




3D-QSAR, ADME-Tox, and molecular docking of semisynthetic triterpene derivatives as antibacterial and insecticide agents

Ossama Daoui¹ · Nouredine Mazoir² · Mohamed Bakhouch³ · Mohammed Salah⁴ · Ahmed Benharref⁵ · Azucena Gonzalez-Coloma⁶ · Souad Elkhatabi¹ · Mohamed El Yazidi⁷ · Samir Chtita⁸ 

Received: 9 November 2021 / Accepted: 3 March 2022 / Published online: 23 March 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

In the present work, 27 triterpene derivatives have been subjected to 3D-QSAR, ADME-Tox, and molecular docking for their insecticidal activity. The selected derivatives are previously semi-synthesized based on compounds obtained from *Euphorbia resinifera* and *Euphorbia officinarum* latex. The in silico studies were used to predict and to evaluate the antibacterial and insecticidal properties of the 3D structure of triterpene derivatives. The 3D-QSAR models are developed using CoMFA and CoMSIA techniques, and they have showed excellent statistical results ($R^2=0.99$; $Q^2=0.672$; $R^2_{\text{pred}}=0.91$ for CoMFA and $R^2=0.97$; $Q^2=0.61$; $R^2_{\text{pred}}=0.94$ for CoMSIA). The results indicate that the built models are able to describe the relationship between the structure of triterpene derivatives and the pLD₅₀ bioactivity. Based on contour maps obtained from CoMFA and CoMSIA models, 38 new molecules are designed and their pLD₅₀ activities are predicted. The drug-like and ADME-Tox properties of the molecule designed are examined and led to the selection of four molecules (**55**, **56**, **59**, **64**) as promising antibacterial and insecticidal agents. Compounds **55**, **56**, **59**, and **64** are able to inhibit the MurE (PDB code: 1E8C) and EcR (PDB code: 1R20) proteins involved in the process of antibacterial and insecticidal activities. This hypothesis is confirmed by the implementation of a molecular docking test. This test predicted the most important referential interactions that occur between the structure of triterpene derivatives and the targeted receptors. Among the four docked molecules, three molecules (**55**, **56**, and **59**) showed greater stability than the reference molecule **16** inside the MurE and EcR receptors pocket. Therefore, the structure of the three new triterpene derivatives can be adopted as reference for the synthesis of antibacterial drugs and also in the development of insecticides.

Keywords *Euphorbia resinifera* · *Euphorbia officinarum* · Triterpenes · Antibacterial · Insecticide · 3D-QSAR · ADME-Tox · Molecular docking

✉ Nouredine Mazoir
mazoirn@gmail.com

✉ Samir Chtita
samirchtita@gmail.com

¹ Laboratory of Engineering, Systems and Applications, National School of Applied Sciences, Sidi Mohamed Ben Abdellah-Fez University, BP Box 72, Fez, Morocco

² Research Unit: Natural Resource Valorizations, Laboratory of Plant Biotechnology and Ecosystem Valorization, Faculty of Sciences, Chouaib Doukkali University, P.O. Box 20, El Jadida, Morocco

³ Laboratory of Bioorganic Chemistry, Department of Chemistry, Faculty of Sciences, University Chouaib Doukkali, P.O. Box 20, El Jadida, Morocco

⁴ Molecular Modeling and Spectroscopy Research Team, Faculty of Science, Chouaib Doukkali University, P.O. Box 20, El Jadida, Morocco

⁵ Laboratory of Biomolecular Chemistry, Natural Substances and Reactivity, URAC 16, Faculty of Sciences Semailia, Cadi Ayyad University, P.O. Box 2390, Marrakech, Morocco

⁶ Instituto de Ciencias Agrarias, CSIC, Serrano 115-dpdo, 28006 Madrid, Spain

⁷ Engineering Laboratory of Organometallic and Molecular Materials and Environment, Faculty of Sciences Dhar EL Mahraz, Sidi Mohamed Ben Abdellah University, P.O. Box 1796, Fez, Morocco

⁸ Laboratory of Analytical and Molecular Chemistry, Faculty of Sciences Ben M'Sik, Hassan II University, P.O. 7955, Casablanca, Morocco

Introduction

Natural products have been used since ancient time as remedies to treat a plenty of illnesses and diseases [1]. They have attracted much attention and have become an important source of new drugs in the last years. Some of the natural products have shown a great interest in the fight against COVID-19 pandemic [2–7].

Terpenoids are secondary metabolites of isopentenyl pyrophosphate oligomers, which include the largest group of plant natural products. Terpenoids is a class of compounds that are divided into subclasses according to the number of its isoprene units (a) and carbon atoms (b) and are identified by the notation a:b, e.g., monoterpenes (2:10), sesquiterpenes (3:15), diterpenes (4:20), sesterpenes (5:25), triterpenes (6:30), carotenoids (8:40), and rubber (> 100: > 500). Among terpenoids compounds, triterpenes are commonly used for medicinal purposes in many countries due to their various pharmacological properties.

In fact, the literature provides a variety of research on the interesting biological activity of oxygenated triterpene compounds. The first studies were based on 7-oxopregnenolone derivatives evaluated as very potential anti-cortisones [8]. In addition, derivatives of 7-oxo-dehydroepiandrosterone have recently been used for their reduction of hyperglycemia or strengthening of the immune system [9]. In the same context, plants of the category *Euphorbiaceae* are the largest genus with species containing cytotoxic, antibacterial, antifungal, antiparasitic, antitumor, and anti-inflammatory triterpenes [10–13]. Derivatives obtained by conventional oxidation, using metalloporphyrin complexes, of triterpenes semi-synthesized from *Euphorbia officinarum* latex [14–16] have demonstrated strong postingestive toxic effects on the insect pest *Spodoptera littoralis* [17, 18]. In addition, some of these derivatives protected tomato plants from *Verticillium dahliae* at low concentrations, elicited H₂O₂, and increased antioxidant enzyme activity suggesting elicitor-like effects [19, 20].

In recent years, research on natural substances has led to the development and discovery of drugs for human use and cytotoxic agents for use as pesticides [21]. In this context, we rely in this work on semisynthetic triterpene derivatives to predict the possibility of using these molecules as a new source in the development of insecticides and also for further application as antibacterial. For this purpose, we used 3D-QSAR techniques because of their wide use in drug discovery, due to the correlation between the quantitative three-dimensional structure of molecules and biological activity [17, 18]. Then, we perform molecular docking and prediction of drug kinetic parameters (ADME-Tox) in silico, in view of their importance in the optimization of the accessibility to a new drug at preclinical phase [22]. In this

paper, a 3D-QSAR study on 27 triterpenes derivatives was used to build the QSAR model, which was generated using comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) [23]. New molecules were designed, and their activities predicted (pLD₅₀) based on the analyses of the contour maps provided by the 3D-QSAR model. Molecular docking studies of semi-synthesized and newly designed compounds were carried out with MurE protein for antibacterial activity and with EcR protein for anti-insecticidal activity in pesticide use on the other hand. It is in order to understand the main structural requirements and to analyze the main interactions between ligands and receptors. The results obtained confirmed the suitability of the semi-synthesized and also designed compounds as antibacterial agents and insecticides. Each designed compound tested by drug-likeness and computational pharmacokinetics (ADME-Tox) parameters.

Material and methods

Preparation of the database

All molecules were carefully sketched and saved as separate Mol2 file in MOL2 format. The SYBYL X2.1.1 software was used to generate three-dimensional molecular structures and to minimize the energy of each 3D structure created with the standard Tripos Powell force field (100 iterations) [24]. The three-dimensional structures of the twenty-seven semi-synthesized triterpenes studied in this work are optimized and their energy minimized by computation of the popular Gasteiger-Hückel atomic partial charges to construct 3D-QSAR models [25]. All molecular structures were analyzed with a distance-dependent buffer function until a root mean square (RMS) deviation of 0.05 kcal/(mol) was reached by the SYBYL X2.1.1 software.

Analysis of the distribution In order to perform more powerful 3D-QSAR models, a cluster analysis based on molecular features was performed [26]. In the present work, we have used a set of 27 compounds, which was divided into subsets, a training containing 20 compounds and a test set enclosing 7 compounds. The test set was used to evaluate the predictive ability of the obtained models. The compounds belonging to the test set were chosen based on a diverse range of lethal dose activities and structural diversities. The structures LD₅₀ and pLD₅₀ of the 27 studied compounds are presented in Table 1. These data were exploited to develop the 3D-QSAR models, comparative molecular field analysis (CoMFA) [27], and comparative molecular similarity indices analysis (CoMSIA) [28]. Using CoMFA and CoMSIA approaches to analyze the physical and chemical properties of the studied molecules by analyzing the contour maps obtained after developing the 3D-QSAR model.

Table 1 Chemical structures of the studied molecules and their lethal dose related to the insecticidal activity

| N° | Structures | LD ₅₀ (μM) | pLD ₅₀ |
|----|------------|-----------------------|-------------------|
| 01 | | 80.87 | 4.092 |
| 02 | | 94.69 | 4.024 |
| 03 | | 100 | 4.000 |
| 04 | | 14.67 | 4.834 |
| 05 | | 14.48 | 4.839 |
| 06 | | 14.05 | 4.852 |
| 07 | | 25.99 | 4.585 |
| 08 | | 67.25 | 4.172 |
| 09 | | 17.69 | 4.752 |
| 10 | | 14.41 | 4.841 |
| 11 | | 50 | 4.301 |
| 12 | | 26.92 | 4.570 |

| | | | |
|----|--|-------|-------|
| 13 | | 2.79 | 5.554 |
| 14 | | 9.04 | 5.044 |
| 15 | | 10.44 | 4.981 |
| 16 | | 0.6 | 6.222 |
| 17 | | 9.9 | 5.004 |
| 18 | | 20.6 | 4.686 |
| 19 | | 11.4 | 4.943 |
| 20 | | 20.6 | 4.686 |
| 21 | | 23.18 | 4.635 |
| 22 | | 17.2 | 4.764 |
| 23 | | 64.2 | 4.192 |
| 24 | | 26.2 | 4.582 |
| 25 | | 27.2 | 4.565 |
| 26 | | 9.7 | 5.013 |
| 27 | | 20.3 | 4.693 |

3D-QSAR studies

The molecular alignment method is an important step in the development of a 3D-QSAR model. The highest lethal dose of compound **16** was chosen as a reference. Figure 1 shows that all 3D molecular structures of the training and test sets were aligned to the common core using the alignment technique [29]. The 3D-QSAR models are also established to predict and explain the lethal dose of the studied compounds. In this work, we have developed 3D-QSAR models by relating the following descriptors: static (S), static (E), hydrophobic (H), hydrogen bond donor (D), and hydrogen bond acceptor (A) with the biological activity (pLD₅₀). The aforementioned molecular descriptors used to develop 3D-QSAR models were obtained by CoMFA and CoMSIA techniques.

The CoMFA model was built based on steric and electrostatic field descriptors, while the CoMSIA model was built based on steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor field descriptors. The realization of the 3D-QSAR patterns was performed as previously reported in the literature [25–29]. All of these analyses were realized using the Tripos force field with a reference spatial grid of 2 Å in all Cartesian directions. We used a carbon atom hybridized sp³ and with a net charge of + 1.0 was used as a source to calculate steric and electrostatic energies. The correction factor was set to the default value of 0.3, which controls the slope of the Gaussian function, and the cutoff energy is set to 30 kcal/mol by default [30].

CoMFA and CoMSIA study

The QSAR CoMFA and CoMSIA models were developed using the partial least square (PLS) algorithm [31]. The built models were then validated using cross-leave-one-out validation method and Bootstrapping methods [32]. Based on the training set, the CoMFA and CoMSIA models are developed based on PLS algorithms. The models selected have high values of the standard correlation coefficient ($R^2 > 0.5$) as well as the cross-validation correlation coefficient ($Q^2 > 0.5$) and low values of the standard error of estimation (SEE) value. The predictive power of the developed 3D-QSAR models is evaluated by the index R^2_{pred} , which is calculated based on the test set.

