1 | −8.6 |
| 33 | −8.3 | −8.2 | −8.1 | −8.0 | −7.9 | −7.8 | −7.7 | −7.7 | −7.5 | −8.3 |
| 34 | −8.2 | −8.1 | −8.1 | −8.1 | −8.0 | −8.0 | −7.9 | −7.9 | −7.8 | −8.2 |
| 35 | −8.3 | −8.2 | −7.9 | −7.8 | −7.8 | −7.8 | −7.7 | −7.6 | −7.6 | −8.3 |
| 36 | −8.1 | −8.1 | −8.0 | −8.0 | −8.0 | −8.0 | −7.9 | −7.8 | −7.7 | −8.1 |
| 37 | −8.8 | −8.5 | −8.5 | −8.2 | −8.2 | −8.2 | −8.0 | −8.0 | −8.0 | −8.8 |
| 38 | −8.8 | −8.6 | −8.2 | −8.2 | −8.2 | −8.2 | −8.1 | −8.1 | −8.1 | −8.8 |
| 39 | −8.2 | −8.2 | −8.2 | −8.1 | −8.1 | −8.1 | −8.1 | −8.0 | −8.0 | −8.2 |
| 40 | −8.3 | −8.3 | −8.2 | −8.2 | −8.2 | −8.2 | −8.2 | −8.1 | −8.1 | −8.3 |

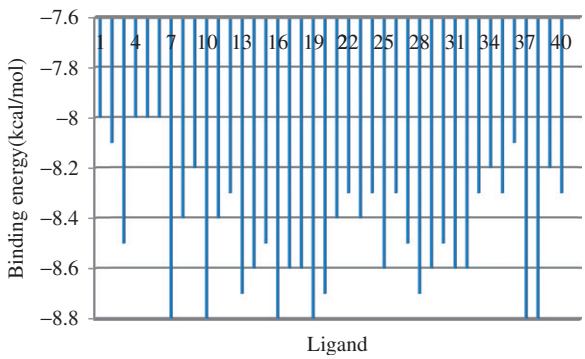


Figure 3. Binding energy (kcal/mol) for the docked ligands.

*Data reduction*

The procedure in the same as described in the Section ''Data reduction''. The new descriptor $SD_{LD50}$ correlates with LD50 as below:

$$LD50 = 0.989 \times SD_{LD50} + 12479.7$$

$$N = 26; \quad R^2 = 0.882; \quad s = 478.864; \quad F = 164.772$$

*QSAR models*

The models were performed on the training set (17 structures in Table 2) and the best results (in decreasing order of $R^2$) are listed below and in Table 10.

(v) Monovariate regression
$LD50 = 12298.6 + 0.986 \times SD_{LD50}$

(vi) Bivariate regression
$LD50 = 12286.2 + 0.989 \times SD_{LD50} + 0.059 \times D3D$

(vii) Three-variate regression
$LD50 = 11832.36 + 1.017 \times SD_{LD50} + 18.889 \times CjDe- -52.112 \times CfDe$

(viii) Five-variate regression
$LD50 = 14921.91 + 1.053 \times SD_{LD50} - 155.286 \times C + 190.495 \times CjDe - 178.9 \times CfDe$

*Model validation*

*Leave-one-out.* The performances in leave-one-out analysis related to the models listed as best in Table 10 are presented in Table 11.

*External validation.* The values $LD50_{calc.}$ for each of the 12 molecules in the test set were chosen based on the lowest energy docking and computed with the same descriptors as in Table 10,

