Regular Article

PHYSICAL CHEMISTRY RESEARCH
Published by the
Iranian Chemical Society
www.physchemres.org
info@physchemres.org

Phys. Chem. Res., Vol. 10, No. 2, 211-223, June 2022
DOI: 10.22036/pcr.2021.293051.1932

An Exploration of the Antiproliferative Potential of Chalcones and Dihydropyrazole Derivatives in Prostate Cancer *via* Androgen Receptor: Combined QSAR, Machine Learning, and Molecular Docking Techniques

O.E. Oyeneyin[a,*], B.S. Obadawo[b], D.S. Metibemu[c], T.O. Owolabi[d], A.A. Olanrewaju[e], S.M. Orimoloye[f], N. Ipinloju[a] and O. Olubosede[g]

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

[b]*Department of Chemistry and Biochemistry, 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*

(Received 17 June 2021, Accepted 24 September 2021)

In this study, the antiproliferative activities of some chalcones and dihydropyrazole derivatives in prostate cancer were investigated via the androgen receptor using QSAR, machine learning, and molecular docking techniques. A total of 30 dichloro substituted chalcones and dihydropyrazole derivatives were collected from the literature and optimized using density functional theory. Genetic function approximation was employed for model development. The developed model was thoroughly validated. Its generalization and predictive capacities were improved with the extreme learning machine (ELM) algorithm. Molecular docking and drug-likeness screening of the compounds were carefully performed. A reduction in the negative coefficient of the descriptor and an increase in the positive coefficient of the descriptor improved bioactivity. An R^2 pred value of 0.737 showed a strong correlation between the experimental and predicted activities. A correlation coefficient of 0.8305 for R^2demonstrated the predictability of the model. The ELM-Sine model showed an improvement of 66.7% and 8.3% in QSAR and ELM-Sig models, respectively. Molecular docking showed the chalcones and dihydropyrazole derivatives to be promising anti-prostate cancer agents, with pi-pi stacking and hydrogen bond interactions favoring the inhibition of the androgen receptor. The lead drugs are drug-like and novel anti-prostate cancer agents.

Keywords: Anticancer properties, Computer-aided drug design, Data science, Extreme learning machine

INTRODUCTION

Prostate cancer (PCa) is the second most common cancer after lung cancer [1]. It is the sixth leading cause of cancer mortality in males globally [2]. According to GLOBOCAN (2018), over 20 million men will be living with PCa by 2024 [3]. The majority of prostate cancer patients are men 65 years and older [4]. Pathological analysis of PCa showed that it is a heterogeneous disease, ranging from inactive to extremely aggressive ones [5]. It is established that PCa is driven by the androgen receptor (AR), which is a member of the nuclear receptor family [6]. Moreover, prostate-specific antigen (PSA), an AR target-

*Corresponding author. E-mail: emmanueltoba90@gmail.com

gene is used as a PCa biomarker. Notwithstanding the controversies surrounding its use, PSA screening is still the only available serum biomarker for both diagnosis and treatment of PCa [7]. Androgen-synthesizing enzyme inhibitors and direct AR antagonists have been approved as two treatment options targeting AR signaling pathways. Moreover, castration-resistant prostate cancer (CRPC) is known to be dependent on androgen-signaling, and almost all pathways promoting CRPC growth converge into the AR. Several approaches have been developed to target AR signaling pathways and inhibit the growth of CRPC. Despite the advantage of estrogen receptor (ER)-targeted therapy for PCa, its serious side-effects, such as venous thromboembolism and estrogenic proliferative actions on PCa cells, have led to the discontinuation of estrogen-based therapy for advanced PCa [8]. Medicinal and pharmaceutical applications of organic compounds have been explored using computational techniques, such as characterization [9-11], molecular docking [12-14], drug-like properties [15-17], quantitative structure-activity relationship (QSAR), and machine learning [18,19]. The present study was designed to explore the antiproliferative potential of chalcones and dihydropyrazole derivatives in PCa via AR. The combination of synthetic and natural compounds has been shown to have anticancer potentials [20,21]. Furthermore, natural compounds have been reported to have inhibitory effects on AR [21]. Chalcones have also been reported to have antiproliferative effects [22]. Dihydropyrazole derivatives are reported to inhibit telomerase activity in neoplasm [23].

This study aimed to investigate the antiproliferative potential of some chalcones and dihydropyrazole derivatives in PCa by AR using combined QSAR, machine learning, and molecular docking techniques. To this end, the drug-like properties were also evaluated.

COMPUTATIONAL METHODOLOGY

QSAR of Dichloro Substituted Chalcones and Dihydropyrazole Derivatives

Data collection. The antiproliferative activities of thirty compounds of dichloro substituted chalcones and dihydropyrazole derivatives were collected from the literature [22]. The structures are presented in Table S1, and their activities were reported in IC_0 (μM) and converted into their corresponding pIC_{50} values (*i.e.*, $logIC_{50} = pIC_{50}$) to fit them into a set of values and a normal distribution curve.

Optimization of Molecular Structure

The 2D structures of the molecules were obtained from ChemDraw [24], transferred to Spartan 14 software [25], and optimized using density functional theory (DFT) with Becke three Lee-Yang-Parr (B3LYP) correlation [26] and 6-31G* basis set. This level of theory was used because it was shown to be consistent with experimental findings [10,11]. The analysis was carried out on the most stable conformer of each molecule derived from a molecular mechanics force field).

Molecular Descriptor Calculation

After removing salts and detecting tautomers from the optimized structures, the 1D, 2D, and 3D descriptors of the compounds were determined using PaDEL-Descriptor software version 2.20 [27]. A total of 1875 descriptors were generated and saved as a Microsoft Excel comma-separated value (CSV) file.

Normalization and Data Pre-Treatment

The calculated descriptors were normalized in the range of 0 to 1 using Eq. (1). This gave each element the same probability to influence and build a successful model at the outset.

$$X = \frac{X_1 - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

In the above equation, X_1 is the descriptor value for each molecule, and X_{min} and X_{max} are the minimum and maximum values for each descriptor, respectively.

