**Chalcones and Dihydropyrazole derivatives demonstrate antiproliferative potentials in prostate cancer - A Combined QSAR, Machine Learning, Molecular Docking and ADMETox Investigation**

O.E. Oyeneyina\*, B.S. Obadawob, D.S. Metibemuc, T.O. Owolabid, A.A. Olanrewajue, S.M. Orimoloyef, N. Ipinlojua and O. Olusayog

a. Theoretical and Computational Chemistry Unit, Department of Chemical Sciences, Adekunle Ajasin University, Akungba-Akoko, Ondo State, Nigeria.

b. Department of Chemistry, University of Toledo, Ohio.

c. Department of Biochemistry, Adekunle Ajasin University, Akungba-Akoko, Ondo State, Nigeria.

d. Department of Physics and Electronics, Adekunle Ajasin University, Akungba-Akoko, Ondo State, Nigeria.

e. Chemistry and Industrial Chemistry Programmes, Bowen University, Iwo, Nigeria.

f. Department of Computer Science, Adekunle Ajasin University, Akungba-Akoko, Ondo State, Nigeria.

g. Department of Physics, Federal University Oye Ekiti, Oye Ekiti, Ekiti State, Nigeria.

\*Corresponding author: emmanueltoba90@gmail.com; oluwatoba.oyeneyin@aaua.edu.ng (+2348030417958)

**Table S1:** Structures of the molecules under investigation

 

  **2-16 17-31**

|  |  |  |  |
| --- | --- | --- | --- |
| MOLECULES | R | MOLECULES | R |
| 2 |  | 17 |  |
| 3 |  | 18 |  |
| 4 |  | 19 |  |
| 5 |  | 20 |  |
| 6 |  | 21 |  |
| 7 |  | 22 |  |
| 8 |  | 23 |  |
| 9 |  | 24 |  |
| 10 |  | 25 |  |
| 11 |  | 26 |  |
| 12 |  | 27 |  |
| 13 |  | 28 |  |
| 14 |  | 29 |  |
| 15 |  | 30 |  |
| 16 |  | 31 |  |

**Table S2:** Generally recommended values of the validation parameters of a built QSAR model

Parameter Definition Recommended value

R2  Coefficient of determination ≥0.6

P(95%) Confidence interval at 95% confidence level <0.05

$Q\_{cv}^{2}$ Cross validation coefficient ≥0.5

R2 -$Q\_{cv}^{2}$ Difference between R2 and $Q\_{cv}^{2} $ <0.3

N(ext & test set) Minimum number of external test set ≥5

$cR\_{p}^{2}$ Coefficient of determination for Y-randomization ≥0.5

**Table S3:** External validation of developed model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Name | ATS1s | VR2\_Dzi | SpMax5\_Bhm | Mi | $$Y\_{exp}\_{test}$$ | $$Y\_{pred}\_{test}$$ |
| 3 | 94.04115 | 7.094497998 | 3.083744 | 7.410015 | 3.75 | 3.669164 |
| 6 | 89.87449 | 9.576671908 | 3.045438 | 7.385456 | 3.85 | 3.752546 |
| 7 | 88.70782 | 9.840120491 | 3.046522 | 7.435682 | 4.51 | 4.125395 |
| 10 | 136.7634 | 15.13777952 | 3.395406 | 7.490422 | 3.86 | 3.651605 |
| 14 | 125.784 | 12.42268676 | 3.395593 | 7.425355 | 3.69 | 3.496997 |
| 16 | 124.6934 | 24.20505641 | 3.543003 | 7.416142 | 4.77 | 4.488785 |
| 21 | 128.7634 | 11.61007426 | 3.394971 | 7.465871 | 3.87 | 3.584192 |
| 24 | 123.0967 | 10.9532925 | 3.394853 | 7.484414 | 3.59 | 3.872006 |
| 28 | 91.39506 | 6.727236349 | 3.266787 | 7.352682 | 3.76 | 3.745752 |

**Table S4:** Calculation of the predicted R2 of developed model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| $$Y\_{pred}\_{test}-Y\_{exp}\_{test}$$ | $$Y\_{pred}\_{test}-Y\_{exp}\_{test}$$ | $$\overbar{Y}\_{training}$$ | $$Y\_{pred}\_{test}-\overbar{Y}\_{training}$$ | $$Y\_{pred}\_{test}-\overbar{Y}\_{training}^{2}$$ |
| -0.08084 | 0.006534 | 3.711 | 0.039 | 0.001521 |
| -0.09745 | 0.009497 | 3.711 | 0.139 | 0.019321 |
| -0.3846 | 0.147921 | 3.711 | 0.799 | 0.638401 |
| -0.20839 | 0.043428 | 3.711 | 0.149 | 0.022201 |
| -0.193 | 0.03725 | 3.711 | -0.021 | 0.000441 |
| -0.28121 | 0.079082 | 3.711 | 1.059 | 1.121481 |
| -0.28581 | 0.081686 | 3.711 | 0.159 | 0.025281 |
| 0.282006 | 0.079527 | 3.711 | -0.121 | 0.014641 |
| -0.01425 | 0.000203 | 3.711 | 0.049 | 0.002401 |
|  |

$\sum\_{}^{}\left(Y\_{pred}\_{test}-Y\_{exp}\_{test}\right) ^{2}= $0.485129$ $

 $\sum\_{}^{}\left(Y\_{pred}\_{test}-\overbar{Y}\_{training}\right)^{2}= $1.845689

