Journal of the Serbian Chemical Society
JSCS@tmf.bg.ac.rs • www.shd.org.rs/JSCS
*J. Serb. Chem. Soc.* 89 (7–8) S213–S217 (2024) *Supplementary material*

SUPPLEMENTARY MATERIAL TO

# Development of 2D and 3D QSAR models of pyrazole derivatives as acetylcholine esterase inhibitors

PUJA MISHRA[1*], SUMIT NANDI[1], ANKIT CHATTERJEE[1], TRIDIB NAYEK[1], SOUVIK BASAK[1], AMIT KUMAR HALDER[1,2]and ARUP MUKHERJEE[3]

*[1]Dr. B.C. Roy College of Pharmacy & Allied Health Sciences, Durgapur, WB, India, [2]LAQV@REQUIMTE/Department of Chemistry and Biochemistry, Faculty of Sciences, University of Porto, Porto, Portugal and [3]Department of Biotechnology, Maulana Abul Kalam Azad University of Technology, WB, India*

*J. Serb. Chem. Soc.* 89 (7–8) (2024) 981–995

## TEXT S1. MODEL VALIDATION PROCEDURES

*Leave-one-out method*

Leave-one-out (LOO), one of the popular cross-validation methods was implemented. The cross-validated $R^2$ is the result from the cross-validation method which is used as a condition of robustness and predictive capability of the model. The cross-validated determination coefficient $R^2$ that is LOO-$Q^2$ is calculated using the following formula

$$Q^2 = 1 - \sum(Y_{obs} - Y_{pred})^2 / \sum(Y_{obs} - \bar{Y})^2$$

In this equation, $\bar{Y}$ denotes average activity value of the training set, $Y_{obs}$ and $Y_{pred}$ signify the values of observed and predicted activity respectively. A high $Q^2$ value ($Q^2$>0.5) is considered as parameter of the high predictive capability of the model.

*External validation*

*Calculation of $R^2_{pred}$ from test set*

In this study, the original dataset was divided into training set and test set. Based on the training set compounds, 2D QSAR model (MLR equation) was developed. The predictive ability of the models was determined by the statistical parameter, like the predictive $R^2$ ($R^2_{pred}$) values, which was derived by the following equation

$$R^2_{pred} = 1 - \sum(Y_{pred(Test)} - Y_{Test})^2 / \sum(Y_{Test} - \bar{Y}_{Training})^2$$

* Corresponding author. E-mail: pujam.phe15@itbhu.ac.in

In this equation, $Y_{pred\ (Test)}$ and $Y_{(Test)}$ represents the predicted and the observed activity values of a certain test set of compounds, respectively, and $\bar{Y}$training denotes the mean activity value of the training set. The value of $R^2$ pred $> 0.6$ indicates a good external predictability of a model.

Here, the multiple linear regression model generated from trainee set was used for the prediction purpose of bioactivity of the test set compounds by forming an equation as

$$\log \text{Predicted activity (y)} = m_1x_1 + m_2x_2 + m_3x_3 + \text{---} + m_nx_n + C$$

where $m_1$, $m_2$, $m_3$, ---, $m_n$ are respective coefficients of descriptors such as $x_1$, $x_2$, $x_3$, -----, $x_n$

A well-known statistical method (metrics) was used to determine the goodness of the fit, the robustness and the internal predictivity of the final QSAR model. The coefficient of determination ($R^2$), adjusted $R^2$ ($R^2_A$) and the MAE (mean absolute error) were used to generate the goodness fit of the model and the internal cross validation coefficient $Q^2_{LOO}$ was used to check the robustness and the internal predictivity. Further, the metrics such as $r_m^2$ and $r_m^2{}_{LOO}$ are used as the internal validation parameters. The external validation parameter $R^2_{Pred}$ is essential to judge the external predictivity of the molecule. The best QSAR model was validated by the inter-correlation (Pearson r) between the two descriptors. A correlation matrix was generated to check the correlation matrix of different descriptors.

