



The Open Medicinal Chemistry Journal

Content list available at: www.benthamopen.com/TOMCJ/

DOI: 10.2174/1874104501711010212



RESEARCH ARTICLE

Ezqsar: An R Package for Developing QSAR Models Directly From Structures

Jamal Shamsara*

Pharmaceutical Research Center, Pharmaceutical Technology Institute, Mashhad University of Medical Sciences, Mashhad, Iran

Received: August 18, 2017

Revised: November 01, 2017

Accepted: November 12, 2017

Abstract:

Background:

Quantitative Structure Activity Relationship (QSAR) is a difficult computational chemistry approach for beginner scientists and a time consuming one for even more experienced researchers.

Method and Materials:

Ezqsar which is introduced here addresses both the issues. It considers important steps to have a reliable QSAR model. Besides calculation of descriptors using CDK library, highly correlated descriptors are removed, a provided data set is divided to train and test sets, descriptors are selected by a statistical method, statistical parameter for the model are presented and applicability domain is investigated.

Results:

Finally, the model can be applied to predict the activities for an extra set of molecules for a purpose of either lead optimization or virtual screening. The performance is demonstrated by an example.

Conclusion:

The R package, ezqsar, is freely available via <https://github.com/shamsaraj/ezqsar>, and it runs on Linux and MS-Windows.

Keywords: Cheminformatics, Lead optimization, MLR, QSAR, R programming language.

1. BACKGROUND

Quantitative Structure Activity Relationship (QSAR) is an old but still applicable method for the various branches of chemistry. It attempts to find a model that can predict the biological activity of chemical compounds using their structural features [1]. There are several modules available in commercial tools [2] (SYBYL, MOE and Schrodinger suite) that make QSAR studies simpler than ever.

2. MATERIALS AND METHODS

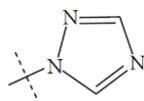
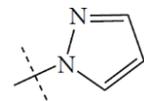
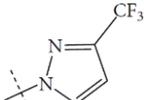
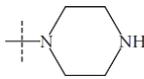
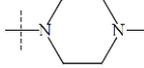
There are also some open source tools available to facilitate the computation of QSAR models. Some of them are collection of individual tools; each performs specific step of a QSAR procedure such as modeling [3] statistical validation [4, 5] and generation of descriptors [6 - 7]. Some of them present some tools that can be used to generate CoMFA-like 3D QSAR models such as, Open3DQSAR or PyCoMFA [8]. CORAL is a freeware that is developed to build QSAR PLS models using a specific set of descriptors so-called SMILES notation based optimal descriptors

* Address correspondence to this authors at the Pharmaceutical Research Center, Pharmaceutical Technology Institute, Mashhad University of Medical Sciences, Mashhad, Iran, Tel: +98 51 38823255, Fax: +98 51 38823251; E-mail: shamsaraj@mums.ac.ir

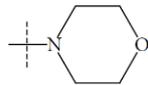
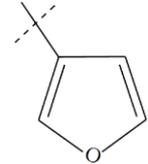
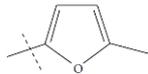
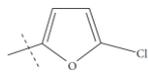
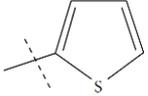
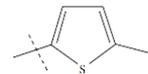
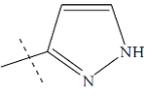
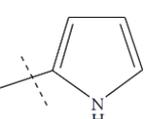
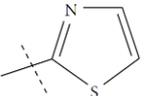
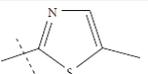
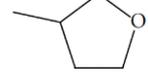
[9 - 10] QSARINS is another standalone freeware that can build QSAR MLR with various functions including data processing and partitioning, validation of the model, prediction of a new compound activity and determination of applicability [11]. However, it does not have any built-in descriptor generation ability. Among the available open source tools, the most similar one to ezqsar is another R-package that is called camb [12]. Some of the features are similar between those packages. They are intended to be used mostly by beginners. They provide a single function that can do the entire job starting with a data set and ends with a complete QSAR model that includes descriptor generation, data processing, modeling, internal and external validity assessment, presentation of a QSAR model with the ability to predict the activities of new structures. The advantages of ezqsar over camb are the applicability domain functionality, easier use and interpretability of the developed model. However, the camb package currently has more training options than ezqsar such as machine learning methods. The advanced users are advised to use caret package to model pre-generated descriptors. It wrapped up QSAR tools in several functions and user can tune several parameters for each one, but ezqsar could be used by advanced users to provide an easy and precise look on the modelability of a data set and prediction of the activity of a test set with estimation of applicability domain. They could also use the power of R scripting to enrich the output plots or automate model building for large number of data set or develop several models for a single data set and discover the best model among them. Finally, the selected descriptors by ezqsar for a reliable QSAR model could be used for mechanistic interpretation of the model.