PLS analysis and test validation

To obtain a statistically valid 3D-QSAR model, we have implemented PLS (partial least squares regression) method to obtain a linear correlation between the observed pLD₅₀ (dependent variable) and the descriptors of CoMFA and CoMSIA (independent variables) separately. During the PLS analysis, the leave-one-out (LOO) cross-validation method

was used to determine the optimal number of components N and the cross-validation correlation coefficient (Q^2). After determining N , the analysis using the cross-validation method was performed to test the overall significance of the model by calculating statistical parameters such as coefficient of determination (R^2), standard error of estimates (SEE), and F -value (Fischer test). To further evaluate the robustness and statistical confidence of the built models, a bootstrapping analysis for 100 cycles was performed. Bootstrapping involves the generation of many new data sets from the original data set [33]. The statistical calculation was performed on each of these bootstrapping samples. The difference between the parameters calculated from the original data set and the average of the parameters calculated from the many bootstrapping samples is a measure of the bias of the original calculations. All validated results are analyzed according to the fact that Q^2 value should be greater than 0.5, which indicates that the probability of association by chance is less than 5% [34]. Similarly, R^2_{pred} values greater than 0.6 was obtained after external validation confirms the reliability of the constructed 3D-QSAR models. The predictive power of the 3D-QSAR models (CoMFA and CoMSIA) was evaluated through a test set of 7 molecules, which were aligned as those in the training set, and their activities were predicted using the model created by the training set. The prediction correlation coefficient r^2_{pred} , the cross-validation coefficient Q^2 , and the root mean square correlation coefficient r^2_m were calculated using the formulas shown in Eqs. (1), (2), and (3).

$$Q^2 = 1 - \frac{\sum_{i=1}^N (Y_{\text{exp}} - Y_{\text{pred}})^2}{\sum_{i=1}^N (Y_{\text{exp}} - \bar{Y}_{\text{exp}})^2} \quad (1)$$

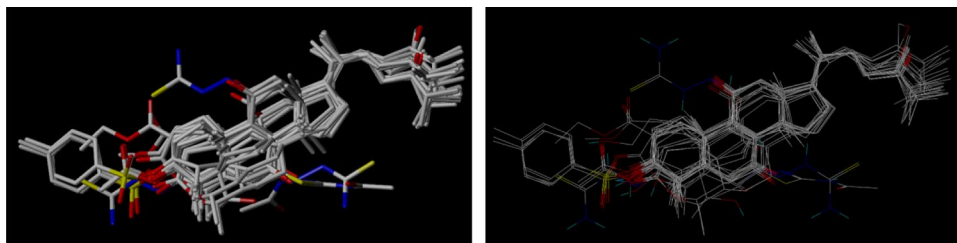
$$R^2_{\text{pred}} = 1 - \frac{\text{PRESS}}{\text{SD}} \quad (2)$$

$$r^2_m = r^2 \times \left[1 - \left| (r^2 - r_0^2) \right| \right] \quad (3)$$

where Y_{pred} , Y_{exp} , and \bar{Y}_{exp} are the predicted, experimental, and mean values of the target property (pLD₅₀), respectively. PRESS is the predicted sum of squared residuals between the predicted and actual activity values for each molecule (i) in the test set, SD is the sum of squared residuals between the biological activities of the test set and the average activity of the molecules in the training set (N), R^2 is the squared correlation coefficient between the observed and predicted activity, and r_0^2 is the squared correlation coefficient between the observed and predicted activity without intercept.

In addition, the performance test of the CoMFA and CoMSIA models developed in this work was carried out by

Fig. 1 Alignment of the data set database



another test called Y-randomization, with which the possibility of obtaining high performance of the models can be refuted randomly (by chance) [35]. In this approach, the values of the dependent variable (biological activity) are randomly distributed, while those of the independent variable (selected descriptors) are kept integer. The random data are then used to build a new QSAR model. Then, to ensure the robustness of this new QSAR model, the squared correlation coefficient (R_r^2) of the randomized model must be less than the squared correlation coefficient (R^2) of the non-randomized model. The difference between the values of the mean square correlation coefficient of the randomized model (R_r^2) and the square correlation coefficient of the non-randomized model (R^2) is reflected in the value of the parameter (cR_p^2) [36]. This measure limits the difference between the values of the squared correlation coefficients of the non-randomized model (R^2) and the randomized model (R_r^2) according to Eq. (4). The value of (cR_p^2) must be more than 0.5 [37].

$${}^cR_p^2 = R \times \sqrt{R^2 - R_r^2} \quad (4)$$

In addition, the results of the CoMFA and CoMSIA models with statistically acceptable efficiency are represented graphically by contour maps. The analysis of these maps allows the prediction of the most important physical and chemical properties of the triterpene derivatives studied in this work, based on the synthesized reference molecule (**16**) that shows the high experimental activity ($pLD_{50} = 6.22$). The structural properties of these molecules allow us to design new molecules and improve their activity (pLD_{50}) and to test the dual use of these molecules as new antibacterial drug agents and also as insecticides.

Drug-likeness properties and ADME-Tox pharmacokinetics in silico

An in silico study was performed to design and to identify new antibacterial and anti-insecticidal agents using 3D-QSAR. The drug-likeness profiles (Adsorption, Distribution, Metabolism, Excretion and Toxicity) were determined for the newly designed compounds. The main reason

for these two profiles is to predict the most important drug properties of the molecules and also to evaluate their pharmacokinetics before their synthesis [38].

The screening of newly designed molecules investigated for drug-like properties is based on the high pLD_{50} values predicted by the 3D-QSAR models. ADME-Tox parameters are predicted only for designed molecules that have successfully passed the drug-likeness screening process. The lethal dose (LD_{50}) values predicted by the 3D-QSAR models for the newly designed molecules should be lower than the values (LD_{50}) predicted by the in silico toxicity test. This is to ensure the safe use of these molecules as non-toxic drugs for humans and lethal for insects.

Molecular docking

In this work, molecular docking simulation is performed using AutoDock Vina software [39], in order to analyze the mechanism of interactions and study the binding modes to obtain a predictive view on the main structural requirements of triterpene derivatives to be considered as promising agents in medicinal use as antibacterials and also as insecticides.

One of the best sources of antibacterial targets is the biochemical pathway of peptidoglycan synthesis. The importance of Mur ligases as a suitable antibacterial therapeutic target was described in a study carried out by Kouidmi et al. [40]. Amide ligases (MurC, MurD, MurE, and MurF) are a convenient catalytic mechanism to develop multi-targeted antibacterial targets while reducing the potential of target resistance development [41]. We are investigating the possibility of using compound **16** and molecules designed from triterpenes as new insecticidal agents. This is done by docking these ligands with the ecdysone receptor (EcR) found in insects, in order to identify the most important types of interactions that take place and predict the most important reference sites in inhibiting the activity of this protein. EcR regulates larval development and promotes insect reproduction [42]. The inhibition of the activity of the EcR protein by designed molecules could be a valid proposal in the fight against the insect *Mythimna separata*, which is harmful to agricultural products.

In this work, the semisynthetic ligand (**16**) that showed the highest pLD₅₀ value and the newly designed ligands are docked into the receptor pocket of MurE (PDB code: 1E8C) [43], also with the ecdysone receptor (EcR) complexed with the pesticide Tebufenozide (BYIO8346) (PDB ID: 1R20) [44]. In this work, ligands and protein were prepared using Discovery Studio; the same software was used to analyze the interactions between ligands and receptors. The 3D network was generated by the AUTOGRID algorithm in MGLTools 1.5.6. The box grid was constructed along the directions $X=60$, $Y=60$, and $Z=60$ within the MurE receptor pocket and $X=40$, $Y=40$, and $Z=40$ within the EcR receptor pocket with a distance of 0.375 Å between the grid points. The network center coordinates are fixed at 10.40 Å, 50.10 Å, and 98.99 Å at the MurE receptor and 59.883103 Å, 29.836276 Å, and 13.911379 Å at the EcR receptor; these coordinates are considered to be the insertion sites of the docked molecules. The obtained molecular docking results were analyzed using the Discovery Studio 2016 software.

Results and discussion

3D-QSAR models

Database division Seven molecules were carefully selected as elements of the test set and the remaining twenty molecules as elements of the training set. The partitioning of the database was performed with respect to structural diversity and gradual biological activity (pLD₅₀).

Molecular alignment All the molecules belonging to the data set were correctly aligned, and the molecule **16** was used as a reference (Fig. 2).

PLS analysis The results of the PLS analysis are shown in Table 2. For a reliable predictive model, the cross-validation coefficient Q^2 which defines the quality of the prediction should be greater than 0.5, while the cross-validated correlation coefficient indicates the accuracy of a QSAR model. The F -test value represents the statistical confidence.

The statistical results presented in Table 2 show that the predictive power of both CoMFA and CoMSIA models is good and statistically significant. In addition, the CoMFA model is more accurate than the CoMSIA model, as evidenced by the low residual MSE (0.052) values, high Q^2 (0.672) and R^2 (0.998) and R^2_{pred} (0.918) value.

As shown in Table 2 of the PLS summary, the CoMFA model shows the following results ($R^2=0.99$, $F=443.73$, and $SEE=0.052$). The CoMSIA model shows also ($R^2=0.97$, $F=60.38$, and $SEE=0.071$).

The high cross-validation coefficients Q^2 (0.672 and 0.534) observed for the CoMFA and CoMSIA models confirm the good correlation between the descriptors of activity of pLD₅₀ and the molecular fields descriptors calculated for all compounds of the training set. Therefore, the results of the CoMFA and CoMSIA models can be considered statistically acceptable by the internal validation test. The success of both models in the internal test did not confirm their performance to predict the activity pLD₅₀ of the molecules outside the training set. Therefore, the robustness of these models was verified by an external test, where these two models applied to the seven molecules of the test set. The high values of the R^2_{pred} coefficient obtained by the external tests of the CoMFA and CoMSIA models (0.918 and 0.94), respectively, indicate the powerful predictive power of these two models. Thus, we can exploit both models to predict the activity of new molecules derived from Triterpene that we can design and improve their anti-insect and antibacterial activity.

In addition, through a bootstrap test of 100 runs, we evaluated the measure of bias of the original database on which the two 3D-QSAR models were built. We performed this procedure to verify the effectiveness of the prepared database to build the obtained CoMFA and CoMSIA models. Through this test, we obtained high R^2_{bs} values (0.96 for CoMFA and 0.97 for CoMSIA) with SEEBs (0.008 and 0.053) for the two models, respectively. These results indicate that there was no significant deviation between the statistical parameters calculated from the original data and the average parameters calculated after 100 runs. Therefore, the database prepared in this work to develop the two models of CoMFA and CoMSIA was adequate, and the high statistical parameters obtained in the internal and external tests are statistically reliable.

The ratios of the coefficients of the molecular field descriptors contributing to the construction of the CoMFA and CoMSIA models. We can notice that the ratios of steric (S) and electrostatic (E) fields descriptors were 38% and 62%, respectively, in the CoMFA model. The contribution ratios of steric (S), electrostatic (E), hydrophobic (H), hydrogen bond donor (D), and hydrogen bond acceptor (A) fields descriptors were 11.9%, 29.5%, 19.3%, 12.8%, and 22.6%, respectively, in the CoMSIA model. The high contribution ratios of the S and E and S , E , H , D , and A field descriptors in the obtained CoMFA and CoMSIA models, respectively, indicate that the biological pLD₅₀ activity of triterpene derivatives structures is strongly related to steric, electrostatic, hydrophobic, and hydrogen bonding properties. Therefore, encoding these properties in the structure of Triterpene derivatives can improve the pLD₅₀ activity of these molecules against insects and bacteria.

Fig. 2 Common substructure (core) used in the alignment, and the structure of the reference molecule **16**

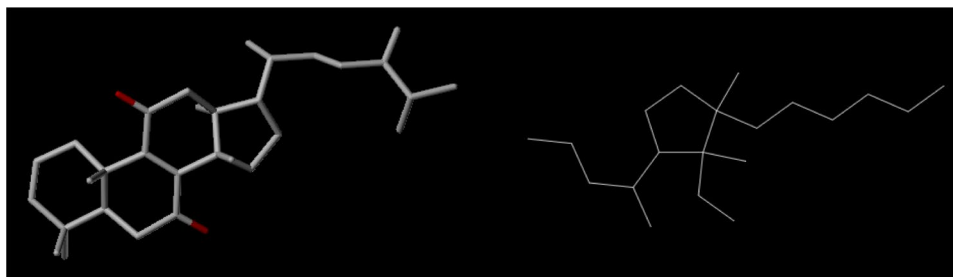


Table 3 gather the experimental and predicted inhibitory activities and residual values of the training and test sets obtained by the CoMFA and CoMSIA models.

The residuals plot for the training and test sets obtained from the CoMFA and CoMSIA models are shown in Fig. 3A, B, respectively. Figure 4 illustrates the correlation plot between the predicted and experimental data for the training and test sets obtained from the CoMFA and CoMSIA models, respectively.

Figure 4 shows a smooth proportional distribution of observed pLD_{50} activity values based on the predicted values via CoMFA and CoMSIA. The linear and regular distribution of the observed pLD_{50} values as a function of the predicted values can be explained by the high R^2 values obtained for the two developed models and the low SEE values, thus confirming the performance of the obtained 3D-QSAR models. To further confirm that the performance of the 3D-QSAR models is not due to random chance, the Y-randomization test was performed five times. The results of this test are presented in Table 4.

From Table 4, we can notice that the Q^2 and R^2 values obtained after five random mixtures are lower than their original model counterparts, and the cR_p values are also higher than 0.5. The strong correlation between the pLD_{50} activity of the structure of each molecule and the three-dimensional molecular field descriptors S, E, H, D, and A is not due to chance.

