

**Journal homepage:** <http://dergipark.ulakbim.gov.tr/jotcsa>



e-ISSN: 2149-0120

---

## **In Silico Modelling of Cytotoxic Behaviour of Anti-Leukemic Compounds on HL-60 Cell Line**

David Ebuka Arthur\*<sup>1</sup>, Adamu Uzairu<sup>1</sup>, Paul Mamza<sup>1</sup>, Stephen Eyije Abechi<sup>1</sup>, and Gideon Shallangwa<sup>1</sup>

<sup>1</sup>Department of Chemistry, Ahmadu Bello University (ABU) Zaria, Kaduna State, Nigeria.

**Abstract:** This research employs multiple linear regression technique in the modelling of some potent anti-leukemic compounds using paDEL molecular descriptor software calculator, to identify the best relationship between the chemical structure and toxicities of the anticancer datasets against some leukemic cell lines (HL-60). Statistical parameters such as  $Q^2$  and  $R^2_{pred}$  (test set) were computed to validate the strength of the model, while Williams plot was used to assess its applicability domain. The mean effects of the molecular descriptors in the models were calculated to illuminate the principal properties of the molecules responsible for their cytotoxicity.

**Keywords:** HL-60 cell lines, leukemia, QSTR, applicability domain,  $\gamma$ -randomization.

**Submitted:** April 01, 2016. **Revised:** April 30, 2016. **Accepted:** May 25, 2016.

**Cite this:** Arthur D, Uzairu A, Mamza P, Abechi S, Shallangwa G. IN SILICO MODELLING OF CYTOTOXIC BEHAVIOUR OF ANTI-LEUKEMIA COMPOUNDS ON HL-60 CELL LINE. Journal of the Turkish Chemical Society, Section A: Chemistry. 2016;3(2):147–158.

**DOI:** 10.18596/jotcsa.40785.

**Correspondence to:** David Ebuka Arthur. E-mail: [hanslibs@myway.com](mailto:hanslibs@myway.com). **Tel:** +2348138325431.

## INTRODUCTION

Leukaemia is normally referred to as the cancer of the blood cells and these cancers do not form solid tumours, instead they form large numbers of abnormal white blood cells which builds up in the blood and bone marrow, crowding out normal blood cells. The low level of normal blood cells thereby makes it harder for the body to get oxygen to its tissues and fight infections. Cancer is a major disease; this can be seen from the WHO statistical report which shows that, at the present time, around one in five of all deaths are because of this disease. Presently the best way for treating metastasised cancer, is chemotherapy but their toxic nature often offers significant limitations.

Quantitative structure-activity relationship (QSAR) and quantitative structure-toxicity relationship (QSTR) make it conceivable to anticipate the activity/toxicity of a given compound as a function of its molecular descriptors. Despite the fact that there is impressive enthusiasm for the use of QSARs to predict toxicity, there seems to be some accompanying limitations. QSTR can be used to predict the toxicity of unknown chemicals in many ways. In industries, these models can be applied to screen new compounds and help with the procedure for designing less toxic but potent drugs.

## MATERIALS AND METHODS

### Experimental Dataset

In the present study, we used a dataset of 112 compounds whose anticancer activity against human acute lymphoblastic leukaemia (HL-60) leukaemia tumour cell line has been previously reported. The dataset along with their NSC number were taken from the drug discovery and development arm of the National Cancer Institute (NCI). Eligible compounds were determined by reviewing and curating the raw data collected from the literature (NCI database). The data encompasses aminopterin and camptothecin derivatives, and colchicine analogues. The anticancer toxicity results are expressed as pLC<sub>50</sub>, which is the lethal concentration for 50% of maximal inhibition of cell proliferation.

