3. Results and Discussion

Cancer is a dangerous disease in which cells grow and divide beyond their normal limits. Currently, the major treatments for cancer include surgery; chemotherapy, and radiation. However, high incidences of undesirable side effects have prompted researchers to search for safer and more effective treatments.

Quantitative Structure Activity Relationship modeling and prediction is the fundamental methodology in an indirect drug design approach and is integral to a rational drug design process. QSAR modeling and analysis of results involves multiple steps requiring in depth knowledge of not only chemistry but also of statistics. The correct interpretation of results from a QSAR model is a key to utility of the entire QSAR exercise.

Computational chemistry represents molecular structures as numerical models and simulates their behavior with the equations of quantum and classical physics. Available programs enable scientists to easily generate and present molecular data including geometries, energies and associated properties (electronic, spectroscopic and bulk). The usual paradigm for displaying and manipulating these data is a table in which compounds are defined by individual rows and molecular properties (or descriptors) are defined by the associated columns. A QSAR attempts to find consistent relationships between the variations in the values of molecular properties and the biological activity for a series of compounds so that these "rules" can be used to evaluate new chemical entities.

The selection of parameters is an important first step in any QSAR study. If the association between the parameter(s) selected and activity is strong, then

activity predictions will be possible. If there is only weak association, knowing the value of the parameter(s) will not help in predicting activity. Thus, for a given study, parameters should be selected which are relevant to the activity for the series of molecules under investigation and these parameters should have values which are obtained in a consistent manner.

To end up with a model of good predictivity is a vital goal of any QSAR study, which is not an easy task, because so many obstacles may face the process starting from the very beginning; that is the choice of the data sets, the training sets, the validation sets and the subsequent computational calculations, i.e. The choice of appropriate methods and software for calculating physiochemical quantities and regression analysis suitable for the chosen sets. In this study a group of anthrapyrazole compounds with the general structure shown in fig (1) and presented with their chemical formula in tables (1 and 6) together with their respective biological activities expressed as optimum dose and inhibitory concentration IC_{50} towards P_{388} and L_{1210} leukemia cell lines respectively is chosen. Considering the given data, this group of compounds was divided into two subgroups, the desoxy subgroup; the one having no substituents in position 7 and 10 and a dihydroxy subgroup the one with a hydroxyl substituent at the 7 and 10 positions, then each of the subgroups was subjected to the same procedures in an attempt to end up with a model of good predictivity. A number of descriptors or physiochemical properties were calculated with the aid of computer software that were used to draw the structure, optimized it and calculate the desire properties, in some instances with more than one method e.g. total energy was calculated using the AM1, PM3 and MNDO methods and all were quoted and tabulated for further analysis. The physiochemical properties that were included in this study were the total energy, dipole moment, heat of formation, molar volume, molar refractivity,

polarizability and Log P. The biological activities results of the regression analysis are shown in the appendixes A &B. From these results, models having good r^2 values (>0.6 or 60%), were chosen for further internal and external validation and as a result 10 models for the desoxy subgroup, all of them relate the optimum dose to two of the physiochemical properties, the regression with $1/IC_{50}$ and $-\text{Log } IC_{50}$ give models of low r^2 values (<.6), so they have been excluded from further analysis or validation, and six models for the dihydroxy subgroup, all of them relate $1/IC_{50}$ to two of the physiochemical properties, the regression with optimum dose and $-\text{Log } IC_{50}$ give models of low r^2 values. In the following paragraphs the findings of the validation process were discussed.

In QSAR modeling, it is very important to validate the relevance of the resulting best QSAR model. One of the most useful methods of validation is the cross-validation method which refers to the use of one or more statistical techniques for internal validation in which different proportions of chemicals are omitted from the training set in order to verify the “internal predictivity”. The internal and external validation was performed to assess the predictivity of the best QSAR model. The internal validation was performed by using the training set compounds and then used to make predictions for the chemicals that were omitted. Cross-validation techniques allow the assessment of internal predictivity, in addition to the robustness of the model (stability of QSAR model parameters). The most important validation is the external validation and when performing external validation, two issues have to be dealt with. One is the number of samples in the external validation set and the other is the procedure for selecting them. It is recommended to use 30% of samples for the external validation of smaller data sets and to keep the same percentage of external samples in boot strapping and external validation.

In many cases r^2 CV are taken as a proof of the high predictive ability of QSAR models. A high value of these statistical characteristic (> 0.5) is considered as a proof of the high predictive ability of the model, although recent reports have proven the opposite. Although a low value of r^2 CV for the training set can indeed serve as an indicator of a low predictive ability of a model, the opposite is not necessarily true. Indeed, the high r^2 CV does not imply automatically a high predictive ability of the model. Thus, the high value of r^2 CV is the necessary condition for a model to have a high predictive power; it is not a sufficient condition. It is proven that the only way to estimate the true predictive power of a model is to test it on a sufficiently large collection of compounds from an external test set. The test set must include no less than five compounds, whose activities and structures must cover the range of activities and structures of compounds from the training set. This application is necessary for obtaining trustful statistics for comparison between the observed and predicted activities for these compounds. Besides high r^2 CV, a reliable model should be also characterized by a high correlation coefficient between the predicted and observed activities of compounds from a test set of molecules that was not used to develop the models.

In this discussion quantities such as the correlation coefficient (r); which is a simple statistical measure of the relationship between a dependent variable y (e.g. an optimum dose in our case) and one or more independent variable(s) x (e.g. the molar volume & the dipole moment) was used. It is given a value from 0 (for no relationship) to 1 (for a perfect fit) (100% in percentage). In QSAR analysis, r can be used as a measure of the statistical fit of a regression-based model, but the preferred form is its squared value (coefficient of determination, see below), often adjusted for the degrees of freedom.

A second quantity is the adjusted correlation coefficient (r^2adj). To understands this one should know that the total variation of any data set is made up of two parts, the part that can be explained by the regression equation and the part that cannot be explained by the regression equation. The coefficient of determination is the amount of dependent variable variance explained by a regression mode. It equals the square of the correlation coefficient R between the experimental response (the dependent variable y) and the predictors (the independent variables x). It represents the explained variance of the model, and is used as a measure of the goodness-of-fit of the model.

It is commonly believed that the closer the value to unity (or 100 in percentages), the better the model. However, it should be noted that r^2 is just a measure of the quality of the fit between model-calculated and experimental values, and it does not reflect the predictive power of the model at all. It is possible that a QSAR model with high r^2 could be a poor predictor.

The third quantity is the standard deviation of the data. It shows how far the activity values are spread about their average. This value provides an indication of the quality of the guess by showing the amount of variability inherent in the data.

3.1 QSAR modeling of Desoxy AP compounds

Regression analysis of desoxy anthrapyrazole compounds was performed using Minitab 16 software. The results are represented in the following tables:

Table (3.1.1): Regression Analysis dose versus E_{tPM3}; H_{fMNDO}

compound No.	Dose	Et_{PM3}	Hf_{MNDO}
1	100	92726.892	189316.5
2	25	95333.13	202727.2
3	400	52137.39	243483.5
4	25	94557.18	198637.2
5	25	113091.5	219367.6
6	50	127386.3	234732.8
7	6.25	145452	258958.2
8	12.5	183497.3	300209.7
9	25	135518.9	252332.4
10	25	102304.8	208241.8
11	25	120562.9	230641
12	50	169066.6	290418

The regression equation is

$$Dose = - 323 - 0.00464\ Et_{PM3} + 0.00395\ Hf_{MNDO}\ ,\ S = 17.2679,\ r^2 = 0.\ 981.$$

Table (3.1.2): Regression Analysis: Dose versus Et_{MNDO}; μ_{PM3}

compound No.	Dose	Et_{MNDO}	μ_{PM3}
1	25	95242.41	1652.696
2	400	136182.3	510.0475
3	25	93431.08	1665.52
4	25	114161.6	1648.098
5	50	127535.2	1908.456
6	6.25	146546.2	1926.868
7	12.5	180591.8	2245.629
8	25	132714.5	1845.86
9	25	103219.3	1417.362
10	25	122015.6	1565.507
11	50	170983.7	2161.546

The regression equation is:

$$Dose = 227 + 0.00205\ Et_{MNDO} - 0.255\ \mu_{PM3}\ ,\ S = 38.4215,\ \ r^2 = 0.\ 981.$$

Table (3.1.3): Regression Analysis: Dose, E_{tPM3} vs MR

compound No.	Dose	Et_{PM3}	MR
1	100	92726.892	94.55
2	25	95333.13	100.19
3	400	52137.39	109.09
4	25	94557.18	101.26
5	25	113091.5	101.26
6	50	127386.3	107.86
7	6.25	145452	112.14
8	12.5	183497.3	121.36
9	25	135518.9	121.36
10	25	102304.8	110.13
11	25	120562.9	115.72
12	50	169066.6	130.26

The regression equation is:

$$Dose = - 307 - 0.00331\ E_{tPM3} + 6.94\ MR\ ,\ S = 74.7089,\ \ r^2 = 0.\ 981.$$

Table (3.1.4): Regression Analysis: Dose, E_{tPM3} vs MV

compound No	Dose	E_{tPM3}	MV
1	100	92726.892	242.9
2	25	95333.13	257.1
3	400	52137.39	296.3
4	25	94557.18	256.0
5	25	113091.5	256.0
6	50	127386.3	277.8
7	6.25	145452	301.0
8	12.5	183497.3	333.1
9	25	135518.9	333.1
10	25	102304.8	295.8
11	25	120562.9	318.9
12	50	169066.6	372.3

The regression equation is:

$$Dose = - 85 - 0.00318\ Et_{PM3} + 1.79\ MV\ \ ,\ \ S = 73.2876,\ r^2 = 0.627.$$

Table (3.1.5): Regression Analysis: Dose, E_{tPM3} vs Polz

compound No	Dose	$E_{t\ PM3}$	polarizability
1	100	92726.892	37.48
2	25	95333.13	39.72
3	400	52137.39	43.24
4	25	94557.18	40.14
5	25	113091.5	40.14
6	50	127386.3	42.75
7	6.25	145452	44.45
8	12.5	183497.3	48.11
9	25	135518.9	48.11
10	25	102304.8	43.66
11	25	120562.9	45.87
12	50	169066.6	51.64

The regression equation is:

$$\text{Dose} = -307 - 0.00331\ Et_{PM3} + 17.5\ \text{polarizability}S = 74.7357,\ r^2 = 0.612$$

Table (3.1.6): Regression Analysis: Dose, MR vs μ_{PM3}

compound No	Dose	MR	μ_{PM3}
1	100	94.55	1642.321
2	25	100.19	1652.696
3	400	109.09	510.0475
4	25	101.26	1665.52
5	25	101.26	1648.098
6	50	107.86	1908.456
7	6.25	112.14	1926.868
8	12.5	121.36	2245.629
9	25	121.36	1845.86
10	25	110.13	1417.362
11	25	115.72	1565.507
12	50	130.26	2161.546

The regression equation is:

$$\text{Dose} = 165 + 2.54\ \text{MR} - 0.227\ \mu_{PM3}\ ,\ S = 61.7437,\ r^2 = 0.735$$

Table (3.1.7): Regression Analysis: Dose, MV vs μ_{PM3}

compound No.	Dose	MV	μ_{PM3}
1	100	242.9	1642.321
2	25	257.1	1652.696
3	400	296.3	510.0475
4	25	256.0	1665.52
5	25	256.0	1648.098
6	50	277.8	1908.456
7	6.25	301.0	1926.868
8	12.5	333.1	2245.629
9	25	333.1	1845.86
10	25	295.8	1417.362
11	25	318.9	1565.507
12	50	372.3	2161.546

The regression equation is:

$$\text{Dose} = 239 + 0.680\ \text{MV} - 0.223\ \mu_{PM3}\ , S = 61.3404, r^2 = 0.739.$$

Table (3.1.8): Regression Analysis: Dose, Log P vs μ_{PM3}

compound No.	Dose	Log P	μ_{PM3}
1	100	1.48	1642.321
2	25	0.70	1652.696
3	400	2.80	510.0475
4	25	0.76	1665.52
5	25	0.76	1648.098
6	50	0.44	1908.456
7	6.25	1.48	1926.868
8	12.5	2.55	2245.629
9	25	2.55	1845.86
10	25	2.85	1417.362
11	25	3.08	1565.507
12	50	4.64	2161.546

The regression equation is:

$$\text{Dose} = 374 + 16.3\ \text{Log P} - 0.204\ \mu_{PM3}\ , S = 63.3502, r^2 = 0.722.$$

Table (3.1.9): Regression Analysis: Dose, Polz vs μ_{PM3}

compound No.	Dose	Polarizability	μ_{PM3}
1	100	37.48	1642.321
2	25	39.72	1652.696
3	400	43.24	510.0475
4	25	40.14	1665.52
5	25	40.14	1648.098
6	50	42.75	1908.456
7	6.25	44.45	1926.868
8	12.5	48.11	2245.629
9	25	48.11	1845.86
10	25	43.66	1417.362
11	25	45.87	1565.507
12	50	51.64	2161.546

The regression equation is:

Dose = 165 + 6.41 polarizability - 0.227 μ_{PM3} , S = 61.7509, r^2 = 0.735.

