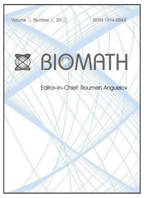
Biomath Forum
BIOMATH
http://www.biomathforum.org/biomath/index.php/biomath/



Descriptor-based Fitting of Lysophosphatidic Acid Receptor 3 Antagonists into a Single Predictive Mathematical Model

Olaposi Idowu Omotuyi, Hiroshi Ueda
Department of Pharmacology and Therapeutic Innovation
University Graduate School of Biomedical Sciences, 852-8521
Nagasaki, Japan
Email: bbis11r104@cc.nagasaki-u.ac.jp

Received: 6 June 2013, accepted: 6 October 2014, published: 16 October 2014

Abstract—**Sixty six diverse compounds previously reported as Lysophosphatidic Acid Receptor (LPA_3) inhibitors have been used to derive a mathematical model based on partial least square (PLS) clustering of 41 molecular descriptors and pIC_{50} values. The pre- and post- cross-validated correlation coefficient (R^2) is 0.94462 (RMSE=0.21390) and 0.74745 (RMSE=0.49055) respectively. Bivariate contingency analysis tools implemented in MOE was used to prune the descriptors and refit the equations at a descriptor-pIC_{50} correlation coefficient of 0.8 cut-off. A new equation was derived with R^2 and RMSE values estimated at 0.88074 and 0.31388 respectively. Both equations correctly predicted the 95% of the pIC_{50} values of the test dataset. Principal component analysis (PCA) was also used to reduce the dimension and linearly transform the raw data; 8 principal components sufficiently account for more than 98% of the variance of the dataset. The numerical model derived here may be adapted for screening chemical database for LPA_3 antagonism.**

Keywords-**upscaling; LPA_3; LPA_3 antagonists; Mathematical Model; PCA; Molecular descriptors**

I. Introduction

Quantitative structure activity relationship (QSAR) allows statistical analysis of experimental data and building of predictive mathematical models from the dataset. The numerical models built using this approach has been successfully implemented in screening of large database of chemical compounds for hit-compound detection [1]. In the presence of experimental dataset [2], the success of QSAR depends on two key factors: array of descriptors that optimally represent the structural parameters required for molecular interaction or reactions [3] and an appropriate statistical learning and validation algorithms [4]. In practice, physical properties descriptors (1D-descriptor), pharmacophore descriptors (2D-descriptors) and geometrical descriptors (3D-descriptors, often requires prior knowledge of target protein binding-pocket) are the most commonly used descriptor types for QSAR modeling [5,6,7]. We seek to answer a single question here, what combination of

Citation: O. Omotuyi, H. Ueda, Descriptor-based Fitting of Lysophosphatidic Acid Receptor 3 Antagonists into a Single Predictive Mathematical Model, Biomath 3 (2014), 1410061, http://dx.doi.org/10.11145/j.biomath.2014.10.061

molecular predictors would numerically and accurately predict the experimental antagonist activities of LPA$_3$ inhibitors? When answered, the mathematical relationship derived from the descriptors will enable screening of chemical databases for compounds exhibiting LPA$_3$ antagonism required for the treatment of diseased conditions such as ovarian cancer [8] and neuropathic pain [9] with LPA$_3$ etiology.

II. STATISTICAL BASIS OF QSAR MODELING USING PARTIAL LEAST SQUARE METHOD

The QSAR/PLS modeling equations and algorithms have been well described in MOE documentations [10]. Given m molecules of a training dataset, suppose that each of the molecules is described by an n-vector of descriptors $x_i = (x_{i1}, ..., x_{in})$, for one of the molecules denoted as i. Let y_i be a representation of the experimental result (pIC_{50}) for a molecule i. A linear model for y (the experimental result) is given by Eq. (1) [11].

$$y = a_0 + a^T X\,, \tag{1}$$

where a_0 is a scalar, and a^T is a n-vector. If each molecule has an importance weight (non-negative) w representing the relative probability that the associated molecule will be encountered, and that the sum of all the weights are designated as W. The mean square error is given as Eq. (2) [12].

$$MSE_{a_0,a} = \frac{1}{w}\sum_{i=1}^{m}\left[y_i - (y = a_0 + a^T X_i)\right]^2 . \tag{2}$$

