STUDIA UBB CHEMIA, LX, 4, 2015 (p. 201-212)
(RECOMMENDED CITATION)

QSAR AND DOCKING STUDY ON INDOLIZINES BY SIMILARITY CLUSTERING

BEATA SZEFLER[a*], TEODORA E. HARSA[b], ALEXANDRA M. HARSA[b]

ABSTRACT. A Quantitative Structure-Activity Relationship study, based on molecular descriptors calculated with correlation weights within the hypermolecule, that mimics the investigated correlational space, was performed on a set of 25indolizines (PubChem database). The best models describing IC50 of this set of indolizines were validated by the leave-one-out procedure, in the external test set and in a new version of prediction by using clusters of similar molecules. The best prediction was provided by the similarity cluster procedure.

Keywords: *Indolizines, QSAR (Quantitative Structure-Activity Relationships), IC50, similarity, Hypermolecule, Docking, Binding energy.*

INTRODUCTION

Indolizines represent a class of heteroaromatic compounds (of pharmacological importance) containing two condensed (5- and 6-memebered) rings bridged by a nitrogen atom. Other names for indolizines in literature include pyrindole, pyrrodine, pyrrolo [1,2-a]pyridine and pyrrocoline. The aromatic indolizine does not occur in nature, but the reduced derivatives are natural products. The biological activities include antimicrobial, antioxidant, anti-inflammatory, anti-tuberculosis, anticonvulsant activity, enzymes inhibition activity and cardiovascular activity (as calcium entry blocker) [1].

Many indolizine derivatives have been isolated from plants, insects, animals, marine lives, and microbes [2]. The importance of indolizines has promoted a variety of synthetic routes in medicinal chemistry (e.g. Tschitschibabin reaction, 1,3-dipolar cycloadditions, or various cyclisation reactions [3]), targeting

[a] *Department of Physical Chemistry, Collegium Medicum, Nicolaus Copernicus University, Kurpińskiego 5, 85-950, Bydgoszcz, Poland.*
[b] *Babes-Bolyai University, Faculty of Chemistry and Chemical Engineering, Arany Janos street 11, Cluj-Napoca, RO-400028, Romania.*
** Corresponding author: beatas@cm.umk.pl*

indolizines with well-defined substitution patterns. Synthetic indolizines are important as histamine H3 receptor antagonists, 5-HT3 receptor antagonists, microtubule inhibitors showing strong antitumor activities [4].

In the present study, a molecular docking analysis has been performed on indolizine derivatives on the proteins 1GA0, 4O0Z, 4O10,then we made a QSAR study to predicting IC50 of indolizine derivatives.

STRUCTURAL MOLECULAR DATA

A set of 25indolizine were taken from PubChem Database [5] (Table 1) and were divided into a training set (15 molecules) and a test set (10 molecules), taken randomly. The property chosen for modeling was IC50 (μM) 15-LO from soybeans) (see Table 1) [6].

Table 1. IC50 (μM) 15-LO from soybeans, Smiles code and CID (PubChem) for 25indolizines