Table (3.1.10): Regression Analysis: Dose, $H_{f\ MNDO}$ vs μ_{PM3}

compound No.	**Dose**	**$H_{f\ MNDO}$**	**μ_{PM3}**
1	100	189316.5	1642.321
2	25	202727.2	1652.696
3	400	243483.5	510.0475
4	25	198637.2	1665.52
5	25	219367.6	1648.098
6	50	234732.8	1908.456
7	6.25	258958.2	1926.868
8	12.5	300209.7	2245.629
9	25	252332.4	1845.86
10	25	208241.8	1417.362
11	25	230641	1565.507
12	50	290418	2161.546

The regression equation is:

Dose = 176 + 0.00129 Hf - 0.248 μ_{PM3} , S = 49.5976. r^2 = 0.829.

3.2 QSAR modeling of Dihydroxy AP compounds

Regression analysis of dihydroxy anthrapyrazole compounds was also performed using Minitab 16 software. The results are represented in the following tables:

Table (3.2.1): Regression Analysis: 1/ IC_{50}, E_{tPM3} vsμ_{AM1}

compound No.	$1/IC_{50}$	E_{tPM3}	μ_{AM1}
13	6666666.67	109456.9	1840.549
14	1282051.28	107180.5	1822.94
15	1369863.01	153711	2169.615
16	1724137.93	106024.3	1816.075
17	625000.00	14103.19	1898.758
18	1351351.35	160329.9	2184.357
19	4347826.09	181441.6	2206.234
20	1960784.31	223491.3	2557.184
21	7692307.69	105365	2078.625
22	21739130.43	130926.9	488.8187
23	1351351.35	144306.2	1789.562
24	1818181.82	149656.2	1901.099

The regression equation is:

$$1/IC_{50} = 20464882 - 10789\ \mu_{AM1} + 32.7\ E_{tPM3}\ ,\ S = 3380243,\ r^2 = 0.735.$$

Table (3.2.2): Regression Analysis: 1/ IC_{50}, MR vs μ_{AM1}

compound No.	$1/IC_{50}$	MR	μ_{AM1}
13	6666666.67	96.25	1840.549
14	1282051.28	101.90	1822.94
15	1369863.01	110.79	2169.615
16	1724137.93	102.96	1816.075
17	625000.00	102.96	1898.758
18	1351351.35	109.56	2184.357
19	4347826.09	113.85	2206.234
20	1960784.31	123.07	2557.184
21	7692307.69	123.07	2078.625
22	21739130.43	111.86	488.8187
23	1351351.35	117.42	1789.562
24	1818181.82	131.96	1901.099

The regression equation is:

$1/IC_{50} = 12754339 - 10091\ \mu_{AM1} + 95485\ MR$ $S = 3652707$, $r^2 = 0.691$.

Table (3.2.3): Regression Analysis: $1/IC_{50}$, MV vs μ_{AM1}

compound No.	$1/IC_{50}$	MV	μ_{AM1}
13	6666666.67	237.4	1840.549
14	1282051.28	251.6	1822.94
15	1369863.01	290.8	2169.615
16	1724137.93	251.1	1816.075
17	625000.00	251.1	1898.758
18	1351351.35	272.3	2184.357
19	4347826.09	295.5	2206.234
20	1960784.31	327.6	2557.184
21	7692307.69	327.6	2078.625
22	21739130.43	290.3	488.8187
23	1351351.35	313.4	1789.562
24	1818181.82	366.8	1901.099

The regression equation is:

$1/IC_{50} = 16158904 + 24782\ MV - 10025\ \mu_{AM1}$,$S = 3659070$, $r^2 = 0.689$.

Table (3.2.4): Regression Analysis: $1/IC_{50}$, Log P vs μ_{AM1}

compound No.	$1/IC_{50}$	Log P	μ_{AM1}
13	6666666.67	1.24	1840.549
14	1282051.28	0.46	1822.94
15	1369863.01	2.55	2169.615
16	1724137.93	0.51	1816.075
17	625000.00	0.51	1898.758
18	1351351.35	0.20	2184.357
19	4347826.09	1.24	2206.234
20	1960784.31	2.30	2557.184
21	7692307.69	2.30	2078.625
22	21739130.43	2.60	488.8187
23	1351351.35	2.84	1789.562
24	1818181.82	4.39	1901.099

The regression equation is:

$1/IC_{50} = 21533904 + 490098\ Log\ P - 9530\ \mu_{AM_1}$ $S = 3747785$, $r^2 = 0.674$.

Table (3.2.5): Regression Analysis: 1/ IC_{50}, Polzarizabilty vs μ_{AM1}

compound No.	$1/IC_{50}$	Polarizability	μ_{AM1}
13	6666666.67	38.15	1840.549
14	1282051.28	40.39	1822.94
15	1369863.01	43.92	2169.615
16	1724137.93	40.81	1816.075
17	625000.00	40.81	1898.758
18	1351351.35	43.43	2184.357
19	4347826.09	45.13	2206.234
20	1960784.31	48.78	2557.184
21	7692307.69	48.78	2078.625
22	21739130.43	44.34	488.8187
23	1351351.35	46.55	1789.562
24	1818181.82	52.31	1901.099

The regression equation is:

$1/IC_{50} = 12768322 - 10091\ \mu AM_1 + 240545$ Polarizability , S = 3653091, r^2 = 0.691.

Table (3.2.6): Regression Analysis: 1/ IC_{50}, $H_{f\ MNDO}$ VS μ_{MNDO}

compound No.	$1/IC_{50}$	$H_{f\ MNDO}$	μ_{MNDO}
13	6666666.67	220723.7	1839.883
14	1282051.28	225510.1	1822.868
15	1369863.01	273923.8	2147.623
16	1724137.93	225658.7	1815.859
17	625000.00	260975.7	1897.517
18	1351351.35	291277.9	2184.18
19	4347826.09	309491.1	2205.99
20	1960784.31	359453	2556.986
21	7692307.69	296792.9	2078.375
22	21739130.43	150636.1	488.915
23	1351351.35	268788.2	1789.349
24	1818181.82	284219.1	1900.813

The regression equation is:

$1/IC_{50} = 17407936 - 15968\ \mu MNDO + 65.0\ H_f$,S = 3411467, r^2 = 0. 73.

The regression equations with high r^2 values (> 0.6), were subjected to further studies to choose the best QSAR model via cross validation of the training and validation sets.

3.3 Internal Validation

The idea behind the internal validation techniques is that, the biological activity of the training set, that is used to generate the QSAR model is recalculated using the model and the calculated values are plotted against the experimental ones (the values used to generate the model). As a result of the regression analysis done, ten models of desoxy compounds and six models of the dihydroxy compounds were chosen for analysis and since a QSAR model is acceptable when it has an r^2 value greater than 0.6 and r^2 (*CV*) greater than 0.5, models that show r^2 values greater or equal to the 0.6 are considered models of good predictivity**.** The results of internal cross validation calculation of the desoxy and dihydroxy models are shown in tables below and their respective graph are shown in Figures

3.3.1 Internal Validation of Desoxy subgroup Models

Model 1

This model relates the optimum dose to total energy calculated using PM3 and the heat of formation using MNDO methods

Dose = - 323 - 0.00464 E_{tPM_3}+ 0.00395 $H_{f\ MNDO}$, r^2 =0.981.

Table (3.3.1.1) internal validation of desoxy AP compounds model 1

Compound No.	Calculated Dose	Experimental Dose
1	5.45	100
2	33.43	25
3	396.84	400
4	22.87	25
5	18.76	25
6	13.22	50
7	24.99	6.25
8	11.4	12.5
9	44.9	25
10	24.86	25
11	28.62	25
12	39.68	50

With $\mathbf{r^2}$ = 0.981 for the training set is a very good indication of the fitness of the model to the data set. The internal validation of this model returns also a good $\mathbf{r^2}$ =0.9212 and the linearity of this relation is shown in the figure below.

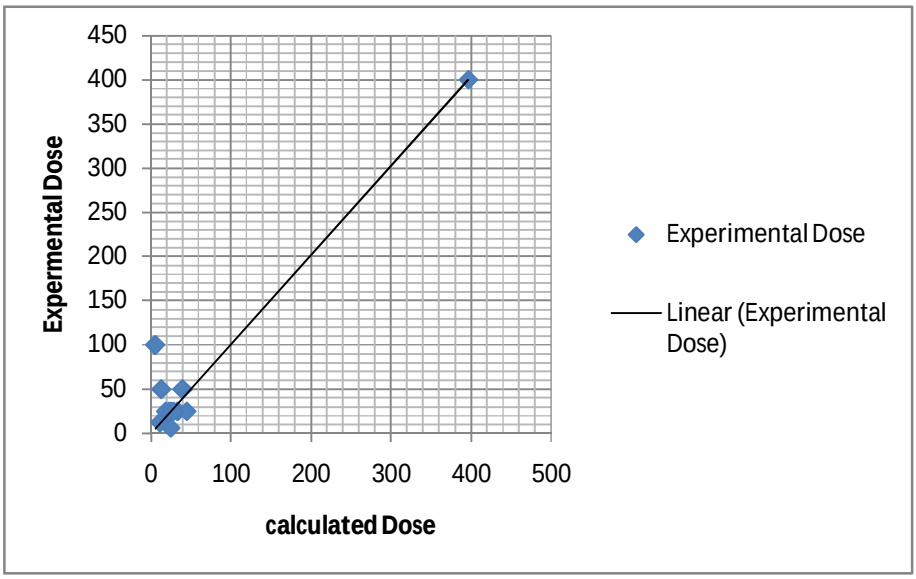

Fig (3.3.1.1) internal validation of model 1 Exp. VS Cal. Dose

Model 2

This model relates the optimum dose to total energies calculated using MNDO and the dipole moments using PM_3 methods

$$\mathbf{Dose = 227 + 0.00205\ E_{t\ MNDO} - 0.255\ \mu_{PM3}}\ , \qquad \mathbf{r^2 = 0.908.}$$

Table (3.3.1.2) internal validation of desoxy AP compounds model 2

Compound No.	Experimental Dose	Calculated Dose
1	25	0.8
2	400	376.11
3	25	-6.17
4	25	40.77
5	50	1.79
6	6.25	36.07
7	12.5	24.58
8	25	28.37
9	25	77.17
10	25	77.93
11	50	26.32

With r^2 = 0.908 for the training set is a very good indication of the fitness of the model to the data. The internal validation of this model returns poor(Fig.3.3.1.2).

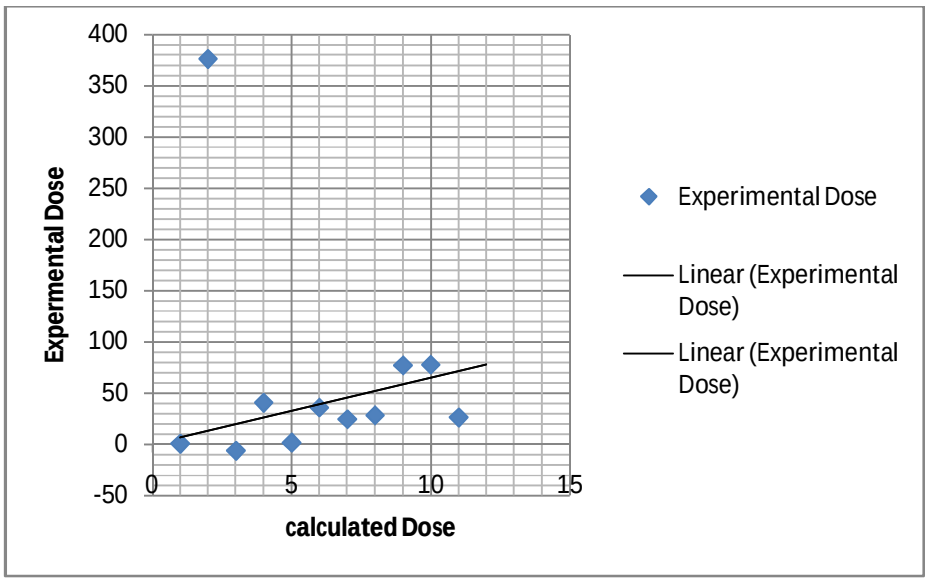

Fig. (3.3.1.2) internal validation of model 2 Exp. VS Cal. Dose

Model 3

This model relates the optimum dose to total energies calculated using PM_3 method and the molar refractivity

$$\textbf{Dose} = -307 - 0.00331\ E_{t\ PM3} + 6.94\ MR\ , \qquad r^2 = 0.613.$$

Table (3.3.1.3) internal validation of desoxy AP compounds model 3

Compound No.	Experimental Dose	Calculated Dose
1	100	42.25
2	25	72.77
3	400	277.51
4	25	82.76
5	25	21.41
6	50	19.9
7	6.25	10.19-
8	12.5	72.14-
9	25	86.67

10	25	118.67
11	25	97.03
12	50	37.39

With r^2 = 0.613 for the training set is a good indication of the fitness of the model to the data. The internal validation of this model returns poor r^2 = 0.245 and the linearity of this relation is shown in Fig.(3.3.1.3).