Here, an open source R (R: a language and environment for statistical computing; R Foundation for Statistical Computing, Vienna, Austria; URL <http://www.R-project.org/>.) package is introduced that could develop Multiple Linear Regression (MLR) QSAR models from 2D or 3D structures and their corresponding activities *via* a single line command. Then, in an iterative process, the QSAR model could be refined by modifications in, for example, the number of selected descriptors and test set selection. Data set selection and preparation is a first step and the most important step in a QSAR study. The structures should be checked if they are retrieved from public databases. Data set should have the least possible experimental uncertainty. Experimental uncertainty arise from systematic error or in case of single point activity determination. The detection of possible experimental uncertainty in the data set may be detected by statistical methods but it is not easy [13 - 15]. The descriptor generation in ezqsar is done using CDK library [16]. It computes 2D and 3D descriptors. They are classified into five groups “topological”, “geometrical” “hybrid”, “constitutional”, and “electronic”. If the input structures are in 3D coordinates, the 3D descriptors will be calculated otherwise, the value for the 3D descriptors would be zero. A list of the all-275 CDK descriptors is presented in Table (1). Now, ezqsar only accepts SDF file as an input and the structures should be verified beforehand regarding specific chirality, protonation state and tautomeric form.

Table 1. Observed and predicted activities of the training, test and new test sets based on the model1. Activities were shown as pIC_{50} (μM). ^a: Test set, ^b: new test set. They are provided as an example data set in the package.

Name	R	Observed pIC_{50}	Predicted pIC_{50} values	Residues
n1		1.89	1.62	0.27
n2		1.82	2.19	-0.37
n3		3.15	2.57	0.58
n4 ^a		1.64	1.93	-0.29
n5 ^b		1.74	1.67	0.07

(Table 1) contd....

Name	R	Observed pIC ₅₀	Predicted pIC ₅₀ values	Residues
n6		2.66	2.82	-0.16
n7 ^a		3.40	4.02	-0.62
n8		4.00	3.72	0.28
n9		4.00	3.94	0.06
n10		3.70	3.78	-0.08
n11		3.70	3.69	0.01
n12		3.00	2.39	0.61
n13 ^b		3.39	3.06	0.33
n14 ^a		3.00	3.08	-0.08
n15		2.92	3.10	-0.17
n16	Methyl	2.07	2.64	-0.58
n17	Ethyl	2.54	2.31	0.22
n18	i-Propyl	2.35	2.38	-0.03
n19	t-Butyl	1.77	1.10	0.67
n20	i-Butyl	2.27	2.28	-0.01
n21	CH ₂ OCH ₃	2.72	2.22	0.50
n22	CF ₃	2.66	2.69	-0.04
n23	Cyclopropyl	2.80	2.46	0.33
n24 ^a	Cyclobutyl	2.64	2.43	0.21
n25 ^b	Cyclohexyl	2.14	2.48	-0.34
n26	Phenyl	2.40	2.62	-0.22
n27		3.52	3.78	-0.26

MLR is a simple, reproducible and easy interpretable method used usually in QSAR. An MLR equation can be expressed generally like the following [17, 18]:

$$Y = a_0 + a_1 \times X_1 + a_2 \times X_2 + a_3 \times X_3 + \dots + a_n \times X_n \quad (1)$$