	Canonical SMILES	CID	IC_{50}
1	C1=C(C=C[N]2C1=CC(=C2C3=CC=CC=C3)C4=CC=CC=C4)C#N	482634	30±2
2	C1=C(C=C[N]2C1=C(C(=C2C3=CC=CC=C3)C4=CC=CC=C4)OC)C#N	10853428	33±2
3	C1=CC=C(C=C1)COC2=C3C=C(C=CN3C(=C2C4=CC=CC=C4)C5=CC=CC=C5)C#N	57399922	31±2
4	[N]12C(=C(C(=C1C=C(C=C2)C#N)CC3=CC=CC=C3)C4=CC=CC=C4)C5=CC=CC=C5	491919	27±2
5	[N]12C(=C(C(=C1C=C(C=C2)C#N)C3=CC=CC=C3)C4=CC=CC=C4)C5=CC=CC=C5	482635	28±1
6	COC1=CC=CC=C1C2=C3C=C(C=CN3C(=C2C4=CC=CC=C4)C5=CC=CC=C5)C#N	10644682	20±1
7	COC1=CC=CC(=C1)C2=C3C=C(C=CN3C(=C2C4=CC=CC=C4)C5=CC=CC=C5)C#N	10668699	20±1
8	C1=CC=C(C=C1)C2=C(N3C=CC(=CC3=C2C=O)C#N)C4=CC=CC=C4	491918	29±1
9	CC(=O)C1=C2C=C(C=CN2C(=C1C3=CC=CC=C3)C4=CC=CC=C4)C#N	491856	23±1
10	C1=CC(=CC=C1C2=C(N3C=CC(=CC3=C2)C#N)C4=CC=C(C=C4)F)F	57394641	30±6
11	O(C1=C4[N](C(=C1C2=CC=CC=C2)C3=CC=CC=C3)C=C(C=C4)C#N)C(=O)C	10066595	31±2
12	CC(C1=C2C=C(C=CN2C(=C1C3=CC=CC=C3)C4=CC=CC=C4)C#N)O	491916	26±2
13	CC1=C2C=C(C=CN2C(=C1C3=CC=CC=C3)C4=CC=CC=C4)C#N	491857	27±2
14	C1=CC=C(C=C1)C2=C(N3C=CC(=CC3=C2C(=O)C4=CC=CC=C4)C#N)C5=CC=CC=C5	491920	23±1
15	C1=CC=C(C=C1)C2=C(N3C=CC(=CC3=C2CO)C#N)C4=CC=CC=C4	491917	26±1
16	C1=CC=C(C=C1)C2=C(N3C=CC(=CC3=C2OS(=O)(=O)C4=CC=CC=C4)C#N)C5=CC=CC=C5	57394621	24±3
17	CN(C)S(=O)(=O)OC1=C2C=C(C=CN2C(=C1C3=CC=CC=C3)C4=CC=CC=C4)C#N	57396377	25±2
18	C1=CC(=CC=C1C2=C(N3C=CC(=CC3=C2)C#N)C4=CC=C(C=C4)Cl)Cl	57392897	33±7
19	C1=C(C=C[N]2C1=C(C(=C2C3=CC=CC=C3)C4=CC=CC=C4)OCC5=CC=C(C=C5)C)C#N	57401703	37±4

	Canonical SMILES	CID	IC_{50}
20	C1=C(C=C[N]2C1=CC(=C2C3=CC=CC=C3)C4=CC=CC=C4)C#N	482634	30±2
21	CC1=CC=C(C=C1)C2=C(N3C=CC(=CC3=C2)C#N)C4=CC=C(C=C4)C	57392898	27±3
22	COC(C1=CC=CC=C1)C2=C3C=C(C=CN3C(=C2C4=CC=CC=C4)C5=CC=CC=C5)C#N	491877	21±3
23	CS(=O)(=O)OC1=C2C=C(C=CN2C(=C1C3=CC=CC=C3)C4=CC=CC=C4)C#N	53855501	22±2
24	CCS(=O)(=O)OC1=C2C=C(C=CN2C(=C1C3=CC=CC=C3)C4=CC=CC=C4)C#N	57403346	28±6
25	C1=CC=C(C=C1)COCOC2=C3C=C(C=CN3C(=C2C4=CC=CC=C4)C5=CC=CC=C5)C#N	57403432	46±4

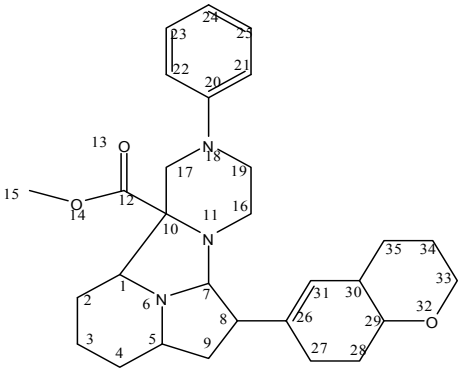
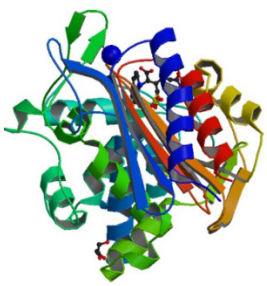
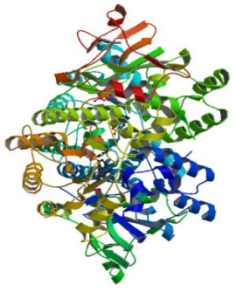


Figure 1. The hypermolecule comprising the common features of the dataset

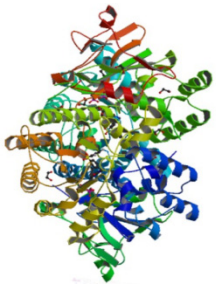
Three proteins: beta lactamase (Figure 2, left), nicotinamide phosphoribosyltransferases (Figure 2, middle, right), were downloaded from RCSB protein data bank with the PDB code-1GA0, 4O0Z, 4O10 [7].