### Geometric optimisation and molecular descriptors calculation

The 2D structures of the compounds were drawn using Chemdraw software and the spatial conformations of the compounds were determined using the Spartan 14 V1.1.4 wavefunction software package, the molecular structures were first minimised by Molecular Mechanics Force Field (MMFF) calculation to remove strain energy before subjecting it to DFT (density functional theory) method for complete geometric optimisation of the structures. Afterwards, we calculated 1875 descriptors (1444 1D, 2D descriptors and 431 3D descriptors) molecular descriptors using the paDEL program such as, atom-type electrotopological state descriptors, McGowan volume, RDF descriptors, *etc.*

### Data Normalisation

The calculated molecular descriptors were normalised via a method preserving the range (maximum and minimum) and introducing the dispersion of the descriptors (standard deviation / variance) before they are converted into a  $N(0,1)$  distribution, making the comparison between descriptors (probabilities calculation) much easier.

### Variable Selection

In this study, genetic algorithm technique was employed as a selection tool to select the most relevant descriptors with respect to an objective function. Genetic algorithm method as a selection tool was written in Matlab 6.5 program and used here. The genetic algorithm (GA) starts with the creation of a population of randomly generated parameter sets. The parameters set used for the GA includes mutation 0.1, crossover 0.9, population 10000, number of generations 1000,  $R^2$  floor limit 50%, and objective function  $R^2/N_{par}$ . Equations were developed between the observed toxicity and the descriptors. The best equation was taken based on statistical parameters such as squared regression coefficient ( $R^2$ ) and leave-one-out cross-validated regression coefficient ( $Q^2_{cv}$ ).

### Data Division

In order to obtain validated QSAR models the dataset was divided into training and test sets. Ideally, this division should be performed such that points representing both training (80% of compounds) and test sets (20% percent of compounds) are distributed within the whole descriptor space occupied by the entire dataset, and each point of the test set is close to at least one point of the training set. This partitioning ensures that a similar principle can be employed for the toxicity prediction of the test set. Kennard-Stone Algorithm will be applied for dividing Dataset into a Training and Test set.

### Model development

MLR is a method used for modelling linear relationship between a dependent variable  $Y$  (pLC50) and independent variable  $X$  (molecular descriptors). MLR is based on least squares: the model is fit such that sum-of-squares of differences of observed and a predicted value is minimized. MLR estimates values of regression coefficients ( $R^2$ ) by applying least squares curve fitting method.

### Evaluation of the QSAR Models

The Developed QSAR models are evaluated using the following statistical measures:  $n$  (Number of compounds in regression);  $K$  (Number of descriptors);  $R^2$  (the squared correlation coefficient);  $F$  test (Fischer's Value) for statistical significance;  $Q^2$  (cross-validated correlation coefficient);  $pred R^2$  ( $R^2$  for external test set). The regression coefficient  $R^2$  is a relative measure of fit by the regression equation. It represents the part of the variation in the observed data is explained by the regression. However, a QSAR model is considered to be predictive, if the following conditions are satisfied:  $R^2 > 0.6$ ,  $q^2 > 0.6$  and  $pred R^2 > 0.5$ . The low standard error of  $pred R^2_{se}$ ,  $Q^2_{se}$  and  $R^2_{se}$  shows absolute quality of the fitness of the model.

### Validation of the QSAR model

The predictive capability of the QSAR equation was determined using the leave-one-out cross-validation method. The cross-validation regression coefficient ( $Q_{cv}^2$ ) was calculated by the following equation:

$$Q_{cv}^2 = 1 - \frac{PRESS}{TOTAL} = 1 - \frac{\sum_{i=1}^n (y_{exp} - y_{pred})^2}{\sum_{i=1}^n (y_{exp} - \bar{y})^2} \quad (1)$$

Where  $y_{pred}$ ,  $y_{exp}$ , and  $\bar{y}$  are the predicted, experimental, and mean values of experimental activity, respectively. Also, the accuracy of the prediction of the QSAR equation was validated by  $F_{value}$ ,  $R^2$ , and  $R^2_{adj}$ . A large F indicates that the model fit is not a chance occurrence. It has been shown that a high value of statistical characteristics is not necessary for the proof of a highly predictive model. Hence, to evaluate the predictive ability of our QSAR model, we used the method described by Roy *et al.* The coefficient of determination in the test set,  $R^2_{test}$ , was calculated using the following equation:

$$R^2_{Test} = 1 - \frac{\sum (Y_{pred_{test}} - Y_{Test})^2}{\sum (Y_{pred_{test}} - \bar{Y}_{Training})^2} \quad (2)$$