Contour maps analysis

The CoMFA contour maps indicates the region in space where aligned molecules can interact favorably or unfavorably with

the receptor. On the 3D structure of the active molecules, the contour maps generated by CoMSIA allow one to encode the physicochemical properties and the areas of potential radicals likely to interact with the target receptors [45]. Contour maps are generated from the ratio of each molecular field descriptor that contributed to the construction of the 3D models of CoMFA and CoMSIA. These contributions are visualized as contour maps in the structure of the examined molecules. These visualizations display hypothetical ratios of favorable (80%) and unfavorable (20%) regions related to predicted biological activity in the three-dimensional structure of the examined molecules. In this work, we examined the structure of molecule **16** that presented the maximum observed pLD_{50} (6.22) value within the triterpene derivative series, and we consider its structure as a reference in the analysis of the generated contour maps.

CoMFA steric interactions Figure 5A–C shows the spatial position of steric contour maps (green and yellow, Fig. 5B) and electrostatic (blue and red, Fig. 5C) in the structure of template molecule **16** obtained by the CoMFA model.

From Fig. 5B, the green contours indicate areas favorable for large groups to increase the activity of the template molecule, while the yellow contours indicate areas unfavorable for large groups to increase the activity. From Fig. 5C, the red contours show the favorable sites for negatively charged groups to improve the activity of pLD_{50} , while the blue contours show the favorable sites for positively charged radicals to improve the structure of pLD_{50} activity of the template molecule structure.

Table 2 The statistical parameters of the CoMFA and CoMSIA models obtained by PLS analysis

| | Q^2 | R^2 | SEE | F | N | R^2_{pred} | R^2_{bs} | SEE_{bs} | Fractions | | | | |
|--------|-------|-------|-------|--------|-----|--------------|------------|------------|-----------|-------|-------|-------|-------|
| | | | | | | | | | S | E | H | D | A |
| CoMFA | 0.672 | 0.998 | 0.052 | 443.73 | 5 | 0.918 | 0.96 | 0.008 | 0.388 | 0.612 | - | - | - |
| CoMSIA | 0.534 | 0.978 | 0.071 | 60.381 | 5 | 0.940 | 0.97 | 0.053 | 0.119 | 0.295 | 0.193 | 0.128 | 0.266 |

Q^2 cross-validation correlation coefficient, N optimal number of components determined by the PLS-LOOCV cross-validation study, SEE standard error of estimation, R^2 conventional correlation coefficient, R^2_{pred} predictive correlation coefficient, R^2_{bs} correlation coefficient after 100 bootstrapping cycles, SEE_{bs} standard error of estimation for 100 bootstrapping cycles.

Table 3 pLD₅₀ values experimental, predicted by CoMFA and CoMSIA analyses and residuals

| Name | pLD ₅₀ | CoMFA | | CoMSIA | |
|------|-------------------|---------------------------|-----------|---------------------------|-----------|
| | | pred (pLD ₅₀) | Residuals | pred (pLD ₅₀) | Residuals |
| 1 | 4.092 | 4.106 | -0.014 | 4.111 | -0.019 |
| 3 | 4.000 | 4.002 | -0.002 | 4.05 | -0.05 |
| 4 | 4.834 | 4.818 | 0.016 | 4.749 | 0.085 |
| 6 | 4.852 | 4.827 | 0.025 | 4.776 | 0.076 |
| 8 | 4.172 | 4.160 | 0.012 | 4.207 | -0.035 |
| 9 | 4.752 | 4.714 | 0.038 | 4.815 | -0.063 |
| 10 | 4.841 | 4.845 | -0.004 | 4.868 | -0.027 |
| 11 | 4.301 | 4.322 | -0.021 | 4.358 | -0.057 |
| 12 | 4.570 | 4.548 | 0.022 | 4.579 | -0.009 |
| 13 | 5.554 | 5.592 | -0.038 | 5.611 | -0.057 |
| 14 | 5.044 | 5.013 | 0.031 | 5.058 | -0.014 |
| 15 | 4.981 | 4.972 | 0.009 | 5.045 | -0.064 |
| 18 | 4.686 | 4.694 | -0.008 | 4.664 | 0.022 |
| 19 | 4.943 | 4.933 | 0.01 | 4.956 | -0.013 |
| 20 | 4.686 | 4.697 | -0.011 | 4.636 | 0.05 |
| 21 | 4.635 | 4.610 | 0.025 | 4.551 | 0.084 |
| 22 | 4.764 | 4.781 | -0.017 | 4.845 | -0.081 |
| 23 | 4.192 | 4.269 | -0.077 | 4.191 | 0.001 |
| 24 | 4.582 | 4.536 | 0.046 | 4.44 | 0.142 |
| 26 | 5.013 | 5.055 | -0.042 | 4.972 | 0.041 |
| 2* | 4.024 | 4.235 | -0.328 | 4.43 | -0.406 |
| 5* | 4.839 | 4.600 | 0.140 | 4.728 | 0.111 |
| 7* | 5.585 | 5.253 | 0.008 | 5.619 | -0.034 |
| 16* | 6.22 | 6.186 | 0.036 | 6.085 | 0.135 |
| 17* | 5.004 | 4.882 | 0.101 | 4.954 | 0.05 |
| 25* | 4.565 | 4.493 | 0.029 | 4.434 | 0.131 |
| 27* | 4.693 | 4.707 | -0.014 | 4.462 | 0.231 |

*Test set items

CoMSIA steric interactions In Fig. 6A–E, it can be seen that the S and E fields of the CoMFA and CoMSIA models are similar and consistent, while the other fields (H, D, A) are distinct.

The colors of the steric (green and yellow, Fig. 6A) and electrostatic (blue and red, Fig. 6B) contour maps reflect the identical structural properties favorable or unfavorable to improved pLD₅₀ activity that are expressed by the CoMFA model.

shows the regions of hydrophobic and hydrophilic fields on the structure of the reference molecule **16**. The yellow contours indicate the favorable positions of the hydrophobic groups to enhance pLD₅₀ activity, while the white contours show the unfavorable positions of the hydrophobic groups but favorable to the hydrophilic groups to enhance pLD₅₀ activity. From (Fig. 6E), the magenta contour maps show the regions favorable to hydrogen bond acceptor groups to

increase pLD₅₀ activity, whereas the red contours show the regions unfavorable to hydrogen bond acceptor groups.

CoMFA contour maps

From In Fig. 5B, the yellow contour immediately surrounding the 3,3-dimethylcyclohex-1-n group indicates that the addition of bulky radicals to this site can decrease the biological activity of molecule **16**. However, the green contour surrounding the aromatic cycle indicates that this site is suitable for the substitution and addition of bulky radicals to improve the anti-insect and antibacterial activity (pLD₅₀). The small green area around the forward methyl radical on the aromatic cycle of 3,3-dimethylcyclohex-1-ene shows the possibility to replace the methyl group with another group to achieve the desirable activity. In parallel, we see that the large green area around the isobutane group located at the right end of the structure shows the possibility to replace the isobutane group with large group. From Fig. 5C, we can see that the contributions of the blue contours are dominant over their red homologue adjacent to the 3,3-dimethylcyclohex-1-ene ring, which means that the insecticidal and antibacterial activity of the triterpene derivatives can be enhanced by the appearance of influential electron donor properties (I+) associated with the 3,3-dimethylcyclohex-1-ene group. As we notice that there is a marked affinity between the red and blue contours near the oxygen atom in the 4-methylcyclohex-2-en-1-one ring, this means that keeping the oxygen atom unmodified at this position is favorable for the conservation of the quantity relationship between the activity of the template molecule and its structure.

CoMSIA contour maps

Through the contour maps shown in Fig. 6 generated from the obtained CoMSIA model, we can notice that the position of the green and yellow contours (Fig. 6A) is very similar to the contours obtained from the CoMFA contour map visualizations, thus the same structural features as presented in the CoMFA analyses are abstracted.

Similarly, the predominance of blue over red electrostatic contours depicted in Fig. 6B greatly supports the results extracted from the steric (Fig. 5C) contours in the CoMFA model, but we can detect the production of a new red contour surrounding an ethene group that was not predicted by the CoMFA model. Thus, the structure of the pLD₅₀ activity of the template molecule structure can be improved by adding radicals on the side of the ethene group that have inductive electronegative effects (-I).

Through the contours of the hydrophobic fields shown in Fig. 6C, we can notice a large profusion of white contours along the structure of the reference molecule. This means

Fig. 3 The residuals plot for the training and test sets obtained from the CoMFA (A) and CoMSIA (B) models

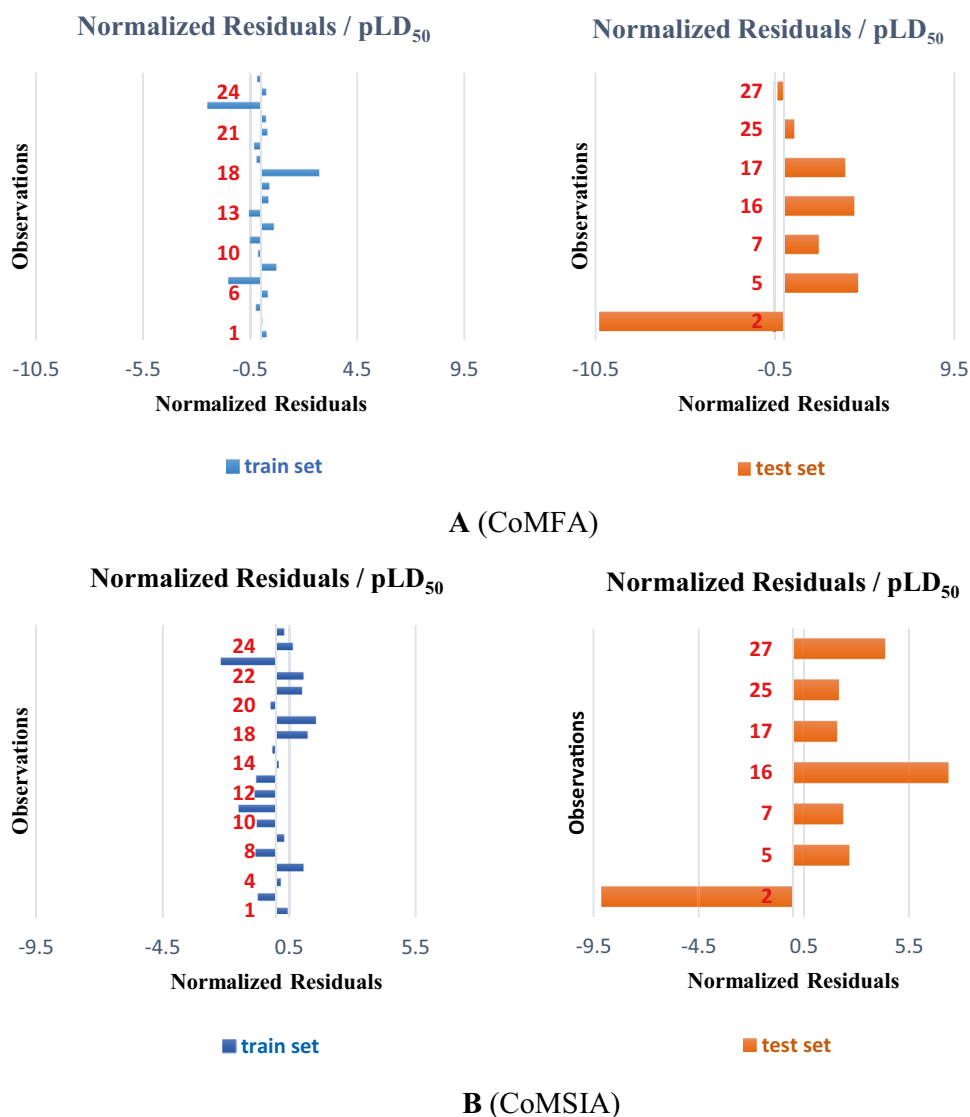


Fig. 4 Experimental versus predicted activity of the training and testing set according to the CoMFA and CoMSIA models

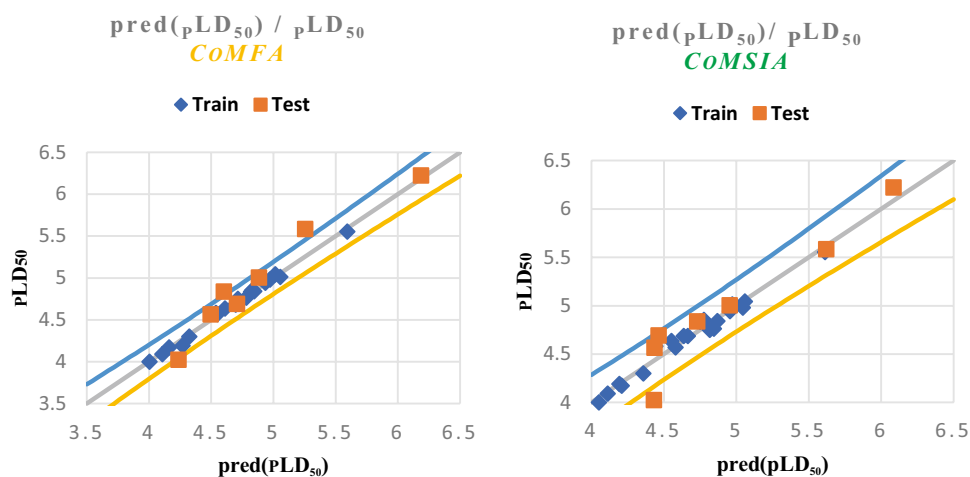


Table 4 The Q^2 and R^2 values after the Y-randomization tests

| Iteration | CoMFA | | | CoMSIA | | |
|-----------|----------------------|----------------------|-------------|----------------------|----------------------|-------------|
| | Q^2_{yrand} | R^2_{yrand} | ${}^cR_p^2$ | Q^2_{yrand} | R^2_{yrand} | ${}^cR_p^2$ |
| Original | 0.672 | 0.986 | - | 0.534 | 0.978 | - |
| 1 | 0.440 | 0.684 | 0.545 | 0.340 | 0.57 | 0.626 |
| 2 | 0.113 | 0.578 | 0.634 | 0.402 | 0.519 | 0.664 |
| 3 | -0.215 | 0.415 | 0.750 | 0.205 | 0.473 | 0.696 |
| 4 | -0.108 | 0.438 | 0.734 | 0.370 | 0.536 | 0.652 |
| 5 | -0.216 | 0.476 | 0.708 | -0.551 | 0.513 | 0.668 |