Data Treatment and Data Division

The normalized data were then subjected to pre-treatment using DTC-QSAR software v 1.0.5 [28]. The data were divided into a training set (70% of the dataset, 21 compounds) for the construction of the model and into a test set (30% of the dataset, 9 compounds) for the external evaluation of the constructed model using the Kennard-Stone algorithm [29].

Model Development

Genetic function approximation (GFA) was employed with biological activities (pIC_{50}) as the dependent variable and physiochemical properties (descriptors) as independent variables [30]. The developed models were based on four descriptors (Table 1).

Internal validation of the model. The training set was validated internally using Material Studio software. The validation parameters are presented below.

Friedman's lack of fit (LOF). Friedman's LOF (Eq. (2)) was used to measure the fitness scores of the models. LOF is defined as follows [31];

$$LOF = \frac{SEE}{\left(1 - \frac{C + dp}{M}\right)^2} \quad (2)$$

where SEE, p, d, c, and M are the standard error of estimation, the total number of descriptors in the model, user-defined smoothing parameter, the number of terms in the model, and the number of compounds in the training set, respectively [32]. A model is considered valid if it has a low SEE value. The value of SEE can be calculated using Eq. (3):

$$SEE \sqrt{\frac{(Y_{\exp} - Y_{pred})^2}{N - P - 1}} \quad (3)$$

where Y_{exp} is experimental activity, Y_{pred} is the predicted activity in the training set, p is the number of descriptors in the model, and n is the number of compounds in the training set.

The correlation coefficient (R^2). This parameter is the most frequently used parameter to evaluate the internal validity of a QSAR model. The closer is the value of R^2 to 1.0, the better is the model generated. R^2 can be expressed as in Eq. (4):

$$R^2 = 1 - \sum \frac{(Y_{\exp} - Y_{pred})^2}{\sum (Y_{\exp} - Y_{training})^2} \quad (4)$$

where $Y_{training}$ is the mean of the experimental activity in the training set.

Adjusted R^2. The R^2 value has to be adjusted because it is not a reliable parameter to examine the stability of the model. The adjusted R^2 is defined as in Eq. (5):

$$R^2_{adj} = \frac{R^2 - P(n-1)}{n - p + 1} \quad (5)$$

where R^2, p and n are as defined above.

The cross-validation coefficient (Q^2_{CV}). The strength of the QSAR model to predict the activity of a new compound was determined using a cross-validation test. The cross-validation coefficient (Q^2_{CV}) is defined as in Eq. (6):

$$((Q^2_{CV}) = 1 - \left\{ \frac{\sum (Y_{pred} - Y_{\exp})^2}{\sum (Y_{\exp} - Y_{training})^2} \right\} \quad (6)$$

where $Y_{training}$, Y_{exp} and Y_{pred} are the mean of the experimental activity, the experimental activity, and the predicted activity in the training set, respectively.

External validation of the model. The developed model was validated externally using the $R^2_{predicted}$ value.

Table 1. List of Descriptors and Their Constructors, Description, and Dimension Used in Developing the QSAR Model

S/No	Name	Description	Dimension
1	ATS1s	Broto-Moreau autocorrelation-lag 1/weighted by I-state	2D
2	VR2_Dzi	Normalized Randic-like eigenvector-based index from Barysz matrix/weighted by first ionization potential	2D
3	SpMax5_Bhm	Largest absolute eigenvalue of Burden modified matrix-n 5/weighted by relative mass	2D
4	Mi	Mean first ionization potentials (scaled on carbon atom)	2D

The $R^2_{predicted}$ value is the most commonly used parameter to validate a developed model. The idea that a satisfactory $R^2_{predicted}$ value indicates that the remaining parameters are also satisfactory is not always true. There are other statistics, such as variance inflation factor (VIF) and mean effect (ME), that can be used to validate a developed model. The closer the value of R^2_{test} is to 1.0, the higher the stability of the model generated. The stability of a model shows its reliability in predicting the activity of a new compound. The R^2_{test} can be calculated by Eq. (7):

$$R^2 = 1 - \sum \frac{(Y_{pred_{test}} - Y_{\exp_{test}})^2}{\sum (Y_{pred_{test}} - \bar{Y}_{training})^2} \quad (7)$$

Statistical Analysis of the Descriptor

Variance inflation factor (VIF). VIF (Eq. (8)) is used to measure the multicollinearity among descriptors and the degree at which one descriptor correlates with the other descriptors.

$$VIF = \frac{1}{1 - R^2} \quad (8)$$

where R^2 is multiple correlation coefficients between the variables within the model. If VIF is equal to 1, it signifies that there is no intercorrelation between any pair of variables. While VIF values ranging from 1 to 5 are considered to be acceptable, VIF values greater than 10 indicate multicollinearity (*i.e.*, the model is unstable) [18,33].

Mean effect (ME). The average effect, or the mean effect (Eq. (9)), shows the correlation between the effect or influence of the given molecular descriptors and the activities of the compounds in the model. The descriptor signs show the direction of their deviation toward the activity of the compounds. That is to say, an increase or decrease in the value of the descriptors will improve the activity of the compounds. The mean effect can be calculated by the following equation:

$$Mean\ effect = \frac{B_j \sum_i^n D_j}{\sum_j^m (B_j \sum_i^n D_j)} \quad (9)$$

where B_j and D_j are the j-descriptor coefficients in the model and the values of each descriptor in the training set, respectively, and m and n stand for the number of molecular descriptors and the number of molecules in the training set, respectively. Therefore, the mean effect of each descriptor used in developing the model was calculated to assess the significance of the model [33].

Evaluation of the applicability domain of the model. This is essential in establishing whether the model can make predictions within the chemical space for which it is developed [32]. The leverage approach developed by Veerasamy *et al*. was employed [34], and it is presented in Eq. (10):

$$h_i = X_i (X^T X)^{-1} X_i^T \quad (10)$$

where h_i is the leverage of a compound, X_i is the matrix of the training set compound of i, X is the m x k descriptor matrix of the training set compound, and X^T is the transpose matrix of X. As a prediction tool, the warning leverage (h*) is the limit of normal values for X outliers and is defined as follows (Eq. (11)):

$$h^* = \frac{3(k+1)}{n} \quad (11)$$

where n and k are the number of descriptors and the training set, respectively.

