Commun. Comput. Chem. Vol.3, No.3, pp.75-102
doi: 10.4208/cicc.2015.v3.n3.2 December 2015

REGULAR ARTICLE

QSAR Study and Molecular Docking of 23-hydroxybetulinic Acid Derivatives as RMGPa and HeLa Cells Inhibitors

Bhaskar Bagchi[1], Shyamal Sharma[1], Abhik Chatterjee[1], Pranab Ghosh[2], and Asim Kumar Bothra[1*]

[1] *Cheminformatics Bioinformatics Lab,Department of Chemistry, Raiganj College (University College) P.O.-Raiganj, Dist. - Uttar Dinajpur, PIN-733134, INDIA*

[2] *Department of Chemistry University of North Bengal*

Received 22 Aug 2015; Accepted (in revised version) 20 Nov 2015

Abstract: A number of 23-hydroxybetulinic acid derivatives were found to be potent glycogen phosphorylase a (GPa) inhibitors. Some derivatives of these triterpenes exhibit anti-tumor activities against a variety of tumor cell lines. In this study, we have constructed two different sets of QSAR equations. One set of QSAR equations predicts inhibitory activity of rabbit muscle glycogen phosphorylase a (RMGPa), which shares considerable sequence similarity with human liver GPa. The other set of equations predicts the antiproliferative activities against HeLa cells. This QSAR study has shown that topological indices and quantum chemical descriptors are the important parameters for determining the activity of 23-hydroxybetulinic acid derivatives. We have also performed Docking study with a number of 23-hydroxybetulinic acid derivatives with RMGPa and it has been found that the important interacting amino acids present in the active site cavity are ILE68, GLN71, GLN72, TYR75, ARG81, TYR155, ARG193, ARG242, ARG310 and SER313. Most of the compounds can form hydrogen bonds with ARG193 and/or ARG310. The unfavorable steric clashes between ligand and the protein and decrease of number of hydrogen bonds lower the inhibitory activity of the ligand.

* Corresponding author. *Email address:* asimbothra@gmail.com Telephone No. +919474441570
Fax No.-+913523242580

AMS subject classifications: 62J05, 92E10, 93A30
Key words: QSAR, Docking, RMGPa, HeLa, 23-hydroxybetulinic acid

Introduction

The breakdown of glycogen into glucose is mediated by glycogen phosphorylase (GP) with the help of a debranching enzyme and plays an important role for controlling hepatic glucose production [1]. GP has three isoforms which are brain, liver, and muscle according to their expression patterns. The muscle isoform supplies energy for muscle contraction. The brain isoform provides glucose during the periods of severe hypoglycemia. In glycogenolysis, liver enzyme plays a rate limiting role and hence it is an attractive target for the treatment of type 2 diabetes [2-5]. GP exists in two interconvertible forms: the phosphorylated high activity glycogen phosphorylase a (GPa) and the dephosphorylated low activity glycogen phosphorylase b (GPb). Allosteric effectors can promote equilibrium between a less active GPb and an active GPa. The active conformation is stabilized by phosphorylation of Ser 14 and binding of AMP [6, 7]. GP contains at least six regulatory sites: glucose analogues at the catalytic site, azasugar inhibitors, lactones at the allosteric site (AMP), caffeine at the purine inhibitor site, indole-2-carboxamide at the indole binding site and cyclodextrins at the glycogen storage site [8,9]. The X-ray analysis indicates that pentacyclic triterpenes bind at the allosteric site [10].

A number of 23-hydroxybetulinic acid derivatives were reported as potent GPa inhibitors [11]. Furthermore, some derivatives of these triterpenes display anti-tumor activities against a variety of tumor cell lines and the mechanism of action may be associated to its effects on the proliferation, migration, cell cycle and apoptosis of tumor cells [12, 13].

Apoptosis or programmed cell death plays an important role in regulating development and homeostasis. In many diseases including cancer, apoptosis is suppressed. Telomerase, a ribonucleoprotein, maintains chromosome lengths by adding telomeres to the chromosome ends repeatedly. Telomerase activation is found in about 90% of tumor tissues, but with very low, almost undetectable activity in somatic cells. Apoptosis is regulated by a number of cellular genes including B cell leukemia/lymphoma 2 (bcl-2). The stable expression of bcl-2 in human cancer cells increases telomerase activity and resistance to apoptosis. The 23-hydroxybetulinic acid induces apoptosis through concurrent inhibition of bcl-2 expression and telomerase activity [14- 17].

Furthermore, 23-hydroxybetulinic acid enhances anti-tumor activity of doxorubicin in vitro and in vivo. The synergism is associated with increase in the doxorubicin concentration in tumor tissue brought about by 23-hydroxybetulinic acid. Hence 23-hydroxybetulinic acid has the prospect to be developed as a novel chemosensitizer [18].

In this communication, we have constructed two different sets of QSAR equations. One set of QSAR equations predicts inhibitory activity of rabbit muscle glycogen phosphorylase a (RMGPa), which shares considerable sequence similarity with human liver GPa [11]. The other set of equations predicts the antiproliferative activities against HeLa cells. We have also performed Docking study with a number of 23-hydroxybetulinic acid derivatives with RMGPa.

Materials and methods

The development of QSAR models are described in the following section as: dataset preparation, parameters and statistical methods.

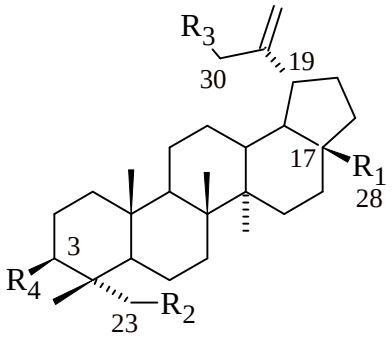
Dataset and parameters

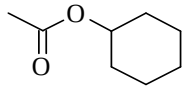
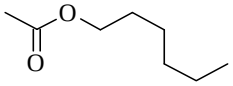
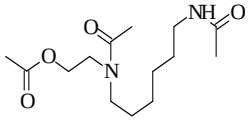
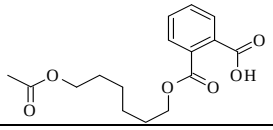
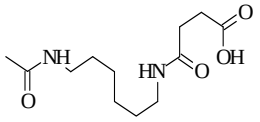
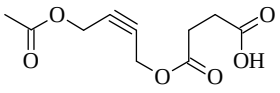
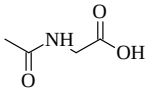
The dataset which is selected from the literature [11, 12] contains 47 compounds of 23-hydroxybetulinic acid derivatives. The biological property of this data set is reported as IC_{50} values. The IC_{50} values were converted into log IC_{50} and taken as the response variable for QSAR modeling. Structural details of the 47 compounds and their biological activity are listed in **Table 1**.

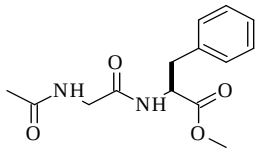
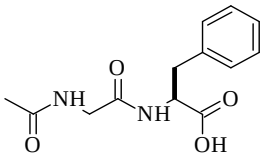
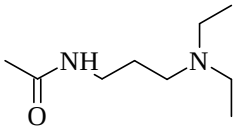
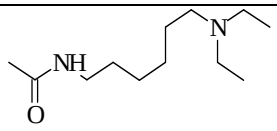
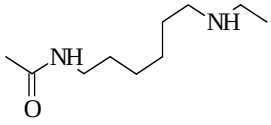
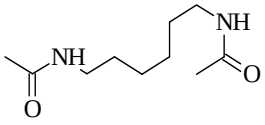
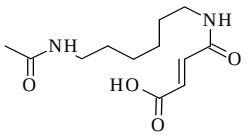
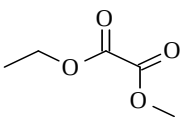
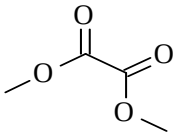
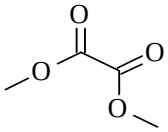
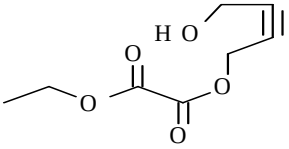
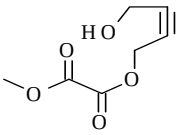
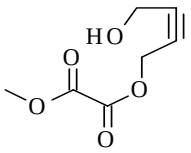
The initial structures of 47 compounds used in this study were constructed by ChemSketch [19]. We attempted several descriptors (data not shown) and it has been found that quantum chemical descriptors (EH, EL, μ), molar refractivity (MR), molar volume (MV) and topological indices such as Structural Information Content (SIC), Complementary Information Content (CIC), Wiener index (W), Harary index (H), Randiac's connectivity index of first order ($\chi 0$) [20-26] can better represent the biological activity of compounds.