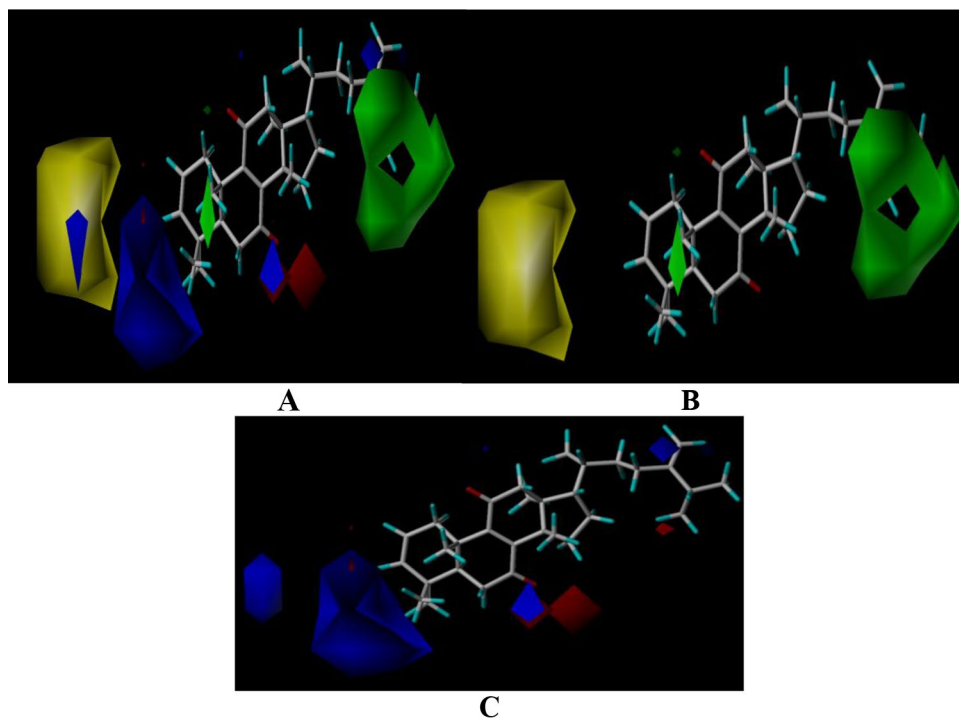
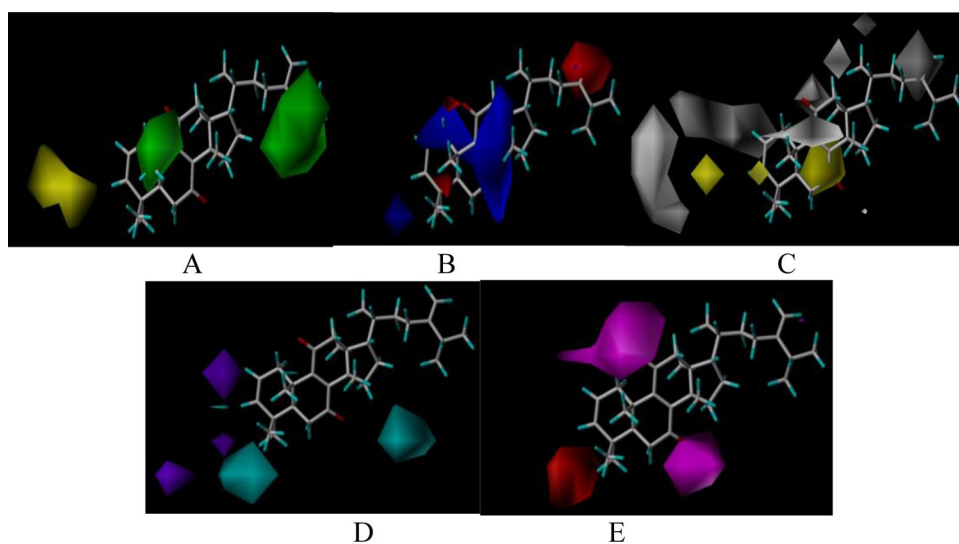
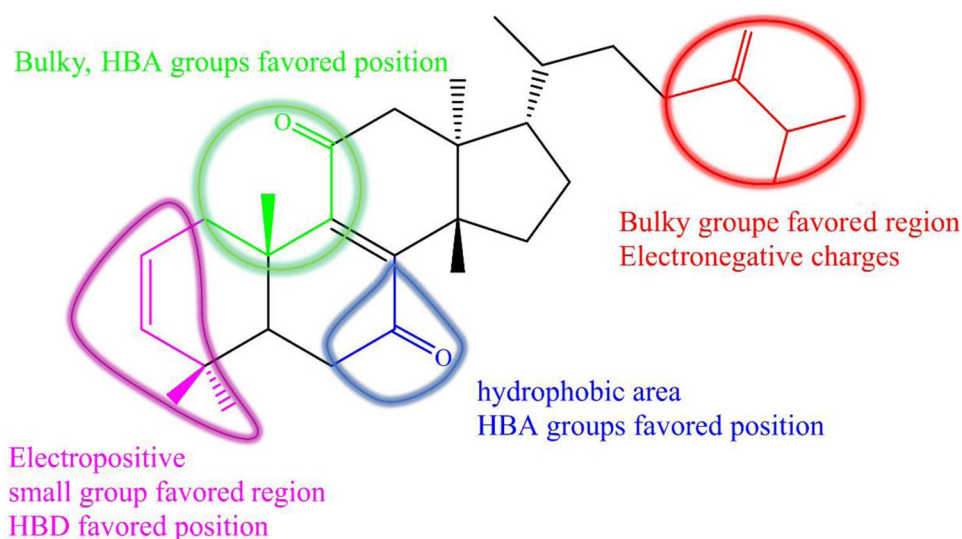
Fig. 5 **A** Combination of steric and electrostatic fields, **B** steric contour maps, and **C** electrostatic contour maps based on the CoMFA model**Fig. 6** **A** Champs stériques, **B** Champs électrostatiques, **C** Champs hydrophobes, **D** Champs donneurs de liaisons H, **E** Champs accepteurs de liaisons H

Fig. 7 Proposed structural modifications of oxo-dehydroepiandrosterone derivatives for the design of new potent and more selective molecules



that the structure of triterpene derivatives has hydrophilic properties, and this property may be adequate for doping and absorption of these molecules by living organisms. Although we can observe a yellow contour surrounded by a segment covering the oxygen atom site in the 4-methylcyclohex-2-en-1-one group, this ring is favorable for hydrophobic radicals to enhance the structure of the activity of the template molecule structure 16.

Through visualization of the contour maps of the molecular field of hydrogen bond donor group (Fig. 6D) and hydrogen bond acceptor (Fig. 6E) interactions, we can notice the presence of spherical contours in the proximity of the 3,3-dimethylcyclohex-1-ene ring, which means that the addition of hydrogen bond donor radicals on the 3,3-dimethylcyclohex-1-ene ring is favorable for the activity of pLD₅₀. On the other hand, we can note the magenta contour positions around the two oxygen atoms on the rings of 4,8-dimethyl-2,3,4,6,7,8-hexahydronaphthalene-1,5-dione; this means that the addition of hydrogen bond acceptor radicals at the dione sites is favorable to improve the biological activity of the triterpene derivative structure against insects and bacteria (Fig. 7).

Design of new compounds

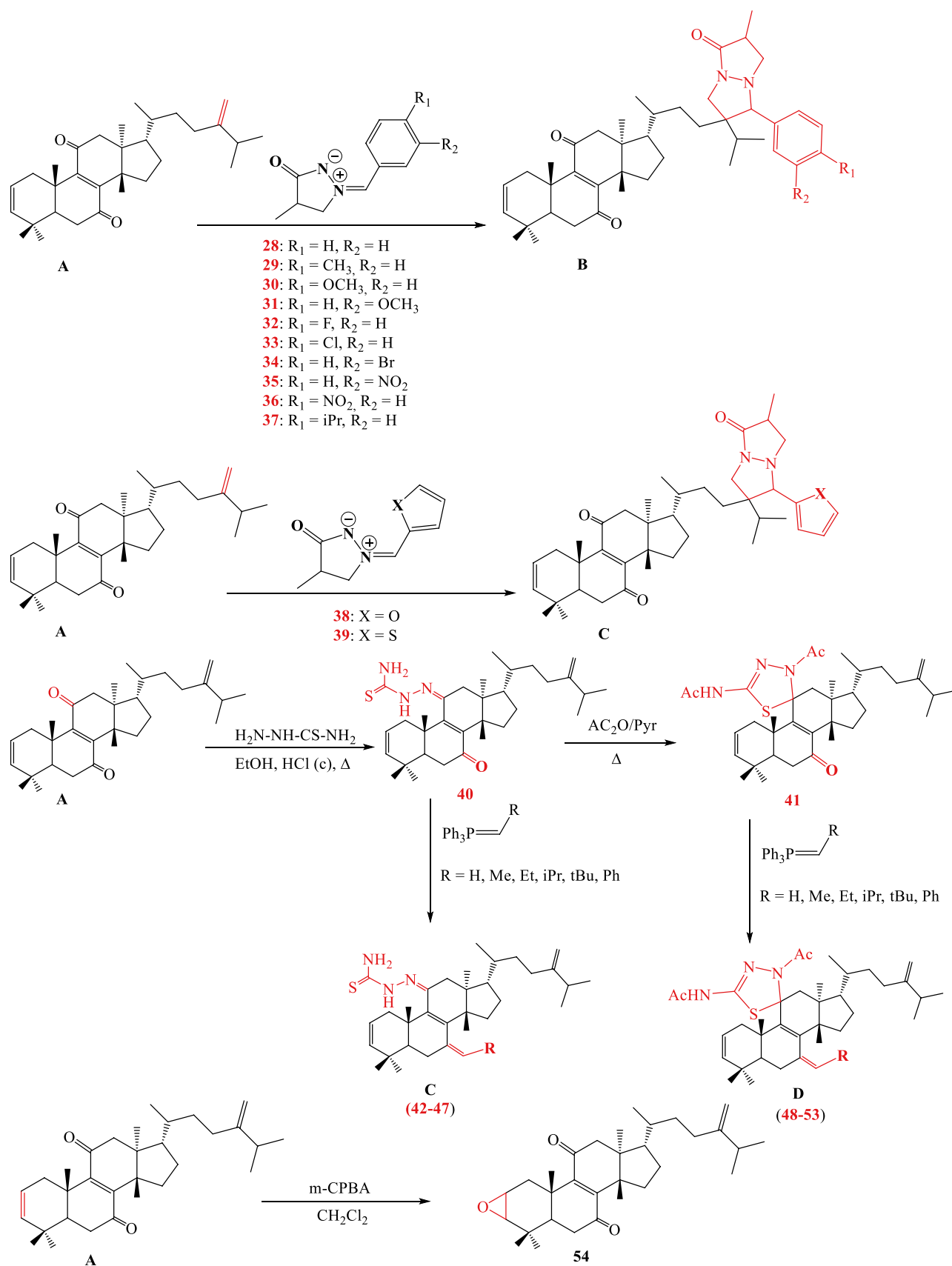
Based on the 3D-QSAR study, we have designed thirty-eight (28–65) new derivatives of 7-oxo-dehydroepiandrosterone by modifying the chemical structure of the reference compound 16. Modification will be introduced based on our expertise in the synthetic chemistry, according to the reactions proposed in schemes 1 and 2.

In the present study, we applied the 3D-QSAR models to predict the pLD₅₀ activity of the proposed new compounds. The proposed new molecules and their predicted pLD₅₀ activities are presented in Table 5.