Model quality assurance. Validation parameters are frequently used to measure the strength, dependability, and predictive ability of a developed QSAR model. The general minimum values required for both internal and external validation parameters used to assess a QSAR model are presented in Table S2) [34].

Extreme learning machine (ELM). Extreme learning machine belongs to a class of feed-forward neural networks characterized with a single hidden layer [35-37]. The algorithm has an excellent ability to approximate patterns, functions, and all intricacies correlating the descriptors with the target model. In addition, this algorithm (*i.e.*, ELM) has a high generalization and predictive ability compared with the traditional single-layer networks [38]. Aside from the hidden layer, the network has the input and output layers with fully connected weights. ELM computes the output weights of the network based on the generalized inverse

matrix theory while the biases and input weights are initialized randomly. This random weight selection introduces the advantage of short training time as well as fast learning speed to the algorithm. A pool of input descriptors (L_P) and the experimental activities (T_C) of the investigated compounds should be considered in such a way that (L_{Pj}, T_{Cj}), where $L_{Pj} = [c_{j1}, c_{j2}, ..c_{jm}]^T \in \Re^m$ and $T_{Cj} = [T_{j1}, T_{j2}, ...T_{jN}] \in \Re^N$ [39]. The connection weights linking the input with the hidden layer are represented in a matrix form in Eq. (12).

$$d = \begin{bmatrix} d_{11} & d_2 & ... & d_{13} \\ d_{21} & d_{22} & ... & d_{23} \\ : & : & ... & : \\ d_{t1} & d_{t2} & ... & d_{tm} \end{bmatrix}_{txm} \tag{12}$$

With $q = [q_1, q_2, ...q_t]^T$ as values for the bias of hidden layer neurons and $g(.)$ as the activation function, the output matrix of the hidden layer H is presented in Eq. (13).

$$\mathbf{H}(d_1, d_2, ..., d_t ..., q_1, q_2, ..q_t, T_1, T_2, ..., T_r) = \begin{bmatrix} g(d_1 T_1 + q_1) & g(d_2 T_1 + q_2) & ... & g(d_t T_1 + q_t) \\ g(d_1 T_2 + q_1) & g(d_2 T_2 + q_2) & ... & g(d_t T_2 + q_t) \\ : & : & : & : \\ g(d_1 T_r + q_1) & g(d_2 T_i + q_2) & ... & g(d_t T_r + q_t) \end{bmatrix}_{xt} \tag{13}$$

The output of the ELM network can be obtained using Eq. (14)

$$T_C = [T_1, T_2, ..., T_r]_{Nxr} \tag{14}$$

where

$$T_i = [T_{1i}, T_{2i}, ..., T_{Ni}] = [\sum_{j=1}^{t} \beta_{j1} g(d_j T_i + q_j), \sum_{j=1}^{t} \beta_{j2} g(d_j T_i + q_j), ..., \sum_{j=1}^{t} \beta_{jN} g(d_j T_i + q_j)]^T \ (i = 1,2,3..,r) \tag{15}$$

and β is the weights linking the hidden layer with the output layer.

The least-square solution of $\underset{\beta}{Minimize} \|\mathbf{H\beta - T_C}\|$ leads to the computation of β, as shown in Eq. (15)

$$\hat{\beta} = H^+ T_C \tag{15}$$

where H^+ represents the Moore-Penrose generalized inverse of H.

Molecular Docking

Protein preparation. The crystal structure of the AR (PDB ID: 2AXA) with a resolution of 1.8 Å was downloaded from the Protein Data Bank [40]. Protein preparation was performed using the protein preparation wizard of Schrodinger Suite 2017-1 [41]. The interfering ligand was removed, bond orders and charges were assigned, and water molecules and all heteroatoms were removed. Furthermore, tautomeric states were generated using EPIK at pH 7.0 ± 2 [42], and the protein was minimized using the OPLS3 force field [43].

Ligands preparation. The optimized structures of the lead compounds (*i.e.*, 7, 15, 16, 18, 30, and 31) and a standard drug (*i.e.*, R-Bicalutamide) were imported and prepared using the ligprep module in Schrodinger Suite 2017-1 *via* OPLS3 force field. The generation of possible ionization and tautomeric states were generated using EPIK at pH 7.0 ± 2.

Molecular docking and visualization. The prepared lead compounds and R-Bicalutamide were docked at the active site of 2AXA to analyze their interactions and binding affinities using the glide docking module of Schrodinger Suite 2017-1[44]. The ligand interactions were depicted using visualization diagrams.

Drug-like properties. The adsorption, diffusion, metabolism, and excretion, parameters were analyzed using QikProp module of Schrodinger Suite 2017-1 [45]. These properties play an important role in determining the safety and efficacy of drugs [46]. Other properties such as molecular weight (MW), the number of hydrogen-bond acceptor (HBA), and hydrogen bond donor (HBD) were also calculated.

RESULTS AND DISCUSSION

QSAR Model Development of Chalcones and Dihydropyrazole Derivatives

GFA was employed to generate QSAR models. The best model was selected based on statistical significance. The selected model was found to meet the recommended standards for a stable and reliable model, as outlined in Table S2.

Model 1

pIC_{50} = (-0.037702307 × ATS1s) + (0.061331393 × VR2_Dzi) + (1.933506556 × SpMax5_Bhm) + (6.184315031 × Mi) – 45.008691509

Descriptors Interpretation

The list, description, and dimension of descriptors used in developing the model are reported in Table 1. The 2D descriptors play a major role in predicting the activity of molecules [47]. The negative coefficient of ATS1s descriptors in the model shows that a decrease in the coefficient of the descriptor will improve the activities (pIC_{50}) of chalcones and dihydropyrazole derivatives, as antiproliferative inhibitors. In other words, the positive coefficients of VR2_Dzi, SpMax5_Bhm, and Mi descriptors signify that an increase in the coefficients of descriptors will invariably increase the activities (pIC_{50}) of the compounds. Therefore, to design potent compounds with high bioactivity, the negative coefficient of the descriptor should be reduced and the positive coefficients of the descriptors should be increased.