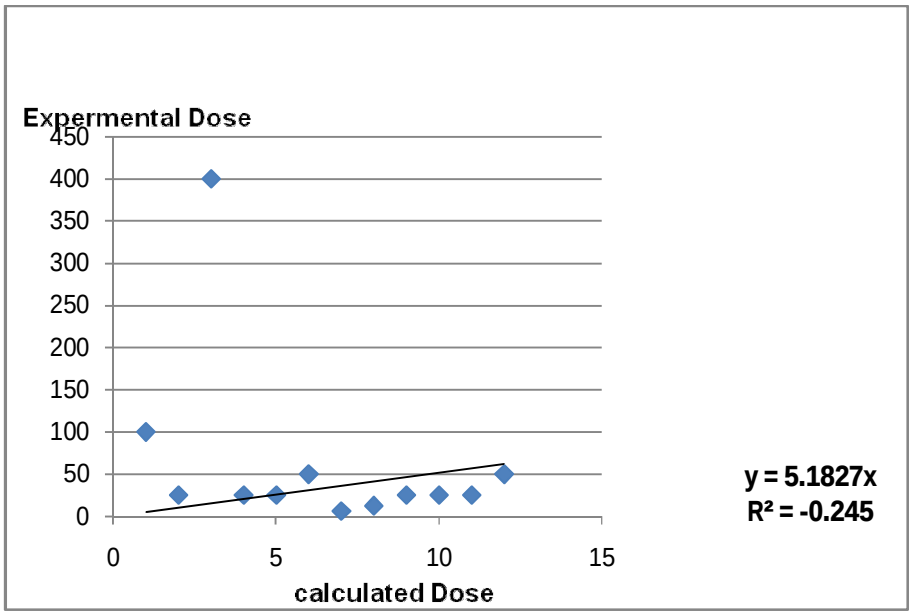

Fig.(3.3.1.3)internal validation of model 3 Exp. VS Cal. Dose

Model 4

This model relates the optimum dose to total energies calculated using PM_3 method and the molar volume

$$\textbf{Dose} = -85 - 0.00318\, E_{t\,PM3} + 1.79\, MV, \quad r^2 = 0.627.$$

Table (3.3.1.4) internal validation of desoxy AP compounds model 4

Compound No.	Experimental Dose	Calculated Dose
1	100	54.92
2	25	72.05
3	400	279.58
4	25	72.55
5	25	13.61
6	50	7.17
7	6.25	-8.75
8	12.5	-72.27
9	25	80.3
10	25	119.15
11	25	102.44
12	50	43.79

With r^2 = 0.627 for the training set is a good indication of the fitness of the model to the data. The internal validation of this model returns poor r^2 = -0.238 and the linearity of this relation is shown in Fig.(3.3.1.4).

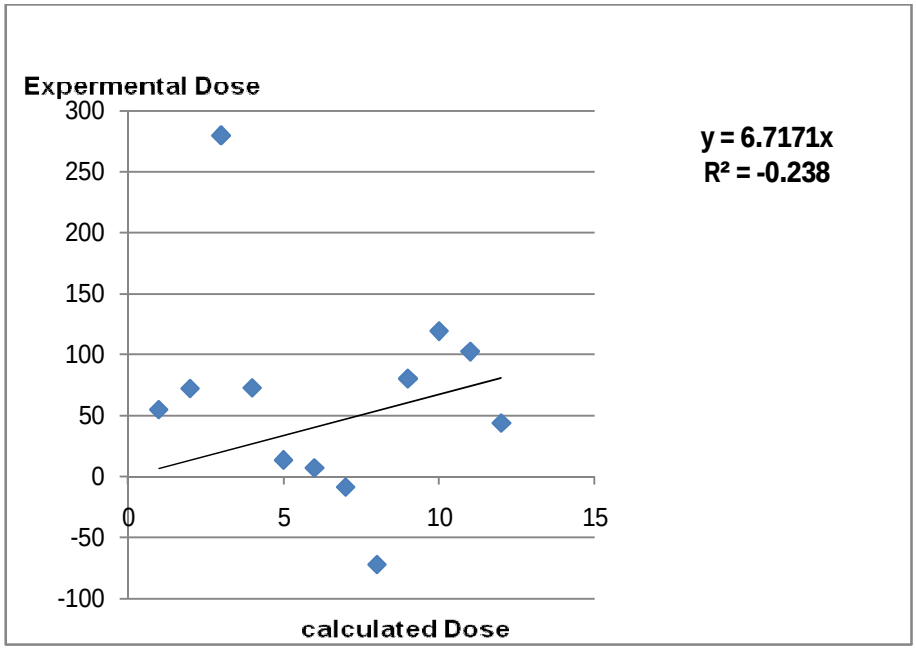

Fig.(3.3.1.4)internal validation of model 4 Exp. VS Cal. Dose

Model 5

This model relates the optimum dose to total energies calculated using PM_3 method and the polarizability

$$\textbf{Dose} = -307 - 0.00331\ E_{t\ PM3} + 17.5\ \textbf{polarizability},\quad r^2 = 0.612.$$

Table (3.3.1.5) internal validation of desoxy AP compounds model 5

Compound No.	Experimental Dose	Calculated Dose
1	100	41.97
2	25	72.55
3	400	547.13
4	25	82.47
5	25	21.12
6	50	19.48
7	6.25	53.48
8	12.5	-72.45
9	25	86.36
10	25	118.42
11	25	96.66
12	50	37.09

With $r^2 = 0.612$ for the training set is a good indication of the fitness of the model to the data. The internal validation of this model returns poor $r^2 = -0.208$ and the linearity of this relation is shown in Fig.(3.3.1.5).

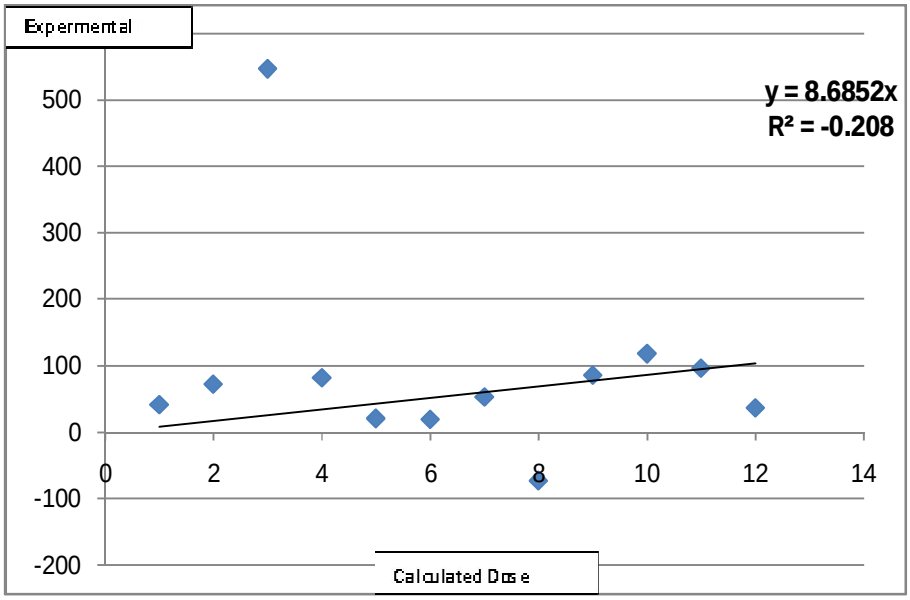

Fig. (3.3.1.5) internal validation of model 5 Exp. VS Cal. Dose

Model 6

This model relates the optimum dose to dipole moments calculated using PM_3 method and the molar refractivity

$$\textbf{Dose} = 165 + 2.54\ MR - 0.227\ \mu_{PM_3}, \qquad r^2 = 0.735.$$

Table (3.3.1.6) internal validation of desoxy AP compounds model 6

Compound No.	Experimental Dose	Calculated Dose
1	100	32.35
2	25	44.32
3	400	326.31
4	25	44.13
5	25	48.08
6	50	5.74
7	6.25	12.44
8	12.5	-36.5
9	25	54.24
10	25	122.99
11	25	103.56
12	50	5.19

With r^2 = 0.735 for the training set is a good indication of the fitness of the model to the data. The internal validation of this model returns poor r^2 = -0.208 and the linearity of this relation is shown in Fig.(3.3.1.6).

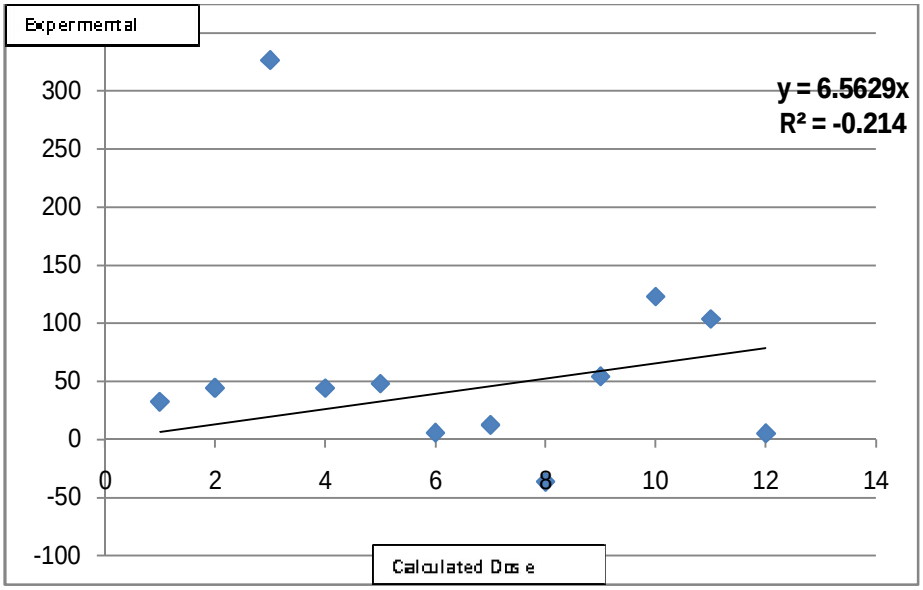

Fig.(3.3.1.6) internal validation of model 6 Exp. VS Cal. Dose

Model 7

This model relates the optimum dose to dipole moments calculated using PM_3 method and the molar volume .

$$\textbf{Dose} = 239 + 0.680\ MV - 0.223\ \mu_{PM3}, \qquad r^2 = 0.739.$$

Table (3.3.1.7) internal validation of desoxy AP compounds model 7

Compound No.	Calculated Dose	Experimental Dose
1	37.93	100
2	45.28	25
3	326.74	400
4	41.67	25
5	45.55	25
6	2.32	50
7	13.99	6.25
8	-35.27	12.5
9	53.88	25
10	124.07	25

11	106.74	25
12	10.14	50

With r^2 = 0.739 for the training set is a good indication of the fitness of the model to the data. The internal validation of this model returns poor r^2 = -0.216 and the linearity of this relation is shown in Fig.(3.3.1.7).