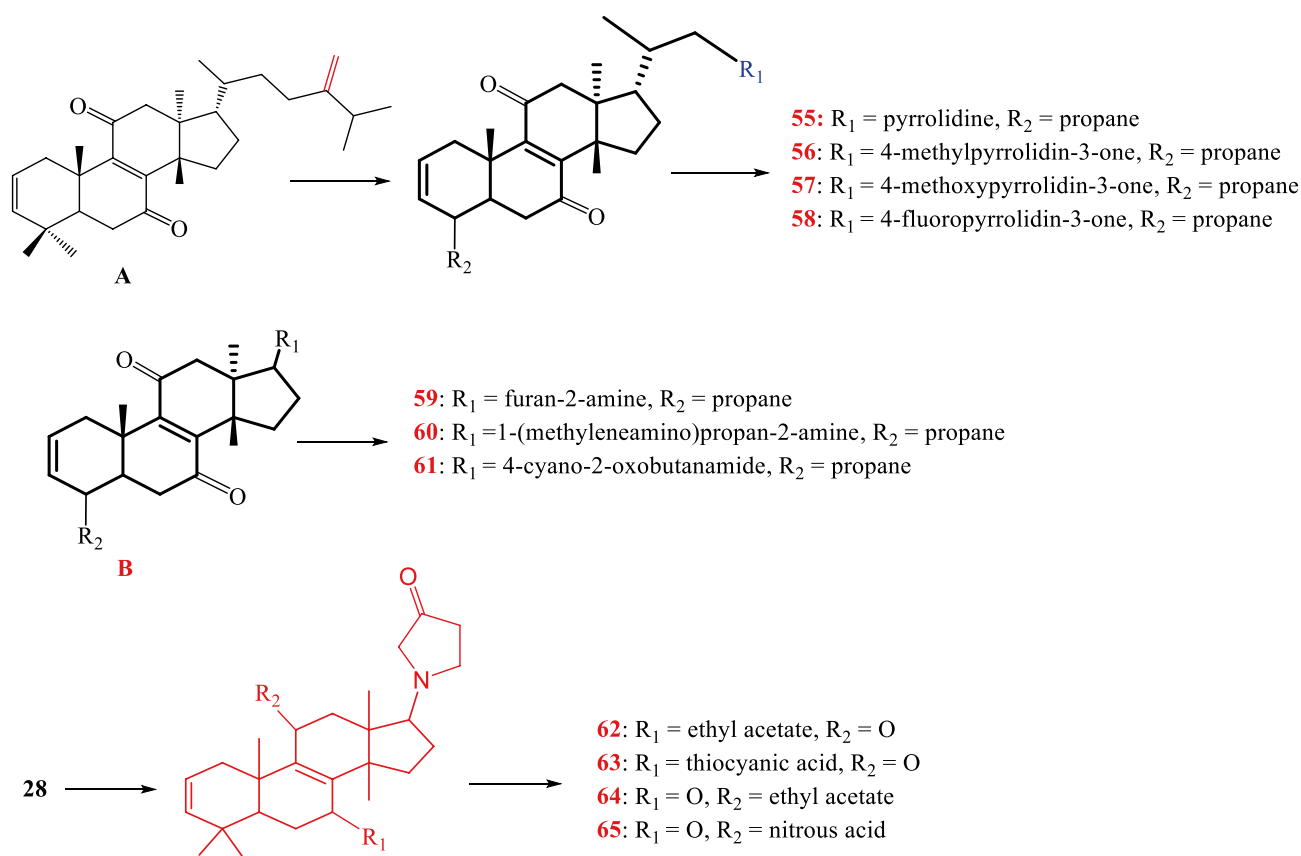
The results of the predictions obtained for pLD₅₀ through the 3D-QSAR models (Table 5) show that the newly designed compounds based on triterpene derivatives have a great ability to kill the *Mythimna separata* insect with the median lethal dose LD₅₀ at the minimum concentration. These designed molecules could be used in the future as new insecticidal agents that are harmless to the environment. We can also test the pharmacokinetics of these molecules and the possibility of using them in the biomedical domain, for example, in the treatment of bacterial infections in humans. In the rest of this study, we first perform the prediction of the pharmacological properties of the new molecules that showed the highest values of pLD₅₀. In order to evaluate the possibility of using these molecules as antibacterial drugs, or that the pLD₅₀ activity at these molecules has a negative effect on the *in silico* pharmacological properties. Secondly, we have tested the possibility of using some designed molecules as insecticides by performing a molecular docking analysis.

ADME-Tox prediction and drug-like character

Evaluation of drug-like properties There are several potential therapeutic agents that do not make it through clinical trials due to their unfavorable absorption, distribution, metabolism, and elimination (ADME) characteristics [46]. To this end, we investigate these properties *in silico* of new molecules designed to test their potential as new drugs. ADME is the most recent method used to find molecules that could drug candidates, which must fit Lipinski et al. [47], Veber et al. [48], and Igan et al. rules [49]. The study of similarity with drugs consists of identifying the properties of a molecule, whether or not it is a candidate for use as a drug antibacterial. In order to select the most candidate molecules for drug use, in a rigorous way, other significant properties such as total polar area (TPSA) and number of rotational



Scheme 1 Proposed synthetic pathways of the designed compounds



Scheme 2 Proposed synthetic pathway of designed compounds 55–65

bonds and molar refractivity were also determined. We performed an evaluation of the pharmacokinetic properties of 17 molecules selected from the 38 newly designed molecules. Molecules that presented high predicted pLD50 values are examined. The evaluation of the drug-likeness properties was performed using SwissADME online server [50], and the obtained outcomes are presented in Table 6.

According to the results in Table 6, all the selected molecules from the designed data set fulfill with all the required rules (“Lipinski’s rules,” “Veber’s rules,” and “Egan’s rules”). They indicate that there is no problem with the oral bioavailability of these compounds, except for the molecules 50 and 53 which did not satisfy Egan’s rules. The results also showed that the molecules (48, 49, 54, 55, 56, 57, 58, 59, 61, 64, and 65) have high absorption capacity, while the molecules 44, 45, 50, 51, 52, and 53 have low absorption capacity. In addition, the synthetic accessibility values for all the proposed compounds are less than 10, which means that these molecules can be easily synthesized. Also, for oral bioavailability which is the partial range of the drug

dose that finally reaches the therapeutic site, quantitatively denoted by % F, the acceptable degree of probability is 55%, which means that the molecule exceeds the five rules have been successfully met. All the proposed molecules obtained a score of 55%, indicating a good bioavailability. Activity artifacts in assays are a major problem for biological screening and medicinal chemistry. These artifacts are often caused by the formation of aggregates or the reactivity of the studied compounds under assay conditions. Several assay disrupting compounds (PAINS) have been identified as potential causes of erroneous or positive results [51]. The PAINS violations for the proposed compounds are shown in Table 6. All compounds showed zero PAINS alert and can be used as reference compounds.

ADME-Tox properties predictions To predict the in silico ADME-Tox properties of the selected molecules (48, 49, 54, 55, 56, 57, 58, 59, 61, 64, and 65) for pharmacological interest, we have used the online tools pkCSM [52] and SwissADME [53]. The results are reported in Table 7.

Table 5 The predicted activities of the newly designed compounds

| N° | Predicted activity (pLD ₅₀) | |
|----|---|--------|
| | CoMFA | CoMSIA |
| 28 | 4.421 | 4.588 |
| 29 | 4.421 | 4.588 |
| 30 | 4.425 | 4.591 |
| 31 | 4.423 | 4.591 |
| 32 | 4.419 | 4.591 |
| 33 | 4.420 | 4.590 |
| 34 | 4.420 | 4.587 |
| 35 | 4.416 | 4.592 |
| 36 | 4.413 | 4.591 |
| 37 | 4.422 | 4.588 |
| 38 | 4.428 | 4.594 |
| 39 | 4.433 | 4.595 |
| 40 | 3.897 | 4.199 |
| 41 | 4.592 | 4.611 |
| 42 | 4.572 | 4.245 |
| 43 | 4.383 | 4.814 |
| 44 | 5.618 | 5.244 |
| 45 | 5.588 | 4.226 |
| 46 | 4.570 | 4.242 |
| 47 | 4.039 | 4.059 |
| 48 | 5.919 | 5.697 |
| 49 | 5.021 | 5.678 |
| 50 | 6.017 | 5.989 |
| 51 | 6.014 | 6.003 |
| 52 | 5.877 | 5.892 |
| 53 | 4.929 | 5.591 |
| 54 | 4.806 | 5.688 |
| 55 | 6.307 | 5.911 |
| 56 | 6.277 | 6.013 |
| 57 | 5.872 | 5.577 |
| 58 | 5.483 | 5.136 |
| 59 | 6.234 | 5.974 |
| 60 | 4.631 | 5.317 |
| 61 | 5.589 | 5.073 |
| 62 | 4.478 | 4.955 |
| 63 | 4.308 | 4.725 |
| 64 | 5.888 | 6.167 |
| 65 | 5.638 | 5.914 |

Based on the results depicted in Table 7, we can conclude that:

- All molecules have shown a high capacity of absorption in the intestine, where absorption is considered satisfactory when it exceeds 30% [52].
- In terms of distribution indicators, the volume of distribution (VD_{ss}) is considered high if its value is higher than 0.45 [54]. The standard value for blood–brain barrier permeability (BBB) is considered good if its value

is greater than 0.3 and poor if $\text{LogBB} < -1$. For the central nervous system (CNS) index, compounds with $\text{LogPS} > -2$ are reputed capable of entering the CNS, whereas compounds with $\text{LogPS} < -3$ are considered unable to enter the CNS [52]. The results obtained indicate that most of the selected compounds have shown the ability to cross the barriers, except for molecule **48** which showed a weak ability to cross the BBB.

- In terms of metabolism, cytochrome P450 (CYP) is an important enzyme for detoxification. CYP enzymes are present in all tissues of the body [55]. This enzyme oxidizes foreign microorganisms to facilitate their excretion. Many drugs are inhibited by the CYP cytochrome, and some can also be activated by it. Inhibitors of this enzyme can disrupt the metabolism of the drug, which may have an opposite effect to that desired. Studies on the ability of compounds to inhibit cytochrome P450 (CYP) enzymes play an important role in determining drug interactions and toxicity. The two isoforms of CYP (2D6 and 3A4) are primarily responsible for drug metabolism [52]. We found that all the designed molecules can be substrates of CYP3A4. In addition, molecules **55**, **56**, **59**, and **64** can be considered as both substrates and inhibitors of CYP3A4.
- Regarding the drug clearance index, which is important in determining drug doses to achieve stability of drug concentrations [52]. In the lower TCL value, it is likely that the drug will be more stable in the body and will reach the therapeutic target before excreted. The results of the prediction of this indicator indicate that the total clearance index of all newly designed molecules is less than 0.5 ($\text{log mL min}^{-1} \text{kg}^{-1}$), thus increasing the possibility that a dose of these molecules will reach the therapeutic target.
- Regarding the toxicity indicator, it is essential to check up whether the predicted compounds are not toxic, as this is important for the selection of drug candidates. The AMES test was used in this study, and this test is widely used to assess the toxicity of compounds [56]. To the extent that all compounds in the database are toxic, it is fortunate that not all proposed molecules are toxic. Table 7 shows that the LD₅₀ values of the new designed particles are between 2.033 and 2.98 (mol/Kg). This indicates that the compounds are only lethal at very high doses as predicted by the oral rat acute toxicity test. The results of the ADME-Tox descriptors predictions obtained for the selected designed compounds have shown good pharmacokinetic properties. Thus, the designed molecules can be proposed in the development of new drugs for pharmacological use as antibacterials.

