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 revealed 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.

© 2017 Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
The data shown here regarding a GQSAR equation development that is used to predict contribution of substituents towards antitubercular potential of benzothiazole dataset.
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 (χ2), 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 1
Table Showing Molecules under Study.

| Mole. No | R   | R1 | R2 | R3 | R4 | R5 |
|---------|-----|----|----|----|----|----|
| 1.      | H   | H  | H  | H  | H  | H  |
| 2.      | H   | Cl | H  | H  | H  | H  |
| 3.      | H   | H  | Cl | H  | H  | H  |
| 4.      | H   | Cl | Cl | H  | H  | H  |
| 5.      | Cl  | H  | H  | H  | Cl | H  |
| 6.      | Cl  | H  | Cl | H  | Cl | H  |
| 7.      | Cl  | Cl | H  | H  | Cl | H  |
| 8.      | Cl  | Cl | Cl | H  | Cl | H  |
| 9.      | Cl  | H  | Cl | H  | Cl | H  |
| 10.     | Cl  | H  | Cl | H  | Cl | H  |
| 11.     | CH3 | H  | H  | H  | H  | H  |
| 12.     | CH3 | H  | Cl | H  | H  | H  |
| 13.     | CH3 | H  | H  | Cl | H  | H  |
| 14.     | CH3 | Cl | H  | Cl | H  | H  |
| 15.     | CH3 | H  | Cl | H  | Cl | H  |
| 16.     | OCH3| H  | H  | H  | H  | H  |
| 17.     | OCH3| Cl | H  | H  | H  | H  |
| 18.     | OCH3| H  | Cl | H  | H  | H  |
| 19.     | OCH3| Cl | H  | Cl | H  | H  |
| 20.     | OCH3| H  | Cl | H  | Cl | H  |
| 21.     | NO2 | H  | H  | H  | H  | H  |
| 22.     | NO2 | Cl | H  | H  | H  | H  |
| 23.     | NO2 | H  | Cl | H  | H  | H  |
| 24.     | NO2 | Cl | H  | Cl | H  | H  |
| 25.     | NO2 | H  | Cl | Cl | H  | H  |

Fig. 1. Molecular Template Utilized for Fragmentation pattern.
2.5. QSAR analysis

Congeneric nature of the dataset is basis prerequisite for any QSAR analysis. Fragment based QSAR is recent methodology where 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:

$$
pMIC: 0.0038 + 2.9110( \pm 0.4296)R1-\text{ChlorinesCount} + 5.5097( \pm 2.0358) R2-\text{MomInertiaX}.
$$

$\chi^2$: 0.8845, $q^2$: 0.6059, F test: 35.45.

### Acknowledgement

The authors are thankful 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-hydroxyacenaphthylene-1-ones, RSC Adv. 3 (2013) 496–507.