Differentiating MSE with respect to the parameters satisfying the normal Eqs (3,4,5,6 &7) solvable by matrix diagonalization:

$$a_0 = y_0 - a^T X_i\,, \tag{3}$$

$$y_0 = \frac{1}{w}\sum_{i=1}^{m}\left[w_i y_i\right], \tag{4}$$

$$x_0 = \frac{1}{w}\sum_{i=1}^{m}\left[w_i x_i\right], \tag{5}$$

$$Sa = b = \frac{1}{W}\sum_{i=1}^{m}\left[w_i y_i (x_i - x_0)\right], \tag{6}$$

$$S = \frac{1}{w}\sum_{i=1}^{m}\left[w_i (x_i - x_0)(x_i - x_0)^T\right]. \tag{7}$$

Starting from the normal equations above, an estimate of a can be computed if columns of the weight matrix (G_A) (Eq. (8)) is obtained through Gram-Schmidt orthogonalization [13] of the vectors generated by Krylov sequence $b, Sb, S^2b, ..., S^{A-1}b$ [14]. The A^{th} PLS coefficient vector is then estimated using Eq. (9).

$$G_A = (g_i, g_2, \ldots, g_A)\,. \tag{8}$$

$$a = G_A(G_A^T S G_A)^{-1} G_A^T b\,. \tag{9}$$

Noting that g_i is the column vectors of length n and A is the degree of the PLS fit; an integer less than or equals n. MOE [10] descriptor calculator was used to generate the numerical representations (a_aro, ASA, ASA_H, a_hyd, SlogP, SlogP_VSA0, SlogP_VSA1, SlogP_VSA2, SlogP_VSA3, SlogP_VSA4, SlogP_VSA5, SlogP_VSA6, SlogP_VSA7, SlogP_VSA8, SlogP_VSA9, SMR_VSA0, SMR_VSA1, SMR_VSA2, SMR_VSA3, SMR_VSA4, SMR_VSA5, SMR_VSA6, SMR_VSA7, a_acc, Kier1, Kier2, Kier3, KierA1, KierA2, KierA3, KierFlex, chi0, chi0v, chi0v_C, chi0_C, chi1, chi1v, chi1v_C, chi1_C, chiral, chiral_u) of the 66 (Supplementary fig. 1) randomly selected LPA$_3$ antagonists retrieved from the European Institute of Bioinformatics dataset (https://www.ebi.ac.uk/chembl/) representing our training dataset (CHEMBL3250). Using the PLS method as described above, Eq. (10) was generated relating the descriptors to the pIC_{50} with a correlation coefficient (R^2) 0.94462 ($RMSE = 0.21390$) (Fig. 1, blue circles and line); when cross validated, R^2 was estimated as 0.74745 $(RMSE = 0.49055)$.

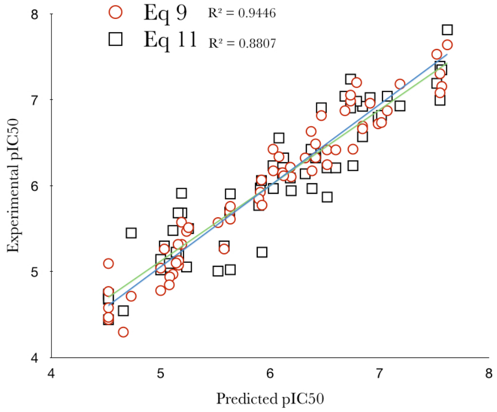


Fig. 1: Scatter plot of the experimental pIC_{50} vs. pIC_{50}-predictions of Eq. (10) (blue) and Eq. (12) (green).