A comparison of experimental activity (pIC_{50}), predicted activity, and the residual of the developed model is presented in Table 2. This comparison was used to examine the external validity of the model, and the results are shown in Tables S3 and S4. The low values of residual recorded confirmed that there was a high correlation between the experimental activities and predicted activities. The value of R^2_{pred} (0.737) signified that the model exceeded the minimum value required for a validation parameter of a developed model (Table S2).

Pearson's correlation, ME, and VIF of the descriptors are presented in Table S5. Low correlation values (≤ 0.5) in *VR2_Dzi*/ Mi/ SpMax5_Bhm descriptors indicate that these descriptors do not correlate with one another while those descriptors with considerably high correlation values are interrelated with each other, such that an increase or decrease in one of the descriptors will affect the other. This parameter is useful in designing compounds using ligand-based drug design (LBDD) since the descriptors of compounds have a great impact in deciding their bioactivity. The ME of the descriptors shown in Table S5 signifies the effect of molecular descriptors on the activity (pIC_{50}) of the compounds. Thus, the order of significance is as follows:

Mi > SpMax5_Bhm > VR2_Dzi > ATS1s

Table 2. A Comparison of the Experimental Activity (pIC_{50}), Predicted Activity (pIC_{50}), and the Residual of the Developed Model

S/No	Experimental activity (pIC_{50})	Predicted activity (pIC_{50})	Residual
2	3.36	3.400	-0.040
3*	3.75	3.669	0.081
4	3.41	3.641	-0.231
5	3.83	3.631	0.199
6*	3.85	3.753	0.097
7*	4.51	4.125	0.385
8	3.6	3.629	-0.029
9	3.53	3.609	-0.079
10*	3.86	3.652	0.208
11	3.45	3.475	-0.025
12	3.59	3.496	0.094
13	3.68	3.701	-0.021
14*	3.69	3.497	0.193
15	4.03	3.980	0.050
16*	4.77	4.489	0.281
17	3.45	3.486	-0.036
18	3.92	3.924	-0.004
19	3.49	3.603	-0.113
20	3.83	3.584	0.246
21*	3.87	3.584	0.286
22	3.58	3.641	-0.061
23	3.69	3.647	0.043
24*	3.59	3.872	-0.282
25	3.88	3.794	0.086
26	3.57	3.564	0.006
27	3.68	3.861	-0.181
28*	3.76	3.746	0.014
29	3.79	3.782	0.008
30	4.08	3.966	0.114
31	4.49	4.526	-0.026

Note. * = test set.

Table 3. Validation Parameters for the Developed Model

S/N	Validation parameter	Value
1	Friedman LOF	0.0701
2	R-squared	0.8305
3	Adjusted R-squared	0.7882
4	Cross validated R-squared	0.7403
5	Significant regression	Yes
6	Significance-of-regression (SOR) F-value	19.6023
7	Critical SOR F-value (95%)	3.0558
8	Replicate points	0
9	Computed experimental error	0.0000
10	Lack-of-fit points	16
11	Min expt. Error for non-significant LOF (95%)	0.0971
12	R^2_{pred}	0.737155

The validation parameters, including the R^2_{pred} value for the test set, for the best model were generated using Material Studio, and the results are presented in Table 3. The plot of experimental activity against the predicted activity for both the training set and the test set is shown in Fig. 1. A high value of correlation coefficient R^2 between the training set and the test set (0.8305) confirmed that the model could successfully predict the activity of a new compound due to the high correlation between the predicted activity and the experimental activity. The randomness of activities on both negative and positive sides of the y-axis shown on the scatter plot between the standardized residual and the experimental activity and demonstrated in Fig. 2 confirms that the developed model is free from systematic error.

To identify outliers and lead compounds in the developed model, the standardized residual for the entire dataset was plotted against the leverages (Figs. 2 and 3). The Williams plot (Fig. 4) shows that two compounds from the test set (15 and 27) and one compound (30) from the training set had leverage values greater than the warning limit (h = 0.714). There was no outlier in the developed model as all compounds fell within ±3.0. Hence, the

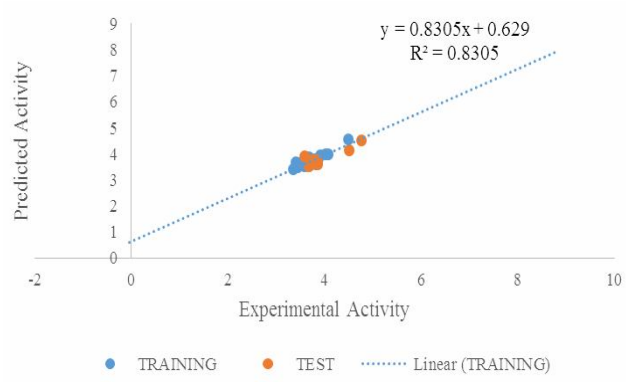

Fig. 1. The plot of predicted activity against experimental activity for both the training set and the test set.

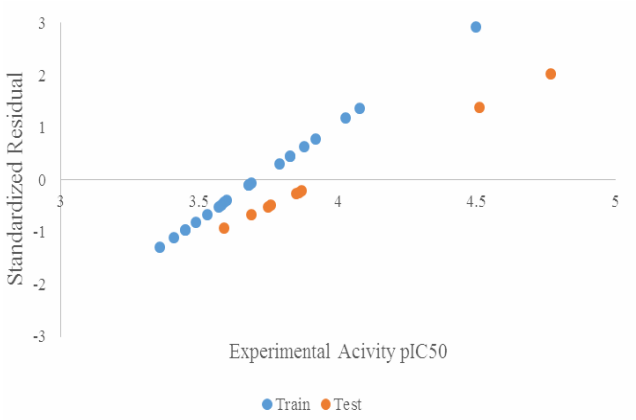

Fig. 2. The plot of the standardized residual against experimental activity (pIC_{50}).