Relying on the aforementioned, several molecules can be synthesized by making modifications on the structure of

Table 6 Drug-likeness properties of the selected compounds

| Entry | ABS | TPSA (A ²) | n-ROTB | MW | LogP | n-OHN acceptors | n-OHNH donors | Lipinski's violations | Veber Violations | Egan Violation | S.A score | %F | Pains alert |
|-----------|------|------------------------|--------|--------|------|-----------------|---------------|-----------------------|------------------|----------------|-------------|------|-------------|
| Rule | - | - | - | < 500 | ≤ 5 | < 10 | < 5 | ≤ 1 | ≤ 1 | ≤ 1 | 1 < SA < 10 | - | - |
| 44 | Low | 40.46 | 5 | 446.75 | 5.19 | 2 | 2 | Yes | Yes | Yes | 5.79 | 0.55 | 0 |
| 45 | Low | 20.23 | 5 | 412.69 | 5.05 | 1 | 1 | Yes | Yes | Yes | 6.40 | 0.55 | 0 |
| 48 | High | 51.21 | 5 | 440.66 | 4.19 | 3 | 0 | Yes | Yes | Yes | 5.95 | 0.55 | 0 |
| 49 | High | 60.44 | 5 | 456.66 | 4.22 | 4 | 0 | Yes | Yes | Yes | 5.85 | 0.55 | 0 |
| 50 | Low | 46.53 | 5 | 446.71 | 4.88 | 3 | 1 | Yes | Yes | No | 5.81 | 0.55 | 0 |
| 51 | Low | 26.30 | 5 | 426.67 | 4.85 | 2 | 0 | Yes | Yes | Yes | 6.00 | 0.55 | 0 |
| 52 | Low | 40.46 | 9 | 430.71 | 4.89 | 2 | 2 | Yes | Yes | Yes | 6.47 | 0.55 | 0 |
| 53 | Low | 46.53 | 10 | 458.72 | 5.45 | 3 | 1 | Yes | Yes | No | 6.28 | 0.55 | 0 |
| 54 | High | 59.06 | 5 | 474.72 | 4.94 | 4 | 1 | Yes | Yes | Yes | 5.96 | 0.55 | 0 |
| 55 | High | 54.45 | 1 | 409.56 | 3.36 | 4 | 0 | Yes | Yes | Yes | 5.07 | 0.55 | 0 |
| 56 | High | 54.45 | 1 | 423.59 | 3.63 | 4 | 0 | Yes | Yes | Yes | 5.27 | 0.55 | 0 |
| 57 | High | 63.68 | 2 | 439.59 | 3.32 | 5 | 0 | Yes | Yes | Yes | 5.26 | 0.55 | 0 |
| 58 | High | 54.45 | 1 | 427.55 | 3.02 | 5 | 0 | Yes | Yes | Yes | 5.31 | 0.55 | 0 |
| 59 | High | 59.64 | 1 | 405.53 | 3.51 | 4 | 0 | Yes | Yes | Yes | 5.72 | 0.55 | 0 |
| 61 | High | 63.68 | 4 | 467.64 | 3.88 | 5 | 0 | Yes | Yes | Yes | 5.73 | 0.55 | 0 |
| 64 | High | 63.68 | 4 | 467.64 | 4.02 | 5 | 0 | Yes | Yes | Yes | 5.73 | 0.55 | 0 |
| 65 | High | 76.04 | 3 | 440.58 | 3.89 | 6 | 0 | Yes | Yes | Yes | 5.47 | 0.55 | 0 |

ABS absorption, TPSA topological polar surface area, n-ROTB number of rotatable bonds, MW molecular weight, MLog P logarithm of partition coefficient of compound between n-octanol and water, n-ON acceptors number of hydrogen bond acceptors, n-OHNH donors number of hydrogen bonds donors, S.A synthetic accessibility, %F bioavailability score

the reference molecule **16** using our expertise in organic synthesis and could be tested to evaluate their biological activity towards different biological targets. In the next section, we will perform a molecular docking assay to study the potential of molecules **16**, **55**, **56**, **59**, and **64** to inhibit one of the most important pathways in bacterial diseases as an example. Thus, through this test, we will discover the most important binding modes that can take place between ligand **16** and the active sites of the target receptors (MurE). The development of new classes of antibacterial agents against carefully selected targets is a high priority task. Through the study of molecular docking, it will help us to identify the type of interactions that will occur between the semi-synthesized ligand **16** and the designed molecules (**55**, **56**, **59**, and **64**) with biological targets that contribute to inhibit bacterial growth.

Molecular docking tests

Bacterial growth inhibitors For this purpose, we have performed molecular docking between ligands **16**, **55**, **56**, **59**, and **64** with the receptor MurE using AutoDock Vina software [57]. As a first step, we removed all water molecules as well as other non-protein elements from the structure of the unbound brute MurE protein. The obtained results are presented in Table 8. The crystal structure of the MurE receptor

is presented in Fig. 8. The modes of interaction obtained for ligands **16**, **55**, **56**, **59**, and **64** are presented in Fig. 9.

The results embedded in Table 8, show that all designed compounds have binding affinity values between -7.9 and -10.0 kcal/mol, while the binding affinity value obtained for the reference molecule **16** is 8.1 kcal/mol. This confirms that the majority of the designed molecules **55**, **56**, and **59** are more stable inside the MurE protein pocket compared to the molecule **16**, except the molecule **64** which is less stable (-7.9 kcal/mol). It is also noted that all the selected designed molecules establish hydrogen bonding and hydrophobic interactions with the MurE receptor. That means that the association of molecules **55**, **56**, **59**, and **64** as well as the molecule **16** with the MurE receptor will change the state of the target protein into a functional state and then trigger a chain reaction that leads to the inhibition of cell lines that cause bacterial growth. From the molecular docking analysis, it is appeared clearly that the designed molecules showed significant binding modes to the MurE receptor, and it confirms the effectiveness of the conducted 3D-QSAR study to determine the sites in the structure of the molecules that control the antibacterial activity.

Inhibitors of insect proliferation “insect Mythemna separata model” The study of the inhibition pathway of the ecdysone

Table 7 In silico predicted ADME properties for the selected compounds

| Properties | | Metabolism | | | | | | | | | | Excretion | | Toxicity | | |
|------------------|-------------------------------|--------------------------|------------------|------------------|----------------------|------------------|-----|-----|------|-----------|-----|-----------|--|----------------------|---------------|---|
| models | Intestinal absorption (human) | Distribution | | BBB permeability | | CNS permeability | | CYP | | Substrate | | Inhibitor | | Total clearance | AMES toxicity | Oral rat acute toxicity (LD ₅₀) |
| | | VD _{ss} (human) | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | |
| Unity | Numeric (%absorbed) | Numeric (Log L) | Numeric (Log BB) | Numeric (Log PS) | Categorical (yes/no) | 2D6 | 3A4 | 1A2 | 2C19 | 2C9 | 2D6 | 3A4 | Numeric (log mL min ⁻¹ kg ⁻¹) | Categorical (yes/no) | | |
| Predicted values | | | | | | | | | | | | | | | | |
| 48 | 99.454 | 0.528 | -2.654 | -1.512 | No | Yes | No | No | No | Yes | No | No | 0.348 | No | | 2.039 |
| 49 | 99.022 | -0.013 | -0.523 | -1.51 | No | Yes | No | No | No | No | No | No | 0.342 | No | | 2.138 |
| 54 | 94.173 | 0.084 | -0.339 | -1.642 | No | Yes | No | No | No | No | No | No | 0.378 | No | | 2.102 |
| 55 | 97.063 | 0.569 | 0.036 | -2.768 | No | Yes | No | No | No | No | No | Yes | 0.417 | No | | 2.754 |
| 56 | 97.302 | 0.584 | -0.079 | -2.745 | No | Yes | No | No | No | No | No | Yes | 0.312 | No | | 2.715 |
| 57 | 97.998 | 0.469 | -0.248 | -2.808 | No | Yes | No | No | No | No | No | No | 0.385 | No | | 2.898 |
| 58 | 96.803 | 0.494 | -0.125 | -2.783 | No | Yes | No | No | No | No | No | No | 0.341 | No | | 2.98 |
| 59 | 96.909 | 0.528 | 0.223 | -2.654 | No | Yes | No | No | No | Yes | No | Yes | 0.024 | No | | 2.522 |
| 61 | 97.61 | 0.44 | -0.308 | -2.792 | No | Yes | No | No | No | No | No | No | 0.264 | No | | 2.884 |
| 64 | 97.067 | 0.426 | -0.413 | -2.78 | No | Yes | No | No | No | No | No | Yes | 0.263 | No | | 3.033 |
| 65 | 94.991 | 0.468 | -0.227 | -2.852 | No | Yes | No | No | No | No | No | No | 0.021 | No | | 2.762 |

Table 8 Interaction modes of ligands **16**, **55**, **56**, **59**, and **64** with the receptor sites

| Ligands | Receptor | Binding affinity (kcal/mol) | Hydrogen-binding interaction | Hydrophobic interaction |
|---------|----------|-----------------------------|------------------------------|--|
| 16 | MurE | −8.1 | Gln88 | Pro361, Leu396, Lys67, Ile400, Ala399 |
| 55 | | −10.0 | Asn265, His214, Tyr213 | His263, His219, Ala223, Leu226, Val189 |
| 56 | | −8.6 | Lys366, Gln88 | Ala66, Lys67 |
| 59 | | −10.4 | Asn265, His214 | His219, Ala223, Val189, Leu226, Tyr213 |
| 64 | | −7.9 | Asn90, Glu91, Gln88 | His263, His43 |

receptor (EcR) by the triterpene derivatives was carried out using the following steps. First, we have re-docked the ligand Tebufenozide with the EcR receptor in order to identify the active sites in the EcR protein pocket, which serve as reference sites for EcR inhibitory activity. Moreover, the re-docking procedure allows us to validate the efficiency of the docking molecular protocol carry out via AutoDock Vina in this work. Second, we have docked the ligands **16**, **55**, **56**, and **59** into the EcR protein pocket and compared the binding energies obtained in each complex and also compared the number of binding interactions obtained between each ligand and the novel active sites predicted in the inhibition activity of EcR. In this study, we consider novel, more stable ligand-interacting active sites in the EcR receptor pocket as the most important reference sites in the inhibition of EcR protein activity.

The docking protocol is validated by re-docking the native ligand (Tebufenozide) into the EcR receptor pocket. The root mean square deviation (RMSD) between the original and re-docked ligand is acceptable in the range of 2 Å [58]. The active sites with which the co-crystallized ligand interacts in the EcR complex before this re-docking are Val416, Met381, Met380, Trp526, Tyr403, Thr343,

Leu420, Ile339, Cys508, Leu511, Tyr408, Met507, and Asn504. Figure 10 shows 3D and 2D visualizations of the EcR protein complexed with the inhibitor Tebufenozide (green color) in chain D of the protein structure (purple) and the interactions between the ligand Tebufenozide and the active amino acid residues in the EcR pocket (PDB code 1R20).

Figure 11 presents a superimposed view of the re-docked conformation (red color) and the original ligand (green color), and the RMSD value between them is 1.97 Å. The clear superimposed between both ligands and also the RMSD value less than 2 indicates the efficiency of the AutoDock Vina algorithms to perform molecular docking protocol with confidence. Thus, we can dock ligands **16**, **55**, **56**, and **59** inside the EcR receptor pocket with the AutoDock Vina software.

From Fig. 12 of the 3D visualization, we can notice that there is a correspondence in the interactions made between the most important active sites of the EcR protein and the original and re-docked ligands.

Figure 13 shows the predictions of the pose docking of ligands **16**, **55**, **56**, and **59** in the EcR receptor pocket and the most important interactions that occurred between these ligands and the reference sites by which the EcR activity responsible for the proliferation of insect *Mythimna separata* is inhibited.

We have summarized the interactions of ligands **16**, **55**, **56**, and **59** with the novel active sites identified by the molecular docking procedure in this work in Table 9.

The outcomes listed in Table 9 indicate clearly that the binding energy between the four studied ligands is close to each other, between −7.9 and −9 kcal/mol. Thus, we consider that the stability of the four molecules in the EcR receptor pocket to be good and close. Table 9 gather the most important active sites that contribute to the inhibition of EcR activity and to the inhibition of the growth of *Mythimna separata*. Finally, in the future, we will be able to complete this study by carrying out in vitro tests to confirm these results, and this study can be used to develop new agents in the fight against insect harmful to the environment, based on the studied triterpene derivatives.

**Fig. 8** The crystal structure of the MurE protein

Fig. 9 2D and 3D docking poses interactions between compounds (X-A), 16; (X-B), 55; (X-C), 56; (X-D), 59; and (X-E), 66 to MurE active sites

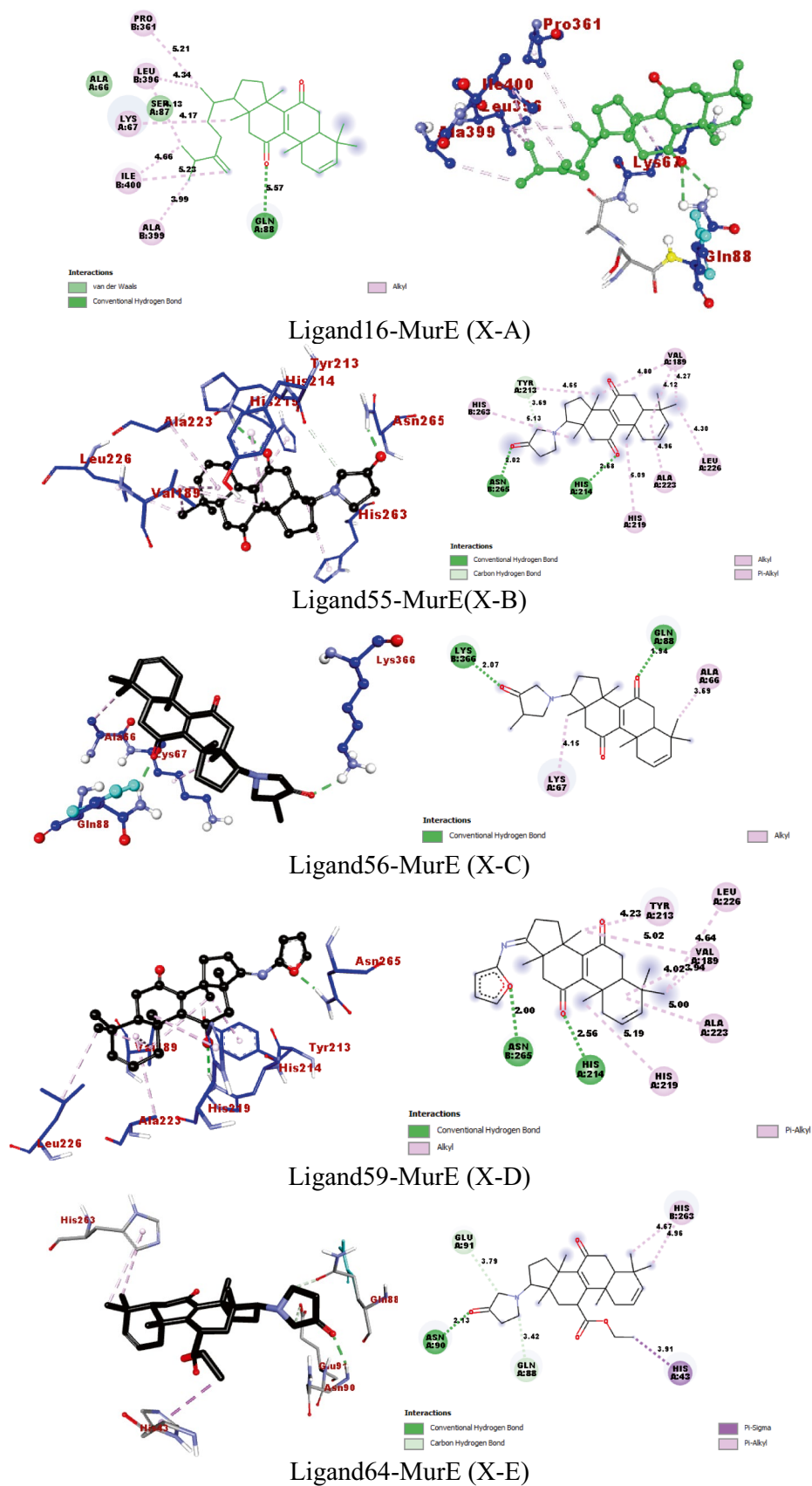


Fig. 10 **A** Original model of EcR protein, **B** and **C** 2D and 3D visualizations of tebufenozide's ligand interactions