where  $Y_{pred_{test}}$  and  $Y_{Test}$  are the predicted value based on the QSAR equation (model response) and experimental activity values, respectively, of the external test set compounds.  $\bar{Y}_{Training}$  is the mean activity value of the training set compounds. Further evaluation of the predictive ability of the QSAR model for the external test set compounds was done by determining the value of  $r_m^2$  by the following equation:

$$rm^2 = r_{test}^2 \left( 1 - \left| \sqrt{r_{test}^2 - r_{test_0}^2} \right| \right) \quad (3)$$

where  $r_{test}^2$  is the square correlation coefficient between experimental and predicted values and  $r_{test_0}^2$  is the squared correlation coefficient between experimental and predicted values without intercept for the external test set compounds.

See Table 1 in Supplementary Files.

### Evaluation of the applicability domain of the model

Evaluation of the applicability domain of the QSAR model is considered an important step to establish that the model is reliable to make predictions within the chemical space for which it was developed. There are several methods for defining the applicability domain of a QSAR model, but we used the most commonly used leverage approach in this study. Leverage of a given chemical compound  $h_i$  is defined as:  $h_i = x_i(X^T X)^{-1} x_i^T$ , where  $x_i$  is the descriptor row-vector of the query compound  $i$ , and  $x$  is the  $n \times k$  descriptor matrix of the training set compounds used to develop the model. As a prediction tool, the warning leverage ( $h^*$ ) which is the limit of normal values for  $X$  outliers, its defined as:  $h^* = 3(p + 1) / n$ , where  $n$  is the number of training compounds, and  $k$  is the number of descriptors in the model. The test compounds with leverages are considered to be reliably predicted by the model. The Williams plot, a plot of standardized residuals vs. leverage values, is used to interpret the applicability domain of the model.

## RESULTS AND DISCUSSION

A QSAR analysis was performed to explore the structure–activity relationship of different 112 compounds with different organic moiety acting as anticancer. In a QSAR study, generally, the quality of a model is expressed by its fitting and prediction ability (see Table 2 in Supplementary Files).