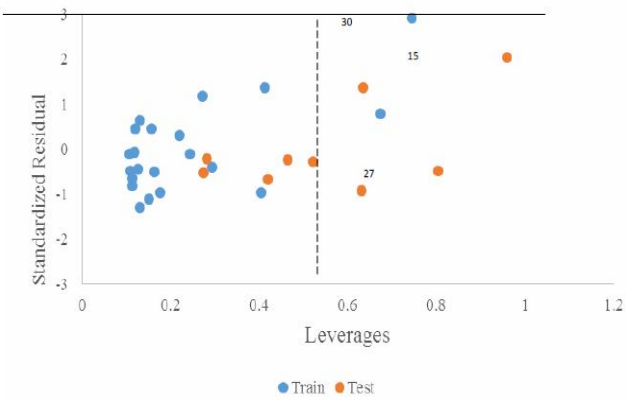

Fig. 3. The Williams plot of standardized residual against leverages.

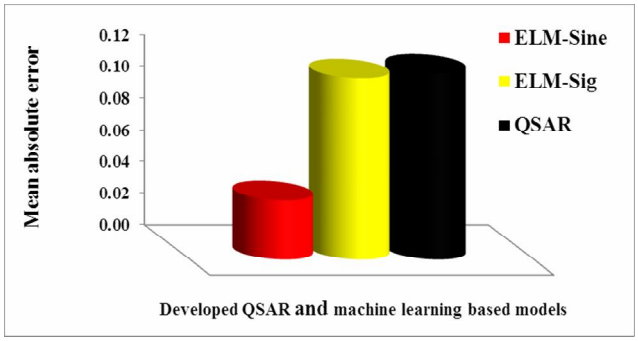

Fig. 4. Mean absolute error of the QSAR, the ELM-Sine, and the ELM-Sig models.

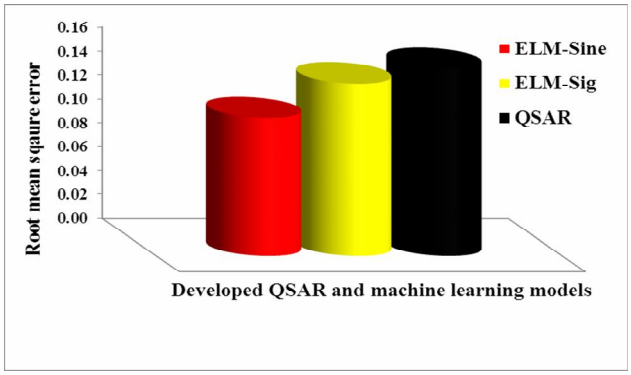

Fig. 5. Root mean square error of the QSAR, the ELM-Sine, and the ELM-Sig models.

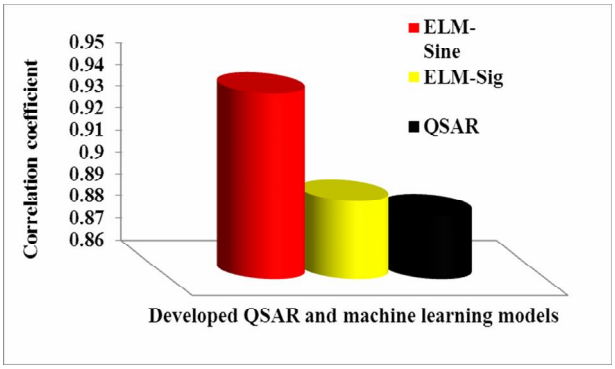

Fig. 6. Correlation coefficient of the QSAR, the ELM-Sine, and the ELM-Sig models.

developed model can be considered reliable and be used as a predictive tool since most leverage values fell within the chemical space.

The EML Model of Chalcones and Dihydropyrazole Derivatives

The antiproliferative effects of chalcones and dihydropyrazole derivatives against PCa were modeled using ELM, and the results of the modeling are presented in Figs. 4, 5 and 6. Figure 4 compares the performance of the developed ELM model with the QSAR model using mean absolute error as a performance evaluator. The developed ELM-Sine model performed better than the developed QSAR and ELM-Sig model, with performance improvement of 66.7% and 8.3%, respectively. A comparison of the QSAR and ELM-Sig models based on root mean square error is presented in Fig. 5, with performance improvement of 25% and 12.5%, respectively. Figure 6 shows a comparison of the developed models using correlation coefficient as a performance criterion. Similarly, Fig. 6 shows the superiority of the ELM-Sine model over the other developed models. It should be noted that each of the developed models was developed using the same number of descriptors to ensure an unbiased comparison. The results revealed a performance improvement of 6.24% for the ELM-Sine model over the QSAR model and 0.75% over the ELM-Sig model. The actual values of the estimates for each of the developed models and the measured activities for each of the investigated compounds are presented in Table S6. In addition, Table S7 presents the values of each of the performance parameters for all the developed models.

Molecular Docking of the Lead Compounds with AR and their Drug-like Properties

The lead compounds and R-Bicalutamide (standard drug) were docked at the receptor active site. The docking results are presented in Table 4 and Fig. 7 and S1-S6. Pi-Pi stacking (attractive and noncovalent interactions between the aromatic rings) with PHE 764 at the receptor active site was conserved in the lead compounds and R-Bicalutamide. Compound 30, with a binding affinity of -9.287 kcal mol^{-1}, shared similar interactions with the standard drug (PHE 764 and ASN 705). The high binding affinity of compound 30, compared with other lead compounds (7, 15, 16, 18 and 31), may be due to the formation of hydrogen bonding (*i.e.*, hydrogen atom located between a pair of other atoms having

Table 4. Binding Affinities and Types of Interactions between the Lead Compounds and R-Bicalutamide and the Receptor, 2AXA

Molecules	Binding affinity (kcal mol^{-1})	Amino acid	Interaction (s)
7	-8.595	PHE 764	Pi-Pi stacking
15	-7.562	PHE 764	Pi-Pi stacking
16	-7.871	PHE 764	Pi-Pi stacking
18	-8.033	PHE 764	Pi-Pi stacking
30	-9.287	PHE 764, ASN 705	Pi-Pi stacking, hydrogen bonding
31	-8.439	PHE 764, TRP 741	Pi-Pi stacking, Pi-Pi stacking
R-Bicalutamide	-8.653	PHE 764, ASN 705, ARG 752	Pi-Pi stacking, hydrogen bonding, hydrogen bonding