Beta lactamase CODE: (1GA0) [7]



Nicotinamide phosphoribosyltransferase CODE: (4O0Z) [7]



Nicotinamide phosphoribosyltransferase CODE: (4O10) [7]

Figure 2. The proteins (beta lactamase, nicotinamide phosphoribosyltransferases) (RCSB PDB CODE: 1GA0, 4O0Z, 4O10).

Using the latest version of AutoDock4.2, the protein molecule is loaded and stored as protein.pdb after assigning hydrogen bonds [8]. The investigative ligand was loaded and their torsions along their rotatable bonds (see Table 2) are assigned and the file was saved as ligand.pdbqt. The grid menu is toggled [9]; after loading protein.pdbqt the map files were selected directly with setting up the grid points appropriate for the searching of ligand within the active site of the protein molecule. In case of protein 4O10 docking was made for two active sides of the enzyme. This way the grid parameter files are created with setting up the map files directly. The docking parameter files were completed by using the Lamarckian genetic algorithm [10].

Table 2. Starting parameters of docking for the investigated ligands

	CID of Ligand	Molecular Weight	Molecular Formula	H-Bond Donor	H-Bond Acceptor	Torsions (rotatable bonds)
1	482634	294.34926	$C_{21}H_{14}N_2$	0	1	4
2	10853428	324.37524	$C_{22}H_{16}N_2O$	0	2	4
3	57399922	400.4712	$C_{28}H_{20}N_2O$	0	2	5
4	491919	384.4718	$C_{28}H_{20}N_2$	0	1	8
5	482635	482635	$C_{27}H_{18}N_2$	0	1	8
6	10644682	400.4712	$C_{28}H_{20}N_2O$	0	2	4
7	10668699	400.4712	$C_{28}H_{20}N_2O$	0	2	3
8	491918	322.35936	$C_{22}H_{14}N_2O$	0	2	4
9	491856	336.38594	$C_{23}H_{16}N_2O$	0	2	7
10	57394641	330.330186	$C_{21}H_{12}F_2N_2$	0	3	10
11	10066595	352.38534	$C_{23}H_{16}N_2O_2$	0	3	7
12	491916	338.40182	$C_{23}H_{18}N_2O$	1	2	8
13	491857	308.37584	$C_{22}H_{16}N_2$	0	1	6
14	491920	398.45532	$C_{28}H_{18}N_2O$	0	2	7
15	491917	398.45532	$C_{28}H_{18}N_2O$	0	2	10
16	57394621	450.50842	$C_{27}H_{18}N_2O_3S$	0	4	3
17	57396377	417.48026	$C_{23}H_{19}N_3O_3S$	0	5	1
18	57392897	363.23938	$C_{21}H_{12}Cl_2N_2$	0	1	3
19	57401703	414.49778	$C_{29}H_{22}N_2O$	0	2	4
20	482634	294.34926	$C_{21}H_{14}N_2$	0	1	8
21	57392898	322.40242	$C_{23}H_{18}N_2$	0	1	5
22	491877	414.49778	$C_{29}H_{22}N_2O$	0	2	2
23	53855501	388.43904	$C_{22}H_{16}N_2O_3S$	0	4	4
24	57403346	402.46562	$C_{23}H_{18}N_2O_3S$	0	4	3
25	57403432	430.49718	$C_{29}H_{22}N_2O_2$	0	3	9

DOCKING RESULTS

Table 3. The final lamarckian genetic algorithm docked state – Binding energy of ligands with the active site of ***Beta lactamase (1GA0)*** during nine conformations