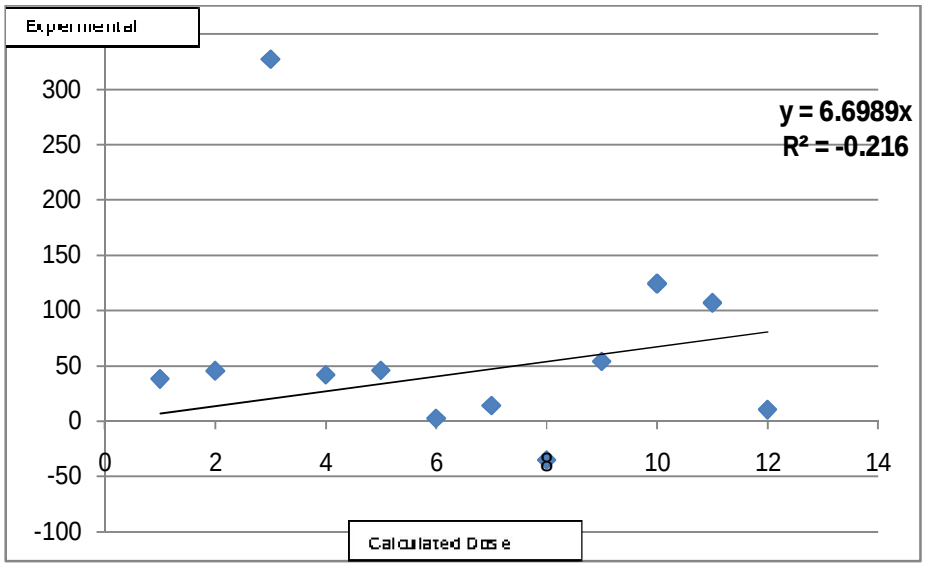

Fig.(3.3.1.7) internal validation of model 7 Exp. VS Cal. Dose

Model 8

This model relates the optimum dose to dipole moments calculated using PM_3 method and the molar refractivity

$$\textbf{Dose} = 374 + 16.3 \log P - 0.204\ \mu_{PM3}\ , \qquad r^2 = 0.722.$$

Table (3.3.8) internal validation of desoxy AP compounds model 8

Compound No.	Calculated Dose	Experimental Dose
1	63.09	100
2	48.26	25
3	315.59	400
4	46.62	25
5	50.18	25

6	-8.15	50
7	5.04	6.25
8	-42.54	12.5
9	39.01	25
10	131.31	25
11	104.84	25
12	8.68	50

With r^2 = 0.722 for the training set is a good indication of the fitness of the model to the data. The internal validation of this model returns poor r^2 = -0.238 and the linearity of this relation is shown in Fig.(3.3.1.8) and only one compound gave a comparable calculated results.

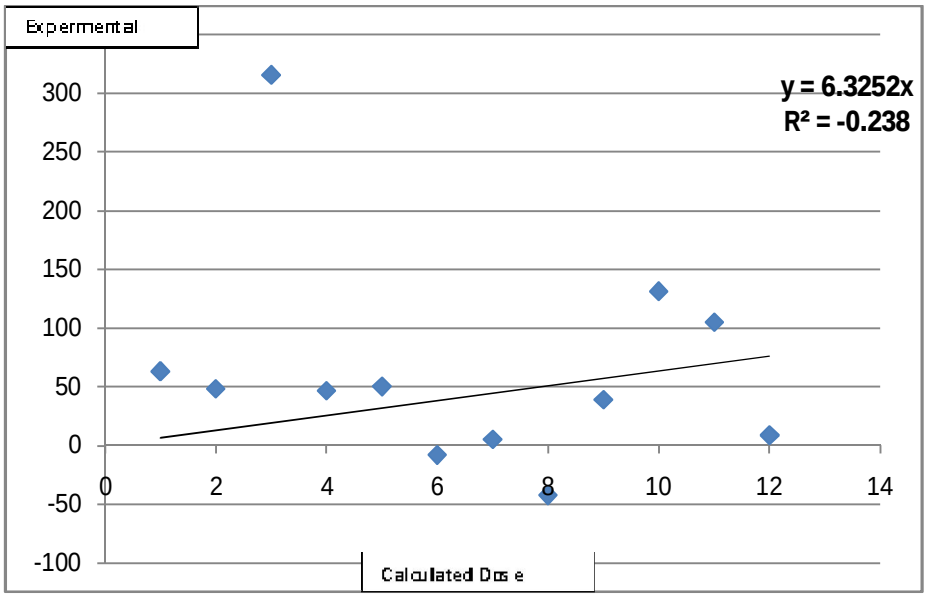

Fig.(3.3.1.8) internal validation of model 8 Exp. VS Cal. Dose

Model 9

This model relates the optimum dose to dipole moments calculated using PM_3 method and the molar refractivity

Dose = 165 + 6.41 polarizability - 0.227 μ_{PM3} , r^2 = 0.735.

Table (3.3.1.9) internal validation of desoxy AP compounds model 9

Compound No.	calculated Dose	Experimental Dose
1	32.44	100
2	44.44	25
3	326.39	400
4	44.22	25
5	48.18	25
6	5.81	50
7	12.53	6.25
8	-36.37	12.5
9	54.37	25
10	123.12	25
11	103.66	25
12	5.341	50

With r^2 = 0.735 for the training set is a good indication of the fitness of the model to the data. The internal validation of this model returns poor r^2 = -0.215 and the linearity of this relation is shown in Fig.(3.3.1.9) and only one compound gave a comparable calculated results.

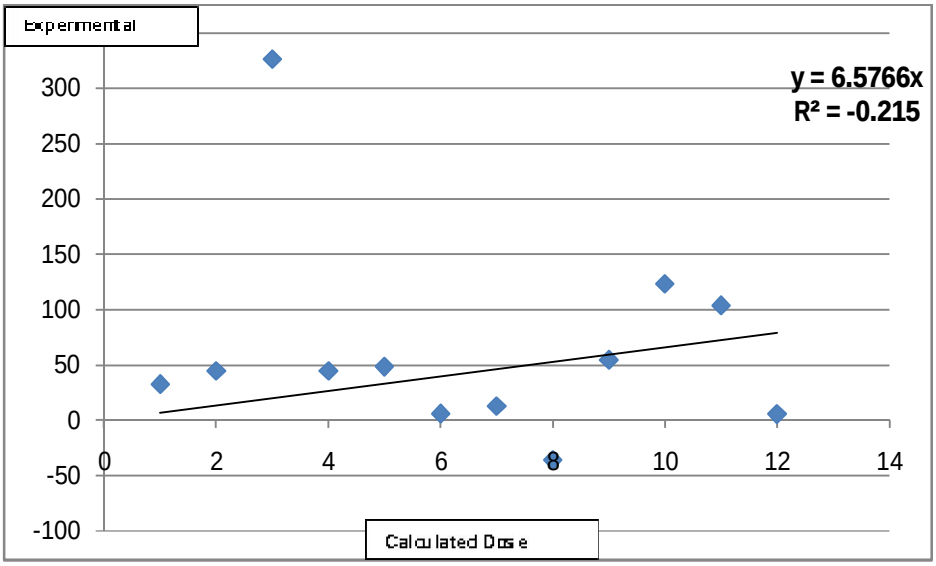

Fig. (3.3.1.9) internal validation of model 9 Exp. VS Cal. Dose

Model 10

This model relates the optimum dose to dipole moments calculated using PM_3 method and the molar refractivity

$$\textbf{Dose} = 176 + 0.00129\, H_f - 0.248\, \mu_{PM3}, \qquad r^2 = 0.829.$$

Table (3.3.1.10) internal validation of desoxy AP model 10

Compound No.	Calculated Dose	Experimental Dose
1	12.92	100
2	27.65	25
3	363.6	400
4	19.19	25
5	50.26	25
6	5.51	50
7	32.19	6.25
8	6.35	12.5
9	43.74	25
10	93.13	25
11	85.28	25
12	14.58	50

With $r^2 = 0.829$ for the training set is a good indication of the fitness of the model to the data. The internal validation of this model returns poor $r^2 = -0.215$ and the linearity of this relation is shown in Fig.(3.3.1.10) and only three compounds gave a comparable calculated results.

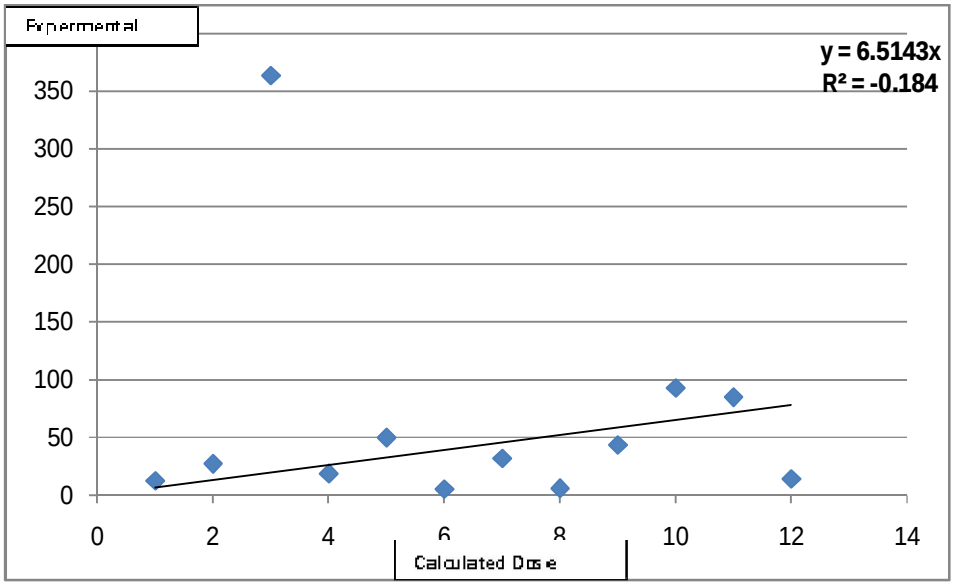

Fig.(3.3.1.10) internal validation of model 10 Exp. VS Cal. Dose

Finally one can say from the ten studied models only model 1 is considered a good model that can be used to predict the activity of desoxy AP subgroup towards the P_{388} leukemia cell lines, no model is suitable to any reasonable degree that can be used to predict the ability of AP desoxy subgroup towards the L_{1210} leukemia cell lines.

3.3.2 Internal Validation of Dihydroxy AP Models

Model 1

This model relates the reciprocal of inhibitory concentration (**$1/IC_{50}$**) to total energy calculated using PM_3 and dipole moments calculated using AM_1 methods

$$\mathbf{1/IC_{50} = 20464882 - 10789\ \mu_{AM1} + 32.7\ E_{t\ PM3}\ ,\ r^2 = 0.735.}$$

Table (3.3.2.1) Internal validation of Dihydroxy AP model 1

Compound No.	experimental $1/IC_{50}$	calculated $1/IC_{50}$
13	6666666.67	4601193.55
14	1282051.28	4386730.88
15	1369863.01	2107222.95
16	1724137.93	4476659.23
17	625000	3688690.24
18	1351351.35	815001.39
19	4347826.09	1116215.5
20	1960784.31	-1708834.12
21	7692307.69	2851833.15
22	21739130.43	18149716.59
23	1351351.35	22925782.32
24	1818181.82	5567960.75

With r^2 = 0.735 which is a good indication of the fitness of this model to the data set. The internal validation of this model returns poor r^2 = 0.2122 and the linearity of this relation is shown in Fig.(3.3.2.1).

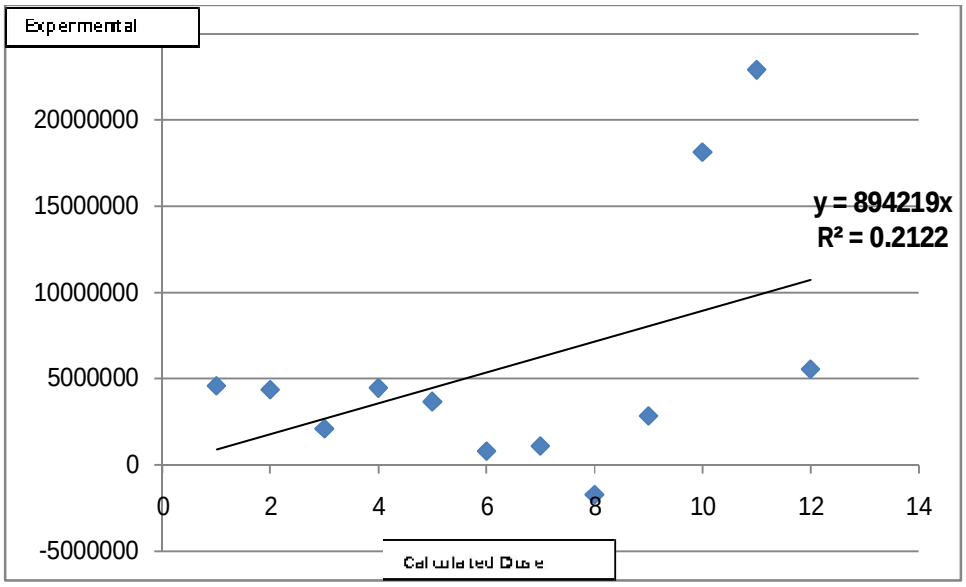

Fig.(3.3.2.1) internal validation of model 1 Exp. VS Cal. Dose

Model 2

This model relates the reciprocal of inhibitory concentration (**$1/IC_{50}$**) to dipole moments calculated using AM_1 method and the molar refractivity

$$\mathbf{1/IC_{50} = 12754339 - 10091\ \mu_{AM1} + 95485\ MR\ ,\quad r^2 = 0.691.}$$

Table (3.3.2.2) internal validation of Dihydroxy AP model 2

Compound No.	Experimental $1/IC_{50}$	calculated $1/IC_{50}$
13	6666666.67	3371790.291
14	1282051.28	4088972.96
15	1369863.01	1439537.185
16	1724137.93	4259461.775
17	625000	3425107.622
18	1351351.35	1173329.113
19	4347826.09	1362198.956
20	1960784.31	-1298865.794
21	7692307.69	3530273.075
22	21739130.43	18502621.6
23	1351351.35	5907717.558
24	1818181.82	6170549.591

With $r^2 = 0.691$ which is a good indication of the fitness of this model to the data set. The internal validation of this model returns poor $r^2 = 0.2122$ and the linearity of this relation is shown in Fig.(3.3.2.2).