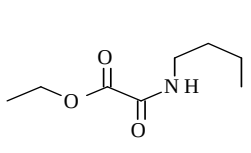
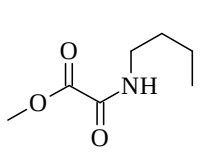
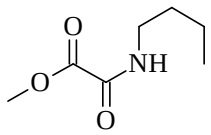
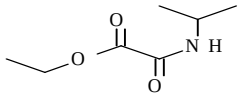
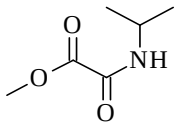
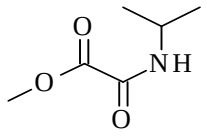
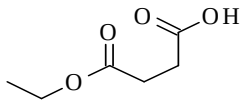
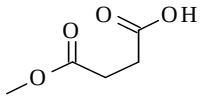
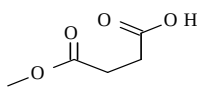
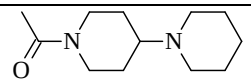
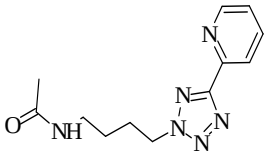
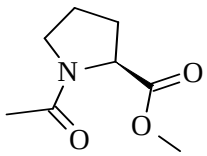
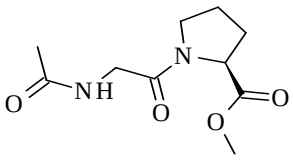
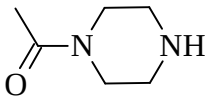
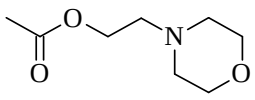
The quantum chemical properties (EH, EL, μ) of the studied molecules have been determined by DFT/B3LYP calculation and the basis set 6-31G* was used. All quantum chemical calculations were performed by Gamess [27]. Molar refractivity (MR) and molar volume (MV) were determined using ChemSketch software. The graph theoretical descriptors such as SIC, CIC, W, H, $\chi 0$ were computed using program written by us in Fortran-77.

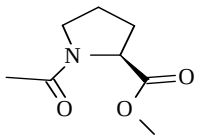
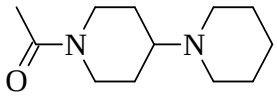
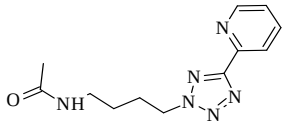
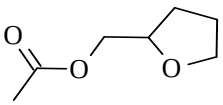
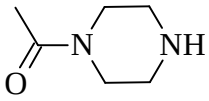
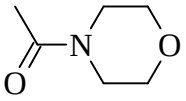
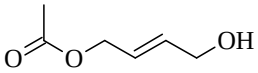
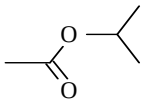
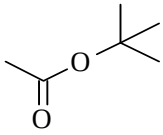
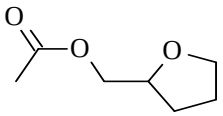
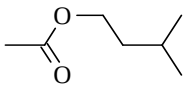
Table 1 Structures and activity of 47 compounds of 23-hydroxybetulinic acid derivatives

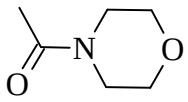
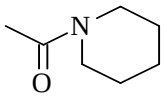
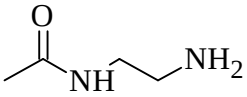
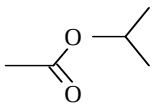
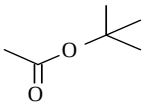
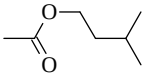
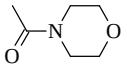
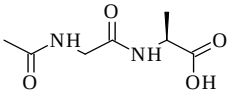
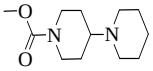
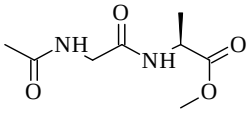


Comp. No.	-R1	-R2	-R3	-R4	IC_{50} (μM) RMGPa	IC_{50} (μM) HeLa
1	-COOH	-OH	-H	-OH	103	46.22
2		-OH	-H	-OH	50.4	-
3		-OH	-H	-OH	69.1	-
4		-OAc	-H	-OAc	96.6	-
5		-OAc	-H	-OAc	69.2	-
6		-OAc	-H	-OAc	194	-
7		-OAc	-H	-OAc	94.5	-
8		-OH	-H	-OH	67.7	-

9		-OH	-H	-OH	40.6	-
10		-OH	-H	-OH	129	-
11		-OAc	-H	-OAc	72.8	-
12		-OAc	-H	-OAc	55.6	-
13		-OAc	-H	-OAc	29.5	-
14		-OAc	-H	-OAc	34.1	-
15		-OAc	-H	-OAc	54.8	-
16			-H		30.1	-
17			-H		19.5	-

18			-H		34.7	-
19			-H		35.9	-
20			-H		97.8	-
21		-OAc	-H	-OH	-	10.80
22		-OAc	-H	-OH	-	17.87
23		-OAc	-OH	-OH	-	13.08
24		-OH	-H	-OH	-	4.84
25		-OAc	-H	-OH	-	18.84
26		-OAc	-H	-OH	-	9.18

27		-OAc	-H	-OH	-	9.78
28		-OH	-H	-OH	-	8.42
29		-OH	-H	-OH	-	46.26
30		-OH	-OH	-OH	-	79.35
31		-OH	-H	-OH	-	7.12
32		-OH	-H	-OH	-	34.29
33		-OH	-H	-OH	-	44.44
34		-OAc	-OH	-OH	-	19.46
35		-OAc	-OH	-OH	-	23.28
36		-OAc	-OH	-OH	-	17.79
37		-OAc	-OH	-OH	-	20.84

38		-OAc	-OH	-OH	-	18.31
39		-OAc	-OH	-OH	-	11.02
40		-OAc	-H	-OH	-	7.39
41		-OH	-OH	-OH	-	22.27
42		-OH	-OH	-OH	-	30.65
43		-OH	-OH	-OH	-	20.85
44		-OAc	-H	-OH	-	7.47
45		-OH	-H	-OH	-	22.10
46	-COOH		-H	-OH	-	12.52
47		-OH	-H	-OH	-	8.27

Statistical methods

Multiple linear regression (MLR) analysis was used to build up QSAR models. Different combinations of parameters were used to develop these models. Statistical qualities of MLR equations were judged by parameters like correlation coefficient (R), square of the correlation coefficient (R^2), cross validated coefficient (R^2_{cv}), standard deviation of the regression (S), Fischer statistics (F) and quality factor (Q). MLR program written by ourselves in Fortran-77 is used.

Molecular docking

The coordinates of RMGPa in complex with CHI (1LWO.pdb) [6] were obtained from the RCSB protein data bank (www.rcsb.org). The 23-hydroxy betulinic acid derivatives were docked into the active pocket of the enzyme by using docking program Autodock 4.0 [28-30]. Initially the structures of the ligands have been optimized with AM1 method and the hydrogen atoms were added to the enzyme. The Lamarckian genetic algorithm (LGA) was applied to look out for the best conformers. A grid map with 80x80x80 points and 0.375 Å spacing was used in Autogrid program to evaluate the binding energies between the inhibitors and RMGPa. The grid centre was set at the active site position 28.672, 0.621 and 53.033 and the default settings were used. For each compound ten docking poses were saved and ranked by binging energy. The lowest energy docking pose of 23-hydroxybetulinic acid was nearly identical to the Asiatic acid and Maslinic acid orientation in crystal structure with GPb [10]. So the pose with highest negative binding energy was selected for analyzing the type of interactions. The binding site was analyzed with molegro molecular viewer software [31].

Results and discussion

The data set of 47 compounds was divided into two groups. The first group of molecules contains 20 compounds having inhibitory activity of RMGPa and the second group of molecules contains 28 compounds with antiproliferative activities against HeLa cells. The 20 compounds of the first group were subdivided into two parts: 15 molecules in training set (1,2,3,4,5,7,8,10,11,12,15,16,18,19,20) and 5 molecules in test set (6,9,13,14 and 17). A cross correlation matrix (Table S1, supporting information) between descriptors and the $\log IC_{50}$ values of training set demonstrates that SIC_1 and EL have positive correlation with activity, CIC_1, μ, lnMR and lnMV shows negative correlation and EH does not show any significant correlation. SIC_1, CIC_1, molecular electronic properties, molar refractivity and molar volume of training compounds are summarized in **Table 2**.