	CID of Ligand	1	2	3	4	5	6	7	8	9	Docked Energy (kcal/mol)
1	482634	-7.5	-7.3	-6.8	-6.8	-6.5	-6.5	-6.3	-6.2	-6.2	-7.5
2	10853428	-6.5	-6.2	-6.1	-6.1	-6.1	-6.0	-5.9	-5.8	-5.7	-6.5
3	57399922	-7.3	-7.3	-7.0	-7.0	-6.9	-6.9	-6.9	-6.8	-6.8	-7.3
4	491919	-8.1	-7.8	-7.6	-7.2	-7.0	-7.0	-6.9	-6.9	-6.8	-8.1
5	482635	-7.5	-7.4	-7.0	-6.9	-6.9	-6.8	-6.7	-6.6	-6.6	-7.5
6	10644682	-6.7	-6.7	-6.6	-6.4	-6.4	-6.2	-6.1	-6.0	-5.9	-6.7
7	10668699	-7.8	-7.7	-7.5	-7.4	-7.3	-7.0	-7.0	-7.0	-6.9	-7.8
8	491918	-7.0	-6.8	-6.5	-6.5	-6.4	-6.1	-6.0	-5.9	-5.8	-7.0
9	491856	-7.5	-7.2	-7.0	-6.9	-6.9	-6.8	-6.7	-6.7	-6.5	-6.5
10	57394641	-7.0	-6.9	-6.9	-6.9	-6.9	-6.8	-6.8	-6.8	-6.8	-7.0
11	10066595	-6.1	-6.0	-5.8	-5.6	-5.6	-5.6	-5.6	-5.5	-5.5	-6.1
12	491916	-7.0	-6.5	-6.4	-6.0	-6.0	-5.9	-5.9	-5.9	-5.9	-7.0
13	491857	-7.8	-7.6	-7.5	-7.3	-7.2	-7.2	-6.8	-6.7	-6.7	-7.8
14	491920	-7.1	-6.9	-6.7	-6.6	-6.4	-6.3	-6.3	-6.3	-6.0	-7.1
15	491917	-7.0	-6.9	-6.8	-6.6	-6.4	-6.2	-6.2	-6.2	-6.1	-7.0
16	57394621	-9.8	-8.9	-8.8	-8.3	-8.2	-8.1	-7.9	-7.6	-7.6	-9.8
17	57396377	-7.6	-7.3	-6.8	-6.6	-6.3	-6.2	-6.1	-6.1	-6.0	-7.6
18	57392897	-7.4	-6.6	-6.5	-6.4	-6.3	-6.1	-6.0	-6.0	-6.0	-7.4
19	57401703	-8.0	-7.8	-7.3	-6.9	-6.3	-6.1	-6.0	-6.0	-6.0	-8.0
20	482634	-7.2	-7.2	-7.1	-7.0	-7.0	-6.7	-6.7	-6.5	-6.4	-7.2
21	57392898	-6.5	-6.4	-6.2	-6.1	-6.1	-5.9	-5.9	-5.8	-5.7	-6.5
22	491877	-7.7	-7.1	-6.7	-6.7	-6.7	-6.7	-6.6	-6.6	-6.5	-7.7
23	53855501	-7.2	-7.2	-7.0	-6.7	-6.6	-6.2	-6.2	-5.9	-5.8	-7.2
24	57403346	-7.4	-7.3	-7.0	-7.0	-6.9	-6.9	-6.9	-6.9	-6.8	-7.4
25	57403432	-7.3	-7.1	-6.9	-6.8	-6.8	-6.8	-6.8	-6.7	-6.7	-7.3

*Num_modesarg (=9) maximum number of binding modes to generate.

The free energy of binding elicited at the vicinity of the active site by the ligands can be found in Tables 3, 4, 5.

Table 3 shows the affinity of Beta lactamase 1GA0 for further ligands, Table 4 the affinity of Nicotinamide phosphoribosyltransferase 4O0Z and Tables 5 of Nicotinamide phosphoribosyltransferase 4O10.

In Table 3, among the ligands, two possess the lowest binding energies: -9.8 and -8.1 kcal/mol while the highest is -6.1 kcal/mol. It seems that the high affinity ligand – enzyme depends of the number of rotatable bonds.

Table 4. The final lamarckian genetic algorithm docked state: binding energy of ligands with the active site of ***Nicotinamide phosphoribosyltransferase (4O0Z)*** during nine conformations

