Journal of Physical Chemistry & Biophysics

Omotuyi and Ueda, J Phys Chem Biophys 2013, 3:3
DOI: 10.4172/2161-0398.1000121

Research Article Open Access

Descriptor-based Fitting of Structurally Diverse LPA1 Inhibitors into a Single predictive Mathematical Model

Olaposildowu Omotuyi* and Hiroshi Ueda

Department of Molecular Pharmacology and Neuroscience, Nagasaki University Graduate School of Biomedical Sciences, 852-8521, Nagasaki, Japan

Abstract

120 structurally diverse compounds previously reported as LPA1 inhibitors have been used to derive a mathematical model based on their descriptors. The pre- and post-cross-validated correlation coefficient (R^2) is 0.79168 (RMSE=0.61459) and 0.70939 (RMSE=0.72938) respectively. Principal component analysis (PCA) was also used to reduce the dimension and linearly transform the raw data. PCA results showed that nine (9) principal components sufficiently accounts for more than 98% of the variance of the dataset with a fitting mathematical equation. Our model accurately predicted ~86% of the compounds tested regardless of their structural diversities.

Keywords: LPA1; Antagonists; Mathematical model; PCA; Descriptors

Introduction

In situations where protein X-ray structure or related structures for template-based homology modeling are unavailable for structure-based virtual screening, computational methods for drug design rely principally on ligand-based approaches. Ligand-based approach depends on at least one known active compound; which serves as the query for searching library of compounds using predefined molecular descriptor parameters [1,2]. Three categories of chemical descriptors have been characterized till date; physical properties descriptors (1D-descriptor), molecular topology and pharmacophore descriptors (2D-descriptors) and geometrical descriptors (3D-descriptors, often requires prior knowledge of target protein binding-pocket) [3-5]. When there are multiple bioactive compounds for a given target, quantitative structure activity relationships (QSARs) method is more beneficial. QSAR method provides predictive mathematical model for biological activities using statistical clustering of multiple descriptors variables [6,7]. We sought to derive a mathematical equation from minimal set of ligand descriptors for set of Lysophosphatidic acid receptor (LPA1) inhibitors. With this equation, we hope to accurately predict the activity of a test set and hopefully used in ligand-based virtual screening for new high-affinity LPA1 antagonists.

Materials and Methods

Here, using Molecular Operating Environment (MOE) [8], multiple descriptors (SlogP (SlogP_VSA0-6), SMR (SMR_VSA0-4), a_acc, ASA, E_stb, a_hyd, and Kier (Kier1-2, KierA1-2)) [8] have been generated for training set of compounds (CHEMBL3819) in order to establish a mathematical equation to model LPA1 inhibition (antagonism). PCA analysis was also conducted to determine the principle components of the equation using scientific vector language (SVL) programming built into the MOE.

Results and Discussion

First, The IC-50 values of 134 unique entries (LPA1 inhibitors) from ChemBL database (CHEMBL3819) were converted to Gibb's free energy of binding using Cheng-Prusoff equation [9] {Equation1} at S<<<Km {Equation 2} approximation at 298K.

$$K_i = \frac{IC_{50}}{1 + \frac{[S]}{K_m}} \qquad \text{(Equation 1)}$$

$$\Delta_r G^\circ = -RT(\ln K_i) = -2.303RT(\log_{10} K_i) \qquad \text{(Equation 2)}$$

The library was randomly and unbiasedly grouped OCHEM server [10] into the training (120 compounds, Supplementary Figure 1) and test (14 compounds) sets. The training set was initially fitted using partial least square (PLS) method into all Chemical descriptors implemented in MOE [8]. The descriptors were pruned in order of their relative importance until a mathematical model (Equation 3) was obtained.

dGExpt = -3.0345 -0.39537 x a_acc -0.02183 x ASA -0.36027 x a_hyd -0.01028 x E_stb +0.64979 x Kier1 +0.21026 x Kier2 +0.08358 x KierA1 -0.47849 x KierA2 +0.03617 x SlogP_VSA0 +0.01945 x SlogP_ VSA1 +0.00494 x SlogP_VSA2 -0.00339 x SlogP_ VSA3 +0.01846 x SlogP_VSA4 +0.05076 x SlogP_VSA6 -0.06603 x SMR_VSA0 -0.05469 x SMR_VSA1 -0.03451 x SMR_VSA2 +0.00294 x SMR_VSA3 -0.01021 x SMR_VSA4 (Equation 3)

This model gives a high probabilistic (r^2=0.79168 with RMSE of 0.61459 dGexpt) Gibb's free energy prediction using minimal set of descriptors (Figure 1). A cross-validated correlation coefficient value of 0.70939 (RMSE = 0.72938) was also obtained for the model.