$$
\begin{aligned}
pIC_{50} = {} & \\
& 3.57363 - 0.25353 \cdot a\_aro - 0.00361 \cdot ASA \\
& + 0.23510 \cdot a\_hyd + 0.05890 \cdot SlogP \\
& - 0.02287 \cdot SlogP\_VSA0 \\
& +0.00032 \cdot SlogP\_VSA1 + 0.03125 \cdot SlogP\_VSA2 \\
& -0.02059 \cdot SlogP\_VSA3 + 0.02954 \cdot SlogP\_VSA4 \\
& +0.07226 \cdot SlogP\_VSA5 + 0.02879 \cdot SlogP\_VSA6 \\
& +0.04687 \cdot SlogP\_VSA7 + 0.03836 \cdot SlogP\_VSA8 \\
& +0.06880 \cdot SlogP\_VSA9 + 0.04912 \cdot SMR\_VSA0 \\
& +0.02536 \cdot SMR\_VSA1 + 0.08743 \cdot SMR\_VSA2 \\
& +0.00289 \cdot SMR\_VSA3 - 0.01524 \cdot SMR\_VSA4 \\
& +0.04694 \cdot SMR\_VSA5 + 0.09067 \cdot SMR\_VSA6 \\
& - 0.01442 \cdot SMR\_VSA7 + 0.18393 \cdot a\_acc \\
& - 0.77650 \cdot Kier1 - 0.43968 \cdot Kier2 \\
& - 0.30735 \cdot Kier3 - 0.43752 \cdot KierA1 \\
& - 0.03578 \cdot KierA2 + 0.76916 \cdot KierA3 \\
& - 0.09573 \cdot KierFlex + 0.00332 \cdot chi0 \\
& + 0.55223 \cdot chi0v + 0.13554 \cdot chi0v\_C \\
& - 0.16530 \cdot chi0\_C + 0.59498 \cdot chi1 \\
& + 0.05911 \cdot chi1v - 0.93262 \cdot chi1v\_C \\
& - 1.22808 \cdot chi1\_C - 0.16986 \cdot chiral \\
& - 0.56204 \cdot chiral\_u.
\end{aligned} \tag{10}
$$

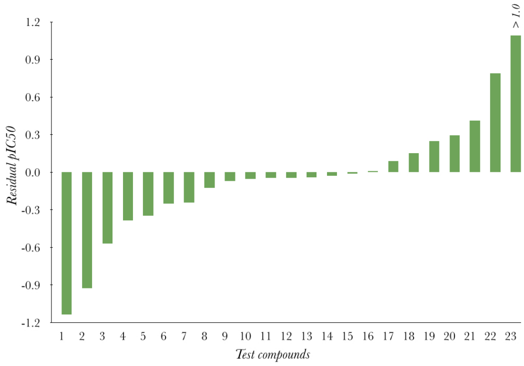


Fig. 2: Bar chart representations of the residual (Experimental pIC_{50}-Predicted pIC_{50} values of the test dataset. Only 1 out of tested compounds (compound 23, see supplementary Fig. 2 for structural details) showed > 1.0 pIC_{50} unit (indication of wrong prediction).

Noting that root mean square error (RMSE) is the square root of MSE function (Eq. (2)) at a given parameter value and the correlation coefficient (R^2) is 1-MSE/YVAR with values raging between 0 and 1 (0= no fit, 1 is perfect fit and YVAR is the sample variance of the y_i values). The predictive suitability of our equation was tested on 23 compounds (Supplementary Fig. 2) with experimentally determined IC_{50} for LPA$_3$ antagonism. If we assume that residual value above 1.0 pIC_{50} unit represents poor fitting. Our data (Fig. 3) suggest that Eq. (10) accurately predicted 22 of the 23 test compounds.

III. Descriptor contingency analysis

To determine the level of significance of each of the descriptors to the overall equation and we performed contingency analysis. The data presented here provides a window of decision on whether pruning of the descriptor set is required. In MOE [10], QSAR-contingency tool performs a bivariate contingency analysis for each descriptor and the experimental activity value and produces a table of correlation coefficients (Eq. (11)) for each descriptor given that X represents a randomly selected molecular descriptor and Y is a randomly selected activity value for a randomly selected sample m, $Var(X)$ and $Var(Y)$, then the covariance of the random variables X and Y is defined to be $Cov(X, Y) = E(XY) - E(X)E(Y)$ [10, 15].

$$R^2 = \frac{[E(XY) - E(X)E(Y)]^2}{Var(X)Var(Y)}\,. \qquad (11)$$

Given that the values of R^2 ranges from 0 to 1, and 1 represents a perfectly linear correlation, we therefore proposed that only descriptors R^2 values ≥ 0.8 are useful and that the descriptors outside this range can be pruned. Our data suggest that 31 out of the original 41 descriptors have R^2 values ≥ 0.8 (Fig. 3, Supplementary Table 1). With the exclusion of the descriptors with unsatisfactory coefficient, QSAR is re-calculated using the residual set of descriptors. New numerical relationship was generated (Eq. (12)) with R^2 (0.88074) and $RMSE$ values (0.31388). The scatter plot of the predicted pIC_{50} and the experimental values for the new Eq. (12) is given in Fig. 1 (green circles and line).