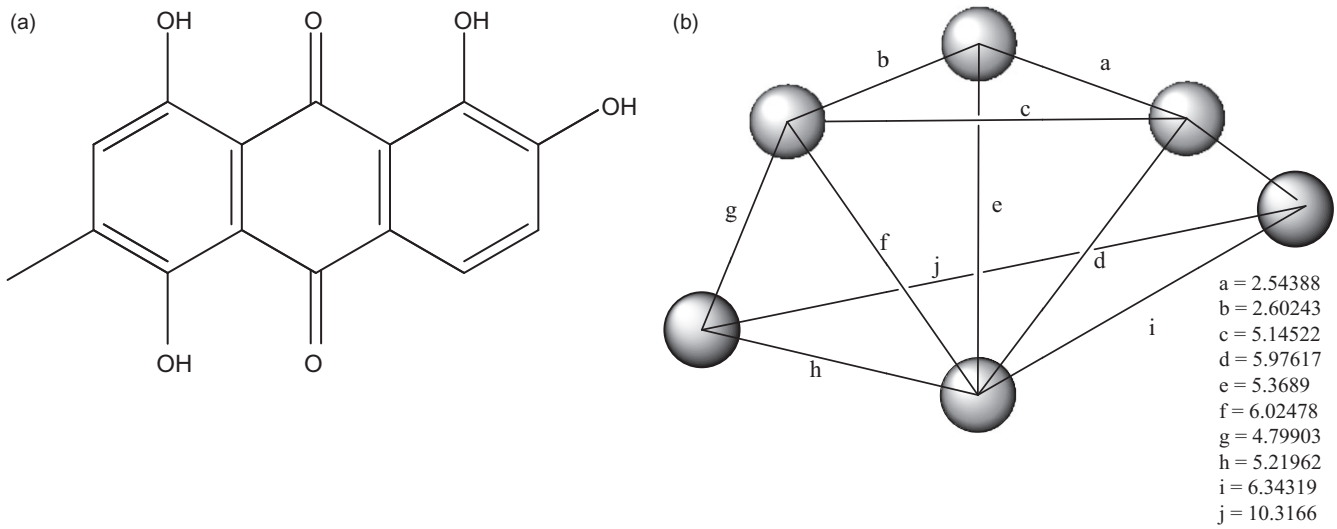


Figure 4. (a): Pharmacophore model for the receptor glycogen synthase kinase-3 beta. (b): Selected data on the pharmacophore model of anthraquinone/3Q3B protein interaction.

Table 4. LD50, sum descriptor and topological indices for the set of 40 anthraquinone derivatives.

| Mol. | LD50 | $SD_{LD50}$ | CjDe | CfDe |
|---|---|---|---|---|
| 1 | 5000 | −7028.6 | 260 | 267 |
| 2 | 5000 | −8140.0 | 309 | 317 |
| 3 | 2500 | −9929.8 | 424 | 436 |
| 4 | 3200 | −9439.3 | 307 | 315 |
| 5 | 3216 | −9314.4 | 307 | 315 |
| 6 | 1110 | −11153.3 | 367 | 378 |
| 7 | 1230 | −10 884.5 | 484 | 500 |
| 13 | 4000 | −9002.3 | 421 | 434 |
| 14 | 2000 | −10 624.0 | 479 | 494 |
| 16 | 2795 | −10 735.0 | 421 | 433 |
| 18 | 5000 | −7611.9 | 422 | 437 |
| 19 | 35 | −12 092.9 | 483 | 498 |
| 20 | 308 | −12 199.7 | 484 | 501 |
| 25 | 1870 | −10 256.7 | 420 | 433 |
| 26 | 1000 | −10 993.7 | 422 | 437 |
| 27 | 2200 | −10 328.4 | 484 | 500 |
| 28 | 2795 | −9297.9 | 420 | 433 |
| 32 | 3950 | −8842.7 | 546 | 563 |
| 33 | 1500 | −11444.5 | 424 | 436 |
| 35 | 2795 | −10280.4 | 365 | 374 |
| 36 | 2795 | −9088.6 | 362 | 373 |
| 38 | 2795 | −9812.2 | 549 | 566 |
| 39 | 2800 | −10 246.8 | 486 | 501 |
| 40 | 2795 | −10 722.6 | 474 | 488 |

entry 10. Data are listed in Table 12 and the monovariate correlation: $LD50 = 0.866 \times LD50_{calc.} + 545.6$; $n = 12$; $R^2 = 0.904$; $s = 477.245$; $F = 95.201$ plotted in Figure 7.

*Similarity cluster validation.* The clusters of similarity in this section were performed by using as leaders the 12 molecules best scored in the docking step, in the same manner as in Section ''Similarity cluster validation''.

The predicted values LD50 are listed in Table 13 and the monovariate correlation: $LD50 = 0.861 \times LD50_{calc.} + 506.19$; $n = 12$; $R^2 = 0.959$; $s = 314.696$; $F = 231.948$ plotted in Figure 8.