These results suggest that the set of descriptors chosen can effectively cluster the minimal structural and molecular parameters required for the predicting relatively small differences in the ligand activity of structurally diverse compounds typifying the training set.

Due to the relatively good mathematical correlation between the descriptors and the estimated free energy of ligand binding, we sought to further study the dataset descriptors long the principle components through the reduction of the dimensionality and linear transformation of the raw data (Principal component analysis (PCA)) [11]. Given the initial 120 compounds (represented as *m*) and for one of the

***Corresponding author:** Omotuyi IO, Department of Molecular Pharmacology and Neurosciences, Nagasaki University Graduate School of Biomedical Sciences, 852-8521, Nagasaki, Japan, E-mail: bbis11r104@cc.nagasaki-u.ac.jp

Received May 31, 2013; **Accepted** July 26, 2013; **Published** July 29, 2013

Citation: Omotuyi O, Ueda H (2013) Descriptor-based Fitting of Structurally Diverse LPA1 Inhibitors into a Single predictive Mathematical Model. J Phys Chem Biophys 3: 121. doi:10.4172/2161-0398.1000121

compounds say '*i*' its descriptors are represented by *n*-vector of real numbers x*i*=(x*i*1,..,xin, where n=1-17). Assuming that each molecule '*i*' has an associated importance weight '*wi*', (non-negative, real number) and that the weights is relative probability that the associated molecule '*xi*' 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 equation (4) where S is a symmetric, semi-definite sample covariance matrix. S can be diagonalized such that S =Q^TDDQ (*Q* is orthogonal, *D* is diagonal-sorted in descending order from top left to bottom right) [12].

$$E(x) \approx \bar{x} = \frac{1}{w}\sum_{i=1}^{m} w_i x_i, \quad Cov(X) \approx S = \frac{1}{w}\sum_{i=1}^{m} w_i x_i x_i^T - \bar{x}\bar{x}^T \quad \text{(Equation 4)}$$

The effect of the each of the principal components (eigenvectors) on the condition and the variance (Supplementary Table 1) shows that nine (9) principal components sufficiently accounts for more than 98% of the variance in the dataset with a fitting mathematical equation (Equation 5). The 3D-scatter plot of the last three principal components (PCA7, PCA8 and PCA9) with respect to free energy is shown in Figure 2; each point in the plot corresponds to a molecule colored according to free energy values.

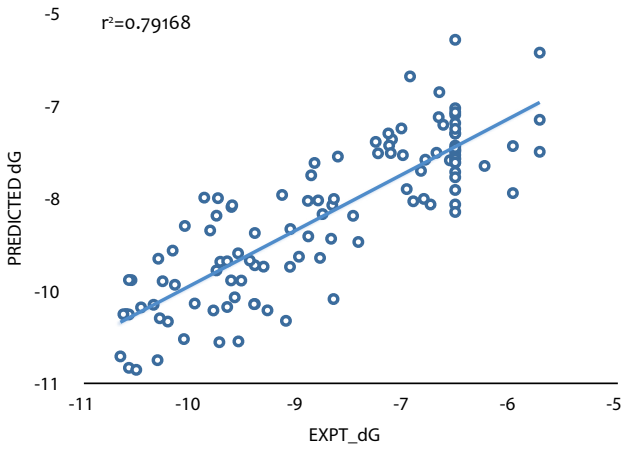

Figure 1: Scatter plot showing the predicted and experimental free energy of LPA1 inhibitors.

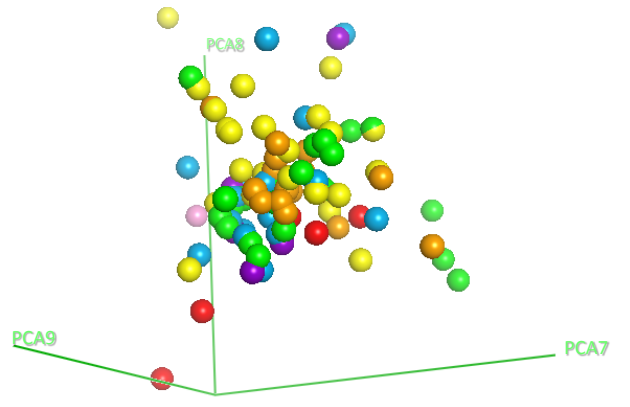

Figure 2: 3D PCA plot of each test compound. Colours represent experimental free energy.