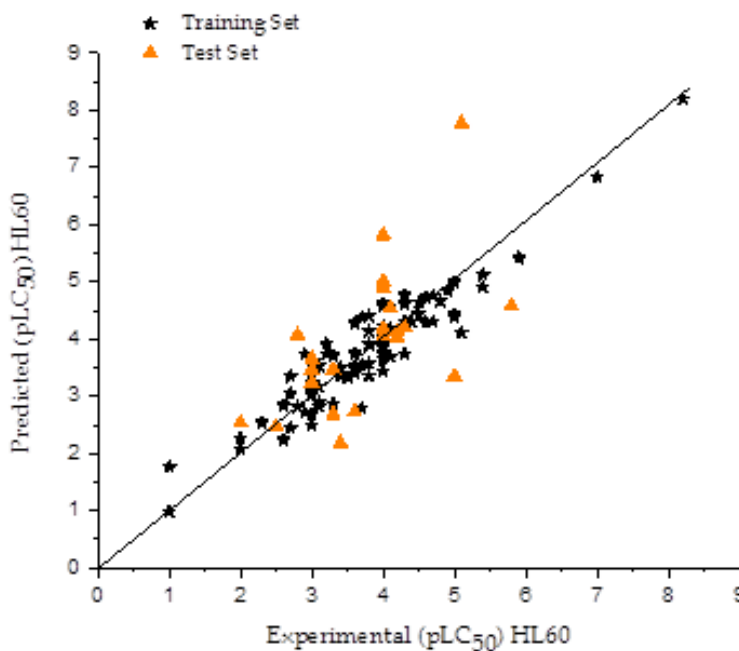
### QSAR on HL-60 cell line dataset

HL-60 cell line

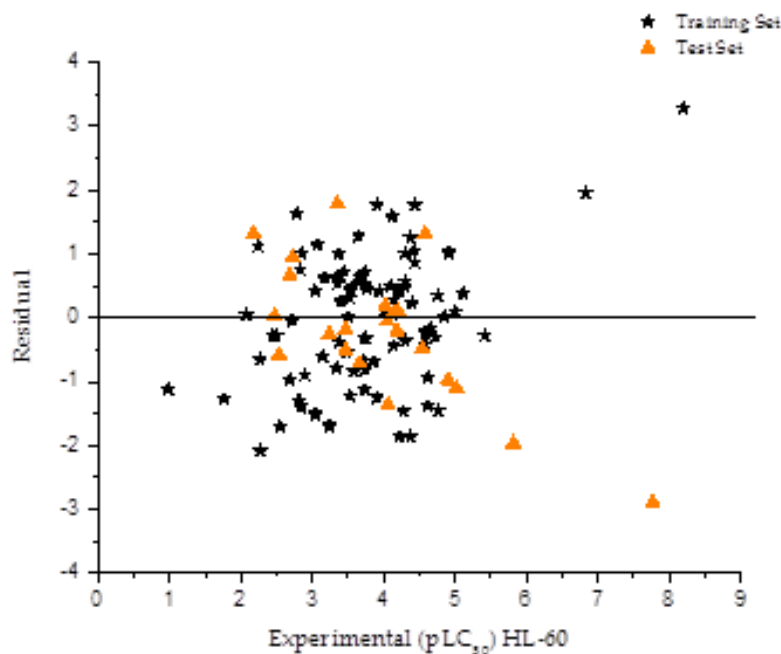
$$pLC_{50} = 4.097(\text{Secondary butyl}) + 0.861(n\text{Acid}) + 1.268(nS) + 2.298(AATSC5i) \\ + 3.315(SHBint6) + 4.039(SaaaC) - 1.584(minHBint7) \\ - 2.476(minHBint8) - 1.112(minaaN) + 2.386(minddssS) \\ + 2.938(WPSA - 1) - 1.684(RDF140s) - 1.070$$

$$N_{train} = 90, R_{train}^2 = 0.873, R_{adjusted}^2 = 0.852, F_{train} = 46.92, Q_{LOO}^2 = 0.826, Outliers > \\ 3.0 = 1, N_{test} = 22$$

$N$  is the number of compounds,  $R^2$  is the squared correlation coefficient,  $Q_{LOO}^2$  is the squared cross-validation coefficients for leave one out,  $F$  is the Fisher F statistic, and RMSE is the root mean square error. The built model was used to predict the test set data, and the prediction results are given in Table 1 (See Supplementary Files). The predicted values for  $pLC_{50}$  for the compounds in the training and test sets for HL-60 leukaemia cell line were plotted against the experimental  $pLC_{50}$  values in Figure 1, while the plot of the residual for the predicted values of  $pLC_{50}$  for both the training and test sets against the experimental  $pLC_{50}$  values of HL-60 cell line is shown in Figure 2. As can be seen the model did not show any proportional and systematic error, because the propagation of the residuals on both sides of zero is random.



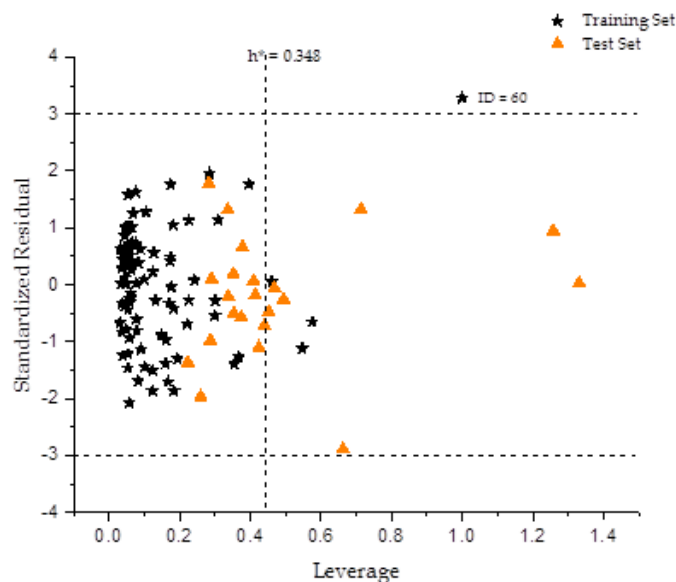
**Figure 1.** The predicted toxicity values (pLC<sub>50</sub>) against the experimental values for the training and test sets of the compounds on HL60 leukaemia cell line.