In the above expression, Y is the dependent variable (here is activity), X_1, X_2, \dots, X_n are independent variables (descriptors) present in the model with the corresponding regression coefficients a_1, a_2, \dots, a_n , respectively, and a_0 is the constant term of the model. The quality of a MLR model is evaluated using the number of metrics as described below [17, 18].

ezqsar_f function uses Leave-one-out (LOO) cross-validation method for cross-validation:

$$Q^2 = 1 - \frac{\Sigma(Y_{obs(train)} - Y_{pred(train)})^2}{\Sigma(Y_{obs(train)} - \bar{Y}_{training})^2} \quad (2)$$

The determination coefficient (R^2) and predictive R^2 (R^2_{pred}) are defined in the following manners:

$$R^2 = 1 - \frac{\Sigma(Y_{obs(train)} - Y_{calc(train)})^2}{\Sigma(Y_{obs(train)} - \bar{Y}_{obs(train)})^2} \quad (3)$$

$$R^2_{pred} = 1 - \frac{\Sigma(Y_{obs(test)} - Y_{pred(test)})^2}{\Sigma(Y_{obs(test)} - \bar{Y}_{training})^2} \quad (4)$$

In equations 2-4, $Y_{obs(train)}$ is the observed activity for the train set, $Y_{pred(train)}$ is the predicted activity of the training set molecules based on the LOO technique, $Y_{calc(train)}$ is the model-derived calculated response for the train set and $\bar{Y}_{obs(train)}$ is the average of the observed response values for the train set, $Y_{obs(test)}$ and $Y_{pred(test)}$ are the observed and predicted activity data for the test set compounds, respectively.

The ability of the model to predict activity of the present and other set can be accessed *via* R^2 and Q^2 , respectively. If the difference between Q^2 and R^2 is more than 0.3, an overtrained model can be implied. The predictivity of the model also can be assessed by R^2_{pred} . Generally, the following criteria roughly indicate a reliable model [17, 19, 20]:

$$\begin{cases} Q^2 > 0.5 \\ R^2 > 0.8 \\ R^2_{pred} > 0.6 \end{cases} \quad (5)$$

The applicability domain is defined as a space constructed (structural, chemical, *etc.*) by the model that plays a crucial role for estimating the applicability of the model for predicting the activity of new compounds [17]. In other word, the prediction of a property of interest (here, activity) of a new compound using a QSAR model is applicable only if it is within the applicability domain of the QSAR model. The final model could be used to predict the activity of an extra set of compounds that are within the applicability domain of the model. This is checked by finding possible outliers of the standardized values of the descriptors (S_{ki}) which are present in the model as well as by calculation of Tanimoto similarity index of a compound fingerprint to the each compound in the train set.

Standardized descriptor values can be calculated as follows, that is a simplified form suggested by Roy K. *et al.* [21]

$$S_{ki} = \frac{|X_{ki} - \bar{X}_i|}{\sigma_{X_i}} \quad (6)$$

Where, K is the number of compounds, i is the number of descriptors, S_{ki} is the standardized descriptor i for compound K (from the training or test set, X_{ki} is the original descriptor i for compound K (from the training or test set), \bar{X}_i is the mean value of the descriptor X_i for the training set compounds only, σ_{X_i} is the standard deviation of the descriptor X_i for the training set compounds only. The above calculation is meant for all descriptor values present in the

model (number of compounds \times number of descriptors).

Tanimoto similarity indexes are calculated as it follows [22 - 24]:

$$\text{Tanimoto (a, b)} = \frac{N_{ab}}{(N_a + N_b) - N_{ab}} \quad (7)$$

where N_{ab} is the number of common “1” bits that occur in both fingerprint a and fingerprint b, N_a is the number of “1” bits in fingerprint a, N_b is the number of “1” bits in fingerprint b. In ezqsar_f, Tanimoto index is computed by fingerprint package.