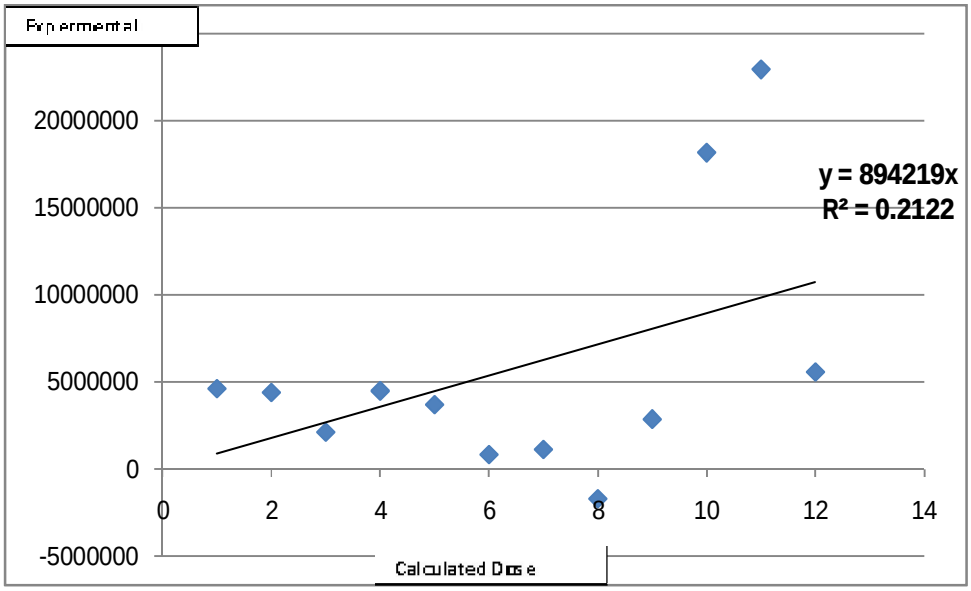

Fig. (3.3.2.2) internal validation of model 2 Exp. VS Cal. Dose

Model 3

This model relates the reciprocal of inhibitory concentration ($1/IC_{50}$) to dipole moments calculated using AM1 method and the molar volume

$$\mathbf{1/IC_{50} = 16158904 + 24782\ MV - 10025\ \mu_{AM1}\ ,\quad r^2 = 0.689.}$$

Table (3.3.2.3) internal validation of Dihydroxy AP model 3

Compound No.	experimental $1/IC_{50}$	calculated $1/IC_{50}$
13	6666666.67	3590647.08
14	1282051.28	4119081.7
15	1369863.01	1615119.23
16	1724137.93	4175512.33
17	625000	3346615.25
18	1351351.35	1008863.68
19	4347826.09	1364489.15
20	1960784.31	-1358282.4
21	7692307.69	3439271.58
22	21739130.43	18452711.13
23	1351351.35	5985223.75
24	1818181.82	6190424.125

With r^2 = 0.689; which is a good indication of the fitness of this model to the data set. The internal validation of this model returns poor r^2 = 0.2122 and the linearity of this relation is shown in Fig (3.3.2.3).

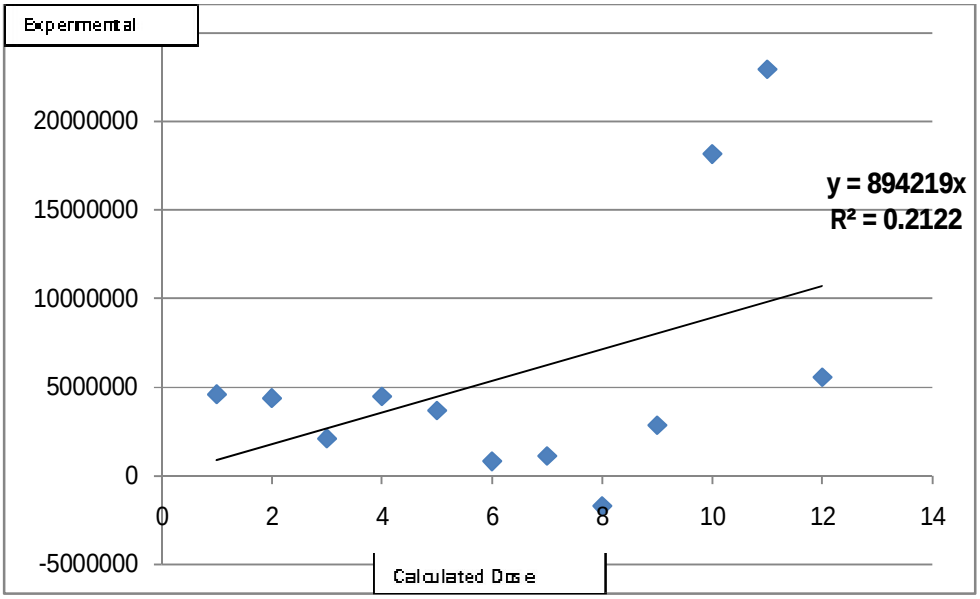

Fig. (3.3.2.3) internal validation of model 3 Exp. VS Cal. Dose

Model 4

This model relates the reciprocal of inhibitory concentration ($\mathbf{1/IC_{50}}$) to dipole moments calculated using AM_1 method and the Log P

$$\mathbf{1/IC_{50} = 21533904 + 490098\ Log\ P - 9530\ \mu_{AM1}\ ,\ r^2 = 0.674.}$$

Table (3.3.2.4) internal validation of Dihydroxy AP model 4

Compound No.	experimental $1/IC_{50}$	calculated $1/IC_{50}$
13	6666666.67	4601193.55
14	1282051.28	4386730.88
15	1369863.01	2107222.95
16	1724137.93	4476659.23
17	625000	3688690.24
18	1351351.35	815001.39
19	4347826.09	1116215.5
20	1960784.31	-1708834.12
21	7692307.69	2851833.15
22	21739130.43	18149716.59
23	1351351.35	22925782.32
24	1818181.82	5567960.75

With r^2 = 0.674; which is a good indication of the fitness of this model to the data set. The internal validation of this model returns poor r^2 = 0.2122 and the linearity of this relation is shown in Fig.(3.3.2.4).

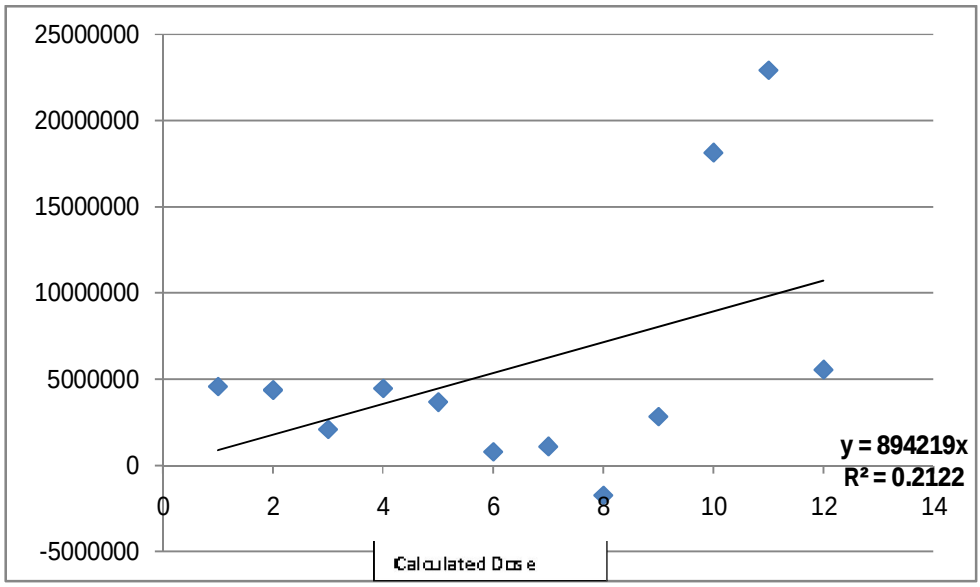

Fig.(3.3.2.4) internal validation of model 4 Exp. VS Cal. Dose

Model 5

This model relates the reciprocal of inhibitory concentration (**$1/IC_{50}$**) to dipole moments calculated using AM_1 method and the polarizability

$1/IC_{50}$ = 12768322 - 10091 μ_{AM1} + 240545 Polarizability, r^2 = 0.691.

Table (3.3.2.5) internal validation of Dihydroxy AP model 5

Compound No.	experimental 1/IC50	calculated 1/IC50
13	6666666.67	3372133.791
14	1282051.28	4088647.01
15	1369863.01	1439473.435

16	1724137.93	4258950.625
17	625000	3424596.472
18	1351351.35	1172844.863
19	4347826.09	1361010.556
20	1960784.31	-1302436.644
21	7692307.69	3526702.225
22	21739130.43	23434087.3
23	1351351.35	5907221.608
24	1818181.82	25351230.95

With r^2 = 0.691; which is a good indication of the fitness of this model to the data set. The internal validation of this model returns poor r^2 = 0.3086 and the linearity of this relation is shown in Fig.(3.3.2.5).

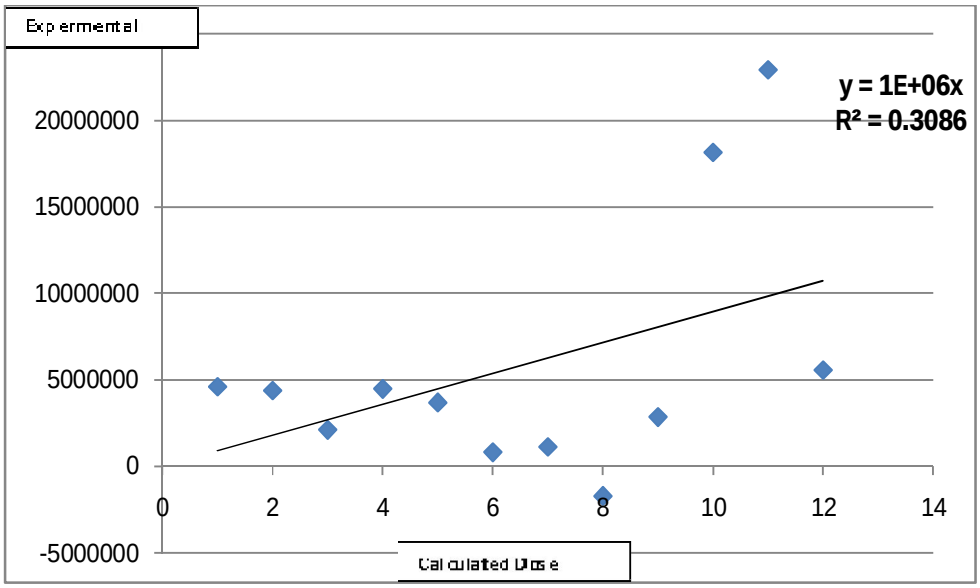

Fig. (3.3.2.5) internal validation of model 5 Exp. VS Cal. Dose

.