Table 2 SIC_1, CIC_1, quantum chemical descriptors, molar refractivity and molar volume of 15 training RMGPa inhibitors

Comp no.	SIC_1	CIC_1	EH (hartree)	EL (hartree)	μ (debye)	MR (cm^3)	MV (cm^3)
1	0.4253	3.6537	-0.2310	0.0089	6.1915	134.70	427.30
2	0.3449	4.3331	-0.2290	-0.0230	3.1764	160.84	501.90
3	0.3411	4.3779	-0.2352	-0.0063	2.2178	162.70	535.20
4	0.3572	4.5137	-0.2289	-0.0299	6.5763	217.62	698.00
5	0.3760	4.3538	-0.2338	-0.0701	5.0497	219.55	680.80
7	0.3823	4.2048	-0.2330	-0.0518	1.8782	192.86	606.90
8	0.4090	3.8272	-0.2255	-0.0302	3.6113	147.59	466.30
10	0.4243	3.8966	-0.2271	-0.0451	1.8798	189.55	582.80
11	0.3602	4.3878	-0.2277	-0.0274	4.1665	192.49	618.80
12	0.3412	4.5893	-0.2160	-0.0229	7.8094	206.39	666.80
15	0.3898	4.2294	-0.2263	-0.0850	3.2123	207.84	651.80
16	0.3809	4.1733	-0.2332	-0.1259	5.0089	183.01	586.90
18	0.3658	4.5019	-0.2298	-0.0881	9.1787	230.42	735.10
19	0.3812	4.3319	-0.2340	-0.0781	8.6982	216.46	689.00
20	0.3800	4.2516	-0.2350	-0.0361	6.2042	196.29	608.70

Among the generated QSAR equations; three equations were finally selected for predicting the inhibitory activity of RMGPa. Three best models are given below:

Model 1

$\log IC_{50} = 0.8954+(2.8856)SIC_1+(3.5920)EL$

Where, N=15, R=0.773, R^2=0.598, R^2_{cv}=0.323, S=0.165, F=8.925, Q=4.685

Model 2

$\log IC_{50} = 3.4496+(4.1769)\ln MR+(-3.7611)\ln MV+(-3.7737)EH+(4.5525)EL+(-0.0256)\mu$

Where, N=15, R=0.864, R^2=0.746, R^2_{cv}=0.256, S=0.172, F=5.287, Q=5.023

Model 3

$\log IC_{50} = -8.4488+(12.1049)SIC_1+(1.1185)CIC_1+(-6.0974)EH+(5.3929)EL+(-0.0388)\mu$

Where, N=15, R=0.946, R^2=0.895, R^2_{cv}=0.735, S=0.181, F=15.343, Q=5.227

By using model number 1, 2 and 3 the theoretical log IC_{50} values of 15 training compounds are given in **Table 3** together with experimental log IC_{50}.

Table 3 List of experimental and predicted $logIC_{50}$ of 15 training RMGPa inhibitors

Comp no.	Experimental $logIC_{50}$	Predicted $logIC_{50}$ (By model 1)	Predicted $logIC_{50}$ (By model 2)	Predicted $logIC_{50}$ (by Model 3)
1	2.0128	2.1546	1.9002	2.0023
2	1.7024	1.8080	1.9600	1.7217
3	1.8395	1.8570	1.8905	1.8909
4	1.9850	1.8187	1.8639	1.9029
5	1.8401	1.7285	1.8688	1.8239
7	1.9754	1.8125	1.9211	1.9504
8	1.8306	1.9671	1.8205	1.8548
10	2.1106	1.9577	2.0098	2.1142
11	1.8621	1.8363	1.8723	1.8980
12	1.7451	1.7977	1.7660	1.7050
15	1.7388	1.7149	1.7548	1.7970
16	1.4786	1.5423	1.4111	1.3783
18	1.5403	1.6345	1.5798	1.5844
19	1.5551	1.7148	1.6357	1.6789
20	1.9903	1.8622	1.9520	1.9039

The model 3 with the R=0.946, R^2 =0.895, R^2_{cv}=0.735, S=0.181, F=15.343, Q=5.227 turns out to be the best fit model. The indices of the 5 test compounds are presented in Table S2, supporting information. Using the model number 3, we calculated the theoretical log IC_{50} values of the test set (R=0.891) are given in **Table 4**.

Table 4 List of experimental and predicted $logIC_{50}$ of 5 test RMGPa inhibitors

Comp no.	Experimental $logIC_{50}$	Predicted $logIC_{50}$ (by Model 3)
6	2.2878	2.1409
9	1.6085	2.0331
13	1.4698	1.8409
14	1.5328	1.9068
17	1.2900	1.7081

The correlation graph of training compounds and the correlation graph of test compounds between experimental log IC_{50} and predicted log IC_{50} (by model 3) are presented in **Figure 1** and **Figure 2** respectively.

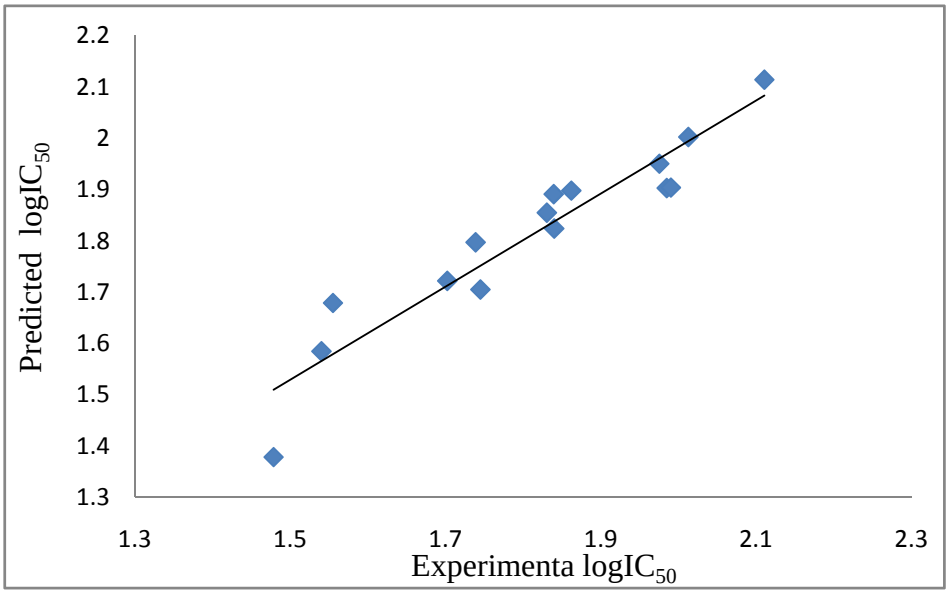


Figure 1: A plot between the predicted and the experimental activities for the training set of RMGPa inhibitors

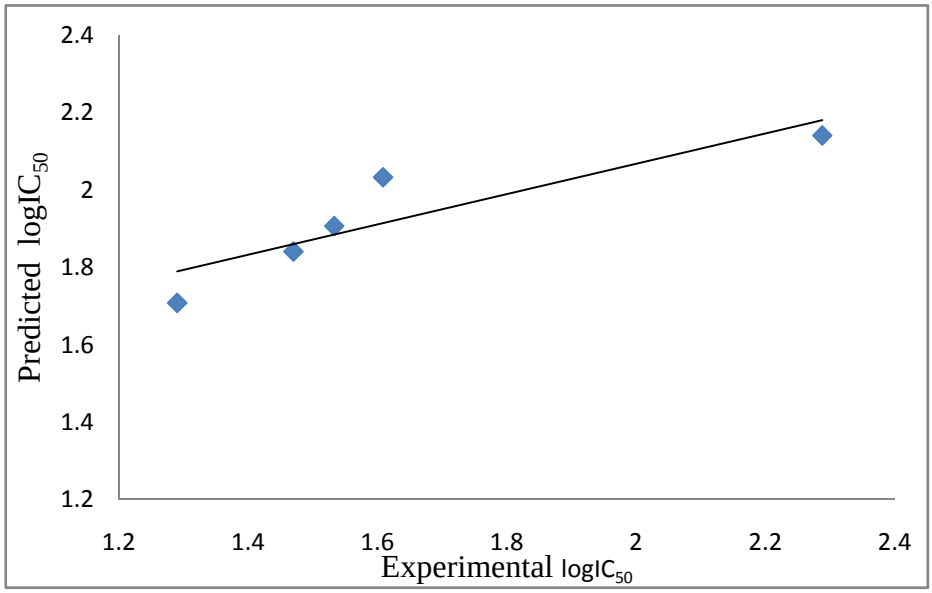


Figure 2: A plot between the predicted and the experimental activities for the test set of RMGPa inhibitors

The second group containing 28 compounds were divided into two parts: 20 molecules in training set (1,21,22,25,26,27,28,32,33,34,35,36,37,38,39,41,42,43,45,46) and test set of 8 molecules (23,24,29,30,31,40,44,47). The correlation matrix of electronic properties and topological indices with cytotoxic activity are presented in Table S3 and S4, supporting information, respectively. Molecular electronic properties, molar refractivity and molar volume of training compounds are summarized in **Table 5** and their topological indices are presented in **Table 6**.