$$\begin{aligned}
&lpIC_{50} = \\
&2.23199 - 0.00516xASA - 0.00516xASA\_H \\
&- 0.48596xa\_hyd - 0.33917xSlogP \\
&-0.05298xSlogP_VSA0 - 0.03967xSlogP_VSA1 \\
&-0.02243xSlogP_VSA2 + 0.01681xSlogP_VSA7 \\
&+ 0.02107xSlogP_VSA9 \\
&-0.00757xSMR_VSA0 - 0.00087xSMR_VSA1 \\
&- 0.00089xSMR_VSA3 \\
&-0.01173xSMR_VSA4 + 0.00955xSMR_VSA5 \\
&- 0.01412xSMR_VSA6 \\
&- 0.02508xSMR_VSA7 - 0.26771xKier1 \\
&+ 0.15306xKier20.56650xKier3 \\
&- 0.30504xKierA2 + 0.98837xKierA3 \\
&- 0.28849xKierFlex + 0.48535xchi0 \\
&+ 0.90693xchi0v + 0.10234xchi0v_C \\
&+ 0.24407xchi0_C + 0.66154xchi1 \\
&+ 0.36006xchi1v - 1.03589xchi1v_C \\
&- 0.62474xchi1_C - 0.36725xa_aro\,. \qquad (12)
\end{aligned}$$

When this equation was used for predicting the pIC_{50} values of the test set, only one compound lies above the 1.0 pIC_{50} unit cutoff (data not shown). Thus, Eq. (12) is less bulky and as accurate as Eq. (10) in predicting LPA$_3$ antagonism.

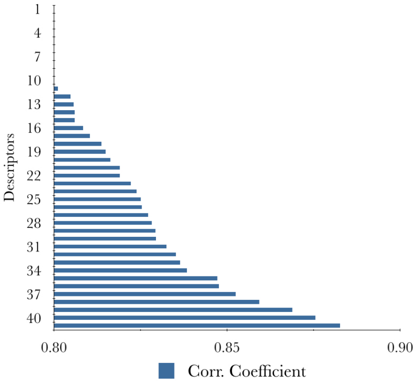


Fig. 3: Bar chart representations of Descriptor-experimental pIC_{50} correlation coefficient. Only 31 out of 41 descriptors lie above 0.8 coefficient cutoff.

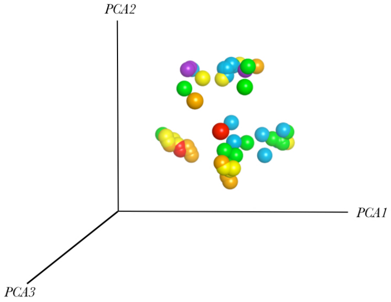


Fig. 4: The $3D$ plot of the first three principal components. Each point represents a compound in the training dataset and each colour represents a distinct cluster of pIC_{50} values.

IV. Principal component analysis of equation

We sought to further study the dataset descriptors along the principle components through the reduction of the dimensionality and linear transformation of the raw data [13]. Given the initial 66 training dataset compounds (represented as m) and for one of the compounds say i its descriptors are represented by n-vector of real numbers $x_i =$

$(x_{i1}, ..., x_{in})$, where $n = 1 - 31$, new Eq. (12). Assuming that each molecule i has an associated importance weight w_i, (non-negative, real number) and that the weights is relative probability that the associated molecule x_i will be encountered (adding up to 1); If W denotes the sum of all the weights then, the eigenvalues and eigenvectors for the final data are estimable from the raw data using Eq. (1). If S is a symmetric, semi-definite sample covariance matrix, S can be diagonalized such that $S = Q^T DDQ$ (Q is orthogonal, D is diagonal-sorted in descending order from top left to bottom right) [13, 14].

$$E(x) \approx \overline{x} = x_0 = \frac{1}{w} \sum_{i=1}^{m} [w_i x_i] \qquad (13)$$

$$Cov(x) \approx S = \frac{1}{w} \sum_{i=1}^{m} [w_i x_i x_i^T - \overline{x}\overline{x}^T]. \qquad (14)$$

The effect of the each of the principal components (eigenvectors) on the condition and the variance shows that nine (8) principal components sufficiently accounts for more than 98% of the variance in the dataset [15]. The $3D$-scatter plot of the first three principal components (PCA1, PCA2 and PCA3) with respect to pIC_{50} values is shown in Fig. (4); each point in the plot corresponds to a dataset molecule colored according to clustered pIC_{50} values.

