Data in Brief 14 (2017) 469–473

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Dataset of 2-(2-(4-aryloxybenzylidene) hydrazinyl) benzothiazole derivatives for GQSAR of antitubercular agents

Amit S. Tapkir [b], Sohan S. Chitlange [a,*], Ritesh P. Bhole [a]

[a] Dr. D. Y. Patil Vidya Pratishthan Society's Dr. D.Y. Patil Institute of Pharmaceutical Sciences & Research, Pimpri, Pune 411018, Maharashtra, India
[b] Progressive Education Society's, Modern College of Pharmacy, Sector 21, Yamunanagar, Nigdi, Pune 411044, Maharashtra, India

ARTICLE INFO

Article history:
Received 4 June 2017
Received in revised form
2 August 2017
Accepted 3 August 2017
Available online 9 August 2017

Keywords:
Antitubercular
Quantitative structure-activity relationship
GQSAR
Benzothiazole

ABSTRACT

Fragment based Quantitative structure activity relationship (QSAR) analysis on reported 25 2-(2-(4-aryloxybenzylidene) hydrazinyl) benzothiazole dataset as antitubercular agents were carried out. Molecules in the current dataset were fragmented into six fragments (R1, R2, R3, R4, R5, R6).Group based QSAR Models were derived using Multiple linear regression (MLR) analysis and selected on the basis of various statistical parameters. Dataset of benzothiazole reveled importance of presence of halogen atoms on is essential requirement. The generated models will provide structural requirements of benzothiazole derivatives which can be used to design and develop potent antitubercular derivatives.

Specifications Table

Subject area	Computational and Insilico Chemistry
More specific subject area	Group Quantitative Structure-Activity Relationship(QSAR)

* Corresponding author.
E-mail address: sschitlange@rediffmail.com (S.S. Chitlange).

http://dx.doi.org/10.1016/j.dib.2017.08.006

Type of data	Equation, Tables,Graphs
How data was acquired	Group based QSAR modelling
Data format	Analysis
Experimental factors	Multiple linear regression GQSAR models for predicting the inhibitory potential of benzothiazole dataset were created. 17 molecules were utilized as training dataset and 8 molecules utilized as test dataset.
Experimental features	Fragment descriptors and pMIC values were utilized in GQSAR analysis via stepwise variable selection method using dataset of 25 molecules.
Data source location	Pharmaceutical chemistry of Laboratory of Progressive Education Society's, Modern College of Pharmacy, Sector 21, Yamunanagar, Nigdi, Pune 411044, Maharashtra, India
Data accessibility	The data is with this article

Value of the data

- Tuberculosis is one of most lethal disease in the current decade; development of potent antitubercular compounds is need of time.
- GQSAR modelling data was developed for predicting structural properties of benzothiazole dataset which are infusing antitubercular activity.
- The GQSAR models generated will be utilized to screen various heterocyclic datasets for antitubercular potency, which will lead to development of novel antitubercular compounds.

1. Data

The data shown here regarding a GQSAR equation development that is used to predict contribution of substituents towards antitubercular potential of benzothiazole dataset.

2. Experimental design, materials and methods

2.1. Data set preparation

Molecular data set for current study were taken from literature reported by Telvekar et al. [1]. All the 24 structures of benzothiazole derivatives were drawn using 2D builder module of Vlife MDS 4.3. These 2D structures were converted into 3D via using V life engine platform. Geometry and structures of 3D molecules were optimized via energy minimization process using Merck molecular force field (MMFF) and Gasteiger charges. A common template which is a representative of the entire molecules under study was prepared with the presence of a dummy atom (X) at the substitution site.

2.2. Calculation of descriptors

The common chemical structure as shown in Fig. 1 was utilized for development of GQSAR model. The molecules in the data set were fragmented in six different fragments (R-R6). The fragmented molecules were incorporated into the QSAR module of V life MDS for calculation of molecular descriptors. Molecular descriptors are nothing but the numerical values which represents physical and chemical information of the molecules. In GQSAR studies descriptors are representation of the physical and chemical behavior of substituents present.