Table 5. Drug-Like Properties of Some of the Lead Compounds

Molecules	M_W	SASA	Donor HB	Accpt HB	QPlogPo/w	Lipinski violations
7	313.13	584.61	0	2	5.31	0
15	266.13	565.87	1	2	4.22	0
16	283.17	579.01	0	2	4.88	0
18	385.27	694.22	0	1	7.29	1
30	356.25	621.21	1	1	7.48	1
31	373.29	626.91	0	1	6.98	1

Note. SASA = solvent accessible surface area (300-1000), M_w = molecular weight (< 500), donorHB = hydrogen bond donor (<5), accptHB = hydrogen bond acceptor (<10), and QPlog Po/w = octanol/ water partition coefficient (2.0-6.5).

a high affinity for electrons) and/or the pi-pi stacking. Compound 30, compared with the other lead compounds, shared two pi-pi stacking interactions at the receptor active site. Compounds 30 and 31, compared with the other lead compounds (7, 15, 16 and 18), had additional hydrogen bonds and pi-pi stacking interactions at the receptor active site, thus having a stronger inhibitory effect on AR. This finding is consistent with those of Pantsar *et al.* [48], who observed that molecular interactions and active site binding were better indicators of inhibition than docking scores. According to Lipinski's rule of five and the results in Table 5, all these six compounds (*i.e.*, 7, 15, 16, 18, 30 and 31) can be used as drug candidates as none of them had more than one violation [46,49,50].

CONCLUSIONS

In this study, the antiproliferative effects of chalcones

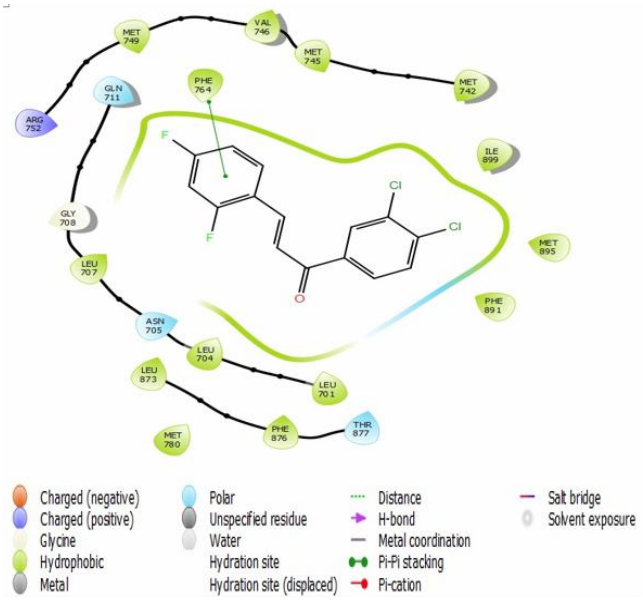

Fig. 7. 7:7 with 2AXA.

and dihydropyrazole derivatives on PCa *via* AR were investigated. The developed model for the antiproliferative activities of chalcones and dihydropyrazole derivatives in PCa was thoroughly validated, and it was found to be robust. A reduction in the negative coefficient of descriptors and an increase in the positive coefficient of descriptors were found to enhance the bioactivities of chalcones and dihydropyrazole derivatives in PCa. The model with an R^2_{pred} of 0.737 and a coefficient R^2 of 0.8305 showed a strong correlation between the experimental activities and the predicted activities and had a good predictive ability. The selected model was further enhanced by the ELM-Sine. Pi-pi stacking and hydrogen bond interactions showed the inhibitory effects of the chalcones and dihydropyrazole derivatives on AR. Accordingly, it can be stated that chalcones and dihydropyrazole derivatives have the potential to serve as anti-prostate cancer agents. A combined QSAR, machine learning, molecular docking, and drug-like screening model adopted in this study revealed six lead compounds, including 7, 15, 16, 18, 30 and 31, which were drug-like and showed anti-prostate cancer activity.

REFERENCES

[1] Rawla, P., Epidemiology of prostate cancer. *World J. Oncol.* **2019**, *10*, 63-89. https://doi.10.14740/wjon1191.

[2] Yousefi, M. S.; Rahimi, S.; Yousefi, S. M.; Hosseini, S.; Mahabadi, A. A.; Abarqui, H. F.; Borujeni, N. N.; Salehiniya, H., The incidence, risk factors, and knowledge about the prostate cancer through worldwide and iran. *World Cancer Res. J.* **2017**, *4*, 1-8. https://doi.10.32113/wcrj_201712_972.

[3] Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R. L.; Torre, L. A.; Jemal, A., Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Canc. J. Clin.* **2018**, *68*, 394-424. https://doi.10.3322/caac.21492.

[4] Tarver, T., Cancer facts and figures, *American Cancer Society (ACS) J. Consum. Health. Intern.* **2012**, *16*, 366-367. https://doi.10.1080/15398285.2012.701177.

[5] Lauer, R. C.; Friend, S. C.; Rietz, C.; Pasqualini, R., Drug design strategies for the treatment of prostate cancer. *Expert Opin Drug Discov.* **2015**, *10*, 81-90. https://doi.10.1517/17460441.2015.978855.

[6] Nevedomskaya, E.; Baumgart, S. J.; Haendler, B., Recent advances in prostate cancer treatment and drug discovery. *Int. J. Mol. Sci.* **2018**, *19*, 1-25. https://doi.10.3390/ijms19051359.

[7] Narayanan, R., Therapeutic targeting of the androgen receptor (AR) and AR variants in prostate cancer. *Asian J. Urol.* **2020**, *7*, 271-283. https://doi.10.1016/j.ajur.2020.03.002.

[8] Veccia, A.; Maines, F.; Kinspergher, S.; Galligioni, E.; Caffo, O., Cardiovascular toxicities of systemic treatments of prostate cancer. *Nat. Rev. Urol.* **2017**, *14*, 230-243. https://doi. 10.1038/ nrurol.2016.273.