$R\_{Pred}^{2}=1- \frac{0.485129}{1.845689}= $0.737155

**Table S5:** Pearson’s correlation matrix, VIF and ME of the descriptors used in the built model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Descriptors | *ATS1s* | *VR2\_Dzi* | *SpMax5\_Bhm* | *Mi* | *VIF* | *ME* |
| ATS1s | 1 | 0.732712 | 0.711697 | 0.73514 | 11.74599 | -0.08976 |
| VR2\_Dzi | 0.732712 | 1 | 0.586512 | 0.251353 | 3.666384 | 0.014118 |
| SpMax5\_Bhm | 0.711697 | 0.586512 | 1 | 0.314264 | 2.608657 | 0.130198 |
| Mi | 0.73514 | 0.251353 | 0.314264 | 1 | 4.490817 | 0.945449 |

**Table S6:** Comparison of the QSAR model with the ELM Sine and Sig models

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| S/No | Measured activity (pIC50) | QSAR (pIC50) | %error | ELM-Sine (pIC50) | % error | ELM\_Sig(pIC50) | % error |
| 2 | 3.36 | 3.40 | 1.20 | 3.36 | 0.00 | 3.44 | 2.38 |
| 3 | 3.75 | 3.67 | 2.15 | 3.75 | 0.00 | 3.75 | 0.11 |
| 4 | 3.41 | 3.64 | 6.79 | 3.84 | 12.61 | 3.73 | 9.47 |
| 5 | 3.83 | 3.63 | 5.19 | 3.83 | 0.00 | 3.73 | 2.65 |
| 6 | 3.85 | 3.75 | 2.53 | 3.85 | 0.00 | 4.00 | 3.88 |
| 7 | 4.51 | 4.13 | 8.53 | 4.51 | 0.00 | 4.46 | 1.14 |
| 8 | 3.60 | 3.63 | 0.79 | 3.60 | 0.00 | 3.60 | 0.08 |
| 9 | 3.53 | 3.61 | 2.24 | 3.53 | 0.00 | 3.65 | 3.37 |
| 10 | 3.86 | 3.65 | 5.40 | 3.86 | 0.00 | 3.76 | 2.48 |
| 11 | 3.45 | 3.47 | 0.71 | 3.45 | 0.00 | 3.61 | 4.77 |
| 12 | 3.59 | 3.50 | 2.61 | 3.59 | 0.00 | 3.76 | 4.84 |
| 13 | 3.68 | 3.70 | 0.56 | 3.68 | 0.00 | 3.66 | 0.45 |
| 14 | 3.69 | 3.50 | 5.23 | 3.69 | 0.00 | 3.53 | 4.35 |
| 15 | 4.03 | 3.98 | 1.25 | 4.03 | 0.00 | 3.95 | 2.09 |
| 16 | 4.77 | 4.49 | 5.90 | 4.77 | 0.00 | 4.74 | 0.72 |
| 17 | 3.45 | 3.49 | 1.06 | 3.45 | 0.00 | 3.39 | 1.78 |
| 18 | 3.92 | 3.92 | 0.11 | 3.92 | 0.00 | 3.98 | 1.45 |
| 19 | 3.49 | 3.60 | 3.25 | 3.49 | 0.00 | 3.64 | 4.28 |
| 20 | 3.83 | 3.58 | 6.42 | 3.84 | 0.14 | 3.62 | 5.41 |
| 21 | 3.87 | 3.58 | 7.39 | 3.83 | 0.99 | 3.62 | 6.38 |
| 22 | 3.58 | 3.64 | 1.70 | 3.54 | 1.17 | 3.60 | 0.62 |
| 23 | 3.69 | 3.65 | 1.16 | 3.69 | 0.00 | 3.64 | 1.24 |
| 24 | 3.59 | 3.87 | 7.86 | 3.59 | 0.00 | 3.54 | 1.35 |
| 25 | 3.88 | 3.79 | 2.21 | 3.88 | 0.00 | 3.60 | 7.19 |
| 26 | 3.57 | 3.56 | 0.18 | 3.57 | 0.00 | 3.64 | 2.00 |
| 27 | 3.68 | 3.86 | 4.92 | 4.10 | 11.33 | 3.93 | 6.80 |
| 28 | 3.76 | 3.75 | 0.38 | 3.76 | 0.00 | 3.88 | 3.07 |
| 29 | 3.79 | 3.78 | 0.22 | 3.79 | 0.00 | 3.81 | 0.50 |
| 30 | 4.08 | 3.97 | 2.80 | 4.08 | 0.00 | 4.04 | 1.05 |
| 31 | 4.50 | 4.53 | 0.57 | 4.69 | 4.27 | 4.25 | 5.59 |
|  | **MAPD** |  | **3.04** |  | **1.02** |  | **3.05** |

**Table S7:** The mean absolute error, root mean-square error and correlation coefficient of the QSAR model with the ELM Sine and Sig models

|  |  |  |  |
| --- | --- | --- | --- |
|  | MAE (pIC50) | RMSE (pIC50) | CC |
| QSAR | 0.12 | 0.16 | 88.87 |
| ELM-Sine | 0.04 | 0.12 | 94.42 |
| ELM-Sig | 0.11 | 0.14 | 89.54 |



**Figure S1:** 15 with 2AXA



**Figure S2:** 16 with 2AXA



**Figure S3:** 18 with 2AXA



**Figure S4:** 30 with 2AXA



 **Figure S5:** 31 with 2AXA



**Figure S6:** R-bicalutamide with 2AXA