## TEXT S2. ENERGY OPTIMIZATION OF SELECTED MOLECULES

The results of Auto energy optimization of the compounds using Avogadro give values, such as 260.085 KJ/mol which is confirmed using Orca (the semi empirical method). The structural optimization using both the software exhibits that there is no significant change in geometry after the various geometry optimization cycle.

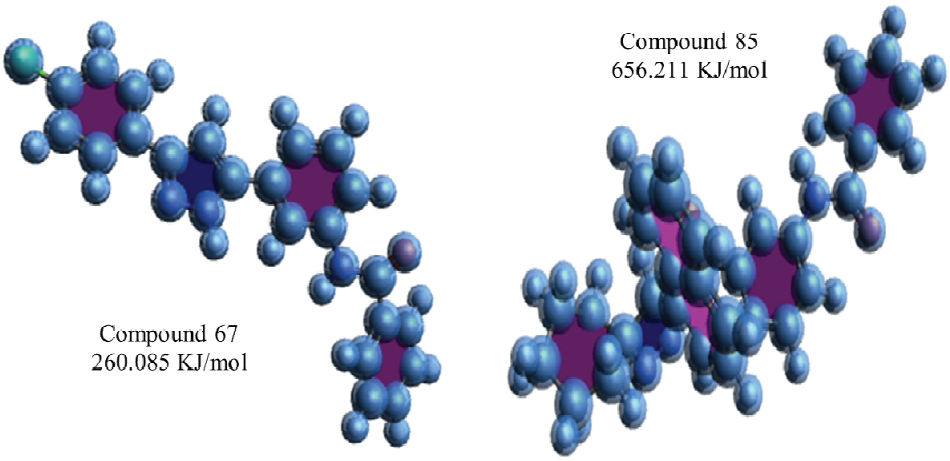


Fig. S1. Geometry optimization using Avogadro by using MM force fields.

QM optimized structures ( eventually with a semi-empirical method)

| | | | |
|---|---|---|---|
| Energy change | -0.0004823076 | 0.0000050000 | (No significant energy change after 24 geometry optimization cycle) |
| RMS gradient | 0.0004785364 | 0.0001000000 | |
| MAX gradient | 0.0034948286 | 0.0003000000 | |
| RMS step | 0.0034607630 | 0.0020000000 | |
| MAX step | 0.0320313071 | 0.0040000000 | |

Diagonalization of the overlap matrix:

| | |
|---|---|
| Smallest eigenvalue | ... 2.478e-04 |
| Time for diagonalization | 0.275 sec |
| Threshold for overlap eigenvalues | 1.000e-08 |
| Number of eigenvalues below threshold | 0 |
| Time for construction of square roots | 0.112 sec |
| Total time needed | 0.395 sec |