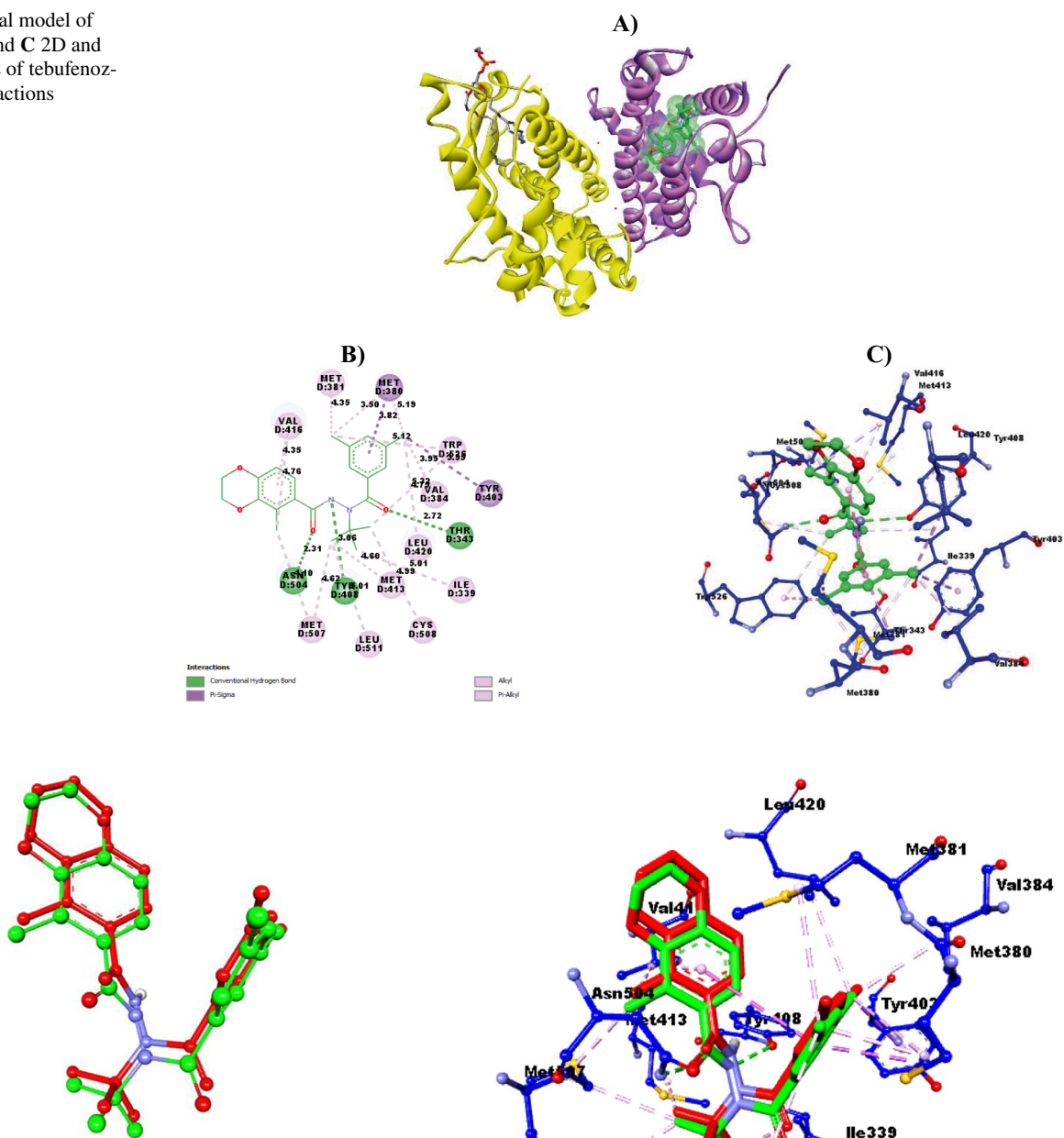


Fig. 11 Re-docking pose with an RMSD value of 1.97 Å (green = original, red = docked)

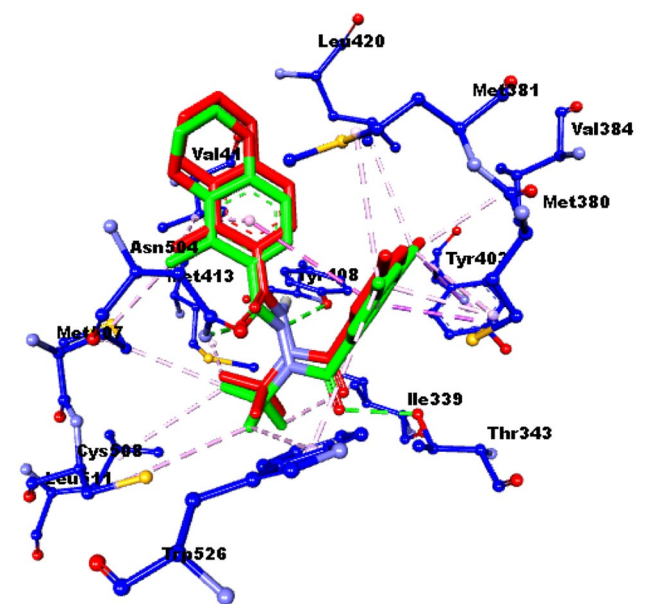
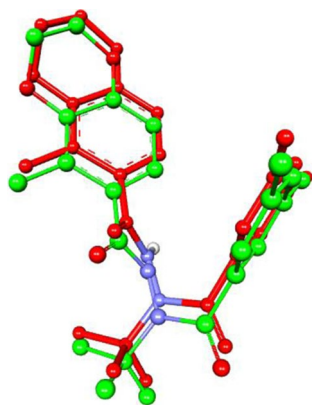


Fig. 12 3D visualization comparison between ligand pose prediction and the crystallographic ligand pose for EcR

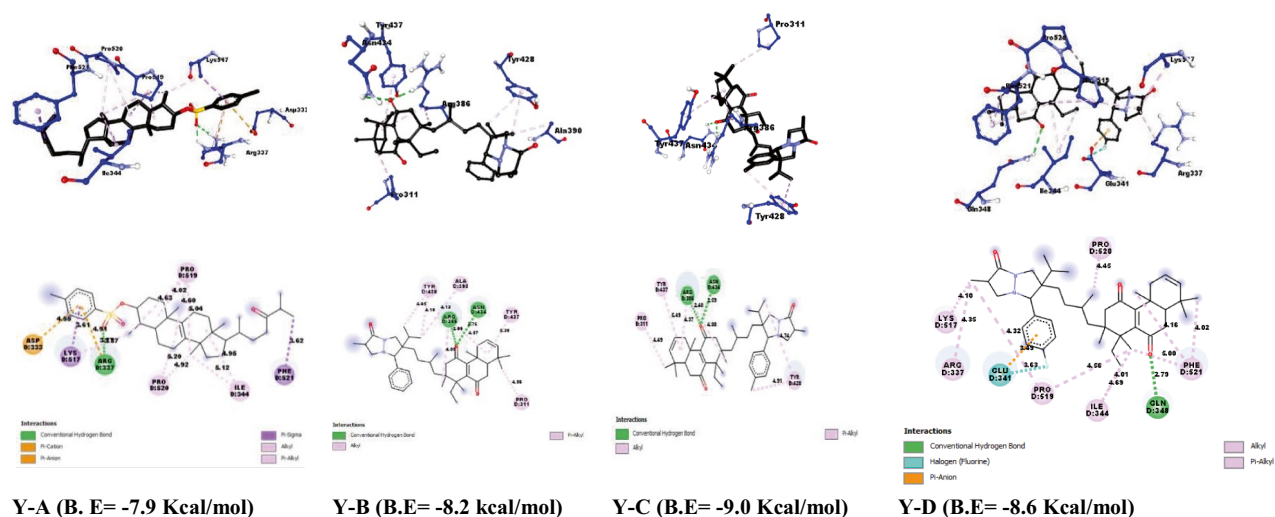


Fig. 13 Interactions of ligands **16** (Y-A), **55** (Y-B), **56** (Y-C), and **59** (Y-D) with the most important active sites in the EcR receptor pocket

Table 9 Docking results of ligands **16**, **55**, **56**, and ligand **59** at receptor novel sites

| Ligands | Binding affinity (kcal/mol) | Hydrogen-binding interaction | Electrostatic interaction | Hydrophobic interaction |
|---------|-----------------------------|------------------------------|---------------------------|--|
| 16 | -7.9 | Arg373 | Asp33, Arg373 | Lys517, Pro520, Ile334, Phe521, Pro519 |
| 55 | -8.2 | Arg386, Asn434 | - | Tyr428, Ala390, Tyr437 and Pro311 |
| 56 | -9.0 | Arg386, Asn434 | - | Pro311, Tyr437, Tyr428 |
| 59 | -8.6 | Gln348 | Glu341 | Lys517, Arg337, Pro519, Ile344, Phe521, Pro520 |

Conclusion

The aim of this study is to improve the antibacterial and insecticidal properties of the studied triterpene derivatives based on the structural modifications. For this reason, we have performed an *in silico* study based on 3D molecular modeling techniques applied on 27 semisynthetic triterpene derivatives. Thus, 3D-QSAR models were developed based on CoMFA and CoMSIA techniques and were checked for their performance by internal and external validations as well as Y-randomization test. Based on the obtained outcomes, we have identified the most favorable structural properties to improve the antibacterial and insecticidal activities of the semi-synthesized triterpene derivatives. In fact, the obtained contour maps, allow us to design 38 new derivatives and to predict their pLD₅₀ against bacteria and insects. Among the 38 designed derivatives, 17 molecules have showed high and promising predicted pLD₅₀.

The 17 new designed molecules were screened *in silico* to predict drug-like and ADME-Tox properties. This was done to evaluate the possibility of using these proposed molecules as promising antibacterial drugs in complement of their use as insecticides. Through the obtained ADME-Tox predictions, the newly designed molecules **55**, **56**, **59**,

and **64** are identified to be favorable candidates for anti-bacterial drugs. In addition, the AMES test has showed the safety of these molecules as oral drugs. Also, the acute toxicity index (LD₅₀) oral rat predictions indicate that the newly designed molecules become toxic above the experimental and predicted LD₅₀ values. This confirms the safety of the four designed molecules as promising antibacterial drugs.

Furthermore, the molecular docking test has showed the most important interactions that occur between the four proposed (**55**, **56**, **59**, and **64**) and a reference molecule **16** with the targeted proteins (MurE and EcR). Hence, we have predicted the reference sites that can be targeted to inhibit the MurE protein (antibacterial activity) and to inhibit the growth and the reproduction of insects by targeting the EcR protein. Rely on the molecular docking test, we could also notice that the designed molecules **55**, **56**, and **59** are more stable with the targeted proteins pocket than the reference molecule **16**. This was verified by the comparison of the binding energies between ligands and receptors into the obtained complexes. Finally, the obtained outcomes in this study indicate clearly that the studied triterpene derivatives have a great chance to become potential antibacterial and insecticidal agents.