Compare the results in Figures 7 and 8 to see: (i) a rather low prediction ($R^2 = 0.904$) by the external test set and (ii) a better prediction ($R^2 = 0.959$) by the same set predicted by the similarity clusters (approaching to the congeneric status), even the test set

Table 5. Topological indices computed for the anthraquinone in Table 1.

| Mol. | SD | Di | D3D | HOMO | Adj | CfDe | CfDi |
|---|---|---|---|---|---|---|---|
| 1 | −17.465 | 378 | 432 | −10.173 | 18 | 267 | 767 |
| 2 | −18.133 | 452 | 519 | −9.915 | 19 | 317 | 916 |
| 3 | −17.601 | 598 | 677 | −9.438 | 21 | 436 | 1248 |
| 4 | −17.494 | 450 | 508 | −9.236 | 19 | 315 | 920 |
| 5 | −17.494 | 450 | 507 | −8.136 | 19 | 315 | 920 |
| 6 | −17.774 | 512 | 578 | −9.489 | 20 | 378 | 1073 |
| 7 | −17.863 | 692 | 786 | −9.277 | 22 | 500 | 1437 |
| 8 | −17.96 | 529 | 608 | −9.499 | 20 | 373 | 1077 |
| 9 | −18.736 | 692 | 784 | −9.425 | 22 | 495 | 1435 |
| 10 | −18.719 | 788 | 895 | −9.384 | 23 | 563 | 1634 |
| 11 | −17.772 | 620 | 700 | −9.114 | 21 | 424 | 1258 |
| 12 | −17.118 | 523 | 597 | −9.439 | 20 | 374 | 1076 |
| 13 | −17.798 | 610 | 698 | −9.292 | 21 | 434 | 1252 |
| 14 | −18.329 | 702 | 803 | −8.891 | 22 | 494 | 1440 |
| 15 | −17.755 | 686 | 779 | −9.279 | 22 | 501 | 1433 |
| 16 | −17.883 | 610 | 696 | −9.204 | 21 | 433 | 1251 |
| 17 | −17.928 | 598 | 677 | −9.215 | 21 | 438 | 1249 |
| 18 | −17.846 | 600 | 685 | −9.193 | 21 | 437 | 1244 |
| 19 | −18.305 | 691 | 784 | −9.287 | 22 | 498 | 1435 |
| 20 | −18.269 | 685 | 777 | −9.281 | 22 | 501 | 1432 |
| 21 | −18.43 | 608 | 691 | −9.448 | 21 | 433 | 1251 |
| 22 | −17.999 | 598 | 677 | −9.313 | 21 | 438 | 1249 |
| 23 | −18.567 | 610 | 694 | −9.235 | 21 | 434 | 1252 |
| 24 | −18.019 | 519 | 590 | −9.229 | 20 | 378 | 1074 |
| 25 | −18.549 | 608 | 693 | −9.212 | 21 | 433 | 1250 |
| 26 | −18.217 | 600 | 685 | −9.314 | 21 | 437 | 1244 |
| 27 | −18.476 | 692 | 784 | −9.284 | 22 | 500 | 1437 |
| 28 | −17.781 | 608 | 694 | −9.313 | 21 | 433 | 1250 |
| 29 | −18.972 | 688 | 779 | −9.217 | 22 | 499 | 1433 |
| 30 | −17.871 | 685 | 778 | −9.36 | 22 | 501 | 1432 |
| 31 | −18.622 | 786 | 893 | −9.382 | 23 | 565 | 1634 |
| 32 | −18.817 | 788 | 895 | −9.289 | 23 | 563 | 1634 |
| 33 | −17.973 | 598 | 671 | −9.426 | 21 | 436 | 1278 |
| 34 | −17.882 | 523 | 594 | −9.447 | 20 | 374 | 1076 |
| 35 | −17.9 | 524 | 596 | −9.458 | 20 | 374 | 1077 |
| 36 | −18.546 | 529 | 605 | −9.314 | 20 | 373 | 1077 |
| 37 | −18.491 | 620 | 712 | −9.473 | 21 | 428 | 1254 |
| 38 | −18.261 | 778 | 880 | −9.173 | 23 | 566 | 1632 |
| 39 | −18.421 | 680 | 765 | −9.081 | 22 | 501 | 1433 |
| 40 | −19.625 | 714 | 818 | −9.334 | 22 | 488 | 1442 |

Table 6. The best models in describing log *P* in the training set of anthraquinone in Table 1.