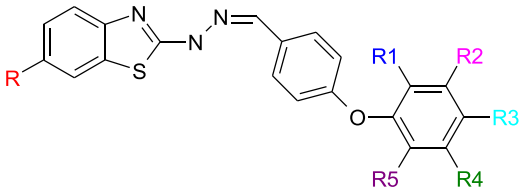

Fig. 1. Molecular Template Utilized for Fragmentation pattern.

Table 1
Table Showing Molecules under Study.

Mole. No	R	R_1	R_2	R_3	R_4	R_5
1.	H	H	H	H	H	H
2.	H	Cl	H	H	H	H
3.	H	H	H	Cl	H	H
4.	H	Cl	H	Cl	H	H
5.	H	H	CH_3	Cl	H	H
6.	Cl	H	H	H	H	H
7.	Cl	Cl	H	H	H	H
8.	Cl	H	H	Cl	H	H
9.	Cl	Cl	H	Cl	H	H
10.	Cl	H	CH_3	Cl	H	H
11.	CH_3	H	H	H	H	H
12.	CH_3	Cl	H	H	H	H
13.	CH_3	H	H	Cl	H	H
14.	CH_3	Cl	H	Cl	H	H
15.	CH_3	H	CH_3	Cl	H	H
16.	OCH_3	H	H	H	H	H
17.	OCH_3	Cl	H	H	H	H
18.	OCH_3	H	H	Cl	H	H
19.	OCH_3	Cl	H	Cl	H	H
20.	OCH_3	H	CH_3	Cl	H	H
21.	NO_2	H	H	H	H	H
22.	NO_2	Cl	H	H	H	H
23.	NO_2	H	H	Cl	H	H
24.	NO_2	Cl	H	Cl	H	H
25.	NO_2	H	CH_3	Cl	H	H

2.3. *Data selection and building G-QSAR model [2–5]*

Generated dataset of 25 benzothiazole derivatives were randomly divided into training set and test set 17 and 8 molecules respectively. Random distribution of training and test set will results into uniform distribution of biological activity across the molecules under study. Multiple linear regression analysis was utilized for development GQSAR models, with number of dependent variable limited to not more than 3 per model (Table 1).

2.4. *Validation of the developed G-QSAR model [6–10]*

Validation is a critical step in the QSAR model development. Validation methods are required for establishing predictability of QSAR model on unseen data and for determination of complexity of QSAR model which is justified by the data under study. Number of methods like the methods of least squares fit (R2), cross validation (Q2), adjusted R2 (R2adj), chi-squared test ($\chi 2$), root mean squared error (RMSE), bootstrapping and scrambling (Y-Randomization) are reported for internal validation of QSAR models. Observed activity of molecules in dataset was expressed in MIC(μg/ml) and converted into pMIC for QSAR analysis. All the molecules in the dataset are having activity (MIC) in the range 1.5–29.00 μg/ml.

Table 2
Table showing observed and predicted activity of selected GQSAR model.

Molecule. No	Observed Activity pMIC (μg/ml)	Predicted Activity pMIC (μg/ml)	Molecule No	Observed Activity pMIC (μg/ml)	Predicted Activity pMIC (μg/ml)
1	2.1	2.7	14	0.9	1.2
2	0.9	0.7	15	0.8	0.8
3	1.2	1.2	16	0.8	0.7
4 #	2.3	1.2	17#	0.7	1.3
5#	1.2	0.9	18#	1.1	0.7
6	1.3	1.6	19	0.6	1.3
7	3.3	4.1	20	0.6	0.8
8	2.8	3.6	21#	0.6	0.7
9	5.6	4.0	22#	0.6	1.1
10	1.5	3.7	23	0.7	0.7
11#	1.0	0.7	24	0.7	1.1
12	1.4	1.2	25#	0.7	0.8
13	0.8	0.7			

Test Set Molecules

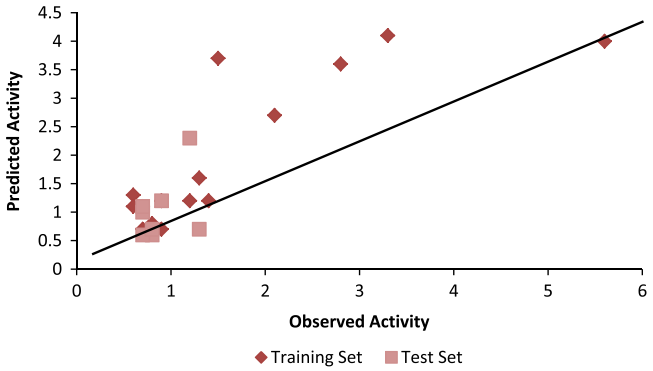

Fig. 2. Figure Showing Correlation Plot for Selected GQSAR model having r^2 0.88.