Table 5 Quantum chemical descriptors, molar refractivity and molar volume of 20 training antiproliferative compounds against HeLa cells

Comp no.	EH (hartree)	EL (hartree)	μ (debye)	MR (cm^3)	MV (cm^3)
1	-0.2310	0.0089	6.1915	134.70	427.30
21	-0.1999	-0.0134	5.1567	191.96	610.10
22	-0.2045	-0.0428	1.6927	202.81	560.70
25	-0.2170	-0.0013	3.8836	166.02	534.90
26	-0.2074	-0.0010	6.0763	175.31	572.80
27	-0.1566	-0.0016	7.3747	173.63	558.80
28	-0.2060	0.0125	7.3125	182.45	571.10
32	-0.2179	-0.0252	3.4896	154.78	490.90
33	-0.1968	-0.0230	2.9099	150.40	476.80
34	-0.2225	0.0001	2.4967	159.80	522.60
35	-0.1948	-0.0245	3.2973	164.44	538.80
36	-0.1963	-0.0375	4.3380	168.64	545.00
37	-0.2111	-0.0179	5.4231	169.07	555.60
38	-0.1693	-0.0068	4.6924	165.82	527.40
39	-0.1731	-0.0013	1.2905	168.82	536.40
41	-0.1986	-0.0012	3.8606	150.30	483.60
42	-0.1978	-0.0255	3.3445	154.94	499.80
43	-0.2032	-0.0367	5.9112	159.56	516.60
45	-0.1892	-0.0141	4.5361	165.07	522.10
46	-0.2008	0.0058	7.5438	5.2403	6.3820

Table 6 Topological indices of 20 training antiproliferative compounds against HeLa cells

Comp no.	SIC_1	CIC_1	ln W	H	χ^0
1	0.4253	3.6537	7.9649	162.1636	24.8970
21	0.3517	4.4462	8.9289	251.9431	34.2775
22	0.4288	3.9028	9.0789	263.7028	37.1059
25	0.3872	4.0716	8.5865	214.7929	30.2943
26	0.3720	4.2252	8.7828	228.4780	32.4156
27	0.3826	4.1372	8.7686	235.9327	32.7419
28	0.3484	4.4273	8.7749	233.0269	31.9930
32	0.3794	4.0679	8.3622	195.8768	28.0099
33	0.3866	4.0112	8.4016	191.0617	28.4325
34	0.3739	4.1325	8.5254	5.3306	30.1730
35	0.3714	4.1765	8.5916	213.8546	31.0957
36	0.3743	4.1747	8.7221	225.2267	31.7085
37	0.3621	4.2655	8.6630	217.3454	31.5872
38	0.3919	4.0402	8.6484	221.8434	31.0014
39	0.3749	4.1707	8.6484	221.8434	31.0014
41	0.3756	4.0734	8.2890	187.6982	27.8885
42	0.3732	4.1177	8.3612	194.8093	28.8112
43	0.3630	4.2135	8.4670	198.6210	29.3028
45	0.4187	3.8536	8.6769	215.3581	31.7504
46	0.3665	4.3284	8.9253	245.4236	34.2775

Among the generated QSAR models; three models were finally selected. Model summary of three best models are given below.

Model 4

$\text{Log } IC_{50}$= 19.6957+(-4.4880)SIC_1+(-0.8479)CIC_1+(0.0427)lnW+(-3.0400)lnH+(0.0885)$\chi 0$

Where, N=28, R=0.859, R^2=0.738, R^2_{cv}=0.288, S=0.197, F=7.887, Q=4.360

Model 5

$\text{Log } IC_{50}$= 13.3115+(-0.1637)lnMR+(-1.8648)lnMV+(-2.2260)EH+(-4.4968)EL+(-0.0005)μ

Where, N=28, R=0.933, R^2=0.870, R^2_{cv}=0.594, S=0.205, F=18.738, Q=4.551

Model 6

Log IC_{50}= 13.4441+(-11.8969)SIC_1+(-2.0004)CIC_1+(-2.7582)EH+(-4.5601)EL+(-0.0010)μ
Where, N=28, R=0.916, R^2=0.839, R^2_{cv}=0.698, S=0.201, F=14.591, Q=4.557

Model 4 and model 5 are constructed by using topological indices and electronic parameters respectively. But the quality of the equation is much increased when we use both the topological and electronic indices simultaneously i.e. Model 6. By using model number 4, 5 and 6 the theoretical log IC_{50} values of 20 training compounds are given in **Table 7** together with experimental log IC_{50}. The indices of the 8 test compounds are given in Table S5, supporting information.

Table 7 List of experimental and predicted logIC_{50} of 20 training antiproliferative compounds against HeLa cells

Comp no.	Experimental logIC_{50}	Predicted logIC_{50} (By model 4)	Predicted logIC_{50} (by Model 5)	Predicted logIC_{50} (By model 6)
1	1.6648	1.7631	1.6839	1.6658
21	1.0334	0.9534	0.9934	0.9731
22	1.2521	1.1862	1.2860	1.2931
25	1.2751	1.2294	1.2467	1.2933
26	0.9628	1.1759	1.0864	1.1369
27	0.9903	1.1337	1.0231	1.0481
28	0.9253	1.0126	0.9836	0.9467
32	1.5352	1.3361	1.5281	1.5054
33	1.6478	1.4666	1.5305	1.4655
34	1.2891	1.3430	1.3031	1.3399
35	1.3670	1.2959	1.2901	1.3166
36	1.2502	1.1867	1.3258	1.3481
37	1.3189	1.2595	1.2338	1.2620
38	1.2627	1.2020	1.1914	1.1929
39	1.0422	1.1676	1.1423	1.1229
41	1.3477	1.4644	1.4096	1.3766
42	1.4864	1.4093	1.4510	1.4256
43	1.3191	1.3629	1.4456	1.4187
45	1.3444	1.3977	1.2881	1.3357
46	1.0976	1.0663	0.9696	0.9452

In these models, N is the number of data points; R is the correlation coefficient between experimental values and calculated values from the equation. R^2 is the square of the correlation coefficient and it measures the goodness of fit of the regression equation. Cross validated coefficient (R^2_{cv}) gives an idea of the performance of the model. S is the standard deviation of the regression. Fischer statistics (F) is a ratio of variances between calculated and observed activity. The larger value of F test signifies the QSAR model. Q is the quality factor. Q value measures predictive power of the QSAR models.

We calculated the theoretical log IC_{50} of the test set (R=0.751) by model number 6 which appeared in **Table 8**.

Table 8 List of experimental and predicted logIC_{50} of 8 test antiproliferative inhibitors against HeLa cells

Comp no.	Experimental logIC_{50}	Predicted logIC_{50} (by Model 6)
23	1.1166	1.2684
24	0.6848	1.1678
29	1.6652	1.3425
30	1.8995	1.4045
31	0.8525	1.3018
40	0.8686	1.2064
44	0.8733	1.3191
47	0.9175	1.3274

The correlation graph of training compounds and the correlation graph of test compounds between experimental log IC_{50} and predicted log IC_{50} (by model 6) are presented in **Figure 3** and **4** respectively.

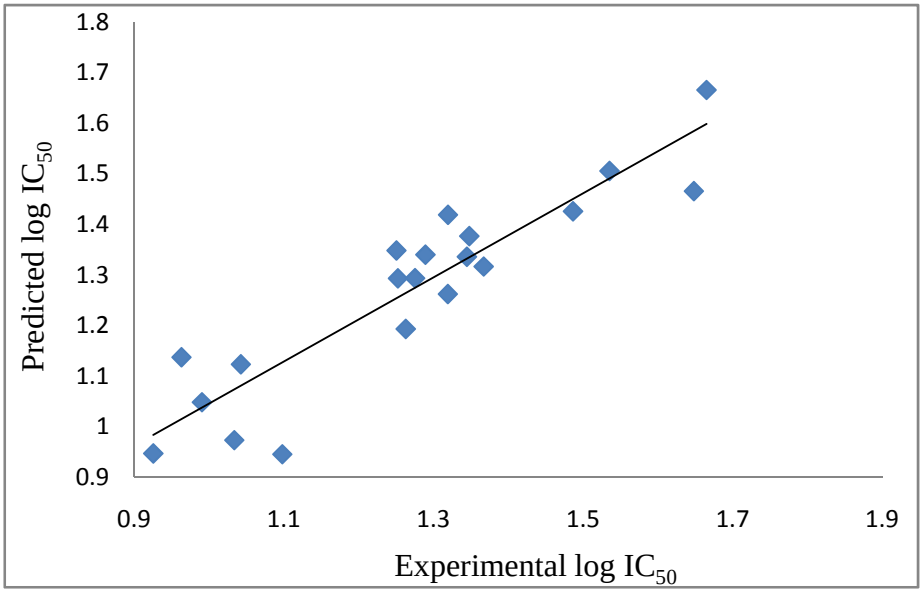


Figure 3: A plot between the predicted and the experimental activities for the training antiproliferative inhibitors against HeLa cells

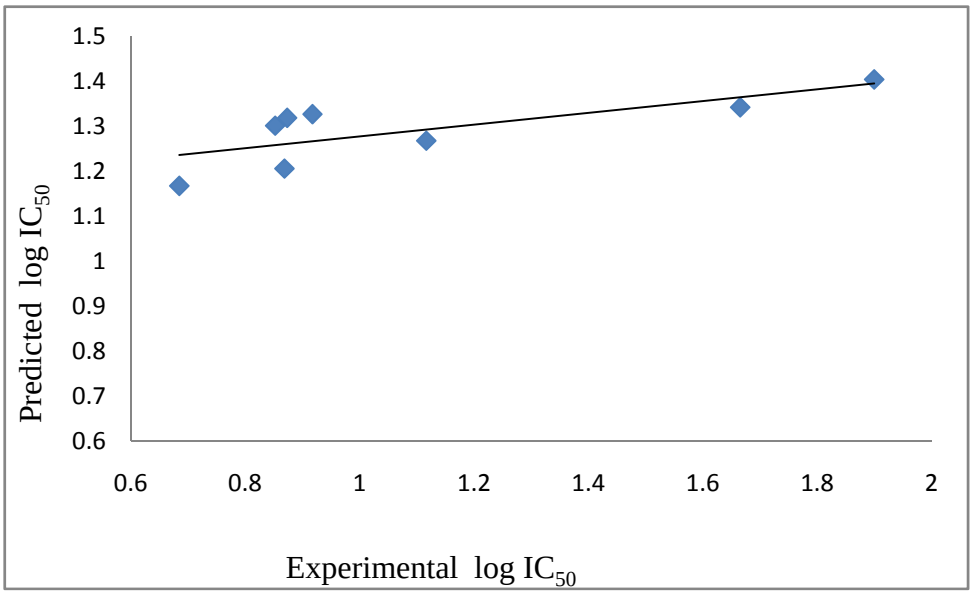


Figure 4: A plot between the predicted and the experimental activities for the test antiproliferative inhibitors against HeLa cells

We also calculated the antiproliferative activity (from compound number 2 to 20) and inhibitory activity of RMGPa (from compound number 21 to 47) by model number 6 and 3 respectively which were not determined experimentally and are shown in **Table 9**.