3. IMPLEMENTATION

The codes were implemented in a package available at github and can be installed and loaded by the following commands in R environment:

- `install.packages("devtools")`
- `devtools::install_github("shamsaraj/ezqsar")`
- `library("ezqsar")` #This will load the package

It depends on four packages: caret, fingerprint, leaps and rcdk.

4. RESULTS AND DISCUSSION

The performance of the ezqsar package in an example data set that is provided by the package after installation was demonstrated in the study. The data set Table (1 and Fig. 1) was taken from a study [25] and a HQSAR model was already available to them [26]. It has a single function called “ezqsar_f”. Like other R functions, one can get help for the function by:

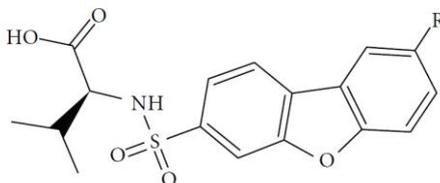


Fig. (1). General structure for the dataset.

- `help("ezqsar_f")`

An overview of the ezqsar_f workflow is demonstrated in Fig. (2). All of the molecules were collected in a single SDF file. Activities were provided in a separate csv file rank ordered same as the SDF file. The activities were expressed as pIC_{50} , however, they also can be expressed as IC_{50} . Alternatively, the activities also can be provided in a SDF file with the header of “ IC_{50} ”. Data set contains 24 MMP-12 inhibitors. There is an extra test set that includes three other MMP-12 inhibitors with similar structures. The example can be run by following the command after loading the installed package, ezqsar:

- `example("ezqsar_f")`

This will run following the script:

- `file1<-system.file("extdata", "molecules-3d.sdf", package = "ezqsar")`
- `file2<-system.file("extdata", "IC50.csv", package = "ezqsar")`
- `file3<-system.file("extdata", "newset-3d.sdf", package = "ezqsar")`
- `model<-ezqsar_f(SDFfile=file1, activityfile=file2, newdataset=file3, testset=c(4,6,12,22))`
- `attributes(model)`

- `print (model$Q2)`
- `print (model$R2)`
- `print (model$test)`
- `print (model$R2_pred)`
- `print (model$Tanimoto_test_sum)`
- `print (model$AD_outlier_test)`
- `print (model$newset)`
- `print (model$Tanimoto_newset_sum)`