**Figure 2.** The Residuals against the Predicted pLC<sub>50</sub> values for the training and test sets of HL-60 leukaemia cell line.

### QSAR model validation

The real usefulness of QSAR models is not just their ability to reproduce known data, verified by their fitting power ( $R^2$ ), but mainly is their potential for predictive application. For this reason, the internal consistency of the training set was confirmed by using leave-one-out (LOO) cross-validation method to ensure the robustness of the model. The high calculated  $Q^2_{LOO}$  value for HL-60 given as 0.826 suggests a good internal validation, other statistical information is presented in Table 2 (See Supplementary Files).



**Figure 3.** The Williams plot, the plot of the standardised residuals versus the Toxicity (pLC50) leverage value for HL-60 dataset.

The leverage values can be calculated for every compound and plotted vs. standardised residuals, and it allows a graphical detection of both the outliers and the influential chemicals in a model. Figure 3, which shows the Williams plot of HL-60 dataset, the applicability domain is established inside a squared area within  $\pm 3$  bound for residuals and a leverage threshold  $h^*$  ( $h^* = 3p^0 / n$ ), where  $p^0$  is the number of model parameters and  $n$  is the number of compounds). It identifies that all the compounds of the training set and test set for HL-60 dataset are inside of this square area, with an exception of one compounds with ID number (60), which was found to be an outlier related to experimental errors of the data set, nine other compounds were not within the applicability domain or the model, though this was found to be as a result of strong differences in their structures with the rest of the dataset. From Figure 3, it could be seen that only one outlier compound with standard residuals  $> 3d$  of the data set. Furthermore, most of the chemicals had a leverage lower than the warning  $h^*$  value of 0.433. The descriptors in the models were subjected to variance inflation factor test (VIF), the test confirms that there is no intercollinerity of descriptors within the model (Table 3; see Supplementary Files), thus each descriptor relays unique information that cannot be correlated with that of other descriptors in the model.

In order to assess the robustness of the model, the Y-randomisation test was applied in this study. Y-randomisation test confirms whether the model is obtained by chance correlation, and is a true structure–activity relationship to validate the adequacy of the training set molecules.

If the activity prediction of the random model is comparable to that of the original equation, the set of observations is not sufficient to support the model. The new QSAR models (after several repetitions) would be expected to have low  $R^2$  and  $Q^2_{LOO}$  values for HL-60 cytotoxicity (Table 4). If the opposite happens, then an acceptable QSAR model cannot be obtained for the specific modelling method and data. The results of Table 3 indicate that an acceptable model is obtained by GA-MLR method, and the model developed is statistically significant and robust.

To examine the relative importance, and the contribution of each descriptor in the model, for each descriptor the value of the mean effect (MF) was calculated. This calculation was performed with the equation below:

$$MF_j = \frac{\beta_j \sum_{i=1}^n d_{ij}}{\sum_j^m \beta_j \sum_i^n d_{ij}} \quad (4)$$

$MF_j$  represents the mean effect for the considered descriptor  $j$ ,  $\beta_j$  is the coefficient of the descriptor  $j$ ,  $d_{ij}$  stands for the value of the target descriptors for each molecule, and  $m$  is the descriptor's number in the model. The MF value indicates the relative importance of a descriptor, compared with the other descriptors in the model. Its sign shows the direction of variation in the toxicity values as a result of the increase (or reduction) of the descriptor values. The mean effect values are presented in Table 3 for HL-60, The descriptors AATSC5i, SHBint6 and WPSA-1 molecular descriptors were found to contribute negatively to the toxicity of the anticancer compounds used in developing the model, and their mean effects were reported as 0.473, 0.403 and 0.254 respectively (See Supplementary Files for the table).