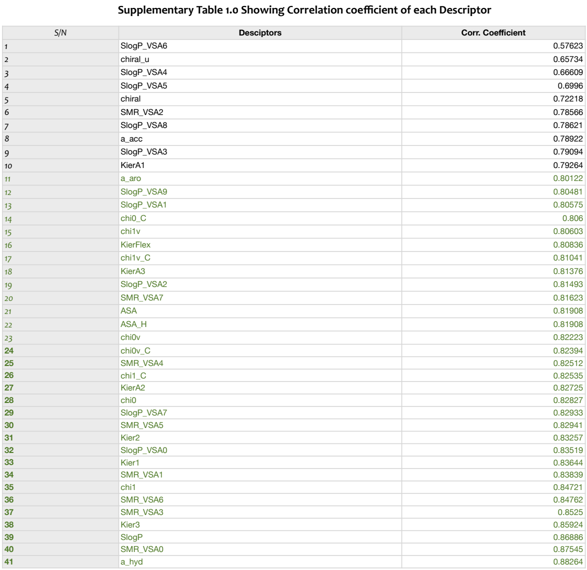
V. Conclusion

Given the good mathematical correlation between the set of descriptors and LPA_3 antagonism, it is not unusual to propose that the equation is prejudiced for those set of compounds with highly related descriptor properties and therefore may not be a universal formula for LPA_3 antagonist screening. That said, it will however capture the compounds with structural properties found within the dataset accurately and therefore may be piped as into ligand-based screening protocol for more successful hit-compound identification.

Acknowledgment

This work was supported by Platform for Drug Discovery, Informatics, and Structural Life Science from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

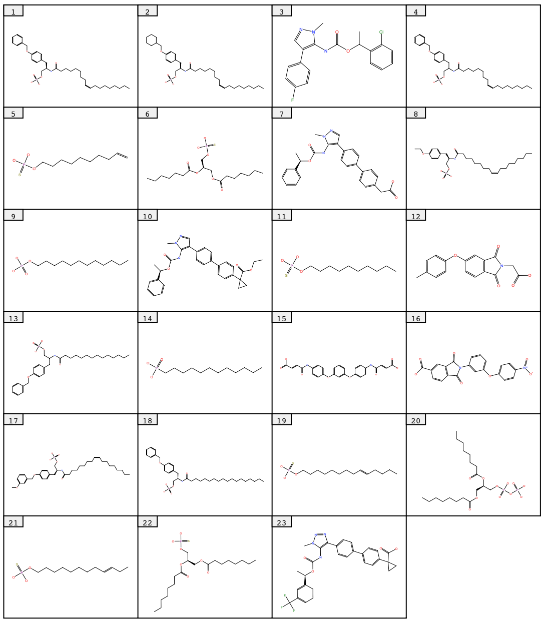
Appendix



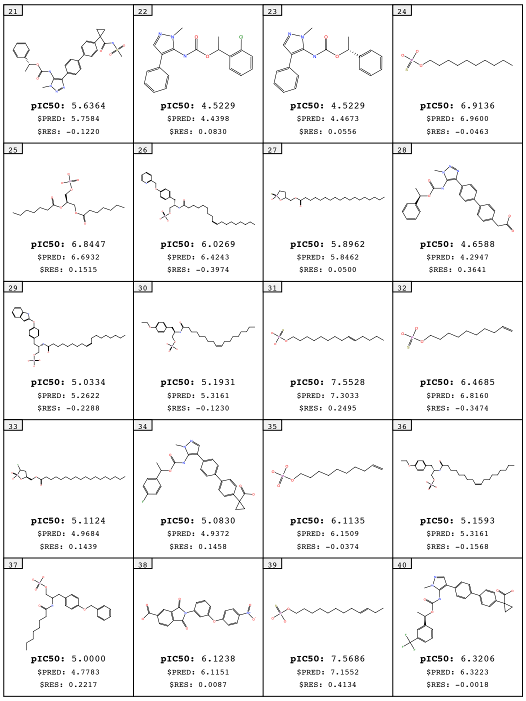
Supplementary Table 1.0 Showing Correlation coefficient of each Descriptor