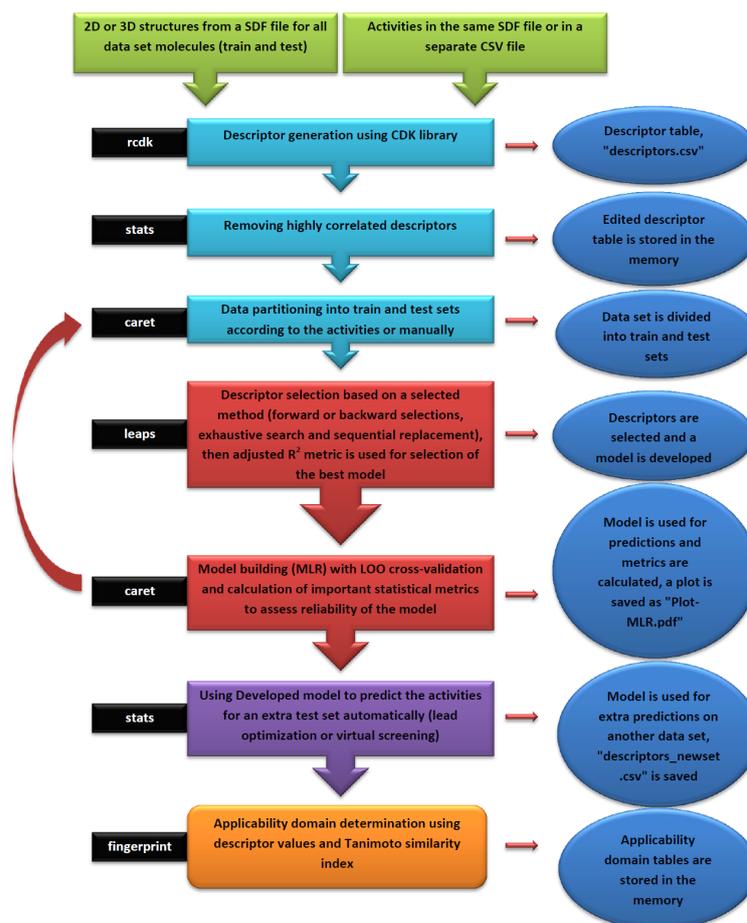
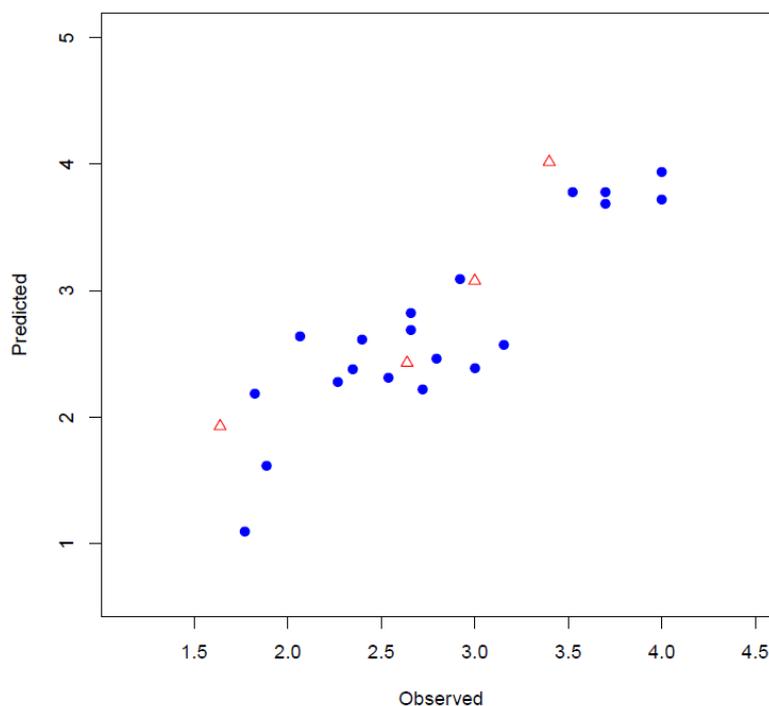


Fig. (2). The workflow of `ezqsar_f` function from `ezqsar` package.

The input files were stored as files 1, 2 and 3 by executing lines 1, 2 and 3. The model is developed by executing line 4. This runs the function with default parameters and stores the results as an object called, `model`. Highly correlated descriptors (correlation coefficient over a defined threshold) can be removed from the descriptor table before further processing. By default, this threshold is 1 (correlated descriptors are not removed), test set ratio is 20% and the selection is based on the activities; descriptors are selected by forward selecting method before final MLR model development. The calculated descriptors are reported as an csv file named “`descriptors.csv`”. A plot (“`Plot-MLR.pdf`”) will also be available in the working directory (can be changed by `setwd()` in R) after a successful run regarding the observed vs predicted activity values for both train and test sets using the developed model Fig. (3). Available outputs are shown after executing line 6 and some of the most important ones were printed out by remaining lines of the example (lines 6 to 13). The statistical metrics for the example data set are shown in (Table 2).

Table 2. Output summary for three models.

Model	Model1	Model2	Model3
Test set compounds	n4,n7,n14,n24	n10, n20, n23, n24	n4, n9, n12, n24
Q2	0.796	0.722	0.858
R2	0.901	0.865	0.941
R2pred	0.708	0.637	0.617
New set compounds and activities	n5=1.74, n13=3.39, n25=2.14	n5=1.74, n13=3.39, n25=2.14	n5=1.74, n13=3.39, n25=2.14
Predicted activities for the new set compounds	n5=1.67, n13=3.06, n25=2.47	n5=1.35, n13=3.00, n25=2.44	n5=1.37, n13=2.78, n25=3.54
predicted residual sum of squares (PRESS)	0.226	0.401	2.480
comment	All of the model descriptors of all compounds have standardized values <3	All of the model descriptors of all compounds have standardized values <3	Standardized value of one of the model descriptors for n25 is very high (= 8.63)

**Fig. (3).** Plot of observed versus predicted activities obtained from model1 for training (blue circles) and test (red triangles) sets.