	CID of Ligand	1	2	3	4	5	6	7	8	9	Docked Energy (kcal/mol)
1	482634	-8.4	-7.9	-7.9	-7.8	-7.6	-7.0	-6.9	-6.9	-6.9	-8.4
2	10853428	-7.3	-7.2	-7.2	-7.0	-6.6	-6.6	-6.6	-6.3	-6.2	-7.3
3	57399922	-8.3	-7.8	-7.7	-7.6	-7.6	-7.6	-7.6	-7.5	-7.4	-8.3
4	491919	-9.4	-9.2	-8.9	-8.8	-8.8	-8.6	-8.6	-8.1	-7.9	-9.4
5	482635	-7.7	-7.7	-7.6	-7.5	-7.5	-7.4	-7.3	-7.3	-7.3	-7.7
6	10644682	-8.8	-7.7	-7.3	-7.3	-7.2	-7.1	-7.0	-6.9	-6.7	-8.8
7	10668699	-7.8	-7.8	-7.7	-7.6	-7.5	-7.3	-7.0	-6.9	-6.9	-7.8
8	491918	-5.2	-5.2	-4.6	-4.5	-3.6	-1.7				-5.2
9	491856	-8.8	-8.2	-8.2	-8.1	-7.9	-7.7	-7.7	-7.4	-7.3	-8.8
10	57394641	-8.1	-8.0	-8.0	-7.9	-7.9	-7.7	-7.7	-7.6	-7.5	-8.1
11	10066595	-7.6	-7.2	-7.2	-7.1	-6.7	-6.6	-6.5	-6.5	-6.3	-7.6
12	491916	-8.3	-8.0	-7.9	-7.4	-7.0	-6.9	-6.7	-6.6	-6.4	-8.3
13	491857	-8.8	-8.6	-8.5	-8.2	-8.1	-8.1	-7.9	-7.9	-7.9	-8.8
14	491920	-8.1	-8.0	-7.8	-7.7	-7.6	-7.4	-7.4	-7.1	-7.0	-8.1
15	491917	-8.4	-8.3	-8.3	-7.7	-7.6	-7.6	-7.6	-7.4	-7.4	-8.4
16	57394621	-9.7	-9.5	-8.9	-8.6	-8.6	-8.6	-8.6	-8.6	-8.6	-9.7
17	57396377	-7.9	-7.8	-7.8	-7.6	-7.5	-7.5	-7.5	-7.5	-7.2	-7.9
18	57392897	-7.7	-7.7	-7.4	-7.3	-7.2	-6.9	-6.8	-6.3	-6.3	-7.7
19	57401703	-8.0	-7.8	-7.8	-7.8	-7.8	-7.8	-7.7	-7.6	-7.5	-8.0
20	482634	-9.7	-9.7	-9.4	-9.1	-8.4	-8.4	-8.3	-8.2	-8.1	-9.7
21	57392898	-7.5	-7.0	-6.7	-6.7	-6.7	-6.6	-6.5	-6.4	-6.3	-7.5
22	491877	-8.1	-7.7	-7.5	-7.5	-7.3	-7.2	-6.4	-6.4	-6.4	8.1
23	53855501	-8.2	-8.1	-8.1	-7.9	-7.8	-7.2	-7.2	-6.9	-6.9	-8.2
24	57403346	-9.0	-8.9	-8.6	-8.4	-8.2	-8.0	-7.	-7.8	-7.8	-9.0
25	57403432	-8.9	-8.7	-8.7	-8.3	-8.3	-8.0	-8.0	-7.9	-7.7	-8.9

Among the ligands in Table 4, the lowest binding energies were: -9.7 and -9.4 kcal/mol while the highest, -5.2 kcal/mol. In case of ligand with CID 491918, only 6 conformations were found.

Table 5. The final lamarckian genetic algorithm docked state – Binding energy of ligands with the active site of ***Nicotinamide phosphoribosyltransferase (4O10)*** during nine conformations