### Elucidation of Descriptors in HL-60 models

By interpreting the descriptors contained in the QSAR model it is possible to gain some insights into factors, which are related to the anti-leukaemic activity. For this reason, an acceptable interpretation of the selected descriptors is given below. The brief presentations of the descriptors in the model are shown in Table 3 (See Supplementary Files for the table). In order to quantitatively determine the relative importance of each descriptor in the model, the value of the mean effect (MF) was calculated for each descriptor. The MF value gives an account of the contribution of each descriptor to the model independent of the rest. Its sign indicates the difference in the direction of the activities as a result increase or decrease of the descriptor values. The mean effect values are presented in Table 3 (See Supplementary Files for the table).



n-Secondary butyl is a 2D descriptor representing the total number of secondary butyl group found in the data set, while n-acid is the total number of acidic groups, which is defined as a two dimensional acid group count descriptor. Their mean effects were recorded to as 0.013 accordingly, though the value is small and even insignificant when compared to some other descriptors in the model. n-S known as number of sulphur atoms was also found to contribute a little more than the previous descriptors and having an overall mean effect of 0.035 on the model. A positive mean effect for these descriptor illustrates that the toxicity increases with decreasing the values of the descriptors, which means that a large number of sulphur atoms, secondary butyl and acid groups will benefit the activity and reduce its toxicity. AATSC5i (Average centered Broto-Moreau autocorrelation - lag 5/weighted by first ionisation potential) belongs to the 2D autocorrelation descriptors. The 2D autocorrelation descriptors have been successfully employed by Fernandez *et al.* (2005). The physico-chemical property for AATSC5i is the ionisation potential of the molecule. Hence increasing the ionisation potential of a molecule increases its AATSC5i value. Mean effect of AATSCi has the positive sign, which indicates that an increase in the ionisation potential of a molecule leads to a decrease in its toxicity.

SHBint6 is defined as the Sum of E-State descriptors of strength for potential hydrogen bonds of path length 6, it a 2D atom type electrotopological state descriptor developed by Hall. The mean effect was reported in Table 3 as 0.403 and increasing its value leads to a decrease in the toxicity (pLC50; see Supplementary Files for the table).

MinHBint7 and minHBint8 are Minimum E-State descriptors of strength for potential Hydrogen Bonds of path length 7 and path length 8. These are atom-type electrotopological state molecular descriptors with mean effect of 0.133 and - 0.182, respectively. The mean effect of minHBint8 is higher than that of minHBint7 and its negative sign indicates that an increase in the hydrogen bonds potential for path length 8 will increase the cytotoxicity (pLC50). WPSA-1 (PPSA-1 \* total molecular surface area / 1000) is a 3D CPSA Descriptor which utilises the total molecular surface area of the molecule to predict the direction of the activity when compared to the descriptor. The value of the mean effect is given as 0.254, this descriptor displays a positive sign, which indicates that toxicity is inversely related to this descriptor. This shows that by decreasing the value of this descriptor will increase toxicities of the compounds.

The final descriptor of the GA-MLR model was the (RDF140s) Radial distribution function - 140 / weighted by relative I-state, which is one of the 3D RDF Descriptors. The descriptor is related to the I-state weight of the molecules, and its toxicity is inversely related to the descriptor since the mean effect is reported in Table 3 as 0.200.

## CONCLUSION

In the present study, a genetic algorithm was used to construct a quantitative relationship between the leukemia HL-60 cell lines cytotoxicity (pLC50) of 112 anti-leukemic compounds and their calculated descriptors. The GA-MLR method resulted in a training set with good statistical significance and acceptable external predictions. Also the AATSC5i, SHBint6 and WPSA-1 molecular descriptors were found to have the highest contribution in the model, their mean effects were reported as 0.473, 0.403 and 0.254, respectively.

Additionally, the proposed model identified and provided some insight into what structural features are related to the toxicity of compounds, as well as how these effects could be used to alter the toxicity of the compounds.