The `ezqsar_f` function has an ability to perform prediction on up to two extra test sets using the developed model. It could be helpful for either lead optimization process or a ligand based virtual screening.

Different test set selection will lead to different descriptor selection and finally a different model. By running `ezqsar` function over different test sets, a model with better statistical metrics can be achieved. Here is a sample script:

- `file1<-system.file ("extdata", "molecules-3d.sdf", package = "ezqsar")`
- `file2<-system.file ("extdata", "IC50.csv", package = "ezqsar")`
- `file3<-system.file ("extdata", "newset-3d.sdf", package = "ezqsar")`
- `for (h in 1:100) {`
- `set.seed(h)# for each seed number a different test set is selected based on the activities`
- `model<-ezqsar_f (SDFfile=file1, activityfile=file2, newdataset=file3)`
- `if ((model$Q2>0.5) & (model$R2>0.8) & (model$R2_pred>0.6)) {`
- `print (h)`

- print (model\$Q2)
- print (model\$R2)
- print (model\$R2_pred)
- print(model\$Model)
- print (model\$Tanimoto_test_sum)
- print (model\$AD_outlier_test)
- print (model\$newset)
- n1<-(abs(as.numeric(model\$newset [1, 2]) - 1.74))^2
- n2<-(abs(as.numeric(model\$newset [2, 2]) - 3.39))^2
- n3<-(abs(as.numeric(model\$newset [3, 2]) - 2.14))^2
- PRESS<-sum(n1,n2,n3)
- print (PRESS)
- print (model\$AD_outlier_newset)
- print(“*****”)
- }
- }
- print (c(1.74,3.39,2.14))# observed activities (pIC₅₀) for the new set compounds
- print (model\$Tanimoto_newset_sum)

By running the script, only the models within the criteria (equation 4) are listed among 100 developed models. To demonstrate the importance of applicability domain, performance of the two model with proper statistical metrics were compared. The results indicated that a QSAR model Table (2, model3), despite good statistical metrics (similar to model2), may fail to predict the activity of even structurally related structures if the descriptor values of the extra test set are significantly out of range (in this case, standardized value of one of the descriptors = 8.5).

CONCLUSION

In conclusion, ezqsar offers a highly useful, easy-to-use tool for QSAR analysis, including calculations of descriptors and applicability domain determination. The package could be extended extending the use of other statistical methods such as Partial Least Squares (PLS) or machine learning methods to develop various QSAR models.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No Animals/Humans were used for studies that are base of this research.

CONSENT FOR PUBLICATION

Not applicable.

CONFLICT OF INTEREST

The author declares no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

This work was supported in part by Mashhad University of Medical Sciences.

REFERENCES

- [1] Lewis, R.A.; Wood, D. Modern 2D QSAR for drug discovery. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, **2014**, 4(6), 505-522. [<http://dx.doi.org/10.1002/wcms.1187>]
- [2] Yee, L.C.; Wei, Y.C. Current Modeling Methods Used in QSAR/QSPR. In *Statistical Modelling of Molecular Descriptors in QSAR/QSPR*, Wiley-VCH Verlag GmbH & Co. KGaA.; **2012**, 1-31.
- [3] Teófilo, R.F.; Martins, J.P.; Ferreira, M.M. Sorting variables by using informative vectors as a strategy for feature selection in multivariate regression. *J. Chemometr.*, **2009**, 23(1), 32-48.