	CID of Ligand	1	2	3	4	5	6	7	8	9	Docked Energy (kcal/mol)
1	482634	-5.8	-5.7	-5.5	-5.4	-5.4	-5.4	-5.3	-5.3	-5.3	-5.8
2	10853428	-5.8	-5.4	-5.3	-5.3	-5.2	-5.1	-5.1	-4.9	-4.9	-5.8
3	57399922	-6.3	-6.2	-6.1	-6.0	-5.9	-5.8	-5.7	-5.7	-5.7	-6.3
4	491919	-6.3	-6.1	-6.1	-5.7	-5.6	-5.4	-5.3	-5.1	-5.0	-6.3
5	482635	-6.5	-6.4	-6.4	-6.3	-6.0	-5.9	-5.9	-5.8	-5.8	-6.5
6	10644682	-6.0	-5.7	-5.6	-5.3	-5.3	-5.3	-5.2	-5.2	-5.2	-6.0
7	10668699	-6.8	-6.3	-6.2	-6.0	-6.0	-5.8	-5.7	-5.7	-5.6	-6.8
8	491918	-4.1	-3.9	-3.8	-3.7	-3.7	-3.4	-2.7	-1.4	-1.2	-4.1
9	491856	-6.4	-6.2	-5.9	-5.8	-5.8	-5.8	-5.6	-5.5	-5.5	-6.4
10	57394641	-6.5	-6.2	-6.2	-6.0	-5.9	-5.9	-5.9	-5.8	-5.8	-6.5
11	10066595	-5.5	-5.4	-5.4	-5.4	-5.3	-5.1	-5.0	-4.9	-4.8	-5.5
12	491916	-5.5	-5.5	-5.3	-5.3	-5.2	-5.1	-5.1	-5.1	-5.1	-5.5
13	491857	-6.6	-6.4	-6.3	-6.3	-6.0	-5.9	-5.8	-5.8	-5.8	-6.6
14	491920	-5.9	-5.7	-5.7	-5.7	-5.7	-5.5	-5.4	-5.4	-5.4	-5.9
15	491917	-5.8	-5.5	-5.5	-5.4	-5.3	-5.3	-5.2	-5.2	-5.2	-5.8
16	57394621	-7.7	-7.3	-7.2	-7.2	-7.1	-7.1	-7.0	-6.9	-6.7	-7.7
17	57396377	-6.4	-6.4	-6.0	-5.9	-5.6	-5.6	-5.6	-5.5	-5.5	-6.4
18	57392897	-5.2	-5.2	-5.1	-5.1	-5.0	-4.9	-4.7	-4.7	-4.6	-5.2
19	57401703	-6.0	-5.9	-5.3	-5.3	-5.2	-5.2	-5.1	-4.9	-4.9	-6.0
20	482634	-6.4	-6.1	-6.1	-5.9	-5.8	-5.8	-5.7	-5.7	-5.4	-6.4
21	57392898	-6.3	-6.0	-5.8	-5.7	-5.4	5.4	-5.3	-5.3	-5.3	-6.3
22	491877	-6.6	-6.4	-6.3	-6.1	-5.8	-5.6	-5.6	-5.6	-5.5	-6.6
23	53855501	-5.8	-5.5	-5.5	-5.4	-5.4	-5.4	-5.4	-5.3	-5.3	-5.8
24	57403346	-6.4	-6.3	-6.1	-6.0	-5.9	-5.9	-5.8	-5.7	-5.7	-6.4
25	57403432	-6.4	-5.9	-5.8	-5.7	-5.7	-5.6	-5.6	-5.6	-5.6	-6.4

The first active site of 4O10 strongly interacts with the ligands, and the energies vary from-7.7and -6.8kcal/ mol to -4.1 kcal/ mol. The highest affinity was -7.7kcal/ mol (for theligandCID57394621) while the smallest energy was -4.1kcal/mol (for the ligand CID491918). For the other ligands, the interaction was quite similar to each other.

COMPUTATIONAL DETAILS

The structures have been optimized at Hartree-Fock HF (6-31g(d,p)) level of theory, by Gaussian 09 [11]. Topological indices have beed computed by TOPOCLUJ software [12]; some of them (Conectivity= C, CS[Sh[Charges]]= Ch,SD), are listed in Table 6.

Table 6. Topological indices computed for the indolizine in case IC50 Table 1

Mol.	IC_{50}(μM)	SD	C.	Ch.
482634	30	-19.374	38	23
10853428	33	-17.745	40	25
57399922	31	-16.665	49	30
57403432	46	-4.228	52	33
491857	27	-22.428	39	24
491919	27	-25.495	49	30
482635	28	-21.307	48	29
491918	29	-22.428	41	25
491920	23	-27.119	51	31
491917	26	-22.428	40	25
57394641	30	-17.874	38	23
482634	30	-22.428	40	25
482634	27	-22.374	40	25
57403346	28	-24.043	46	29
57394621	24	-26.336	54	33
10066595	31	-18.45	43	27
10644682	20	-28.727	50	31
10668699	20	-29.375	50	31
491856	23	-22.428	42	26
491916	26	-22.423	41	26
57392897	33	-17.874	40	25
491877	21	-25.898	51	32
53855501	22	-24.043	45	28
57396377	25	-24.332	47	30
57401703	37	-11.519	51	32

QSAR models (for IC50)

The models were performed on the training set (the 15 structures in Table 1) and the best results are listed below and in Table 7 [13].