S/N	Desciptors	Corr. Coefficient
1	SlogP_VSA6	0.57623
2	chiral_u	0.65734
3	SlogP_VSA4	0.66609
4	SlogP_VSA5	0.6996
5	chiral	0.72218
6	SMR_VSA2	0.78566
7	SlogP_VSA8	0.78621
8	a_acc	0.78922
9	SlogP_VSA3	0.79094
10	KierA1	0.79264
11	a_aro	0.80122
12	SlogP_VSA9	0.80481
13	SlogP_VSA1	0.80575
14	chi0_C	0.806
15	chi1v	0.80603
16	KierFlex	0.80836
17	chi1v_C	0.81041
18	KierA3	0.81376
19	SlogP_VSA2	0.81493
20	SMR_VSA7	0.81623
21	ASA	0.81908
22	ASA_H	0.81908
23	chi0v	0.82223
24	chi0v_C	0.82394
25	SMR_VSA4	0.82512
26	chi1_C	0.82535
27	KierA2	0.82725
28	chi0	0.82827
29	SlogP_VSA7	0.82933
30	SMR_VSA5	0.82941
31	Kier2	0.83257
32	SlogP_VSA0	0.83519
33	Kier1	0.83644
34	SMR_VSA1	0.83839
35	chi1	0.84721
36	SMR_VSA6	0.84762
37	SMR_VSA3	0.8525
38	Kier3	0.85924
39	SlogP	0.86886
40	SMR_VSA0	0.87545
41	a_hyd	0.88264

Scatter plot of the experimental pIC_{50} vs. pIC_{50}-predictions of Eq.(10) (blue) and Eq. (12).