[9] Adejoro, I. A.; Oyeneyin, O. E.; Adeboye, O. O.; Obaleye, J. A., Characterization of a novel polymeric Zinc(II) complex containing the anti-malarial Quinine as ligand: A theoretical approach (semi-empirical and DFT methods). *Amer. J. Sci. Ind. Res.* **2013**, *4*, 111-122. https://doi.org/10.5251/ajsir.2013.4.1.111.122.

[10] Odewole, O. A.; Ibeji, C. U.; Oluwasola, H. O.; Oyeneyin, O. E.; Akpomie, K. G.; Ugwu, C. M.; Bakare, T. E., Synthesis and anti-corrosive potential of Schiff bases derived 4-nitrocinnamaldehyde for mild steel in HCl medium: Experimental and DFT studies. *J. Mol. Str.* **2021**, *1223*, 129214. https://doi.org/10.1016/j.molstruc.2020.129214.

[11] Kumer, A.; Sarker, N.; Paul, S., The thermo physical, HOMO, LUMO, vibrational spectroscopy and QSAR study of morpholinium formate and acetate Ionic Liquid Salts using computational method. *Turk. Comput. Theor. Chem.* **2019**, *3*, 59-68. https://doi.org/10.33435/tcandtc.481878.

[12] Nath, A.; Kumer, A.; Khan, W., Synthesis, computational and molecular docking study of some 2,3-dihydrobenzofuran and its derivatives. *J. Mol. Str.* **2021**, *1224*, 129225. https://doi.org/ 10.1016/j.molstruc.2020.129225.

[13] Hoque, M. M.; Kumer, A.; Hussen, S., Theoretical evaluation of 5,6-diaroylisoindoline-1,3-dione as potential carcinogenic kinase PAK1 inhibitor: DFT calculation, molecular docking study and ADMET prediction. *Int. J. Adv. Bio. Biomed. Res.* **2021**, *9*, 77-

104. https://doi.org/10.22034/ijabbr.2021.45696.

[14] Olanrewaju, A. A.; Ibeji, C. U.; Oyeneyin, O. E., Biological evaluation and molecular docking of some newly synthesized 3d-series metal(II) mixed-ligand complexes of fluoro-naphthyl diketone and dithiocarbamate. *SN Appl. Sci.* **2020**, *2*, 678. https://doi.org/10.1007/s42452-020-2482-0.

[15] Nath, A.; Kumer. A.; Zaben, F.; Khan, M., Investigating the binding affinity, molecular dynamics, and ADMET properties of 2,3-dihydrobenzofuran derivatives as an inhibitor of fungi, bacteria, and virus protein. Beni-Suef Univ. *J. Bas. Appl. Sci.* **2021**, *10*, 36. https://doi.org/10.1186/s43088-021-00117-8.

[16] Kumer, A.; Khan, A., The effect of alkyl chain and electronegative atoms in anion on biological activity of anilinium carboxylate bioactive ionic liquids and computational approaches by DFT functional and molecular docking. *Heliyon*. **2021**, *7*, e07509. https://doi.org/10.1016/j.heliyon.2021.e07509.

[17] Sarkar, M.; Nath, A.; Kumer, A.; Mallik, C.; Akter, F.; Moniruzzaman, M.; Ali, M. A., Synthesis, molecular docking screening, ADMET and dynamics studies of synthesized 4-(4-methoxyphenyl)-8-methyl-3,4,5,6,7,8-hexahydroquinazolin-2(1H)-one and quinazolinone derivatives. *J. Mol. Str.* **2021**, *1244*, 130953. https://doi.org/10.1016/j.molstruc.2021.130953.

[18] Oyeneyin, O. E.; Obadawo, B. S.; Olanrewaju, A. A.; Owolabi, T. O.; Gbadamosi, F. A.; Ipinloju, N.; Modamori, H. O., Predicting the bioactivity of 2-alkoxycarbonylallyl esters as potential antiproliferative agents against pancreatic cancer (MiaPaCa-2) cell lines: GFA-based QSAR and ELM-based models with molecular docking. *J. Genet. Eng. Biotechnol.* **2021**, *19*, 1-15. https://doi.org/10.1186/s43141-021-00133-2.

[19] Oyeneyin, O. E.;Odadawo, S. O.; Orimoloye, S. M.; Akintemi, E. O.; Ipinloju, N.; Asere, A. M.; Owolabi, T. O., Prediction of inhibition activity of BET bromodomain inhibitors using grid search based extreme learning machine and molecular docking. *Lett. Drug Des. Discov.* **2021**, *18*, 1-11. https://doi.org/10.2174/1570180818666210521215433.

[20] Akinloye, O. A.; Akinloye, D. I..; Lawal, M. A.; Shittu, M. T.; Metibemu, D. S., (2020). Terpenoids from azadirachta indica are potent inhibitors of akt: Validation of the anticancer potentials in hepatocellular carcinoma in male wistar rats. *J. Food Biochem*. **2020**, e13559 https://doi.10.1111/jfbc.13559.

[21] Fontana, F.; Raimondi, M.; Marzagalli, M.; Di Domizio, A.; Limonta, P., Natural compounds in prostate cancer prevention and treatment: Mechanisms of action and molecular targets. *Cells*. **2020**, *9*, 460. https://doi.10.3390/cells9020460.

[22] Shaik, A. B.; Bhandare, R. R.; Nissankararao, S.; Edis, Z.; Tangirala, N. R.; Shahanaaz, S.; Rahman, M. M., Design, facile synthesis and characterization of dichloro substituted chalcones and dihydropyrazole derivatives for their antifungal, antitubercular and antiproliferative activities. *Molec*. **2020**, *25*, 1-16. https://doi.org/10.3390/molecules25143188

[23] Wang, Y.; Cheng, F. X.; Yuan, X. L.; Tang, W. J.; Shi, J. B.; Liao, C. Z.; Liu, X. H., Dihydropyrazole derivatives as telomerase inhibitors: structure-based design, synthesis, SAR and anticancer evaluation *in vitro* and *in vivo*. *Europ. J. Medicin. Chem.* **2016**, https://doi.10.1016/j.ejmech.2016.02.009.