(i) Monovariate regression

$$IC50 = 48.475 + 0.923 \times SD$$

(ii) Bivariate regression

$$IC50 = 47.767 + 0.911 \times SD + 3.129 \times Ch$$

(iii)Three-variate regression

$$IC50 = 47.354 + 0.927 \times SD - 0.105 \times HOMO + 2.688 \times Ch$$

Table 7. Best models in describing IC50 in the training set of indolizines

	Descriptors	R^2	Adjust. R^2	St. Error	F
1	**SD**	**0.923**	0.915	1.530	155.683
2	**Ch**	0.047	0.027	5.381	0.635
3	**IP_{max}**	0.119	0.051	5.174	1.750
4	**IE_{max}**	0.079	0.008	5.288	1.121
5	**SD, Ch**	**0.935**	0.924	1.459	86.696
6	**SD, IE_{max}**	0.928	0.916	1.537	77.556
7	**SD, Adj**	0.926	0.914	1.560	75.170
8	**SD, C**	0.925	0.913	1.569	74.223
9	**SD, D3D**	0.925	0.912	1.572	73.859
10	**SD, HOMO**	0.924	0.911	1.580	73.119
11	**SD, C, Ch**	**0.945**	0.930	1.406	62.929
12	**SD, Adj, C**	0.939	0.922	1.483	56.176
13	**SD, De, IE_{max}**	0.934	0.915	1.544	51.558
14	**SD, D3D, De**	0.927	0.907	1.621	46.399
15	**SD, HOMO, Adj**	0.927	0.907	1.622	46.338
16	**SD, C, D3D**	0.925	0.905	1.639	45.361

Leave-one-out

The performances in leave-one-out analysis related to the models listed as best in Table 7 are presented in Table 8 [14].

Table 8. Leave-one-out analysis for best IC50 models

	Descriptors	Q^2	R^2-Q^2	St. $Error_{loo}$	F_{loo}
1	**SD**	**0.863**	0.06	2.038	82.057
5	**SD, Ch**	**0.880**	0.055	1.908	95.519
11	**SD, C, Ch**	**0.894**	0.051	1.792	109.920

External Validation

The values IC50 for the test set of indolizines (10 molecules) molecules best scored in the docking step, were calculated by using the best equation in Table 7, entry 5.Data are listed in Table 9 and the monovariate correlation: $IC50 = 0.781 \times IC50_{calc.} 7.418$; n=10; R^2=0.897; s=2.027; F=69.333is plotted in Figure 3 [15].

Table 9. Calculated values of IC50 for the molecules in the test set (Table 1)

Molecules	IC50	$IC50_{calc.}$
10066595	31	30.49
10644682	20	21.78
10668699	20	20.67
491856	23	27.78
491916	26	27.69
57392897	33	32.05
491877	21	23.94
53855501	22	27.06
57396377	25	27.04
57401703	37	37.19

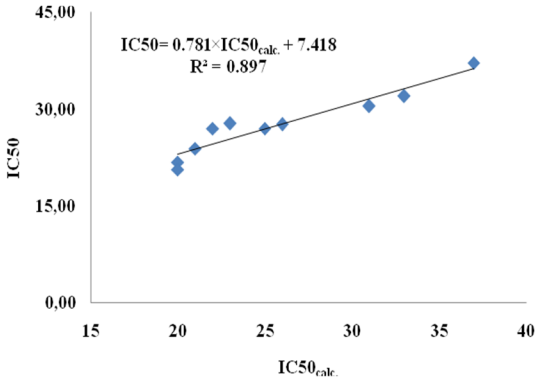


Figure 3. The plot IC50 vs. $IC50_{calc.}$ for the test set (external validation)

Similarity Cluster Validation

Validation can be performed by calculating IC50 for the molecules in the test set with equations learned on clusters of similarity: each of the 10 molecules is the leader in its own cluster best scored in the docking step, selected by (2D) similarity among the 15 structures of the initial learning set [16]. The values $IC50_{calc.}$for each of the 10 molecules in the test set were computed by 10 new equations (the leader being left out) with the same descriptors as in eq. 11, Table 7. Data are listed in Table 10 and the monovariate correlation: n=10; R^2=0.955; s=1.329; F=171.735 is plotted in Figure 4 [17].