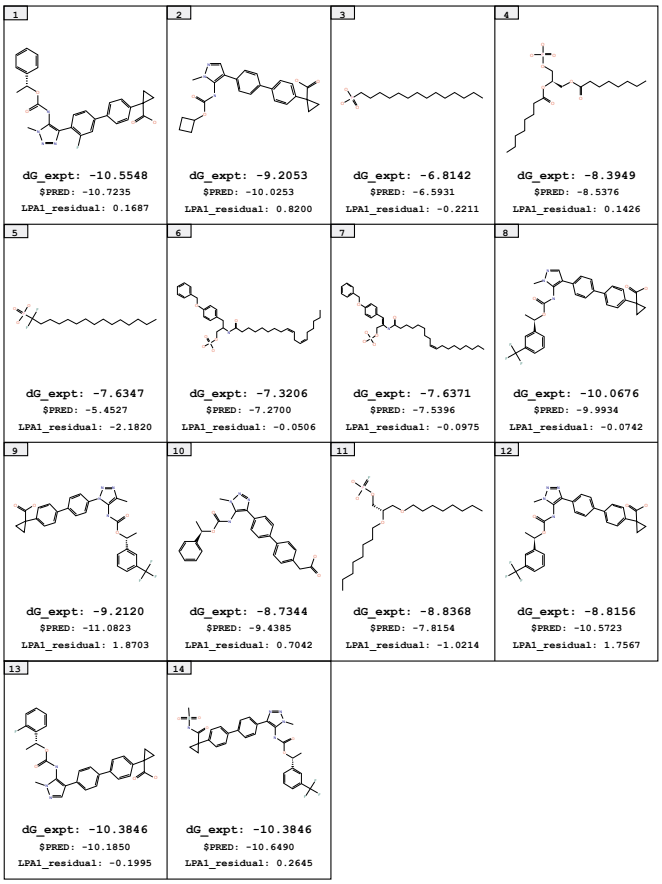

Figure 3: Structures of test compounds showing the experimental and predicted free energies (dG) of binding and residual values (dG_expt --- dG_predicted).

PCA9 = 5.53218413e-001 -1.47174139e-003 X ASA -5.28867555e-004 X E_stb - 9.64502253e-003 X Kier1 +2.92612997e-002 X Kier2 -9.05227786e-004 X KierA1+2.57936088e-002 X KierA2 +4.04361621e-002 X SMR_VSA0-2.37125484e-002 X SMR_VSA1 +5.03998977e-002 X SMR_VSA2 +8.13078695e-003 X SlogP_VSA0 - 1.03630885e-002 X SlogP_VSA1 -5.72337043e-002 X SlogP_VSA2 -1.64177905e-003 X SlogP_VSA3 -7.55989243e-002 X SlogP_VSA4 -1.02026342e-002 X SlogP_VSA6 +1.79553609e-001 X a_acc -3.68295238e-002 X a_hyd (Equation 5)

When equation 3 was used to predict the Gibb's free energy of the test set, it predicted accurately (residual free energy > +1.0) ~86% of the compounds regardless of their structural diversities (Figure 3).

Conclusion

Given the predictive finesse of this mathematical model, there is a question to be answered and two areas of potential applications to be exploited. Will this model sufficiently predict more chemically diverse compounds? If the yes, then we can predict a more robust interrelationship between statistics and Computer-Aided Drug Discovery in the future. Also, descriptor-based mathematical model screening may be piped as confirmatory steps following structure-based screening 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.

References

1. Ballester PJ (2011)Ultrafast shape recognition: method and applications. Future MedChem 3: 65–78.
2. Bohacek RS, McMartin C, Guida WC (1996) The art and practice of structure-based drug design: a molecular modeling perspective. Med Res Rev 16: 3-50.
3. Shoichet BK, Kuntz ID, Bodian DL (2004) Molecular docking using shape descriptors. Journal of Computational Chemistry 13: 380-397.
4. Morris RJ, Najmanovich RJ, Kahraman A, Thornton JM (2005) Real spherical harmonic expansion coefficients as 3D shape descriptors for protein binding pocket and ligand comparisons. Bioinformatics 21: 2347-2355.
5. Goldman BB, Wipke WT (2000) QSD quadratic shape descriptors. 2. Molecular docking using quadratic shape descriptors (QSDock). Proteins 38: 79-94.
6. Roy PP, Leonard JT, Roy K (2008) Exploring the impact of size of training sets for the development of predictive QSAR models. Chemometrics and Intelligent Laboratory Systems 90: 31-42.
7. Helguera AM, Pérez-Garrido A, Gaspar A, Reis J, Cagide F,et al. (2013) Combining QSAR classification models for predictive modeling of human monoamine oxidase inhibitors. Eur J Med Chem 59: 75-90.
8. Molecular Operating Environment (MOE) (2012) Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7.
9. Cheng Y, Prusoff WH (1973) Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. BiochemPharmacol 22: 3099-3108.
10. Sushko I, Novotarskyi S, Körner R, Pandey AK, Rupp M, et al. (2011) Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. J Comput Aided Mol Des 25: 533-554.
11. Abdi H, Williams LJ (2010) Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics 2: 433-459.
12. Wasserman Larry (2004) All of Statistics: A Concise Course in Statistical Inference. ISBN 0-387-40272-1.