| | Descriptors | $R^2$ | Adjust. $R^2$ | St. Error | *F* |
|---|---|---|---|---|---|
| 1 | SD | **0.935** | 0.932 | 0.137 | 330.631 |
| 2 | D3D | 0.229 | 0.195 | 0.472 | 6.83 |
| 3 | Di | 0.218 | 0.184 | 0.475 | 6.421 |
| 4 | Adj | 0.175 | 0.139 | 0.488 | 4.875 |
| 5 | SD, D3D | **0.938** | 0.932 | 0.167 | 166.436 |
| 6 | SD, CfDe | 0.938 | 0.932 | 0.137 | 165.511 |
| 7 | SD, De | 0.937 | 0.932 | 0.137 | 165.256 |
| 8 | SD, Adj | 0.937 | 0.932 | 0.137 | 165.077 |
| 9 | SD, C | 0.937 | 0.931 | 0.136 | 163.616 |
| 10 | SD, HOMO | 0.935 | 0.929 | 0.14 | 158.383 |
| 11 | SD, Adj, Di | **0.939** | 0.931 | 0.138 | 108.685 |
| 12 | SD, C, D3D | 0.939 | 0.931 | 0.138 | 108.493 |
| 13 | SD, C, Di | 0.939 | 0.931 | 0.139 | 108.254 |
| 14 | SD, Adj, D3D | 0.939 | 0.93 | 0.139 | 107.901 |
| 15 | SD, Di, HOMO | 0.939 | 0.93 | 0.139 | 107.814 |
| 16 | SD, Di, CfDi | 0.938 | 0.929 | 0.14 | 106.199 |
| 17 | SD, Adj, D3D, CfDe | **0.943** | 0.931 | 0.138 | 82.489 |
| 18 | SD, C, Di, HOMO | 0.94 | 0.929 | 0.14 | 79.251 |
| 19 | SD, C, Di, De | 0.939 | 0.927 | 0.142 | 77.695 |

The bold values show the best result.

Table 7. Leave-one-out analysis for best log *P* models in Table 6.

| | Descriptors | $Q^2$ | $R^2-Q^2$ | St. $Error_{loo}$ | $F_{loo}$ |
|---|---|---|---|---|---|
| 1 | SD | **0.925** | 0.01 | 0.148 | 281.724 |
| 5 | SD, D3D | **0.924** | 0.014 | 0.148 | 280.667 |
| 11 | SD, Adj, Di | **0.921** | 0.018 | 0.151 | 268.824 |
| 17 | SD, Adj, D3D, CfDe | **0.916** | 0.027 | 0.155 | 252.011 |

The bold values show the best result.

Table 8. Calculated values of log *P* for the molecules in the test set (mass fragments) Table 1.

| Molecules | log *P* | log $P_{calc.}$ |
|---|---|---|
| 1 | 3.4 | 3.69 |
| 2 | 3 | 2.92 |
| 4 | 3.2 | 3.64 |
| 5 | 3.9 | 3.64 |
| 8 | 2.9 | 3.12 |
| 11 | 3.2 | 3.39 |
| 12 | 3.9 | 4.06 |
| 13 | 3.1 | 3.32 |
| 14 | 2.7 | 2.77 |
| 16 | 3.1 | 3.23 |
| 17 | 3.1 | 3.13 |
| 18 | 3.1 | 3.23 |
| 24 | 3.2 | 3.01 |
| 26 | 2.5 | 2.80 |
| 36 | 2.7 | 2.45 |

has been chosen the one with the lowest docking energies. This result put our approach in a favorable light and demonstrates its utility in QSAR studies.

## Conclusions

A set of 40 anthraquinone, downloaded from the PubChem database, was submitted to a QSAR study, the modeled property/activity being log *P* and LD50. The set was split into a learning set and a test set, used in the model (external) validation. Also, the validation was made by a new version of prediction by using similarity clusters.

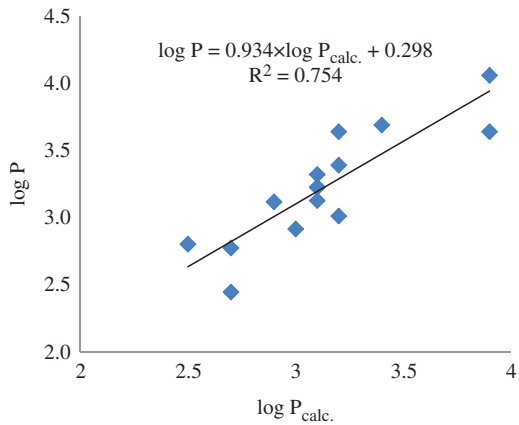


Figure 5. The plot log *P* versus log $P_{calc.}$ for the test set (mass fragments, external validation).