Table 10. Calculated values of IC50 by similarity clusters, for the molecules in the test set

Molecules	IC50	$IC50_{calc.}$
10066595	31	30.47
10644682	20	22.59
10668699	20	21.75
491856	23	26.63
491916	26	27.43
57392897	33	32.81
491877	21	23.91
53855501	22	26.12
57396377	25	27.14
57401703	37	34.76

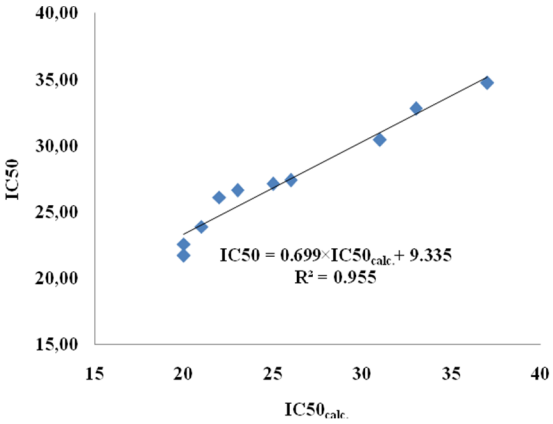


Figure 4. The plot IC50 vs. $IC50_{calc.}$ by similarity clusters

CONCLUSIONS

A set of 25indolizine, downloaded from the PubChem database, was submitted to a QSAR study. The set was split into a learning set (15 molecules) and a test set (10 molecules), the test set being used for the validation of the models.

Beta lactamase and nicotinamide phosphoribosyltransferase, has been investigated for its potential binding affinity with selective indolizine derivatives. The docking result of the study of 25 molecules demonstrated that the binding energies when we use Beta lactamase were in the range of -9.8 kcal/mol to -6.1 kcal/mol, when use nicotinamide phosphoribosyltransferase were in the range of -9.7 kcal/mol to -5.2 kcal/mol.

ACKNOWLEDGMENTS

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 postoctoral programs - support for increasing research competitiveness in the field of exact Sciences”.

REFERENCES

[1]. G.S. Singh, E.E. Mmatli, *Eur. J. Med. Chem*., **2011**, *46*, 5237-5257.
[2]. T. Przewloka, S. Chen, Zh. Xia, H. Li, Sh. Zhang,D. Chimmanamada, E. Kostik, D. James, K. Koya, L. Sun, *Tetrahedron Letters*, **2007**, *48*, 5739–5742.
[3]. I.V. Seregin, A.W. Schammel, V. Gevorgyan, *Tetrahedron,* **2008**, *64*, 6876–6883.
[4]. S.S. Juang, M. Chang, L.F. Wang, J.L. Han, C.H. Ong, *Tetrahedron*, **2005**, *61*, 1693-1697.
[5]. PubChem database, accessed 10.05. **2014**.
[6]. G.S. Singh, E.E. Mmatli, *European Journal of Medicinal Chemistry*, **2011**, 465237-5257.
[7]. H.M. Berman, K. Henrick, H. Nakamura, Announcing the worldwide Protein Data Bank Nature Structural Biology, **2003**, *10*, 12, 98.
[8]. O. Trott, A.J. Olson, *Journal of Computational Chemistry*, **2010**, 31, 455-461.
[9]. K. Dhananjayan, K. Kalathil, A. Sumathy, P. Sivanandy, *Der Pharma Chemica,* **2014**, *6*, 2, 378-387.
[10]. R. Abagyan, M. Totrov, *Curr Opin Chem Biol*., **2001**, *5*: 375-82.

[11]. Gaussian 09, Gaussian Inc C.T. Wallingford, Revision A.1 M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A.Montgomery, J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd,E. Brothers, K.N. Kudin, V.N. Staroverov, Kobayashi R., Normand J., Raghavachari K., Rendell A., Burant J.C., Iyengar S.S., Tomasi J., Cossi M., Rega N., N.J. Millam,M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Ö. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, **2009**.
[12]. O. Ursu, M.V. Diudea, TopoCluj software program. Babes-Bolyai University Cluj, **2005**.
[13]. J.G. Topliss, R. J. Costello, *J. Med. Chem.* **1972,** 15, 1066-1068.
[14]. L. Jäntschi, LOO Analysis (LOO: leave one out), Academic Direct Library of software, 2005. Available at: *http://l.academicdirect.org/Chemistry/SARs/MDF_SARs/loo/*
[15]. T.E. Harsa, A.M. Harsa, B. Szefler, *Cent. Eur. J. Chem.*, **2014**, *12*, 365-376.
[16]. S.D. Bolboacă, L. Jäntschi, and M.V. Diudea, *Current Computer-Aided Drug Design*, **2013**, *9*, 2, 195-205.
[17]. A.M. Harsa, T.E. Harsa, S.D. Bolboacă, M.V. Diudea, *Current Computer-Aided Drug Design*, **2014**, *2*, 10, 115-128.