Table 9 Predicted Antiproliferative activity/ RMGPa inhibitory activity of studied compounds

Comp. No.	HeLa IC_{50}(μM)	RMGPa IC_{50}(μM)	Comp. No.	HeLa IC_{50}(μM)	RMGPa IC_{50}(μM)
2	25.48	-	25	-	90.72
3	20.12	-	26	-	63.61
4	8.44	-	27	-	29.51
5	16.63	-	28	-	57.57
6	11.34	-	29	-	86.53
7	22.99	-	30	-	43.87
8	47.69	-	31	-	89.11
9	26.87	-	32	-	56.37
10	27.02	-	33	-	47.91
11	13.50	-	34	-	91.14
12	7.88	-	35	-	44.26
13	12.88	-	36	-	37.83
14	12.61	-	37	-	48.47
15	22.63	-	38	-	42.42
16	59.79	-	39	-	56.57
17	36.50	-	40	-	39.14
18	12.97	-	41	-	51.11
19	17.23	-	42	-	41.03
20	16.79	-	43	-	29.49
21	-	53.46	44	-	43.88
22	-	114.14	45	-	67.81
23	-	67.81	46	-	61.92
24	-	68.55			

Docking

The binding energies of 47 docked compounds are given in Table S6, supporting information, and are ranges between -4.58 and -13.2kcal/mol. The docking study shows both polar (GLN7, ARG10, LYS11, SEP14, ARG69, GLN71, GLN72, TYP74, TYR75, GLU76, ARG81, TYR155, LYS191, ARG193, GLU195, THR240, ARG242, ARG306, ARG309, AGR310, LYS312, SER313, SER314) and non polar (ILE13, VAL15, LEU18, ILE63, VAL64, TRP67, ILE68, PHE196,

ASP227) amino acids make important interactions to the inhibitors.

In our study, we used different topological and quantum chemical indices to obtain phenogram based on Unweighted Pair Group Method with Arithmetic mean (UPGMA) of 47 compounds. The phenogram of the compounds is presented in **Figure 5**. Compounds having similar molecular properties are in the same clade though in some cases their binding energies and activity are different.

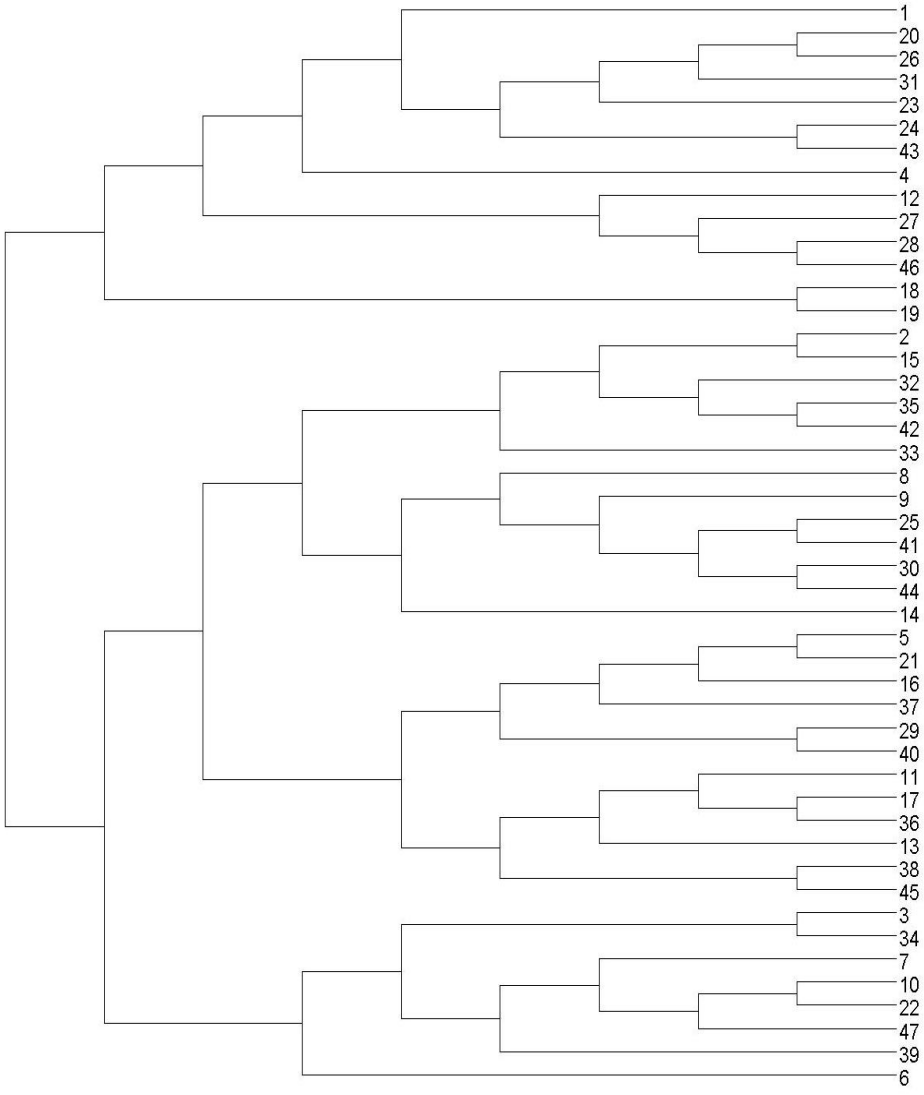


Figure 5: Phenogram using Unweighted Pair Group Method with Arithmetic Mean (UPGMA) of 47 compounds of 23-hydroxy betulinic acid derivatives

Compound 1, 20, 26, 31, 23, 24, 43, 4 are in same clade. Ligand 1 (23-hydroxy betulinic

acid) was used as a model drug (**Figure 6a**) as a reference for other compounds and have three hydrogen bonds. The carboxylate group at C-28 forms two hydrogen bonds at 1.743Å and 1.821Å with ARG310 and ARG 242 respectively. The –OH group at C-3 also forms another hydrogen bond with ILE68 (1.988 Å). In this clade, compound 20 and 26 remain as a pair and they have almost same binding energy. The binding energy of 20 and 26 are -7.38 and -7.56kcal/mol respectively. Compound 24 and 43 have different binding energies though they remain as a pair. Their predicted IC_{50} values are 68.50μM and 29.49μM respectively. In case of ligand 43 (**Figure 6b**), –OH group present at C-30 and the ester group at C-28 form two hydrogen bonds with SER313 (2.181 Å) and ARG310 (1.933 Å) respectively. Ligand 24 also forms two hydrogen bonds with ARG81 and GLN71 but at the same time the N-substituted amide group at C-28 may increase the chance of steric bumps with ARG309, which weakens the interaction between ligand and enzyme causing decrease in bioactivity (**Figure 6c**).

Compound 12, 27, 28, 46 are in same clade in which 46 has highest negative binding energy. In case of ligand 46 (**Figure 6d**), the carboxylate group at C-28 forms hydrogen bond with ARG193 (1.937Å). Compound 18 and 19 remain as a pair in the next clade and have comparable binding energy. Compound 2, 15, 32, 35, 42, 33 are present in same clade. Though compound 2 and 15 are in a pair, their binding energies are different. The IC_{50} values of ligand 2 and 15 are 50.4μM and 54.8μM respectively. Compound 2 lies well inside the protein (**Figure 6e**) and the –OH group at C-23 and C=O group at C-28 make two hydrogen bonds with GLN72 (2.675Å) and ARG310 (2.015Å) respectively. Although the amide group at C-28 of ligand 15 makes important interactions with the enzyme but the major part of the ligand is outside the protein (**Figure 6f**). Compound 35 and 42 are in a pair in same clade and they have almost same binding energy. In another clade, compound 8, 9, 25, 41, 30, 44, 14 are present and their binding energies are almost as same as the other.

Compound 5, 21, 16, 37, 29, 40 are in same clade. Compound 16 has highest negative binding energy than the other compounds and forms one hydrogen bond with ARG310 (**Figure 6g**). In the next clade compound 17 and 36 are in a pair. The binding energy of 36 is more negative than 17. In ligand 36, the –OH group at C-3 and O atom at C-23 form hydrogen bonds with ARG310 at 2.002Å and 2.223Å respectively (**Figure 6h**). The acetyl group at C-23 also forms hydrogen bond with ARG242 (1.87Å). Ligand 17 forms hydrogen bonds with ARG69, GLN71 and TYR155. In spite of the steric bump formation between the long chains at C-3 with ARG310, this compound possesses good inhibitory activity due to hydrogen bonds (**Figure 6i**). Compound 3, 34 and 10, 22 are in pairs and they have comparable binding energy.