## REFERENCES

- [1] Kwon H-C, Kim S-H, Kim J-S, Han H, Sook Roh M, Han J-Y, et al. Establishment and characterization of an STI571-resistant human myelogenous leukemia cell line, SR-1. *Cancer Genetics and Cytogenetics*. 2004 Oct;154(1):52–6. DOI: 10.1016/j.cancergencyto.2004.01.009.
- [2] Monga M, Sausville E. Developmental Therapeutics Program at the NCI: molecular target and drug discovery process. *Leukemia*. 2002;16:520–6. URL: [https://www.researchgate.net/profile/Manish\\_Monga2/publication/11407712\\_Developmental\\_therapeutics\\_program\\_at\\_the\\_NCI\\_molecular\\_target\\_and\\_drug\\_discovery\\_process/links/546f54740cf2d67fc0310d3b.pdf](https://www.researchgate.net/profile/Manish_Monga2/publication/11407712_Developmental_therapeutics_program_at_the_NCI_molecular_target_and_drug_discovery_process/links/546f54740cf2d67fc0310d3b.pdf).
- [3] Bhat KS, Poojary B, Prasad DJ, Naik P, Holla BS. Synthesis and antitumor activity studies of some new fused 1,2,4-triazole derivatives carrying 2,4-dichloro-5-fluorophenyl moiety. *European Journal of Medicinal Chemistry*. 2009 Dec;44(12):5066–70. DOI: <http://dx.doi.org/10.1016/j.ejmech.2009.09.010>.
- [4] Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA: A Cancer Journal for Clinicians*. 2011 Mar;61(2):69–90. DOI: 10.3322/caac.20107.
- [5] Yap CW. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of Computational Chemistry*. 2011 May;32(7):1466–74. DOI: 10.1002/jcc.21707.
- [6] Anonymous. 2013 ESC Guidelines on cardiac pacing and cardiac resynchronization therapy: The Task Force on cardiac pacing and resynchronization therapy of the European Society of Cardiology (ESC). Developed in collaboration with the European Heart Rhythm Association (EHRA). *European Heart Journal*. 2013 Aug 1;34(29):2281–329.
- [7] Waller CL, Bradley MP. Development and Validation of a Novel Variable Selection Technique with Application to Multidimensional Quantitative Structure–Activity Relationship Studies. *Journal of Chemical Information and Computer Sciences*. 1999 Mar;39(2):345–55. DOI: 10.1021/ci980405r.
- [8] Gomes AR, Santos Sanches I, Aires de Sousa M, Castañeda E, de Lencastre H. Molecular Epidemiology of Methicillin-Resistant *Staphylococcus aureus* in Colombian Hospitals: Dominance of a Single Unique Multidrug-Resistant Clone. *Microbial Drug Resistance*. 2001 Mar;7(1):23–32. DOI: 10.1089/107662901750152729.
- [9] Ahmad S, Gromiha MM. Design and training of a neural network for predicting the solvent accessibility of proteins. *Journal of Computational Chemistry*. 2003 Aug;24(11):1313–20. DOI: 10.1002/jcc.10298.

[10] Caballero J, Fernandez M. Artificial Neural Networks from MATLAB; in Medicinal Chemistry. Bayesian-Regularized Genetic Neural Networks (BRGNN): Application to the Prediction of the Antagonistic Activity Against Human Platelet Thrombin Receptor (PAR-1). *Current Topics in Medicinal Chemistry*. 2008 Dec 1;8(18):1580–605. DOI: 10.2174/156802608786786570.

[11] Galvao R, Araujo M, Jose G, Pontes M, Silva E, Saldanha T. A method for calibration and validation subset partitioning. *Talanta*. 2005 Oct 15;67(4):736–40. DOI: 10.1016/j.talanta.2005.03.025.

[12] Kennard RW, Stone LA. Computer Aided Design of Experiments. *Technometrics*. 1969 Feb;11(1):137–48. DOI: 10.1080/00401706.1969.10490666.

[13] Golbraikh A, Tropsha A. Beware of  $q^2$ ! *Journal of Molecular Graphics and Modelling*. 2002 Jan;20(4):269–76. DOI: 10.1016/S1093-3263(01)00123-1.

[14] Roy K, Kar S, Das RN. Background of QSAR and Historical Developments. In: *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment* [Internet]. Elsevier; 2015 [cited 2016 Jun 1]. p. 1–46. Available from: <http://linkinghub.elsevier.com/retrieve/pii/B9780128015056000016>.