Model 6

This model relates the reciprocal of inhibitory concentration ($1/IC_{50}$) to dipole moments calculated using MNDO method and the heat of formation calculated using MNDO method.

$$1/IC_{50} = 17407936 - 15968\ \mu_{MNDO} + 65.0\ H_f\ ,\ r^2 = 0.73.$$

Table (3.3.2.6) internal validation of Dihydroxy AP model 6

Compound No.	experimental 1/IC50	calculated 1/IC50
13	6666666.67	2375724.756
14	1282051.28	49475851.37
15	1369863.01	919738.936
16	1724137.93	3080114.988
17	625000	4071805.044
18	1351351.35	1464013.26
19	4347826.09	2299609.18
20	1960784.31	-57571.448
21	7692307.69	3511982.5
22	21739130.43	19392287.78
23	1351351.35	6306844.168
24	1818181.82	5529995.516

With $r^2 = 0.73$; which is a good indication of the fitness of this model to the data set. The internal validation of this model returns poor $r^2 = 0.2122$ and the linearity of this relation is shown in Fig.(3.3.2.6).

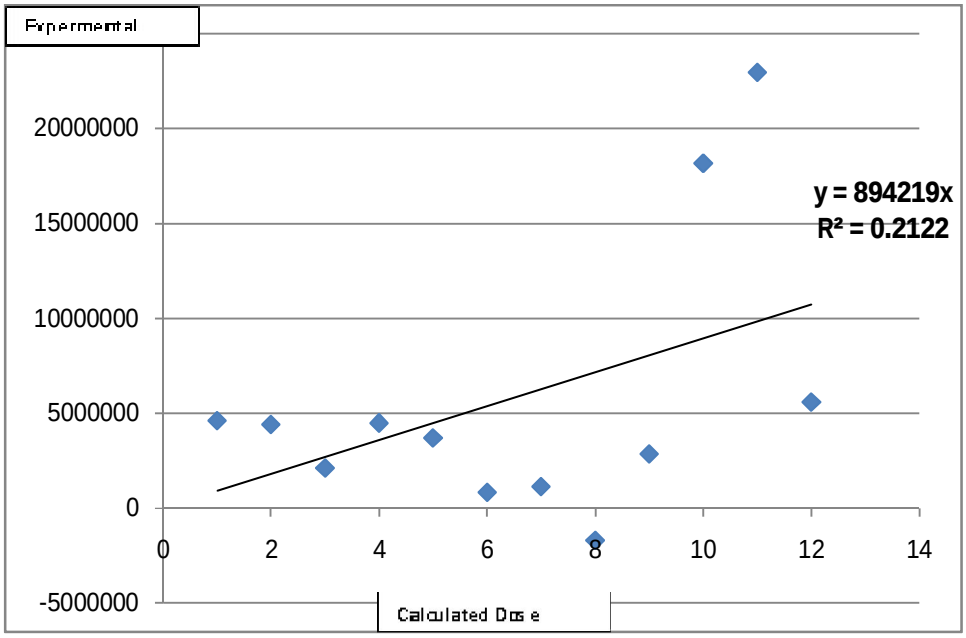

Fig. (3.3.2.6) internal validation of model 6 Exp. VS Cal. Dose

3.3.3 External validation

3.3.3.1 External validation of Desoxy AP compounds QSAR models

A group composed of six anthrapyrazole compounds were chosen with biological activities (expressed as optimum dose) not less than the least biological activity of any of the compounds of the training set and not more than the biological activity any of the compounds having the highest value of the biological activity of the training set i.e. the biological activity values for the desoxy anthrapyrazole compounds are within the range between 6.25 and 400 optimum dose for P_{383} leukemia cell line. The optimum dose values of the test set were calculated using the chosen models to examine their predict ability, then the calculated values were graphed versus the experimental values and the R^2 values were quoted.

Model 1

This model was generated by regression analysis of the optimum dose versus total energy values calculated using PM_3 method and the heat of formation calculated by the MNDO method. The model shows considerably high R^2 values (0.981).

$$\textbf{Dose} = -323 - 4.64 \times 10^{-3} E_t + 3.95 \times 10^{-3} H_f$$

Compounds No. were chosen as a test set and the model was used to calculate the respective dose values (table below).

Table (3.3.3.1) the Experimental and the Calculated Dose values of the test set for model 1

Experimental Dose	Calculated Dose
100	35.34
200	50.26
200	93.33
12.5	42.11
25	57.62
25	34.03

The linearity of the model was shown graphically in the figure below:

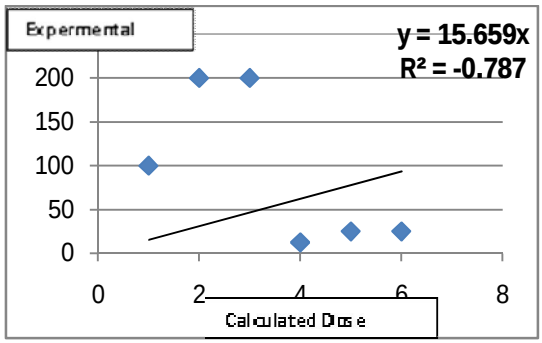

Fig (3.3.3.1) experimental dose VS calculated dose for desoxy AP model 1

Considering the findings of the external validation, table (3.3.3.1) and Fig. (3.3.31), although only one comparable result was found, but the goodness of the r^2 value of 0.787, makes this model, so far a considerably a good enough for prediction purposes.

The $\mathbf{r^2}$ of the external validation is 0.787; which is greater than the accepted value for the cross validation $\mathbf{r^2}$ value (> 0.5). Since the results of any QSAR model is just an estimative tool and no model can predict the exact values, this model is considered a good model for prediction of AP compounds towards the P_{388} leukemia cell lines.

Model 2

This model was generated by regression analysis of the optimum dose versus total energy values calculated using MNDO method and the dipole moment calculated by the PM_3 method. The model shows a considerably high $\mathbf{R^2}$ values (**R-Sq = 90.8% R-Sq (adj)** and a good s- value (**S = 38.4215**).

$$\mathbf{Dose = 227 + 2.5 \times 10^{-3} E_t - 2.55 \times 10^{-1} \mu_{PM3}}$$

Table (3.3.3.2) the Experimental and the Calculated Dose values of the test set for model 2

experimental Dose	**calculated Dose**
100	28.47
200	40.77
200	115.83
12.5	24.8
25	85.93
25	97.57

The same test set was used to examine the linearity of the model(Fig.3.3.3.2)

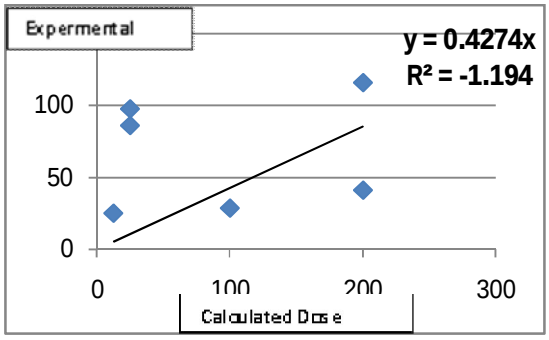

Fig.(3.3.2.2) experimental dose VS calculated dose for desoxy AP model 2

The findings of the external validation, table (3.3.2.2) and figure (3.3.2.2), no comparable results are found, and a poor r^2 value for the external validation of -1.194 was quoted. The minus sign indicates no correlation, so this model cannot be used for prediction purposes.

Model 3

This model was generated by regression analysis of the optimum dose versus total energy values calculated using PM_3 method and the molecular refractivity. The model shows a good $\mathbf{R^2}$ values (**R-Sq = 61.3% R-Sq (adj) = 52.7%)** and a little pit high s- value (**S = 74.7089**).

$$\mathbf{Dose = - 307 - 3.31 \times 10^{-3} E_t + 6.94\ MR}$$

Table (3.3.2.3) the Experimental and the Calculated Dose values of the test set for model 3

Experimental Dose	Calculated Dose
100	-6.54
200	-23.07
200	5.5
12.5	64.22
25	-17.77
25	194.17

The same test set was used to examine the linearity (fig.3.3.2.3).

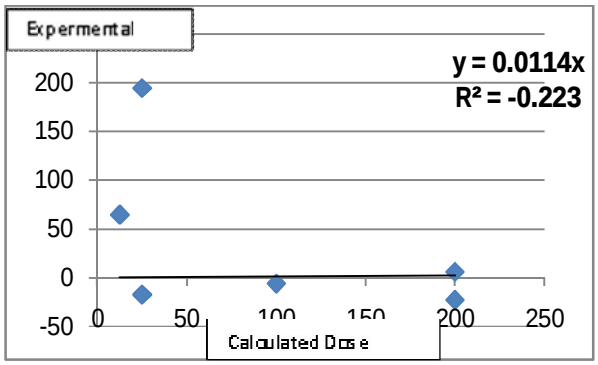

Fig.(3.3.2.3) experimental dose VS calculated dose for desoxy AP model 3

The findings of the external validation, table (3.3.2.3) and Fig.(3.3.2.3), no comparable results were found and a poor r^2 value for the external validation of 0.223 was quoted, so this model cannot be used for prediction purposes.

Model 4

This model was generated by regression analysis of the optimum dose versus total energy values calculated using PM_3 method and the molar volume .The model shows a good **R^2** values (**R-Sq = 62.7% R-Sq (adj) = 54.4%)** and a little pit high s- value (**S = 73.2876**).

$$\textbf{Dose} = -85 - 3.18 \times 10^{-3} E_t + 1.79\ MV$$

Table (3.3.2.4) the Experimental and the Calculated Dose values of the test set for model 4

Experimental Dose	Calculated Dose
100	9.61
200	-9.84
200	20.56
12.5	62.07
25	-11.31
25	184.28

The same test set was used to examine the linearity of the model Fig.(3.3.2.4

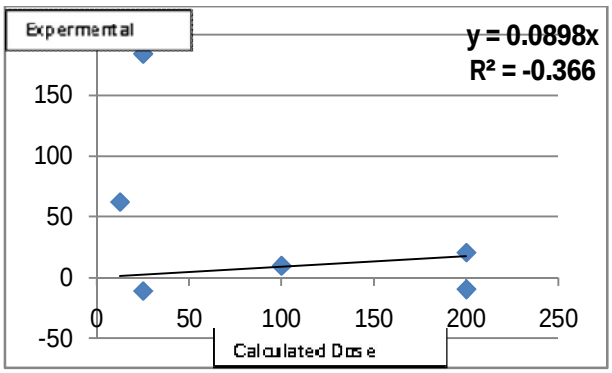

Fig.(3.3.2.4) experimental dose VS calculated dose for desoxy AP model

The findings of the external validation, table(3.3.3.4) and Fig.(3.3.2.4), no comparable results were found and a poor r^2 value for the external validation of -0.366 was quoted, so this model cannot be used for prediction purposes.

Model 5

This model was generated by regression analysis of the optimum dose versus total energy values calculated using PM_3 method and the polarizability. The model shows a good $\mathbf{R^2}$ values (**R-Sq = 61.2% R-Sq (adj) = 52.6%)** and a little pit high s- value (**S = 74.7357**).

$$\textbf{Dose} = -85 - 3.18 \times 10^{-3} E_{t\ PM3} + 1.79\ \textbf{polarizability}$$

Table (3.3.2.5) the Experimental and the Calculated Dose values of the test set for model 5

Experimental Dose	Calculated Dose
100	71.03
200	-25.43
200	5.11
12.5	63.92
25	-18.22
25	193.68

The same test set was used to examine the linearity of the model(Fig3.3.2.5).

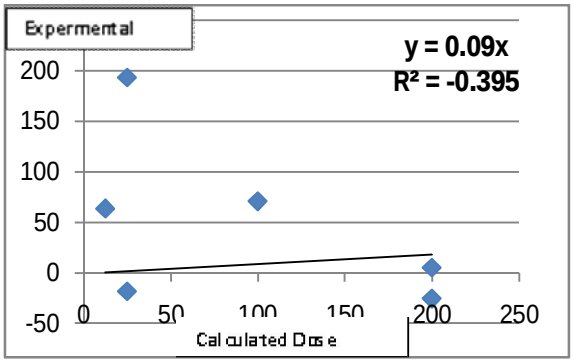

Fig.(3.3.2.5) experimental dose VS calculated dose for desoxy AP model 5

The findings of the external validation, table (3.3.2.5) and Fig.(3.3.2.5), no comparable results were found, and a poor r^2 value for the external validation of -0.395 was quoted, so this model cannot be used for prediction purposes.