Table 9. Calculated values of log *P* by similarity clusters, for the molecules in the test set (mass fragments) (Table 1).

| Molecules | log *P* | log $P_{calc.}$ |
|---|---|---|
| 1 | 3.4 | 3.47 |
| 2 | 3 | 2.91 |
| 4 | 3.2 | 3.35 |
| 5 | 3.9 | 4.06 |
| 8 | 2.9 | 3.04 |
| 11 | 3.2 | 3.32 |
| 12 | 3.9 | 4.00 |
| 13 | 3.1 | 3.29 |
| 14 | 2.7 | 2.75 |
| 16 | 3.1 | 3.19 |
| 17 | 3.1 | 3.13 |
| 18 | 3.1 | 3.20 |
| 24 | 3.2 | 3.13 |
| 26 | 2.5 | 2.65 |
| 36 | 2.7 | 2.69 |

Glycogen synthase kinase 3 beta has been investigated for its potential binding affinity with selective anthraquinone derivatives. The docking test of the studied anthraquinones have shown binding energies in the range of −8.8 kcal/mol to −8 kcal/mol.

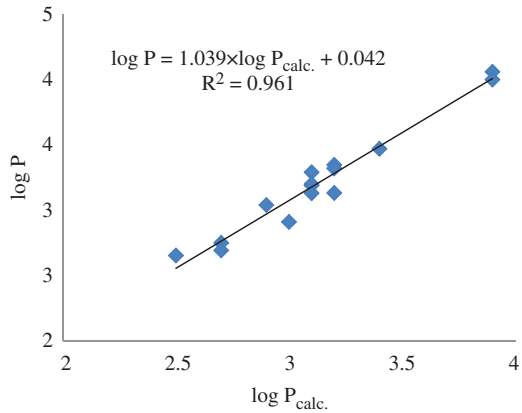


Figure 6. The plot log $P$ versus log $P_{calc.}$ by similarity clusters (mass fragments).

Table 10. The best models in describing LD50 in the training set of anthraquinone in Table 2.

| | Descriptors | $R^2$ | Adjust. $R^2$ | St. Error | $F$ |
|---|---|---|---|---|---|
| 1 | $SD_{LD50}$ | **0.937** | 0.933 | 376.367 | 223.866 |
| 2 | CjDe | 0.36 | 0.317 | 1201.22 | 8.449 |
| 3 | CfDe | 0.358 | 0.315 | 1203.472 | 8.362 |
| 4 | HOMO | 0.011 | 0.005 | 1493.48 | 0.169 |
| 5 | $SD_{LD50}$, D3D | **0.937** | 0.928 | 389.534 | 104.495 |
| 6 | $SD_{LD50}$, Di | 0.937 | 0.928 | 389.456 | 104.54 |
| 7 | $SD_{LD50}$, De | 0.937 | 0.928 | 389.431 | 104.554 |
| 8 | $SD_{LD50}$, Adj | 0.937 | 0.933 | 376.367 | 223.866 |
| 9 | $SD_{LD50}$, C | 0.937 | 0.928 | 389.166 | 104.706 |
| 10 | $SD_{LD50}$, CjDe, CfDe | **0.952** | 0.941 | 353.239 | 86.056 |
| 11 | $SD_{LD50}$, De, D3D | 0.938 | 0.924 | 401.168 | 65.748 |
| 12 | $SD_{LD50}$, De,CjDi | 0.937 | 0.923 | 403.545 | 64.925 |
| 13 | $SD_{LD50}$, De, Di | 0.943 | 0.93 | 384.287 | 72.041 |
| 14 | $SD_{LD50}$, Di, D3D | 0.943 | 0.93 | 384.287 | 72.04 |
| 15 | $SD_{LD50}$, C, CjDe, CfDe | **0.953** | 0.937 | 362.101 | 61.515 |
| 16 | $SD_{LD50}$, D3D, CjDi, De | 0.946 | 0.928 | 388.737 | 52.976 |
| 17 | $SD_{LD50}$, De, Di, D3D | 0.945 | 0.927 | 393.789 | 51.549 |

The bold values show the best result.