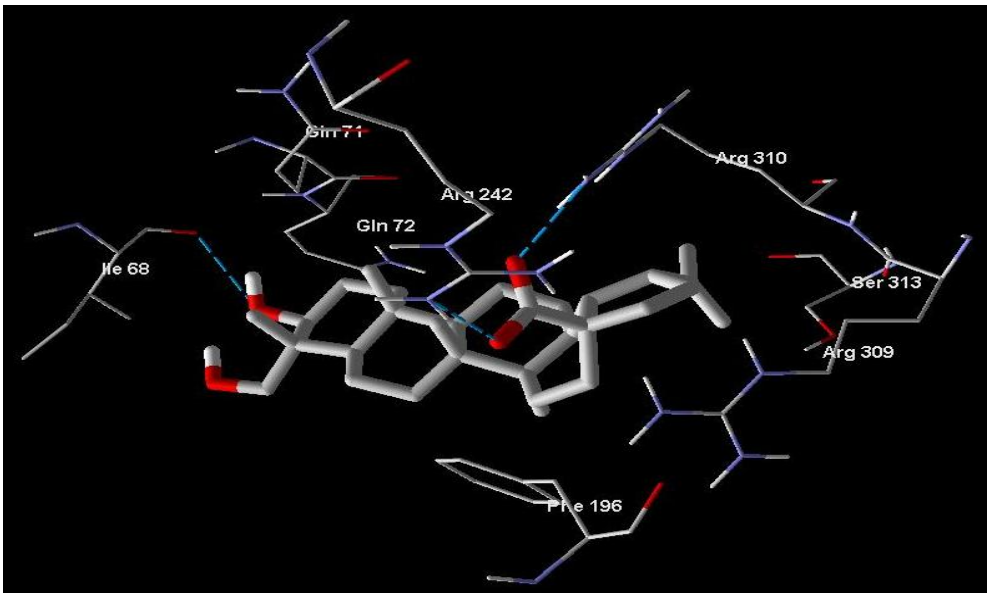


Figure 6a

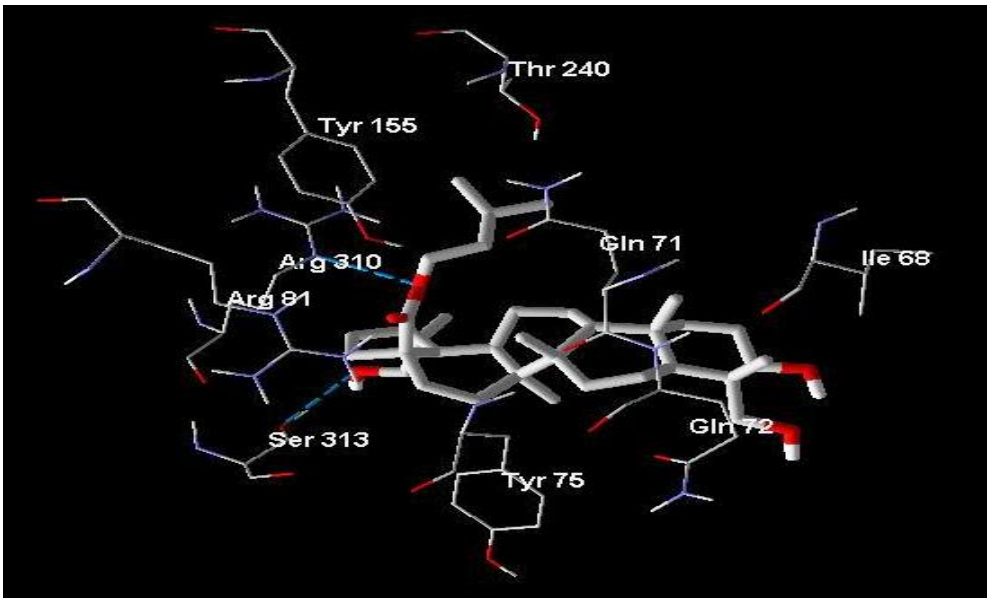


Figure 6b

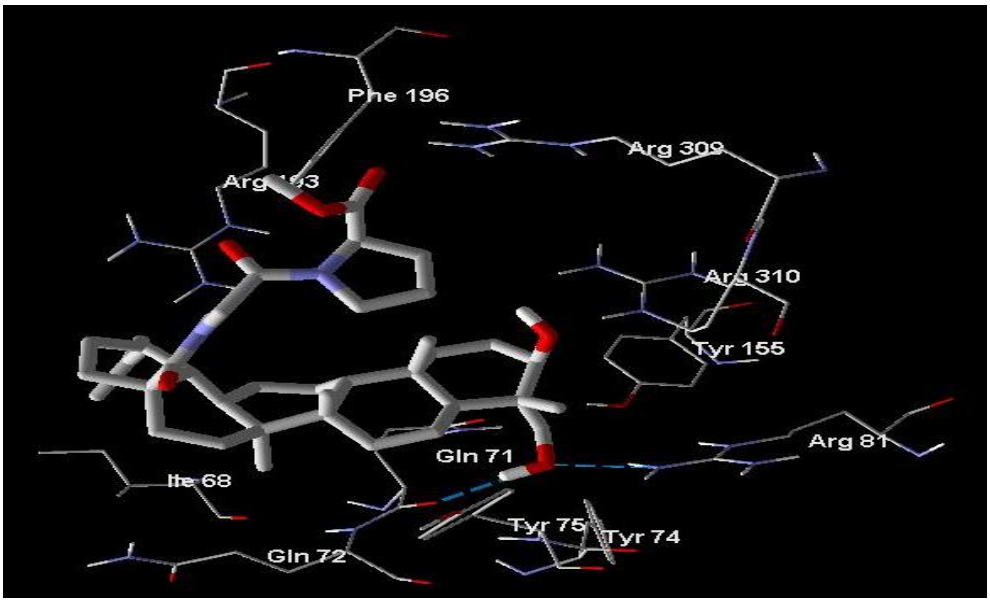


Figure 6c

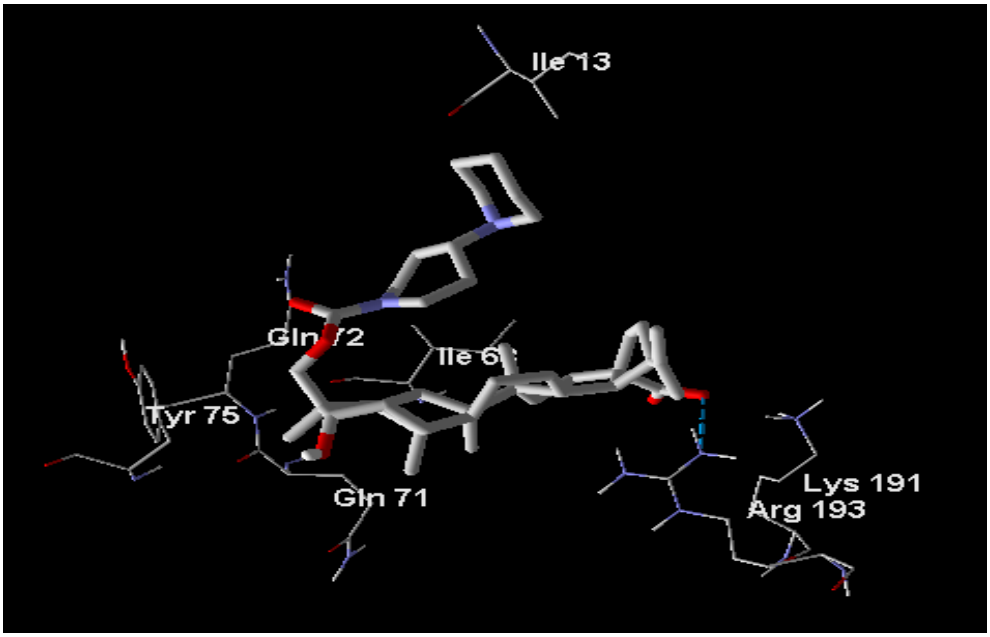


Figure 6d

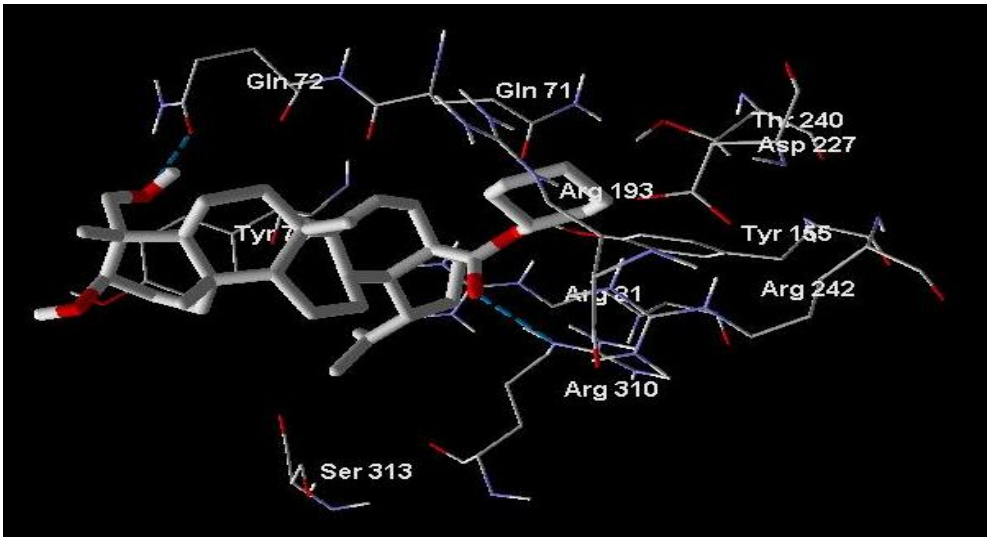


Figure 6e

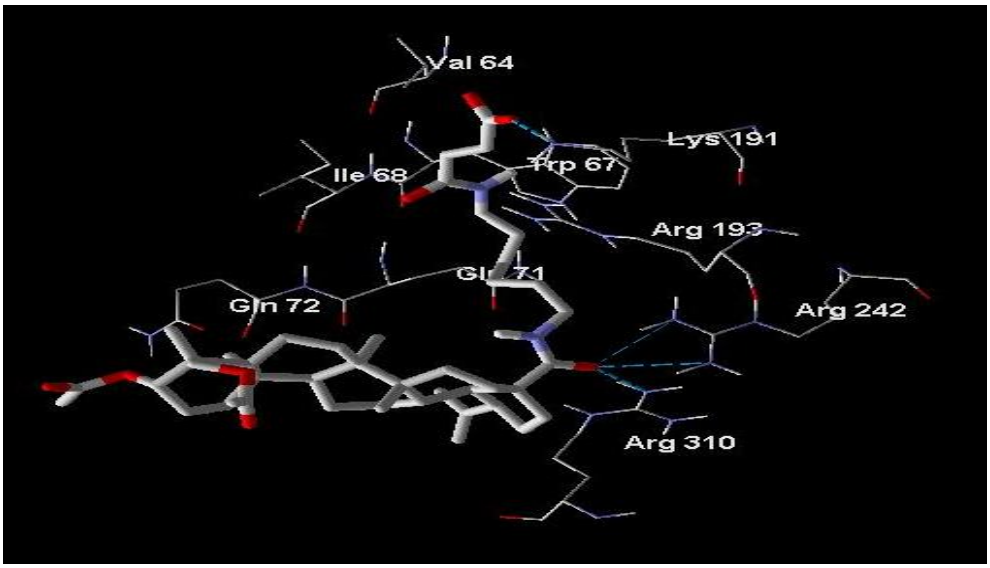


Figure 6f

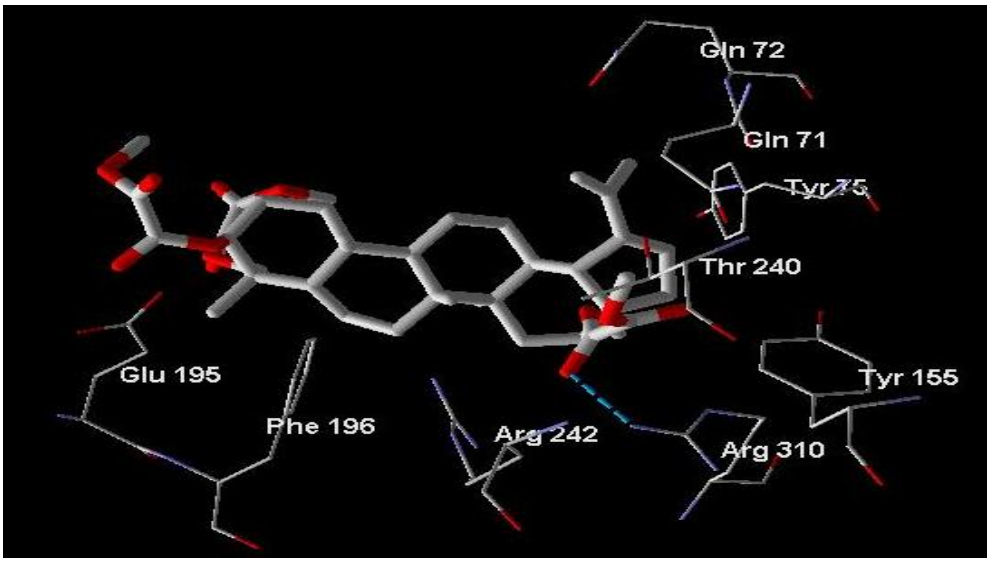


Figure 6g

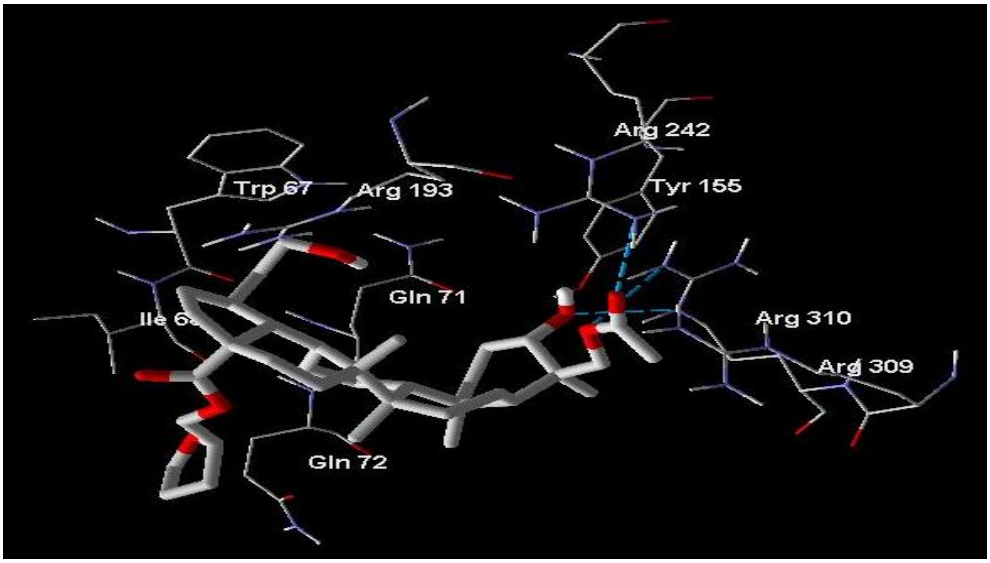


Figure 6h

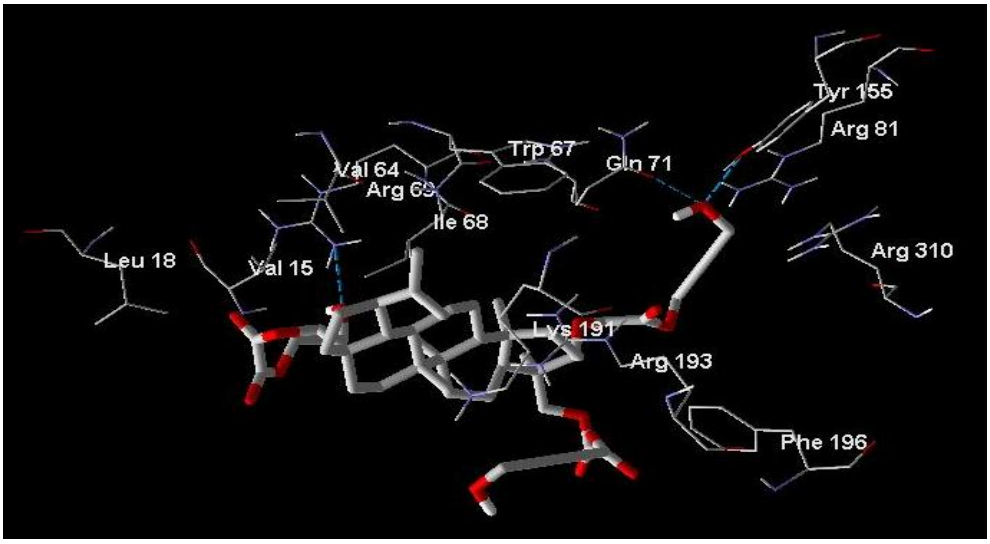


Figure 6i

Figure 6: Poses of different ligands in the active site of rabbit muscle glycogen phosphorylase a (RMGPa). (6a) Docked conformation of ligand 1 along with the important amino acid residues of RMGPa. (6b) Docked conformation of ligand 43 along with the important amino acid residues of RMGPa. (6c) Docked conformation of ligand 24 along with the important amino acid residues of RMGPa. (6d) Docked conformation of ligand 46 along with the important amino acid residues of RMGPa. (6e) Docked conformation of ligand 2 along with the important amino acid residues of RMGPa. (6f) Docked conformation of ligand 15 along with the important amino acid residues of RMGPa. (6g) Docked conformation of ligand 16 along with the important amino acid residues of RMGPa. (6h) Docked conformation of ligand 36 along with the important amino acid residues of RMGPa. (6i) Docked conformation of ligand 17 along with the important amino acid residues of RMGPa.