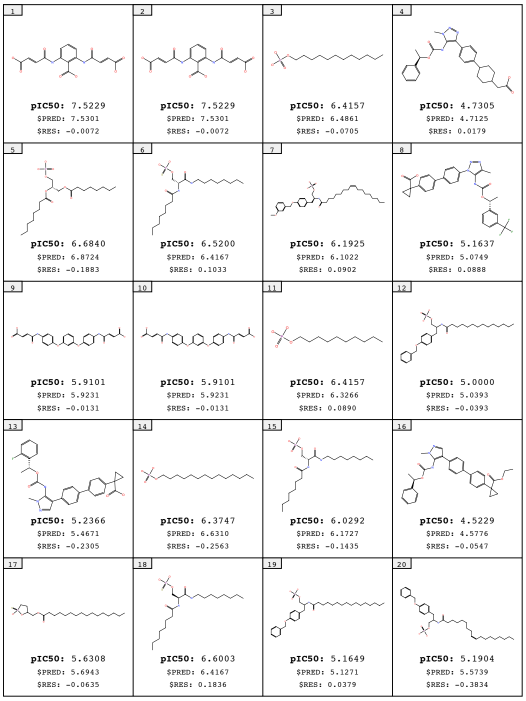


LPA3 INIHIBITORS: TRAINING SET FOR QSAR MODELING



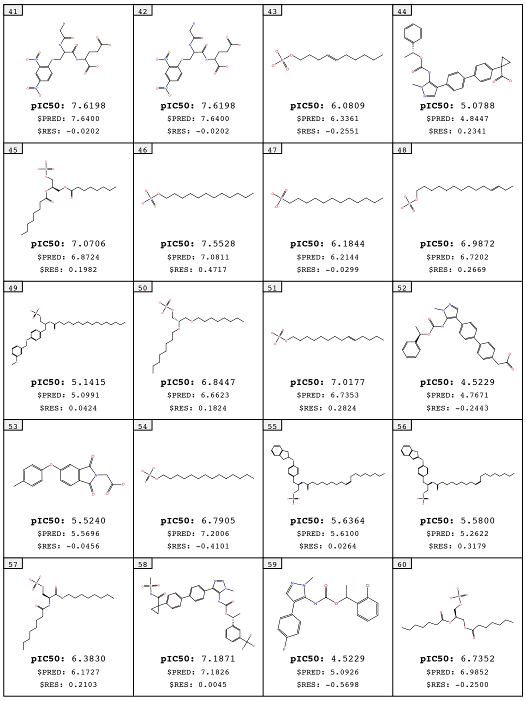
21 pIC50: 5.6364 $PRED: 5.7584 $RES: -0.1220	22 pIC50: 4.5229 $PRED: 4.4398 $RES: 0.0830	23 pIC50: 4.5229 $PRED: 4.4673 $RES: 0.0556	24 pIC50: 6.9136 $PRED: 6.9600 $RES: -0.0463
25 pIC50: 6.8447 $PRED: 6.6932 $RES: 0.1515	26 pIC50: 6.0269 $PRED: 6.4243 $RES: -0.3974	27 pIC50: 5.8962 $PRED: 5.8462 $RES: 0.0500	28 pIC50: 4.6588 $PRED: 4.2947 $RES: 0.3641
29 pIC50: 5.0334 $PRED: 5.2622 $RES: -0.2288	30 pIC50: 5.1931 $PRED: 5.3161 $RES: -0.1230	31 pIC50: 7.5528 $PRED: 7.3033 $RES: 0.2495	32 pIC50: 6.4685 $PRED: 6.8160 $RES: -0.3474
33 pIC50: 5.1124 $PRED: 4.9684 $RES: 0.1439	34 pIC50: 5.0830 $PRED: 4.9372 $RES: 0.1458	35 pIC50: 6.1135 $PRED: 6.1509 $RES: -0.0374	36 pIC50: 5.1593 $PRED: 5.3161 $RES: -0.1568
37 pIC50: 5.0000 $PRED: 4.7783 $RES: 0.2217	38 pIC50: 6.1238 $PRED: 6.1151 $RES: 0.0087	39 pIC50: 7.5686 $PRED: 7.1552 $RES: 0.4134	40 pIC50: 6.3206 $PRED: 6.3223 $RES: -0.0018

LPA3 INIHIBITORS: TRAINING SET FOR QSAR MODELING



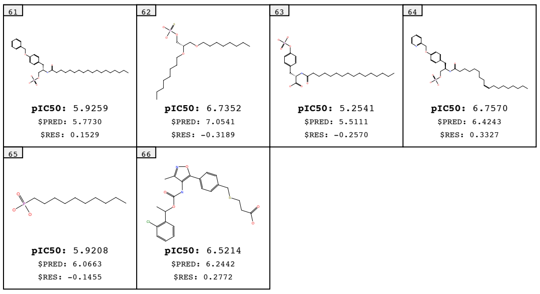
1 pIC50: 7.5229 $PRED: 7.5301 $RES: -0.0072	2 pIC50: 7.5229 $PRED: 7.5301 $RES: -0.0072	3 pIC50: 6.4157 $PRED: 6.4861 $RES: -0.0705	4 pIC50: 4.7305 $PRED: 4.7125 $RES: 0.0179
5 pIC50: 6.6840 $PRED: 6.8724 $RES: -0.1883	6 pIC50: 6.5200 $PRED: 6.4167 $RES: 0.1033	7 pIC50: 6.1925 $PRED: 6.1022 $RES: 0.0902	8 pIC50: 5.1637 $PRED: 5.0749 $RES: 0.0888
9 pIC50: 5.9101 $PRED: 5.9231 $RES: -0.0131	10 pIC50: 5.9101 $PRED: 5.9231 $RES: -0.0131	11 pIC50: 6.4157 $PRED: 6.3266 $RES: 0.0890	12 pIC50: 5.0000 $PRED: 5.0393 $RES: -0.0393
13 pIC50: 5.2366 $PRED: 5.4671 $RES: -0.2305	14 pIC50: 6.3747 $PRED: 6.6310 $RES: -0.2563	15 pIC50: 6.0292 $PRED: 6.1727 $RES: -0.1435	16 pIC50: 4.5229 $PRED: 4.5776 $RES: -0.0547
17 pIC50: 5.6308 $PRED: 5.6943 $RES: -0.0635	18 pIC50: 6.6003 $PRED: 6.4167 $RES: 0.1836	19 pIC50: 5.1649 $PRED: 5.1271 $RES: 0.0379	20 pIC50: 5.1904 $PRED: 5.5739 $RES: -0.3834