Table 11. Leave-one-out analysis for best LD50 models in Table 10.

| | Descriptors | $Q^2$ | $R^2-Q^2$ | St. $Error_{loo}$ | $F_{loo}$ |
|---|---|---|---|---|---|
| 1 | $SD_{LD50}$ | **0.919** | 0.018 | 426.602 | 170.922 |
| 5 | $SD_{LD50}$, D3D | **0.911** | 0.026 | 449.032 | 152.712 |
| 11 | $SD_{LD50}$, CjDe, CfDe | **0.914** | 0.038 | 439.322 | 160.312 |
| 17 | $SD_{LD50}$, C,CjDe, CfDe | **0.911** | 0.042 | 449.206 | 152.682 |

The bold values show the best result.

Table 12. Calculated values of LD50 for the molecules in the test set (partial charges).

| Mol. | LD50 | $LD50_{calc.}$ |
|---|---|---|
| 3 | 2500 | 2586.67 |
| 7 | 1230 | 1850.72 |
| 13 | 4000 | 3709.78 |
| 14 | 2000 | 2484.99 |
| 16 | 2795 | 1980.80 |
| 18 | 5000 | 4816.41 |
| 19 | 35 | 688.14 |
| 20 | 308 | 366.56 |
| 25 | 1870 | 2448.36 |
| 28 | 2795 | 3423.44 |
| 32 | 3950 | 4365.44 |
| 38 | 2795 | 3185.28 |

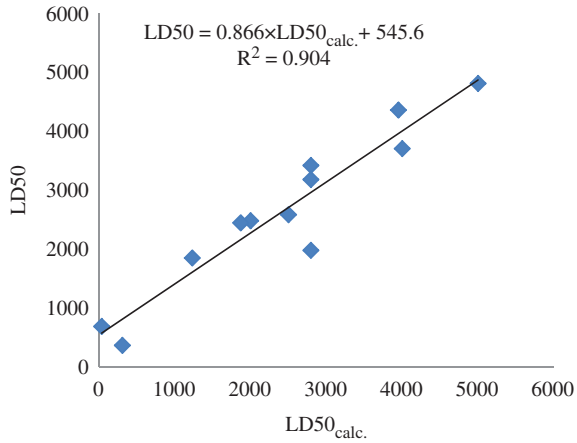


Figure 7. The plot LD50 versus $LD50_{calc.}$ for the test set (partial charges, external validation).

Table 13. Calculated values of LD50 by similarity clusters, for the molecules in the test set (partial charges).

| Mol. | LD50 | $LD50_{calc.}$ |
|---|---|---|
| 3 | 2500 | 2528.503 |
| 7 | 1230 | 1622.526 |
| 13 | 4000 | 3463.02 |
| 14 | 2000 | 2377.399 |
| 16 | 2795 | 2796.373 |
| 18 | 5000 | 4810.405 |
| 19 | 35 | 594.5576 |
| 20 | 308 | 366.6446 |
| 25 | 1870 | 2432.604 |
| 28 | 2795 | 3328.103 |
| 32 | 3950 | 4081.971 |
| 38 | 2795 | 2872.895 |

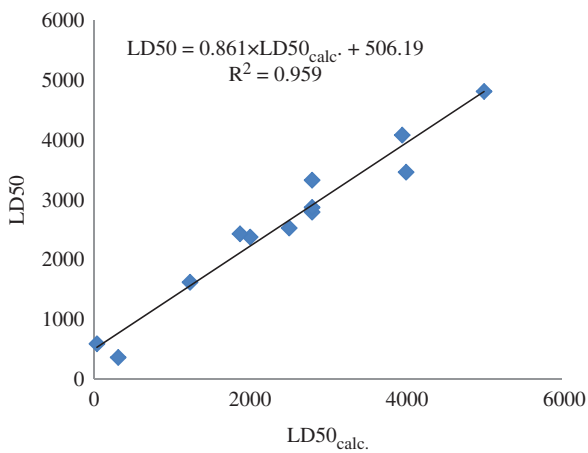


Figure 8. The plot LD50 versus $LD50_{calc.}$ by similarity clusters (partial charges).

The excellent prediction of LD50 obtained by the clusters built on the basis of docking study (leaders being those molecules with the highest affinity to 3Q3B protein) enabled us to suggest the toxicity of anthraquinones is given (with high probability) by the interaction of these molecules with 3Q3B protein.