In conclusion, this QSAR study has shown that topological indices (e.g. SIC, CIC) and quantum chemical descriptors (e.g. EH, EL, μ) are the important parameters for determining the activity of 23-hydroxybetulinic acid derivatives. Model 3 and model 6 are the best equation for predicting the inhibitory activity of RMGPa and the antiproliferative activities against HeLa cells respectively and these QSAR models may be used in prediction of activity of designed compounds. The docking study shows that the important interacting amino acids present in the active site are ILE68, GLN71, GLN72, TYR75, ARG81, TYR155, ARG193, ARG242, ARG310 and SER313. Most of the ligands can form hydrogen bonds with ARG193 and/or ARG310. The –OH group at C-3 and C-23 can increase the hydrogen bond interaction

between ligands and enzyme. However, acetylation or esterification of the –OH group at the C-3 and C-23 not only decreases the number of hydrogen bond, but may also increase the unfavorable steric clashes. Thus binding energy may decrease. Large substituent at C-17 may increase the chance of steric bumps, thus lowering the inhibitory activity of the ligand.

References

[1] M. BOLLEN, S. KEPPENS AND W. STALMANS, Specific features of glycogen metabolism in the liver, Biochem J, 336 (1998), pp. 19-31.

[2] H. G. HERS, The Control of Glycogen Metabolism in the Liver, Annu Rev Biochem, 45 (1976), pp. 167-190.

[3] W. STALMANS, The Interaction of Liver Phosphorylase a with Glucose and AMP, Eur J Biochem, 49 (1974), pp. 415-427.

[4] W. PIMENTA, N. NURJHAN, P. A. JANSSON, M. STUMVOLL, J. GERICH AND M. KORYTKOWSKI, Glycogen: its mode of formation and contribution to hepatic glucose output in postabsorptive humans, Diabetologia, 37 (1994), pp. 697-702.

[5] M. K. HELLERSTEIN, R. A. NEESE, P. LINFOOT, M. CHRISTIANNSEN, S. TURNER AND A. LETSCHER, Hepatic gluconeogenic fluxes and glycogen turnover during fasting in humans. A stable isotope study, J Clin Invest, 100 (1997), pp. 1305-1319.

[6] N. G. OIKONOMAKOS, E. D. CHRYSINA, M. N. KOSMOPOULOU AND D. D. LEONIDAS, Crystal structure of rabbit muscle glycogen phosphorylase a in complex with a potential hypoglycaemic drug at 2.0 A resolution, Biochim Biophys Acta, 1647 (2003), pp. 325– 332.

[7] V. L. RATH, M. AMMIRATI, D. E. DANLEY, J. L. EKSTROM, E. M. GIBBS, T. R. HYNES, A. M. MATHIOWETZ, R. K. MCPHERSON, T. V. OLSON, J. L. TREADWAY AND D. J. HOOVER, Human liver glycogen phosphorylase inhibitors bind at a new allosteric site, Chem Biol, 7 (2000), pp. 677-682.

[8] R KURUKULASURIYA, J. T. LINK, D. J. MADAR, Z. PEI, S. J. RICHARDS J. J. ROHDE, A. J. SOUERS AND B. G. SZCZEPANKIEWICZ, Potential drug targets and progress towards pharmacologic inhibition of hepatic glucose production, Curr Med Chem, 10 (2003), pp. 123-153.

[9] Z. LIANG, L. ZHANG, L. LI, J. LIU, H. LI, L. ZHANG, L. CHEN, K. CHENG, M. ZHENG, X. WEN, P. ZHANG, J. HAO, Y. GONG, X. ZHANG, X. ZHU, J. CHEN, H. LIU, H. JIANG, C. LUO AND H. SUN, Identification of pentacyclic triterpenes derivatives as potent inhibitors against glycogen phosphorylase based on 3D-QSAR studies, Eur J Med Chem, 46 (2011), pp. 2011-2021.

[10] X. WEN, H. SUN, J. LIU, K. CHENG, P. ZHANG, L. ZHANG, J. HAO, L. ZHANG, P. NI, S. E. ZOGRAPHOS, D. D. LEONIDAS, K. M. ALEXACOU, T. GIMISIS, J. M. HAYES AND N. G. OIKONOMAKOS, Naturally occurring pentacyclic triterpenes as inhibitors of glycogen phosphorylase: synthesis, structure-activity relationships, and X-ray crystallographic studies, J Med Chem, 51 (2008), pp. 3540–3554.

[11] P. ZHU, Y. BI, J. XU, Z. LI, J. LIU, L. ZHANG, W. YE AND X. WU, Terpenoids. III: Synthesis and biological evaluation of 23-hydroxybetulinic acid derivatives as novel inhibitors of glycogen phosphorylase, Bioorg Med Chem Lett, 19 (2009), pp. 6966–6969.

[12] P. LAN, J. WANG, D. M. ZHANG, C. SHU, H. H. CAO, P. H. SUN, X. M. WU, W. C. YE AND W. M. CHEN, Synthesis and antiproliferative evaluation of 23-hydroxybetulinic acid derivatives, Eur J Med Chem, 46 (2011), pp. 2490-2502.

[13] Y. BI, J. XU, F. SUN, X. WU, W. YE, Y. SUN AND W. HUANG, Synthesis and biological activity of 23-hydroxybetulinic acid C-28 ester derivatives as antitumor agent candidates, Molecules, 17 (2012), pp. 8832-8841.

[14] W. DING, M. SUN, S. LUO, T. XU, Y. CAO, X. YAN AND Y. WANG, A 3D QSAR Study of Betulinic Acid Derivatives as Anti-Tumor Agents Using Topomer CoMFA: Model Building Studies and Experimental Verification, Molecules, 18 (2013), pp. 10228-10241.

[15] S. FULDA AND K. M. DEBATIN, Betulinic Acid Induces Apoptosis Through a Direct Effect on Mitochondria in Neuroectodermal Tumors, Med Pediatric Oncol 35 (2000), pp. 616–618.

[16] Z. N. JI, W. C. YE, G. G. LIU AND W. L. WENDY HSIAO, 23-Hydroxybetulinic Acid-Mediated Apoptosis is Accompanied by Decreases in Bcl-2 Expression and Telomerase Activity in HL-60 Cells, Life Sci, 72 (2002), pp. 1–9.

[17] M MANDAL AND R KUMAR, Bcl-2 Modulates Telomerase Activity, J Biol Chem, 272 (1997), pp. 14183–14187.

[18] Y ZHENG, F. ZHOU, X. WU, X. WEN, Y. LI, B. YAN, J. ZHANG, G. HAO, W. YE AND G. WANG, 23-Hydroxybetulinic acid from Pulsatilla chinensis (Bunge) Regel synergizes the antitumor activities of doxorubicin in vitro and in vivo, Journal of Ethnopharmacol, 128 (2010), pp. 615–622.

[19] ACD/ChemSketch version 12.01 Advanced Chemistry Development, Inc., Toronto, Ontario, 2009

[20] E. EROGLU, Some QSAR Studies for a Group of Sulfonamide Schiff Base as Carbonic Anhydrase CA II Inhibitors, Int J Mol Sci, 9 (2008), pp. 181-197.

[21] L. B. KIER, Use of molecular negentropy to encode structure governing biological activity, J Pharm Sci, 69 (1980), pp. 807-810.

[22] S. C. BASAK, D. K. HARRISS AND V. R. MAGNUSON, Comparative study of lipophilicity versus topological molecular descriptors in biological correlations, J Pharm Sci, 73 (1984), pp. 429-437.

[23] H. WIENER, Structural Determination of Paraffin Boiling Points, J Am Chem Soc, 69 (1947), pp. 17-20.

[24] D. PLAYSIC, S. NIKOLIC, N. TRINAJSTIC AND Z. MIHALIC, On the Harary Index for the Characterization of Chemical Graphs, J Math Chem, 12 (1993), pp. 235-250.

[25] M. RANDIAC, On the characterization of Molecular branching, J Am Chem Soc, 97 (1975), pp. 6609-6615.

[26] L. B. KIER AND L. H. HALL, Molecular connectivity in structure-activity analysis, Research studies press: Letchworth, Hertfordshire, U K (1986).

[27] M. W. SCHMIDT, K. K. BALDRIDGE, J. A. BOATZ, S. T. ELBERT, M. S. GORDON AND J. H. JENSEN, GAMESS Version= 24 Mar 2007 (R1) from Iowa State University, J Comput Chem, 14 (1993), pp. 1347-1363.

[28] G. M. MORRIS, R. HUEY, W. LINDSTROM, M. F. SANNER, R. K. BELEW, D. S. GOODSELL AND A. J. OLSON, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, J Comput Chem, 30 (2009), pp. 2785–2791.

[29] R. HUEY, G. M. MORRIS, A. J. OLSON AND D. S. GOODSELL, Software News and Update A Semiempirical Free Energy Force Field with Charge-Based Desolvation, J Comput Chem, 28 (2007), pp. 1145–1152.

[30] R. HUEY, D. S. GOODSELL, G. M. MORRIS AND A. J. OLSON, Grid-based hydrogen bond potentials with improved directionality, Lett Drug Des Discov, 1 (2004), pp. 178–183.

[31] R. THOMSEN AND M. H. CHRISTENSEN, MolDock: a new technique for high-accuracy molecular docking, J Med Chem, 49 (2006), pp. 3315–3321.