Model 6

This model was generated by regression analysis of the optimum dose versus dipole moment values calculated using PM_3 method and the molar refractivity. The model shows a good $\mathbf{R^2}$ values (**R-Sq = 73.5% R-Sq (adj) = 67.7%)** and a little pit high s- value (**S = 61.7437**).

$$\textbf{Dose} = 165 + 2.54\ \textbf{MR} - 0.227\ \mu_{PM3}$$

Table (3.3.2.6) the Experimental and the Calculated Dose values of the test set for model 6

Experimental Dose	Calculated Dose
100	32.333
200	31.8
200	68.08
12.5	44.09
25	-32.41
25	-62.35

The same test set was used to examine the linearity of the model Fig. (3.3.2.6)

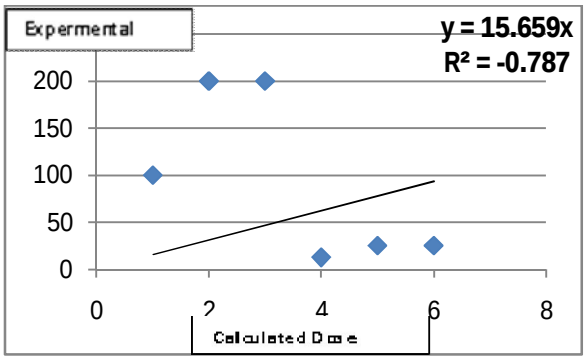

Fig. (3.3.2.6) experimental dose VS calculated dose for desoxy AP model 6

The findings of the external validation, table (3.3.2.6) and Fig.(3.3.2.6), no comparable results were found and a poor r^2 value for the external validation of -0.747 was quoted, so this model cannot be used for prediction purposes.

Model 7

This model was generated by regression analysis of the optimum dose versus dipole moment values calculated using PM_3 method and the molar volume. The model shows a good $\mathbf{R^2}$ values (**R-Sq = 73.9% R-Sq (adj) = 68.1%)** and a little pit high s- value (**S = 61.3404**).

$$\mathbf{Dose = 239 + 0.68\ MV - 0.223\ \mu_{PM}}$$

Table (3.3.2.7): the Experimental and the calculated dose values of the test set for model 7

Experimental Dose	Calculated Dose
100	38.49
200	36.65
200	74.6
12.5	44.92
25	23.29
25	-59.19

The same test set was used to examine the linearity of the model Fig. (3.3.2.7)

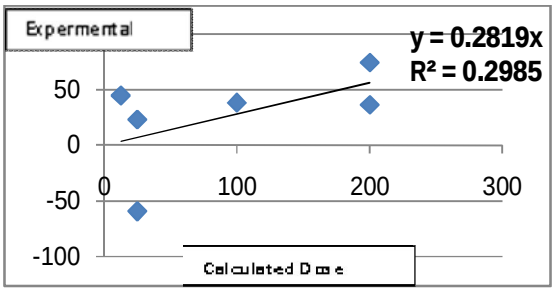

Fig.(3.3.2.7) experimental dose VS calculated dose desoxy AP model 7

The findings of the external validation, table (3.3.2.7) and Fig.(3.3.2.7), a poor r^2 value for the external validation of 0.2985 was quoted, so this model cannot be used for prediction purposes.

Model 8

This model was generated by regression analysis of the optimum dose versus dipole moment values calculated using **PM_3** method and the Log P. The model shows a good **R^2** values (**R-Sq = 72.2% R-Sq (adj) = 66.0%)** and a little pit high s- value (**S = 63.3502**).

Dose = 374 + 16.3 Log P - 0.204 μ_{PM3}

Table (3.3.2.8) the Experimental and the Calculated Dose values of the test set for model 8

Experimental Dose	Calculated Dose
100	64.1
200	64.52
200	102.32
12.5	56.32
25	16.13
25	-70.08

The same test set was used to examine the linearity of the model

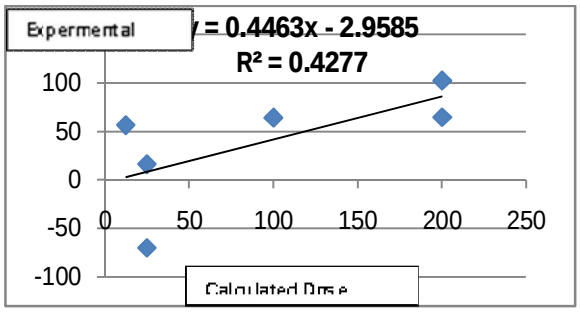

Fig.(3.3.2.8) experimental dose VS calculated dose desoxy AP model 8

The findings of the external validation, table (3.3.2.8) and graph (3.3.2.8), a poor r^2 value of 0.4277 was quoted, so this model cannot be used for prediction purposes.

Model 9

This model was generated by regression analysis of the optimum dose versus dipole moment values calculated using PM_3 method and the polarizability. The model shows a good $\mathbf{R^2}$ values (**R-Sq = 73.5% R-Sq (adj) = 67.7%)** and a little pit high s- value (**S = 61.7509**).

$$\textbf{Dose} = 165 + 6.41\ \textbf{polarizability} - 0.227\ \mu_{PM3}$$

Table (3.3.2.9) the Experimental and the Calculated Dose values of the test set for model 9

Experimental Dose	Calculated Dose
100	35.34
200	50.26
200	93.33
12.5	42.11
25	57.62
25	34.03

The same test set was used to examine the linearity of the model

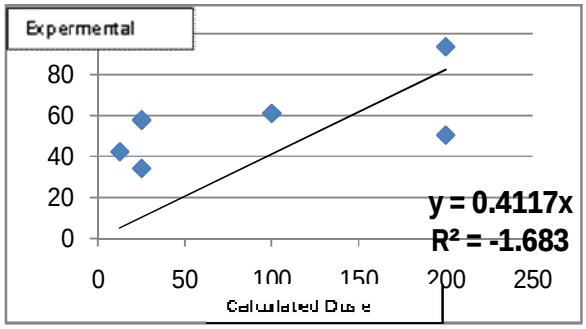

Fig.(3.3.2.9) Experimental dose VS Calculated dose desoxy AP model 9

The findings of the external validation, table (3.3.2.9) and Fig.(3.3.2.9) a poor r^2 value of -1.683 was quoted, so this model cannot be used for prediction purposes.

Model 10

This model was generated by regression analysis of the optimum dose versus dipole moment values calculated using **PM_3** method and the heat of formation MNDO. The model shows a good **R^2** values (**R-Sq = 82.9% R-Sq (adj) = 79.1%)** and a little pit high s- value (**S = 49.5976**).

$$\textbf{Dose} = 176 + 1.29 \times 10^{-3} H_f - 2.48 \times 10^{-1} \mu_{PM3}$$

Table (3.3.2.10) the Experimental and the Calculated Dose values of the test set for model 10

Experimental Dose	Calculated Dose
100	35.34
200	50.26
200	93.33
12.5	42.11
25	57.62
25	34.03

The same test set was used to examine the linearity of the model

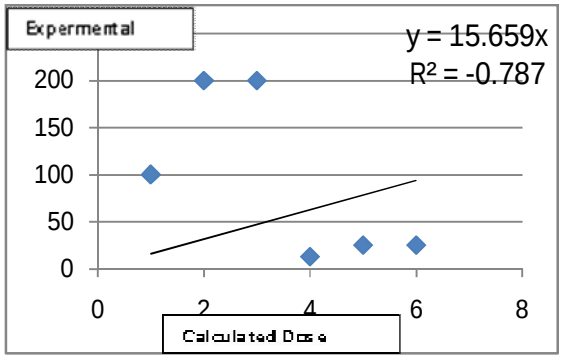

Fig. (3.3.2.10) Experimental dose VS Calculated dose desoxy AP model 10

The findings of the external validation, table (3.3.2.10) and Fig.(3.3.2.10) and a poor r^2 value of -.787 was quoted, so this model cannot be used for prediction purposes.

3.3.3.2 External validation of the Dihydroxy AP compounds QSAR models

A group composed of six dihydroxy anthrapyrazole compounds were chosen with a biological activities (expressed as $1/IC_{50}$) not less than the biological activity of the least compound of the training set and not exceeding the biological activity of the compound having the highest value of the biological activity of the training set i.e. the biological activity values for the dihydroxy anthrapyrazole compounds are within the range between 625000 and 2.173913043×10^7 reciprocal of the inhibitory concentration for L_{1210} leukemia cell line. The reciprocal of the inhibitory concentration values of the test set were calculated using the chosen models to examine their prediction, then the calculated values were graphed versus the experimental values and the R^2 values were quoted.

Model 1

This model was generated by regression analysis of the reciprocal of the inhibitory concentration values versus dipole moment values calculated using AM_1 method and the total energy using PM_3. The model shows a good $\mathbf{R^2}$ values (**R-Sq = 73.5% R-Sq (adj) = 67.6%)** and an s- value (**S = 3380243**).

$$\mathbf{1/IC_{50} = 20464882 - 10789\ \mu_{AM1} + 32.7\ E_{t_{PM3}}}$$

The same test set was used to examine the linearity of the model and the calculated dose values (table below) were shown graphically versus the experimental ones in the figure below:

Table (3.3.3.2.1) the Experimental and the calculated $1/IC_{50}$ values of the test set for Dihydroxy AP model 1

Experimental $1/IC_{50}$	**Calculated $1/IC_{50}$**
2222222	1629262.734
1282051	4126553.631
454545.5	-643764.761
2272727	3284999.613
1041667	4892733.068
4545455	17064503.59

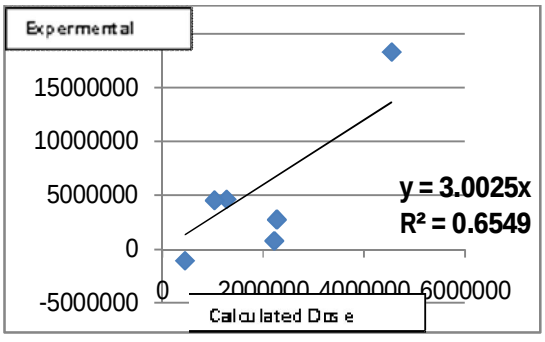

Fig. (3.3.3.2.1) Experimental dose VS Calculated $1/IC_{50}$ of dihydroxy AP model 1

Model 2

This model was generated by regression analysis of the reciprocal of the inhibitory concentration values versus dipole moment values calculated using **AM_1** method and the molar refractivity. The model shows a good **R^2** values (**R-Sq = 69.1% R-Sq (adj) = 62.2%)** and an s- value (**S = 3652707**).

$$\mathbf{1/IC50 = 12754339 - 10091\ \mu_{AM1} + 95485\ MR}$$

The same test set was used to examine the linearity of the model and the calculated dose values (table below) were shown graphically versus the experimental ones in the figure below:

Table (3.3.3.2.2) the Experimental and the calculated $1/IC_{50}$ values of the test set for dihydroxy AP model 2

Experimental $1/IC_{50}$	**Calculated $1/IC_{50}$**
2222222	752472.356
1282051	4628304.049
454545.5	-1097182.979
2272727	2740697.997
1041667	4524841.232
4545455	18289534.43

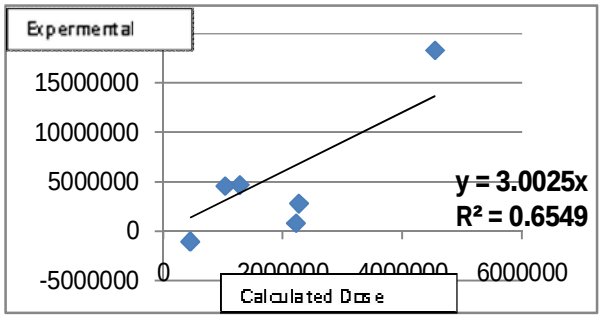

Fig.(3.3.3.2.2) Experimental dose VS Calculated $1/IC_{50}$ of dihydroxy AP model 2

Considering the findings of the external validation, table (3.3.3.2.2) and Fig.(3.3.3.2.2), although only one comparable result was found, but the goodness of the r^2 value of 0.6549, makes this model, so far a considerably a good enough for prediction purposes.