## Acknowledgements

The authors acknowledge to the referees for the valuable suggestions.

## Declaration of interest

This paper is a result of a doctoral research made possible by the financial support of the Sectoral Operational Programme for Human Resources Development 2007–2013, co-financed by the European Social Fund, under the project POSDRU/159/1.5/S/137750 – ''Doctoral and postdoctoral programs – support for increasing research competitiveness in the field of exact Sciences''.

## References

1. Thomson RH. Naturally occurring quinones IV. London: Springer; 1996.
2. Blum MS, Hilker M. Chemical protection of insect eggs. In: Hilker M, Meiners T, eds. Chemoecology of insect eggs and egg deposition. Berlin, Oxford: Blackwell Publishing; 2002:61–90.
3. Matasyoh JC, Dittrich B, Schueffler A, Laatsch H. Larvicidal activity of metabolites from the endophytic *Podospora* sp. against the malaria vector *Anopheles gambiae*. Parasitol Res 2011;108:561–6.
4. Izhaki I. Emodin: a secondary metabolite with multiple ecological functions in higher plants. New Phytol 2002;155:205–17.
5. Dragos D, Heghes A, Medeleanu M, Vlaia V, et al. Topological similarity/dissimilarity indicators: application to cytochrome P450 inhibition by alcohols. TMJ 2004;54:128–34.
6. Ho DC, Kwang SL, Tae BK, No KT. Quantitative structure-activity relationship (QSAR) study of new fluorovinyloxyacetamides. Bull Korean Chem Soc 2001;22:4.
7. Meylan WM, Howard PH. Estimating log *P* with atom/fragments and water solubility with log *P*. Perspect Drug Discov 2000;19:67–84.
8. Lyman WJ, Reehl WF, Rosenblatt DH. Handbook of chemical property estimation methods: environmental behaviour of organic compounds. Washington, DC: American Chemical Society; 1990.
9. Balaban AT, Chiriac A, Motoc I, Simon Z. Steric fit in QSAR, lectures notes in chemistry. Berlin: Springer; 1980.
10. Duda-Seiman C, Duda-Seiman D, Dragos D, et al. Design of anti HIV ligands by means of Minimal Topological Difference (MTD) method. Int J Mol Sci 2006;7:537–55.
11. The RCBS Protein data bank. Available from: http://www.rcsb.org/pdb.
12. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 2010;31:455–61.
13. Jubie S, Kalirajan R, Pavankumar Y. Design, synthesis and docking studies of a novel ciprofloxacin analogue as an antimicrobial AGENT. E-J Chem 2012;9:980–7.
14. Sanner MF. Python: a programming language for software integration and development. J Mol Graph Model 1999;17:57–61.
15. Dhananjayan K, Kalathil K, Sumathy A, Sivanandy P. A computational study on binding affinity of Bio-flavonoids on the crystal structure of 3-hydroxy-3-methyl-glutaryl-CoA reductase – an insilico molecular docking approach. Der Pharma Chemica 2014;6:378–87.
16. Nagy Cs.L, Diudea MV. Nano Studio software package. Cluj: Babes-Bolyai University; 2009.
17. Frisch MJ, Trucks GW, Schlegel HB, et al. Gaussian 09, Revision A.1. Wallingford (CT): Gaussian Inc; 2009.
18. Wiener H. Structural determination of the paraffin boiling points. J Am Chem Soc 1947;69:17−20.
19. Ursu O, Diudea MV. TOPOCLUJ software program. Cluj: Babes-Bolyai University; 2005.
20. Hawkins DM, Basak SC, Mills D. Assessing model fit by cross-validation. J Chem Inf Comp Sci 2003;43:579–86.
21. Bolboacă SD, Jäntschi L, Diudea MV. Molecular design and QSARs with molecular descriptors family. Curr Comput Aided Drug Des 2013;9:195–205.
22. Jäntschi L. LOO Analysis (LOO: leave one out), Academic Direct Library of software; 2005. Available from: http://l.academicdirect.org/Chemistry/SARs/MDF_SARs/loo/.
23. Harsa TE, Harsa AM, Szefler B. QSAR of caffeines by similarity cluster prediction. Cent Eur J Chem 2014;12:365–76.
24. Harsa AM, Harsa TE, Bolboaca S, Diudea MV. QSAR in flavonoids by similarity cluster prediction. Curr Comput Aided Drug Des 2014;10:115–28.