[15] Eriksson L, Jaworska J, Worth A, Cronin M, McDowell R, Gramatica P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environ Health Perspect*. 2003 Aug;111(10):1361–75. URL: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1241620/pdf/ehp0111-001361.pdf>.

[16] Ojha PK, Mitra I, Das RN, Roy K. Further exploring  $rm^2$  metrics for validation of QSPR models. *Chemometrics and Intelligent Laboratory Systems*. 2011 May;107(1):194–205. DOI: 10.1016/j.chemolab.2011.03.011.

[17] Golbraikh A, Shen M, Xiao Z, Xiao Y-D, Lee K-H, Tropsha A. Rational selection of training and test sets for the development of validated QSAR models. *Journal of Computer-Aided Molecular Design*. 2003 Feb;17(2):241–53.

[18] Gramatica P, Giani E, Papa E. Statistical external validation and consensus modeling: A QSPR case study for Koc prediction. *Journal of Molecular Graphics and Modelling*. 2007 Mar;25(6):755–66. DOI: 10.1016/j.jmglm.2006.06.005.

[19] Schultz TW, Netzeva TI, Roberts DW, Cronin MTD. Structure–Toxicity Relationships for the Effects to *Tetrahymena pyriformis* of Aliphatic, Carbonyl-Containing,  $\alpha,\beta$ -Unsaturated Chemicals. *Chemical Research in Toxicology*. 2005 Feb;18(2):330–41. DOI: 10.1021/tx049833j.

[20] OECD. Guidance document on the validation of (quantitative) structure-activity relationships [(Q)SAR] models. *OECD Series on Testing and Assessment No 69 ENV/JM/MONO (2007) 2*. 2007;154.

[21] Riahi S, Pourbasheer E, Ganjali MR, Norouzi P. Investigation of different linear and nonlinear chemometric methods for modeling of retention index of essential oil components: Concerns to support vector machine. *Journal of Hazardous Materials*. 2009 Jul 30;166(2-3):853-9. DOI: 10.1016/j.jhazmat.2008.11.097.

[22] Zhu X, Kruhlak NL. Construction and analysis of a human hepatotoxicity database suitable for QSAR modeling using post-market safety data. *Toxicology*. 2014 Jul;321:62-72. DOI: 10.1016/j.tox.2014.03.009.

[23] Todeschini R, Consonni V. *Molecular descriptors for chemoinformatics*. Weinheim: Wiley-VCH; 2009. (Methods and principles in medicinal chemistry). ISBN: 978-3-527-31852-0.

## HL-60 Hücre Dizisi Üzerinde Anti-Lösemi Bileşiklerinin Sitotoksik Davranışının İn Siliko Modellenmesi

David Ebuka Arthur, Adamu Uzairu, Paul Mamza, Stephen Eyije Abechi, Gideon Shallangwa

**Öz:** Bu çalışma çoklu lineer regresyon tekniğini bazı kuvvetli anti-lösemik bileşiklerin modellenmesinde kullanarak paDEL moleküler tarif edici yazılım hesap makinesi ile bazı lösemik hücre sıralarına (HL-60) karşı antikanser veri setlerinin kimyasal yapısı ve zehirlilikleri arasındaki en iyi ilişkiyi tanımlamayı amaçlamaktadır.  $Q^2$  ve  $R^2_{pred}$  (test seti) gibi istatistik parametreler modelin kuvvetini doğrulamak için hesaplanmıştır, Williams eğrisi de bunların uygulanma alanlarını değerlendirmek için kullanılmıştır. Modellerdeki moleküler tanımlayıcıların ortalama etkisi, bunların zehirliliklerinden sorumlu olan moleküllerin birincil özelliklerine ışık tutmak için hesaplanmıştır.

**Anahtar kelimeler:** HL-60 hücre sırası, lösemi, QSTR, uygulama alanı,  $\gamma$ -rasgeleleştirme.

**Gönderilme:** 1 Nisan 2016, **Düzeltilme:** 30 Nisan 2016, **Kabul:** 25 Mayıs 2016.