[24] Mills, N., ChemDraw Ultra 10.0. *J. Amer. Chem. Soc.* **2016**, *128*, 13649-13650. https://doi.10.1021/ja0697875.

[25] SPARTAN 14', build 1.01. Irvine (CA), **2014**.

[26] Becke, A. D., Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648-5652.https://doi.org/10.1063/1.464913.

[27] Yap, C. W., PaDEL-Descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem*. **2011**, *32*, 1466-1474. https://doi.10.1002/jcc.21707.

[28] Obadawo, B. S.; Oyeneyin, O. E.; Anifowose, M. M.; Fagbohungbe, K. H.; Amoko, J. S., QSAR evaluation of C-8-tert-butyl substituted as potent anti-enterovirus agents. *Sci. Lett.* **2020**, *8*, 28-35.

[29] Kennard, R. W.; Stone, L. A., Computer aided Design of experiments. *Technom*. **1969**, *11*, 137-148.

https://doi.10.1080/00401706.1969.10490666.

[30] Khaled, K .F., Modeling corrosion inhibition of iron in acid medium by genetic function approximation method: A QSAR model. *Corros. Sci.* **2011**, *53*, 3457-3465. https://doi.10.1016/j.corsci.2011.01.035.

[31] Friedman, J. H., Multivariate adaptive regression splines. *Annals Stat.* **1991**, *19*, 1-67. https://doi.10.1214/aos/1176347963.

[32] Bansal, R.; Karthikeyan, C.; Moorthy, N. S. H. N.; Trivedi, P., QSAR analysis of some phthalimide analogues based inhibitors of HIV-1 integrase. *Arkivoc.* **2007**, *15*, 66-81. https://doi.10.3998/ark.5550190.0008.f08.

[33] Ikwu, F. A.; Shallangwa, G. A.; Mamza, P. A.; Uzairu, A., In silico studies of piperazine derivatives as potent anti-proliferative agents against PC-3 prostate cancer cell lines. *Heliyon* **2020**, *6*, e03273. https://doi.10.1016/j.heliyon.2020.e03273.

[34] Veerasamy, R.; Rajak, H.; Jain, A.; Sivadasan, S.; Varghese, C. P.; Agrawal, R. K., Validation of QSAR models-strategies and importance. *Int. J. Drug Des. Discov.* **2011**, *3*, 511-519.

[35] Bin Huang, G.; Zhu, Q. Y.; Siew, C. K., Extreme learning machine: Theory and applications," *Neurocomp.* **2006**, *70*, 449-501. https://doi.10.1016/j.neucom.2005.12.126.

[36] Cao, Y.; Wakil, K.; Alyousef, R.; Jermsittiparsert, K.; Si, L., Application of extreme learning machine in behavior of beam to column connections. *Struct.* **2020**, *25*, 861-867. https://doi.10.1016/j.istruc.2020.03.058.

[37] Owolabi, T. O.; Rahman, A. A., Prediction of band gap energy of doped graphitic carbon nitride using genetic algorithm-based support vector regression and extreme learning machine. *Symmet.* **2021**, *13*, 411, https://doi.10.3390/sym13030411.

[38] Shamsah, S. M. I.; Owolabi, T. O., Modeling the maximum magnetic entropy change of doped manganite using a grid search-based extreme learning machine and hybrid gravitational search-based support vector regression. *Cryst.* **2020**, *10*, 310, https://doi.10.3390/cryst10040310.

[39] Owolabi, T. O., Extreme learning machine and swarm- based support vector regression methods for predicting crystal lattice parameters of pseudo-cubic/cubic perovskites extreme learning machine and swarm based support vector regression methods for predicting crystal lat. *J. Appl. Phys.* **2020**, *147*, 245107, https://doi.10.1063/5.0008809.

[40] https://www.rcsb.org/structure/2AXA.

[41] Sastry, W.; Adzhigirey, G. M.; Day, M.; Annabhimoju, T.; Sherman, R., Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comp. Aid. Molec. Des.* **2013**, *27*, 221-234. https://doi.10.1007/s10822-013-9644-8.

[42] Shelley, J. C.; Cholleti, A.; Frye, L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M., Epik: a software program for pK. *J. Comp. Aid. Molec. Des.* **2007**, *21*, 681-691. https://doi.10.1007/s10822-007-9133-z.

[43] Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W., Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J. Chem. Theor. Comput.* **2010**, *6*, 1509-19. https://doi.10.1021/ct900587b.

[44] Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T., Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Medic. Chem.* **2006**, *49*, 6177-6196. https://doi.10.1021/jm051256o.

[45] https://www.schrodinger.com/qikprop.

[46] Lipinski, C. A., Rule of five in 2015 and beyond: Target and ligand structural limitations, ligand chemistry structure and drug discovery project decisions. *Adv. Drug Deliv. Rev.* **2016**, *101*, 34-41. https://doi.10.1016/j.addr.2016.04.029.

[47] Metibemu, D. S.; Oyeneyin, O. E.; Omotoyinbo, D. E.; Adeniran, O. Y.; Metibemu, A. O.; Oyewale, M. B., Molecular docking and quantitative structure activity relationship for the identification of novel phyto-inhibitors of matrix metalloproteinase-2. *Sci. Lett.* **2020**, *8*, 61-68. https://doi.10.47262/sl/8.2.132020005.

[48] Pantsar, T.; Poso, A., Binding affinity *via* docking: Fact and fiction. *Molecules.* **2018**, *23*, 1899.

https://doi.org/10.3390/molecules23081899.

[49] Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **2001**, *46*, 3-26. https://doi.10.1016/s0169-409x(00)00129-0.

[50] Metibemu, D. S., 3D-QSAR and molecular docking approaches for the identification of novel phyto-inhibitors of the cyclin-dependent kinase 4. *Sci. Lett.* **2021**, *9*, 42-48. https://doi.10.47262/SL/9.2.132021007.