- [http://dx.doi.org/10.1002/cem.1192]
- [4] Baumann, D.; Baumann, K. Reliable estimation of prediction errors for QSAR models under model uncertainty using double cross-validation. *J. Cheminform.*, **2014**, *6*(1), 47. [http://dx.doi.org/10.1186/s13321-014-0047-1] [PMID: 25506400]
- [5] Roy, K.; Ambure, P. The “double cross-validation” software tool for MLR QSAR model development. *Chemom. Intell. Lab. Syst.*, **2016**, *159*(Suppl. C), 108-126. [http://dx.doi.org/10.1016/j.chemolab.2016.10.009]
- [6] Dong, J.; Cao, D-S.; Miao, H-Y.; Liu, S.; Deng, B-C.; Yun, Y-H.; Wang, N-N.; Lu, A-P.; Zeng, W-B.; Chen, A.F. ChemDes: an integrated web-based platform for molecular descriptor and fingerprint computation. *J. Cheminform.*, **2015**, *7*, 60. [http://dx.doi.org/10.1186/s13321-015-0109-z] [PMID: 26664458]
- [7] Tetko, I.V.; Gasteiger, J.; Todeschini, R.; Mauri, A.; Livingstone, D.; Ertl, P.; Palyulin, V.A.; Radchenko, E.V.; Zefirov, N.S.; Makarenko, A.S.; Tanchuk, V.Y.; Prokopenko, V.V. Virtual computational chemistry laboratory--design and description. *J. Comput. Aided Mol. Des.*, **2005**, *19*(6), 453-463. [http://dx.doi.org/10.1007/s10822-005-8694-y] [PMID: 16231203]
- [8] Tosco, P.; Balle, T. Open3DQSAR: a new open-source software aimed at high-throughput chemometric analysis of molecular interaction fields. *J. Mol. Model.*, **2011**, *17*(1), 201-208. [http://dx.doi.org/10.1007/s00894-010-0684-x] [PMID: 20383726]
- [9] Veselinović, A.M.; Veselinović, J.B.; Živković, J.V.; Nikolić, G.M. Application of SMILES notation based optimal descriptors in drug discovery and design. *Curr. Top. Med. Chem.*, **2015**, *15*(18), 1768-1779. [http://dx.doi.org/10.2174/1568026615666150506151533] [PMID: 25961525]
- [10] Toropov, A.A.; Rallo, R.; Toropova, A.P. Use of quasi-SMILES and monte carlo optimization to develop quantitative feature Property/Activity relationships (QFPR/QFAR) for nanomaterials. *Curr. Top. Med. Chem.*, **2015**, *15*(18), 1837-1844. [http://dx.doi.org/10.2174/1568026615666150506152000] [PMID: 25961527]
- [11] Gramatica, P.; Chirico, N.; Papa, E.; Cassani, S.; Kovarich, S. QSARINS: A new software for the development, analysis, and validation of QSAR MLR models. *J. Comput. Chem.*, **2013**, *34*(24), 2121-2132. [http://dx.doi.org/10.1002/jcc.23361]
- [12] Murrell, D.S.; Cortes-Ciriano, I.; van Westen, G.J.; Stott, I.P.; Bender, A.; Malliavin, T.E.; Glen, R.C. Chemically aware model builder (camb): An R package for property and bioactivity modelling of small molecules. *J. Cheminform.*, **2015**, *7*(1), 45. [http://dx.doi.org/10.1186/s13321-015-0086-2] [PMID: 26322135]
- [13] Wenlock, M.C.; Carlsson, L.A. How experimental errors influence drug metabolism and pharmacokinetic QSAR/QSPR models. *J. Chem. Inf. Model.*, **2015**, *55*(1), 125-134. [http://dx.doi.org/10.1021/ci500535s] [PMID: 25406036]
- [14] Zhao, L.; Wang, W.; Sedykh, A.; Zhu, H. Experimental errors in QSAR modeling sets: What we can do and what we cannot do. *ACS Omega*, **2017**, *2*(6), 2805-2812. [http://dx.doi.org/10.1021/acsomega.7b00274] [PMID: 28691113]
- [15] Roy, K.; Ambure, P.; Aher, R.B. How important is to detect systematic error in predictions and understand statistical applicability domain of QSAR models? *Chemom. Intell. Lab. Syst.*, **2017**, *162*(Suppl. C), 44-54. [http://dx.doi.org/10.1016/j.chemolab.2017.01.010]
- [16] Steinbeck, C.; Hoppe, C.; Kuhn, S.; Floris, M.; Guha, R.; Willighagen, E.L. Recent developments of the chemistry development kit (CDK) - an open-source java library for chemo- and bioinformatics. *Curr. Pharm. Des.*, **2006**, *12*(17), 2111-2120. [http://dx.doi.org/10.2174/13816120677585274] [PMID: 16796559]
- [17] Roy, K.; Kar, S.; Das, R. Statistical methods in QSAR/QSPR. *A Primer on QSAR/QSPR Modeling*; Springer International Publishing, **2015**, pp. 37-59.
- [18] Roy, K.; Kar, S.; Das, R. QSAR/QSPR Modeling: Introduction. *A Primer on QSAR/QSPR Modeling*; Springer International Publishing, **2015**, pp. 1-36.
- [19] Veerasamy, R.; Rajak, H.; Jain, A.; Sivadasan, S.; Varghese, C.P.; Agrawal, R.K. Validation of QSAR Models - Strategies and Importance. *Int. J. Drug Des. Discovery*, **2011**, *2*(3), 511-519.
- [20] Todeschini, R.; Ballabio, D.; Grisoni, F. Beware of Unreliable Q(2)! A Comparative Study of Regression Metrics for Predictivity Assessment of QSAR Models. *J. Chem. Inf. Model.*, **2016**, *56*(10), 1905-1913. [http://dx.doi.org/10.1021/acs.jcim.6b00277] [PMID: 27633067]
- [21] Roy, K.; Kar, S.; Ambure, P. On a simple approach for determining applicability domain of QSAR models. *Chemom. Intell. Lab. Syst.*, **2015**, *145*, 22-29. [http://dx.doi.org/10.1016/j.chemolab.2015.04.013]
- [22] Pearlman, R.S.; Smith, K. *Novel software tools for chemical diversity. In 3D QSAR in drug design*; Springer, **1998**, pp. 339-353.
- [23] Ma, C.; Wang, L.; Xie, X-Q. GPU accelerated chemical similarity calculation for compound library comparison. *J. Chem. Inf. Model.*, **2011**, *51*(7), 1521-1527. [http://dx.doi.org/10.1021/ci1004948] [PMID: 21692447]