Model 3

This model was generated by regression analysis of the reciprocal of the inhibitory concentration values versus dipole moment values calculated using AM_1 method and the molar volume. The model shows a good $\mathbf{R^2}$ values (**R-Sq = 68.9% R-Sq (adj) = 62.0%)** and an s- value (**S = 3659070**).

$$\mathbf{1/IC_{50} = 16158904 + 24782\ MV - 10025\ \mu_{AM1}}$$

The same test set was used to examine the linearity of the model and the calculated dose values (table below) were shown graphically versus the experimental ones in the figure below:

Table (3.3.3.2.3) the Experimental and the calculated $1/IC_{50}$ values of the test set for dihydroxy AP model 3

Experimental $1/IC_{50}$	**Calculated $1/IC_{50}$**
2222222	23013605.2
1282051	4649193.675
454545.5	-25078730.48
2272727	22954128.4
1041667	22778176.2
4545455	18320129.1

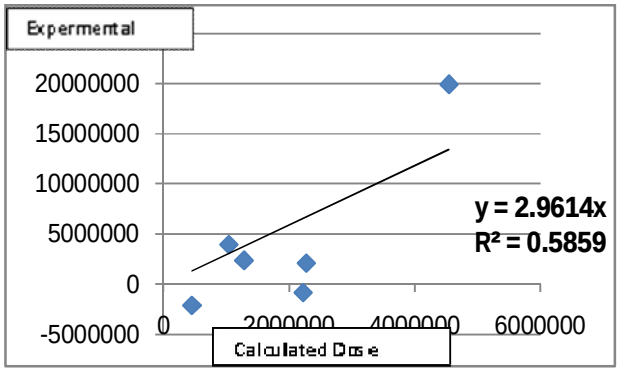

Fig.(3.3.3.2.3) Experimental dose VS Calculated $1/IC_{50}$ of dihydroxy AP model 3

Considering the findings of the external validation, table (3.3.3.2.3) and Fig.(3.3.3.2.3), although only one comparable result was found, but the goodness of the r^2 value of 0.5859, makes this model, also good enough and comparable to the above model for prediction purposes.

Model 4

This model was generated by regression analysis of the reciprocal of the inhibitory concentration values versus dipole moment values calculated using **AM_1** method and Log P. The model shows a good **R^2** values (**R-Sq = 67.4% R-Sq (adj) = 60.2%)** and an s- value (**S = 3747785**).

$1/IC_{50} = 21533904 + 490098$ Log P - 9530 μ_{AM1}

The same test set was used to examine the linearity of the model and the calculated dose values (table below) were shown graphically versus the experimental ones in the figure below:

Table (3.3.3.2.4) the Experimental and the calculated $1/IC_{50}$ values of the test set for dihydroxy AP model 4

Experimental $1/IC_{50}$	**Calculated $1/IC_{50}$**
2222222	23165930.34
1282051	4890667.75
454545.5	-2056575.09
2272727	2623402.51
1041667	4473303.38
4545455	18260400.16

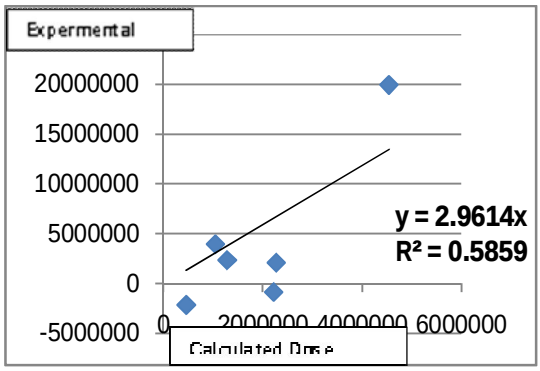

Fig.(3.3.3.2.4) Experimental dose VS Calculated $1/IC_{50}$ of dihydroxy AP model 4

Considering the findings of the external validation, table (3.3.3.2.4) and Fig.(3.3.3.2.4), although two comparable result was found, but the goodness of the r^2 value of 0.6549, makes this model, so far a considerably a good enough for prediction purposes.

Model 5

This model was generated by regression analysis of the reciprocal of the inhibitory concentration values versus dipole moment values calculated using **AM_1**

method and the polarizability. The model shows a good $\mathbf{R^2}$ values (**R-Sq = 69.1% R-Sq (adj) = 62.2%)** and an s- value (**S = 3653091**).

$$\mathbf{1/IC_{50} = 12768322 - 10091\ \mu_{AM1} + 240545\ Polarizability}$$

The same test set was used to examine the linearity of the model and the calculated dose values (table below) were shown graphically versus the experimental ones in the figure below:

Table (3.3.3.2.5) the Experimental and the calculated $1/IC_{50}$ values of the test set for dihydroxy AP model 5

Experimental $1/IC_{50}$	**Calculated $1/IC_{50}$**
2222222	752123.206
1282051	4622248.999
454545.5	-1100142.279
2272727	2703655.547
1041667	4084144.232
4545455	18290713.08

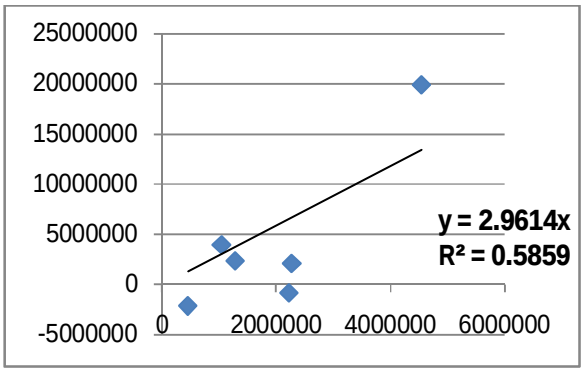

Fig.(3.3.3.2.5) Experimental dose VS Calculated $1/IC_{50}$ of dihydroxy AP model 5

Considering the findings of the external validation, table (3.3.3.2.5)and

Fig.(3.3.3.2.5), although two comparable result was found, but the goodness of the r^2 value of 0.58559, makes this model, so far a considerably a good enough for prediction purposes.

Model 6

This model was generated by regression analysis of the reciprocal of the inhibitory concentration values versus dipole moment values calculated using MNDO method and the heat of formation calculated using MNDO method. The model shows a good $\mathbf{R^2}$ values (**R-Sq = 73.0% R-Sq (adj) = 67.0%)** and an s- value (**S = 3411467**).

$$\mathbf{1/IC50 = 17407936 - 15968\ \mu_{MNDO} + 65.0\ H_{f\ MNDO}}$$

The same test set was used to examine the linearity of the model and the calculated dose values (table below) were shown graphically versus the experimental ones in the figure below:

Table (3.3.3.2.6) the Experimental and the calculated $1/IC_{50}$ values of the test set for dihydroxy AP model 6

Experimental $1/IC_{50}$	**Calculated $1/IC_{50}$**
2222222	-813350.596
1282051	2386896.348
454545.5	-2102145.8
2272727	2112184.56
1041667	3966917.12
4545455	19925362.35

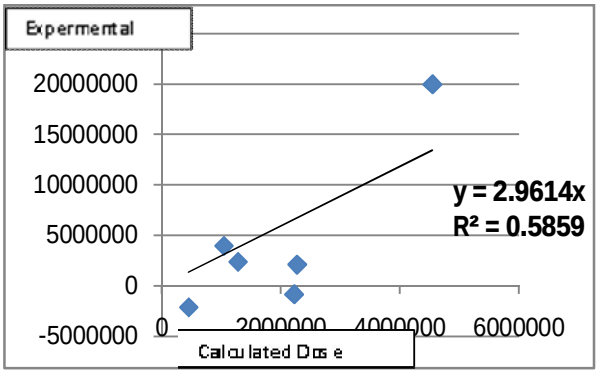

Fig.(3.3.3.2.6) Experimental dose VS Calculated $1/IC_{50}$ of dihydroxy AP model 6

Considering the findings of the external validation, table (3.3.3.2.6) and Fig. (3.3.3.2.6), although only one comparable result was found, but the goodness of the r^2 value of 0.5859, makes this model, good enough for prediction purposes.

3.4 Choosing the best model for desoxy subgroup

To make the comparison easy we tabulate the information concerning each of the models in the table below.

Table (3.4) comparison of the desoxy models results

Model	r^2	r^2adj	r^2CV_{int}	r^2CV_{ext}	S	No. of predicted compound (internal set)	No. of predicted compound (external set)
No.1	**0.921**	**0.981**	**0.921**	**0.787**	**17.27**	**7**	**1**
No.2	**0.908**	**0.885**	**-0.182**	**-1.194**	**38.42**	**2**	**0**
No.3	**0.613**	**0.527**	**0.245**	**0.223**	**74.71**	**2**	**0**
No.4	**0.627**	**0.544**	**-0.238**	**-0.366**	**73.29**	**0**	**4**
No.5	**0.612**	**0.526**	**-0.208**	**-0.395**	**74.74**	**2**	**0**
No.6	**0.735**	**0.627**	**-0.208**	**-0.747**	**61.74**	**3**	**0**
No.7	**0.739**	**0.681**	**0.216**	**0.2985**	**61.34**	**0**	**1**
No.8	**0.722**	**0.660**	**-0.238**	**0.4277**	**63.35**	**1**	**1**
No.1	**0.735**	**0.677**	**-0.215**	**-1.683**	**61.75**	**1**	**2**
No.10	**0.829**	**0.791**	**-0.215**	**-0.787**	**49.60**	**3**	**1**

It is now very obvious that only that have the highest r^2, adj r^2, r^2CV_{int} and r^2CV_{ext} values and a higher number of predicted compounds is the model No.1, the rest of the models show negative r^2CV_{int} and r^2CV_{ext}; which indicates no correlation, except No.3; which has no negative r^2CV_{int} and r^2CV_{ext} values but these values are less than (0.5). Other models, like model No.4 have high number of the predicted compounds but is not more than a coincidence because r^2CV_{int} and r^2CV_{ext} both are negative.

3.5 Choosing the best model for dihydroxy subgroup

To make the comparison easy we tabulate the information concerning each of the models in the table below.

Table (3.5) comparison of the Dihydroxy models results

Model	r^2	r^2 adj	r^2CV_{int}	r^2CV_{ext}	S	No. of predicted compound (internal set)	No. of predicted compound (external set)
No.1	**0.735**	**0.676**	**0.2122**	**0.6549**	**3380243**	**1**	**1**
No.2	**0.691**	**0.622**	**0.2122**	**0.6549**	**3652707**	**3**	**1**
No.3	**0.689**	**0.620**	**0.2122**	**0.5859**	**3659070**	**2**	**2**
No.4	**0.674**	**0.602**	**0.2122**	**0.6549**	**3747785**	**3**	**2**
No.5	**0.691**	**0.622**	**0.3086**	**0.6589**	**3653091**	**4**	**1**
No.6	**0.730**	**0.760**	**0.2122**	**0.6549**	**3411467**	**3**	**1**

It is now very obvious that all the models have high r^2, adj, and r^2CV_{ext} values and a same value for r^2CV_{int} except model No. 5 which has a little bit higher value and still less than the recommended value (>0.5).

Although a high value of r^2 CV alone is necessary condition for a model to have a high predictive power; it is not a sufficient condition. It is proven that the only way to estimate the true predictive power of a model is to test it on a

sufficiently large collection of compounds from an external test set. The external set must include no less than five compounds, whose activities and structures must cover the range of activities and structures of compounds from the training set. This application is necessary for obtaining trustful statistics for comparison between the observed and predicted activities for these compounds. Besides high r^2 CV, a reliable model should be also characterized by a high correlation coefficient between the predicted and observed activities of compounds from a test set of molecules that was not used to develop the models. The model that meats such requirements is model No.4

3.6 Conclusion

Since its conception in the 1960s, a keen interest for QSAR has been observed in the drug discovery area to enable the design of safe and potent drug candidates. During drug discovery and development phases, pharmacodynamics and pharmokinetic profiles of molecules can be derived using QSAR models. A great deal has been achieved in recent years in terms of developing and harmonizing the formats which report the results of QSAR methods. This has been an essential step towards ensuring the reproducibility of predictions and transparency in their interpretation. Furthermore, an increasing and arguably overwhelming array of different computational tools are being developed to implement QSAR methods. An increasing number of these tools are being made freely available, and in some cases they are also open to the scientific community for further development. However, additional efforts are still required to extend the applicability domains and the accuracies of the underlying models and to create user-guided workflows that facilitate their integrated use. Another important challenge remains in developing a common understanding of how best to integrate multiple predictions and existing experimental data in weight of- evidence

approaches. The way forward will be to develop a general framework that encourages transparency as well as carefully documented case studies that show how the framework has been applied to specific chemicals for specific regulatory purposes. Any attempt to develop a rigid set of acceptance criteria is unlikely to be productive because this ignores the context-dependent nature of the regulatory decision-making process. Already, it is possible to generate a huge amount of information by simply pressing a button, but this does not replace the need for expert interpretation and consensus in the regulatory use of QSAR methodology.

Referring to the above discussion it was concluded that the best model that can be used to predict the biological activity of desoxy AP subgroup towards the P_{388} leukemia cell lines is the model No. 1.

When considering the dihydroxy subgroup it was concluded that the best model that can be used to predict the biological activity of dihydroxy AP subgroup towards the L_{1210} leukemia cell lines is the model No. 4.

3.7 Recommendations:

- Trials of other QSAR models can be built using other programmes.
- To predict QSAR models for chemical compounds, other than anthrapyrazole following same methods used in this study
- Trying of 3D-QSAR techniques ,can be also mentioned as a recommendation at the end of this thesis.