International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://www.theaspd.com/ijes.php

Exploring Molecular Docking And Quantitative Structure-Activity Relationships Of Oridonin Derivatives As Inhibitors Of Akt Protein Kinase

Maher Wassof[1], M.H. Fatemi [2]*, Zahra Pahlavan Yali[3]
[1,2,3]Chemometrics Laboratory, Department of Chemistry, University of Mazandaran, Babolsar, IRAN. maherwassof666@gmail.com[1], mhfatemi@umz.ac.ir[2], zpahlevanyali@yahoo.com[3]

Abstract
In this work hybrid molecular docking quantitative structure activity relationship (QSAR) methodology is used to modeling and predict the inhibitory activities of some Oridonin derivatives to ward kinase B(AKT) protein. Data set consist of inhibitory activities of 57 Oridonin derivatives (as EC_{50} in μM^{-1}), which can be used in treatment of type 2 leukemia. After docking of these derivative's to B protein (AKT), the most stable structure of ligands are chosen and frozen, to calculate molecular descriptors. In the next step prescreened of descriptors are done and stepwise feature selection methods used to select the most relevel descriptors. Then the selected descriptors are used to developing multiple linear regression (MLR) and support vector machine (SVM) models. The statistical parameters of these model are; R^2 of 0.62 and 0.80; SE of 1.44 and 1.33 for MLR, and for SVM models R^2 of 0.89 and 0.82; SE of 0.44 and 0.79 for training and test sets respectively. Comparison between these valves of and other statistics reveals the superiority of SVM over MLR models. In the next step virtual screening based on the lead derivatives is outperformed to identify new efficient candidate based on ADME properties and docking studies.
Keywords: *Kinase B Protein, Inhibitory constant, Molecular descriptor, Quantitative structure – activity relationship, Molecular docking, Cancer.*

1. INTRODUCTION

Leukemia is a heterogeneous group of hematologic (blood) malignancies that arise from the dysfunctional proliferation of developing leukocytes (white blood cells) [1]. According to the World Health Organization (WHO) reports in 2021, were an estimated 508796 people living with Leukemia. In the United States, with 61090 estimated new cases in that year, which making it the 10th most common cancer in the United States [2].The disease is classified as acute or chronic based on the rapidity of proliferation, and as microcytic, or lymphocytic based on the cell of origin. The environmental risk factors play an important role in the development of *Leukemia* [3]. Protein kinase B (PKB), which is also known as Akt, is the collective name of a set of three serine/threonine-specific protein kinases that play key roles in multiple cellular processes such as; glucose metabolism, apoptosis, cell proliferation, transcription, and cell migration [4].There are three different genes that encode isoforms of protein kinase B, which are: *AKT1*, *AKT2*, and *AKT3* that can encode the Ras-related C3 botulinum toxin substrate (RAC alphe), beta, and gamma serine/threonine protein kinases, respectively . The terms of PKB and Akt may refer to the products of all three genes collectively, but sometimes are used to refer to PKB alpha and Akt1 alone. Today treatment of *Leukemia* can be done by chemotropic or by using some PKB chemo-therapeutic inhibitor such as vincristine, prednisone or dexamethasone, asparaginase and methotrexate. Significant advancements in treatment of cancer have been made in the past decade through the identification of new compounds and the investigation of their underlying molecular mechanisms, with the goal of addressing human cancers and other diseases. Microbes, plants, and animals yield a wide variety of natural products with diverse structures and biological properties, offering researchers promising opportunities to create new molecular compounds for human treatments [5].Out of the 112 pioneering drugs approved by the FDA from 1999 to 2013 [6], 31 are derived from natural pharmacophores (28%). The important thing to mention is that the 2015 Nobel Prize in Physiology or

Medicine emphasized the important role of natural substances (like Artemisinin) in fighting serious parasitic infections [7].