2.5. QSAR analysis

Congeneric nature of the dataset is basis prerequisite for any QSAR analysis. Fragment based QSAR is recent methodology were complex structures can be analyzed. 30 different G-QSAR models were generated and best one of them are selected on basis of the statistical values like r^2, q^2, pred_r^2, F-test and standard error. The predicted activity data via QSAR models was in accordance with the observed biological activity with small variations which were clearly identified in the correlation plot of different model (Table 2 and Fig. 2). Selected model is given by.

pMIC: 0.0038 + 2.9110(± 0.4296)R1-ChlorinesCount + 5.5097(± 2.0358) R2-MomInertiaX.

r^2: 0.8845, q^2: 0.6059, F test: 35.45.

Acknowledgement

The authors are thank full to Vlife Sciences for providing software for study.

Transparency document. Supporting information

Transparency data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2017.08.006.

References

[1] V.N. Telvekar, V.K. Bairwa, K. Satardekar, A. Bellubi, Novel 2-(2-(4-aryloxybenzylidene) hydrazinyl)benzothiazole derivatives as anti-tubercular agents, Bioorganic Med. Chem. Lett. 22 (2012) 649–652.
[2] P.B. Choudhari, M.S. Bhatia, 3D QSAR, docking studies and pharmacophore modelling of selected factor Xa inhibitors, Med. Chem. Res 21 (2012) 1427–1432.
[3] M.S. Bhatia, K.D. Pakhare, P.B. Choudhari, S.D. Jadhav, R.P. Dhavale, N.M. Bhatia, Pharmacophore modeling and 3D QSAR studies of aryl amine derivatives as potential lumazine synthase inhibitors, Arab J. Chem. 10 (2017) S100–S104.
[4] P.B. Choudhari, M.S. Bhatia, S.D. Jadhav, Pharmacophore modelling, quantitative structure activity relationship (QSAR) and docking studies of pyrimidine analogs as potential calcium channel blockers, J. Korean Chem. Soc. 57 (2013) 99–103.
[5] C. Vats, J. Dhanjal, S. Goyal, N. Bharadvaja, A. Grover, Computational design of novel flavonoid analogues as potential AChE inhibitors: analysis using group-based QSAR, molecular docking and molecular dynamics simulations, J. Str. Chem. 26 (2014) 467–476.
[6] M.D.S. VLife, Molecular Design Suite. In., 3.0 edn, Vlife Sciences Technologies Pvt. Ltd., Pune, India, 2004.
[7] M.P. Gonzalez, C. Teran, L. Saiz-urraand, M. Teijeira, Variable selection methods in QSAR: an overview, Curr. Top. Med. Chem. 8 (2008) 1606–1607.
[8] A. Golbraikhand, A. Tropsha, Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection, J. Com. Mol. Des. 16 (2002) 357–359.
[9] A.D. Abdullahi, A.M. Abdualkader, N.H.A. Samat, F. Mohamed, B.Y. Muhammad, H.A. Mohammed, A. Aljarbou, A. Kasmuri, Novel insight into the structural requirements of P70S6K inhibition using Group-based Quantitative Structure Activity Relationship (GQSAR), J. Appl. Pharm. Sci. 4 (2014) 16–24.
[10] N.V. Lakshmi, P.M. Sivakumar, D. Muralidharan, M. Doble, P.T. Perumal, Expeditious synthesis, antibacterial activity evaluation and GQSAR studies of 3-bisoxindoles, 2-oxindolyl-2-hydroxyindan-1,3-diones and 2-oxindolyl-2-hydroxyacenaphthylen-1-ones, RSC Adv. 3 (2013) 496–507.