**Table SI.** Trainee set with chosen descriptors of 2D-QSAR

| Sr. No. | Experimental $pIC_{50}$ | Predicted $pIC_{50}$ | nBM | C-001 | Mv | CATS2D_08_DL | C-014 |
|---|---|---|---|---|---|---|---|
| 43 | 5.393 | 5.713 | 24 | 0 | 0.6853 | 4 | 0 |
| 44 | 5.713 | 5.517 | 24 | 0 | 0.7045 | 4 | 0 |
| 46 | 5.711 | 5.665 | 24 | 0 | 0.7238 | 3 | 0 |
| 47 | 5.665 | 5.521 | 24 | 0 | 0.7091 | 4 | 0 |
| 49 | 5.68 | 5.68 | 24 | 0 | 0.6942 | 4 | 0 |
| 50 | 5.589 | 5.589 | 24 | 0 | 0.6942 | 4 | 0 |
| 54 | 5.627 | 5.073 | 24 | 0 | 0.6988 | 4 | 0 |
| 55 | 5.495 | 5.06 | 24 | 0 | 0.6988 | 4 | 0 |
| 56 | 5.215 | 5.627 | 24 | 0 | 0.6991 | 4 | 1 |
| 57 | 5.417 | 5.495 | 25 | 0 | 0.7037 | 4 | 0 |
| 58 | 5.405 | 5.215 | 25 | 0 | 0.7037 | 4 | 0 |
| 59 | 4.828 | 5.417 | 24 | 1 | 0.6738 | 4 | 0 |
| 60 | 4.771 | 5.405 | 24 | 1 | 0.6738 | 4 | 0 |
| 66 | 5.705 | 4.828 | 21 | 0 | 0.6469 | 4 | 0 |
| 67 | 6.333 | 4.81 | 21 | 0 | 0.6607 | 4 | 0 |
| 68 | 5.706 | 7.721 | 21 | 0 | 0.6607 | 3 | 0 |
| 69 | 5.877 | 5.705 | 21 | 0 | 0.6745 | 3 | 0 |
| 70 | 5.943 | 6.333 | 21 | 0 | 0.664 | 4 | 0 |
| 71 | 5.752 | 5.877 | 21 | 0 | 0.664 | 4 | 0 |
| 72 | 6.023 | 5.943 | 21 | 0 | 0.6533 | 4 | 0 |
| 73 | 5.789 | 5.752 | 21 | 0 | 0.6533 | 4 | 0 |
| 74 | 5.635 | 6.023 | 21 | 0 | 0.6415 | 4 | 0 |
| 75 | 5.548 | 5.789 | 21 | 0 | 0.6415 | 4 | 0 |
| 76 | 5.521 | 5.635 | 21 | 0 | 0.6367 | 4 | 0 |
| 77 | 5.838 | 5.548 | 21 | 0 | 0.6586 | 4 | 0 |
| 78 | 5.716 | 5.521 | 21 | 0 | 0.6586 | 4 | 0 |
| 79 | 5.65 | 5.716 | 21 | 0 | 0.6595 | 4 | 1 |
| 81 | 5.707 | 5.65 | 22 | 0 | 0.6608 | 4 | 0 |
| 82 | 5.329 | 5.75 | 21 | 1 | 0.6403 | 4 | 0 |
| 83 | 5.23 | 5.707 | 21 | 1 | 0.6403 | 4 | 0 |
| 84 | 5.146 | 5.23 | 21 | 2 | 0.6289 | 6 | 0 |
| 85 | 4.535 | 5.146 | 26 | 0 | 0.6567 | 4 | 0 |

**Table SII.** Test set with chosen descriptors (arranged in descending order of $pIC_{50}$ values)

| Sr. No. | Experimental $pIC_{50}$ | Predicted $pIC_{50}$ | C-001 | Mv | CATS2D_08_DL | C-014 |
|---|---|---|---|---|---|---|
| 45 | 5.517 | 5.393 | 24 | 0 | 0.7045 | 3 |
| 48 | 5.521 | 5.711 | 24 | 0 | 0.7091 | 4 |
| 51 | 5.073 | 4.818 | 24 | 0 | 0.6746 | 4 |
| 52 | 5.06 | 4.771 | 24 | 0 | 0.6746 | 4 |
| 53 | 4.818 | 4.556 | 24 | 0 | 0.6657 | 4 |
| 61 | 4.81 | 5.706 | 24 | 2 | 0.6547 | 6 |
| 62 | 4.556 | 5.838 | 29 | 0 | 0.6937 | 4 |
| 80 | 5.75 | 5.329 | 22 | 0 | 0.6608 | 4 |

**Table SIII.** Elaboration on chosen descriptors

| Sr.No. | Descriptors | Class | Description |
|---|---|---|---|
| 1. | nBM | constitutional descriptors | number of multiple bond |
| 2. | C-001 | atom centred fragments | $CH_3R/CH_4$ |
| 3. | Mv | constitutional descriptors | mean atomic vander waals volume (scaled on carbon atom) |
| 4. | CATS2D_08_DL | pharmacophore descriptors | CATS2D donor-lipophilic at lag 8 |
| 5. | C-014 | atom centred fragments | $CX_4$ |