LPA3 INIHIBITORS: TRAINING SET FOR QSAR MODELING



41 pIC50: 7.6198 $PRED: 7.6400 $RES: -0.0202	42 pIC50: 7.6198 $PRED: 7.6400 $RES: -0.0202	43 pIC50: 6.0809 $PRED: 6.3361 $RES: -0.2551	44 pIC50: 5.0788 $PRED: 4.8447 $RES: 0.2341
45 pIC50: 7.0706 $PRED: 6.8724 $RES: 0.1982	46 pIC50: 7.5528 $PRED: 7.0811 $RES: 0.4717	47 pIC50: 6.1844 $PRED: 6.2144 $RES: -0.0299	48 pIC50: 6.9872 $PRED: 6.7202 $RES: 0.2669
49 pIC50: 5.1415 $PRED: 5.0991 $RES: 0.0424	50 pIC50: 6.8447 $PRED: 6.6623 $RES: 0.1824	51 pIC50: 7.0177 $PRED: 6.7353 $RES: 0.2824	52 pIC50: 4.5229 $PRED: 4.7671 $RES: -0.2443
53 pIC50: 5.5240 $PRED: 5.5696 $RES: -0.0456	54 pIC50: 6.7905 $PRED: 7.2006 $RES: -0.4101	55 pIC50: 5.6364 $PRED: 5.6100 $RES: 0.0264	56 pIC50: 5.5800 $PRED: 5.2622 $RES: 0.3179
57 pIC50: 6.3830 $PRED: 6.1727 $RES: 0.2103	58 pIC50: 7.1871 $PRED: 7.1826 $RES: 0.0045	59 pIC50: 4.5229 $PRED: 5.0926 $RES: -0.5698	60 pIC50: 6.7352 $PRED: 6.9852 $RES: -0.2500

LPA3 INIHIBITORS: TRAINING SET FOR QSAR MODELING



References

[1] A.M. Helguera, A. Prez-Garrido, A. Gaspar, J. Reis, F. Cagide, D. Vina, M. Cordeiro, F. Borges, Combining QSAR classification models for predictive modeling of human monoamine oxidase inhibitors, Eur J Med Chem. 2013 ;59:75-90. http://dx.doi.org/10.1016/j.ejmech.2012.10.035

[2] P.P. Roy, J.T. Leonard, K. Roy, Exploring the impact of size of training sets for the development of predictive QSAR models. Chemometrics and Intelligent Laboratory Systems 90 2008 (1): 31-42.

[3] R. Todeschini, V. Consonni. ”Molecular Descriptors for Chemoinformatics” (2 volumes), 2009 Wiley-VCH. http://dx.doi.org/10.1002/9783527628766

[4] T. Scior, J.L. Medina-Franco, QT. Do, K. Martnez-Mayorga, J.A. Yunes-Rojas, P. Bernard. ”How to recognize and workaround pitfalls in QSAR studies: a critical review”. Curr Med Chem. 2009; 16 (32):4297-313.

[5] B.K. Shoichet, I.D. Kuntz, D.L. Bodian. ”Molecular docking using shape descriptors”. Journal of Computational Chemistry 13; 2004 (3): 380-397

[6] R.J. Morris, J. Najmanovich, A. Kahraman, J.M. Thornton. ”Real spherical harmonic expansion coefficients as 3D shape descriptors for protein binding pocket and ligand comparisons”. Bioinformatics 21; 2005 (10): 2347-55.

[7] B.B. Goldman, W.T. Wipke. ”QSD quadratic shape descriptors. Molecular docking using quadratic shape descriptors (QSDock)”. Proteins 38; 2000 (1): 79-94.

[8] P. Wang, X.H. Wu, W.X. Chen, B.E. Shan, Q. Guo. ”Expression of lysophosphatidic acid receptor in human ovarian cancer cell lines 3AO, SKOV3, OVCAR3 and its significance” Di Yi Jun Yi Da Xue Xue Bao. 2005 25(11):1422-4, 1431.

[9] H. Ueda, H. Matsunaga, O.I. Omotuyi, J. Nagai, ”Lysophosphatidic acid: chemical signature of neuropathic pain”. Biochim Biophys Acta. 2013; 1831(1):61-73.http://dx.doi.org/10.1016/j.bbalip.2012.08.014

[10] Molecular Operating Environment (MOE), 2012.10; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite 910, Montreal, QC, Canada, H3A 2R7, 2012.

[11] M.J. Wichura, ”The coordinate-free approach to linear models”. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press. pp. xiv+199. ISBN 978-0-521-86842-6. 2006. MR 2283455

[12] D. Wackerly, W. Scheaffer. ”Mathematical Statistics with Applications” (7 ed.). Belmont, CA, USA: Thomson Higher Education. ISBN 0-49538508-5. 2008