Author contribution Ossama Daoui, writing, original draft preparation, conceptualization, and methodology; Noureddine Mazoir: data curation, methodology, writing, original draft preparation, and supervision; Mohamed Bakhouch, validation, visualization, and original draft preparation; Mohammed Salah, visualization; Ahmed Benharref, visualization; Azucena Gonzalez-Colouma, visualization; Souad Elkhatabi, visualization and supervision; Mohamed El Yazidi, visualization and supervision; Samir Chtita, visualization, investigation, validation, conceptualization, methodology, and software.

Data availability All data used in this work are private.

Code availability The codes used in this work are not available.

Declarations

Conflict of interest The authors declare no competing interests.

References

- Ya'u Ibrahim Z, Uzairu A, Shallangwa G, Abechi S (2020) Molecular docking studies, drug-likeness and in-silico ADMET prediction of some novel β -amino alcohol grafted 1,4,5-trisubstituted 1,2,3-triazoles derivatives as elevators of P53 protein levels. *Sci Afr* 10:e00570
- Fouedjou RT, Chtita S, Bakhouch M, Belaidi S, Ouassaf M, Djoumbissie LA, Tapondjou LA, Qais FA (2021) Cameroonian medicinal plants as potential candidates of SARS-CoV-2 inhibitors. *J Biomol Struct Dyn* 1–15. <https://doi.org/10.1080/07391102.2021.1914170>
- Belhassan A, Zaki H, Chtita S, Alaqrbeh M, Alsakhen N, Benlyas M, Lakhlifi T, Bouachrine M (2021) Camphor, artemisinin and sumac phytochemicals as inhibitors against COVID-19: computational approach. *Comput Biol Med* 136:104758
- Ouassaf M, Belaidi S, Chtita S, Aouidate A, Lanez T, Abul Qais F (2021) Combined molecular docking and dynamics simulations studies of natural compounds as potent inhibitors against SARS-CoV-2 main protease. *J Biomol Struct Dyn*. <https://doi.org/10.1080/07391102.2021.1957712>
- Chtita S, Belhassan A, Bakhouch M, Taourati AI, Aouidate A, Belaidi S, Moutaabbid M, Belaouad S, Bouachrine M, Lakhlifi T (2021) QSAR study of unsymmetrical aromatic disulfides as potent avian SARS-CoV main protease inhibitors using quantum chemical descriptors and statistical methods. *Chemom Intell Lab Syst* 210(15):104266
- Chtita S, Belhassan A, Aouidate A, Belaidi S, Bouachrine M, Lakhlifi T (2021) Discovery of potent SARS-CoV-2 inhibitors from approved antiviral drugs via docking and virtual screening, combinatorial chemistry & high throughput screening 24(3):441–454
- Belhassan A, Zaki H, Chtita S, Lakhlifi T, Bouachrine M (2020) Molecular docking analysis of N-substituted Oseltamivir derivatives with the SARS-CoV-2 main protease. *Bioinformation* 16(5):404–410
- Marshall CW, Ray RE, Laos I, Riegel B (1957) 7-Keto Steroids. II. 1 Steroidal 3β -Hydroxy- Δ^5 -7-ones and Δ^3 , 5-7-Ones. *J Am Chem Soc* 79(23):6308–6313
- Reich IL, Reich HJ, Kneer N, Lardy H (2002) Ergosteroids V: preparation and biological activity of various D-ring derivatives in the 7-oxo-dehydroepiandrosterone series. *Steroids* 67(3–4):221–233
- Toume K et al (2012) Cycloartane triterpenes and ingol diterpenes isolated from *Euphorbia neriifolia* in a screening program for death-receptor expression-enhancing activity. *Planta Med* 78(12):1370–1377
- Sun Y, Gao L, Tang M, Feng B, Pei Y, Yasukawa K (2018) Triterpenoids from *Euphorbia maculata* and their anti-inflammatory effects. *Molecules* 23(9):2112
- Somani H, Malik S, Kumbhat S, Joshic V, Khunt RC, Parkasha V (2017) Isolation and characterisation of triterpenoids from bioactive fraction of latex of *Euphorbia caducifolia* Haines (Family Euphorbiaceae). *Chem Biol Interface* 7(4):236–244
- Yu C-X et al (2020) Eupulcherol A, a triterpenoid with a new carbon skeleton from *Euphorbia pulcherrima*, and its anti-Alzheimer's disease bioactivity. *Org Biomol Chem* 18(1):76–80
- Benharref A, Lavergne JP (1985) Triterpenes issus des latex des euphorbes cactoides marocaines, E Resinifera E *Echinus* 965–972
- Daoubi M, Benharref A, Hernandez-Galan R, Macías-Sánchez AJ, Collado IG (2004) Two novel steroids from *Euphorbia officinarum* latex. *Nat Prod Res* 18(2):177–181
- Daoubi M et al (2007) Isolation of new phenylacetylating derivatives that reactivate HIV-1 latency and a novel spirotriterpenoid from *Euphorbia officinarum* latex. *Bioorg Med Chem* 15(13):4577–4584
- Mazoir N, Benharref A, Bailén M, Reina M, González-Coloma A (2008) Bioactive triterpene derivatives from latex of two *Euphorbia* species. *Phytochemistry* 69(6):1328–1338
- Mazoir N, Benharref A, Vaca L, Reina M, González-Coloma A (2020) Optimization of insecticidal triterpene derivatives by biomimetic oxidations with hydrogen peroxide and iodosobenzene catalyzed by MnIII and FeIII porphyrin complexes. *Chem Biodivers* 17(9):e2000287
- Smaili A et al (2019) Semisynthetic triterpenes derived from *Euphorbia officinarum* as plant growth promoters and inducers of disease resistance. *J Plant Growth Regul* 38(1):262–272
- Smaili A et al (2017) Triterpene derivatives from *Euphorbia* enhance resistance against *Verticillium* wilt of tomato. *Phytochemistry* 135:169–180
- Bailen M, Daoubi M, Benharref A, Martinez-Diaz RA, Gonzalez-Coloma A (2016) New bioactive semi synthetic derivatives of 31-norlanostenol and obtusifoliol from *Euphorbia officinarum*. *Nat Prod Commun* 11:733–738
- Daoui O, Elkhatabi S, Chtita S, Elkhallabi R, Zgou H, Benjelloun AT (2021) QSAR, Molecular docking and ADMET properties in silico studies of novel 4, 5, 6, 7-tetrahydrobenzo [D]-thiazol-2-Yl derivatives derived from dimedone as potent anti-tumor agents through inhibition of C-Met receptor tyrosine kinase, *Heliyon* 7(7):e07463
- Sobhia ME, Bharatam PV (2005) Comparative molecular similarity indices analysis (CoMSIA) studies of 1,2-naphthoquinone derivatives as PTP1B inhibitors. *Bioorg Med Chem* 13:2331–2338
- Clark M, Cramer RD, Van Opdenbosch N (1989) Validation of the general purpose tripos 5.2 force field. *J Comput Chem* 10(8):982–1012
- Aparoy P, Suresh GK, Kumar Reddy K, Reddanna P (2011) CoMFA and CoMSIA studies on 5-hydroxyindole-3-carboxylate derivatives as 5-lipoxygenase inhibitors: generation of homology model and docking studies. *Bioorg Med Chem Lett* 21(1):456–462
- Golbraikh A, Tropsha A (2000) Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *Mol Divers* 5(4):231–243
- Cramer RD, Patterson DE, Bunce JD (1988) Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc* 110(18):5959–5967
- Klebe G, Abraham U, Mietzner T (1994) Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J Med Chem* 37(24):4130–4146
- Borisa A, Bhatt H (2015) 3D-QSAR (CoMFA, CoMFA-RG, CoMSIA) and molecular docking study of thienopyrimidine and thienopyridine derivatives to explore structural requirements for aurora-B kinase inhibition. *Eur J Pharm Sci* 79:1–12
- Xue CX, Cui SY, Liu MC, Hu ZD, Fan BT (2004) 3D QSAR studies on antimalarial alkoxylated and hydroxylated chalcones by CoMFA and CoMSIA. *Eur J Med Chem* 39(9):745–753

31. Zhu W et al (2005) QSAR analyses on ginkgolides and their analogues using CoMFA, CoMSIA, and HQSAR. *Bioorg Med Chem* 13(2):313–322
32. Cramer RD III, Bunce JD, Patterson DE, Frank IE (1988) Cross-validation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. *Quant Struct-Act Relatsh* 7(1):18–25
33. Wang L, Politis DN (2021) Asymptotic validity of bootstrap confidence intervals in nonparametric regression without an additive model. *Electronic Journal of Statistics* 15:392–426
34. Tetko IV, Tanchuk VYu, Villa AEP (2001) Prediction of n-octanol/water partition coefficients from PHYSPROP database using artificial neural networks and E-state indices. *J Chem Inf Comput Sci* 41(5):1407–1421
35. Roy K, Mitra I (2011) On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design. *Comb Chem High Throughput Screen* 14(6):450–474
36. Roy K, Ambure P, Aher RB (2017) How important is to detect systematic error in predictions and understand statistical applicability domain of QSAR models?. *Chemom Intell Lab Syst* 162:44–54
37. Rücker C, Rücker G, Meringer M (2007) y -Randomization and its variants in QSPR/QSAR. *J Chem Inf Model* 47:2345–2357
38. Ouassaf M, Belaidi S, Khamouli S, Belaidi H, Chtita S (2021) Combined 3D-QSAR and molecular docking analysis of thienopyrimidine derivatives as *Staphylococcus aureus* inhibitors. *Acta Chim Slov* 68(2):289–303
39. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31:455–461
40. Kouidmi I, Levesque RC, Paradis-Bleau C (2014) The biology of Mur ligases as an antibacterial target. *Mol Microbiol* 94(2):242–253
41. Lugtenberg EJJ, Van Schijndel-van Dam A (1972) Temperature-sensitive mutants of *Escherichia coli* K-12 with low activities of the l-alanine adding enzyme and the d-alanyl-d-alanine adding enzyme. *J Bacteriol* 110(1):35–40
42. Mello TRP et al (2014) Developmental regulation of ecdysone receptor (EcR) and EcR-controlled gene expression during pharate-adult development of honeybees (*Apis mellifera*). *Front Genet* 5:445
43. Gordon EJ, Chantala L, Dideberg O (2000) Structure of MurE the UDP-N-acetylmuramyl tripeptide synthetase from *E. coli*
44. Billas IML, Iwema T, Garnier JM, Mitschler A, Rochel N, Moras D (2003) Crystal structure of the ligand-binding domains of the heterodimer EcR/USP bound to the synthetic agonist BYI06830
45. Cao H, Zhang H, Zheng X, Gao D (2007) 3D QSAR studies on a series of potent and high selective inhibitors for three kinases of RTK family. *J Mol Graph Model* 26(1):236–245
46. Szakács G, Váradi A, Özvegy-Laczka C, Sarkadi B (2008) The role of ABC transporters in drug absorption, distribution, metabolism, excretion and toxicity (ADME-Tox). *Drug Discov Today* 13(9–10):379–393
47. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 23(1–3):3–25
48. Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, Kopple KD (2002) Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem* 45(12):2615–2623
49. Egan WJ, Merz KM, Baldwin JJ (2000) Prediction of drug absorption using multivariate statistics. *J Med Chem* 43(21):3867–3877
50. Daina A, Michielin O, Zoete V (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep* 7(1):1–13
51. El-Shoukrofy MS, Abd El Razik HA, AboulWafa OM, Bayad AE, El-Ashmawy IM (2019) Pyrazoles containing thiophene, thienopyrimidine and thienotriazolopyrimidine as COX-2 selective inhibitors: design, synthesis, in vivo anti-inflammatory activity, docking and in silico chemo-informatic studies. *Bioorganic Chem* 85:541–557
52. Pires DEV, Blundell TL, Ascher DB (2015) pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J Med Chem* 58(9):4066–4072
53. Kalantzi L, Goumas K, Kalioras V, Abrahamsson B, Dressman JB, Reppas C (2006) Characterization of the human upper gastrointestinal contents under conditions simulating bioavailability/bioequivalence studies. *Pharm Res* 23(1):165–176
54. Daina A, Zoete V (2016) A boiled-egg to predict gastrointestinal absorption and brain penetration of small molecules. *Chem Med Chem* 11(11):1117
55. Ferraz ERA et al (2011) Differential toxicity of disperse red 1 and disperse red 13 in the Ames test, HepG2 cytotoxicity assay, and *Daphnia* acute toxicity test. *Environ Toxicol* 26(5):489–497
56. Stead AG, Hasselblad V, Creason JP, Claxton L (1981) Modeling the Ames test, mutation research/environmental mutagenesis and related subjects 85:13–27
57. Seeliger D, de Groot BL (2010) Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J Comput Aided Mol Des* 24(5):417–422
58. Shivanika C, Kumar D, Ragunathan V, Tiwari P, Sumitha A (2020) Molecular docking, validation, dynamics simulations, and pharmacokinetic prediction of natural compounds against the SARS-CoV-2 main-protease. *J Biomol Struct Dyn*. <https://doi.org/10.1080/07391102.2020.1815584>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.