- [24] Bajusz, D.; Rácz, A.; Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.*, **2015**, 7, 20.
[<http://dx.doi.org/10.1186/s13321-015-0069-3>] [PMID: 26052348]
- [25] Wu, Y.; Li, J.; Wu, J.; Morgan, P.; Xu, X.; Rancati, F.; Vallese, S.; Raveglia, L.; Hotchandani, R.; Fuller, N.; Bard, J.; Cunningham, K.; Fish, S.; Krykbaev, R.; Tam, S.; Goldman, S.J.; Williams, C.; Mansour, T.S.; Saiah, E.; Sypek, J.; Li, W. Discovery of potent and selective matrix metalloprotease 12 inhibitors for the potential treatment of chronic obstructive pulmonary disease (COPD). *Bioorg. Med. Chem. Lett.*, **2012**, 22(1), 138-143.
[<http://dx.doi.org/10.1016/j.bmcl.2011.11.046>] [PMID: 22153340]
- [26] Shamsara, J.; Shahr-Sadr, A. A predictive HQSAR model for a series of tricycle core containing MMP-12 inhibitors with dibenzofuran ring. *Int. J. Med. Chem.*, **2014**. 2014
[<http://dx.doi.org/10.1155/2014/630807>]

© 2017 Jamal Shamsara.

This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Public License (CC-BY 4.0), a copy of which is available at: <https://creativecommons.org/licenses/by/4.0/legalcode>. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.