One types of herbal extracted chemicals which can be used in treatment of *Leukemia* cancers are Ordines.The discovery of Oridonin in *Isodon* plants and its use as a traditional herbal, remedy has been demonstrated in China and Japan as anti-cancer agent [8]. Although it has shown potential in terms of safety and efficacy in cancer treatment, its relatively modest potency, limited solubility in water, low bioavailability, and unclear mechanisms of action have significantly hindered its further preclinical development and clinical applications [9].To overcome these limitations and develop improved drug candidates with enhanced activity, several derivatives of Oridonin have been created and synthesized. As a notable milestone the Oridonin –derived compound, L-alanine-(14-Oridonin) ester trifluoroacetate (2, HAO472) has recently progressed to a phase I clinical trial (CTR20150246): in China, conducted by Hengrui Medicine Co. Ltd, for the treatment of acute myelogenous Leukemia (https://seer.cancer.gov/statfacts/html/leuks.html). In this work, they try to provide a concise overview of the biological and pharmacological studies on Oridonin, as well as summarize the recent medicinal chemistry developments involving novel Oridonin analogues, with the objective of appreciating the therapeutic potential and value of the Oridonin scaffold as an exciting platform for drug discovery.

Substantial progress has been achieved in identifying and designing of new agents and conducting relevant molecular mechanistic studies for the treatment of human cancers and other diseases in the past decade [10]. Since developing of new drugs is a costly and time consuming process therefore developing of theoretical model to designing and modeling of new drug candidate is very important and necessary .One of these method is quantitative structure–activity relationship (QSAR) technique. In QSAR the structure features of molecules are correlated with their biological activities quantitively. In hybrid docking –QSAR these molecular features are calculated from molecular structures after docking the drugs candidate (ligand) to targets protein. By using the molecular docking method, it is possible to achieve a structure that is close to the real spatial arrangement of the ligand in binding the active site of the protein. Then, by these molecular features (molecular descriptor, combined hybrid QSAR models are created to predict and calculate the desired activity [9]. In the upcoming research, the effective interactions of some Oridonin derivatives as Akt inhibiters will be investigated using the molecular docking method and QSAR strategy. Then the results are used to introducing more effective drugs candidates [11].

2. MATERIAL AND METHODS

2.1. Data set

To date, hundreds of Oridonin derivatives have been synthesized in order to improve their solubilityies and bioavailabilityies as well as their inhibitory activities [12]. Some of them has shown improved efficacy against proliferation and fibrosis compared to Oridonin [13]. In this work data set consist of 57 Oridonin derivatives which their which their Akt effective concentration of were reported by [14]. The chemical structures of dataset are shown in Table 1. Furthermore their half maximal effective concentration (*EC50*) are indicated in Table 2. The values of *EC50* were ranged from 0.1 to 8.1μM^{-1} for compounds 1 and 47, respectively. In order to divide data set to tanning and test sets, all compounds were sorted according to their *EC50* values, and then the test set was chosen from this list by desired distance from each other. By using this procedure, 47 molecules were considered as the training set for model development and 10 compounds were selected as the test set for evaluating the predictability of develop the models (Table 2). In the next step, the chemical structure of the molecules were drawn by Hyperchem package (version 7.5) and optimized by employing molecular mechanics and semi-empirical (AM_1) methods. Then the optimized structures were converted from **.hin* format to **.pdb* using the Open Babel program and further transformed to *pdbqt* using the *PYRX* package to use as inputs for docking studies [15].

2.2. Molecular docking

Molecular docking is a computational technique used to predict the preferred structure and orientation of one molecule, typically a small ligand or drug, when it binds to a target protein or receptor [16] [17]. This

process helps to understanding the interactions between molecules and is widely used in drug discovery and design. The molecular docking computations were performed using Auto Dock 4.2 by using flexible ligand-rigid protein docking strategy [18]. Several X-ray crystal structures illustrating the binding of AKT with inhibitors are available in the Protein Data Bank (*PDB*). For molecular docking, the X-ray crystal structure of *PDK1* (*PDB ID: 1O6L*) was retrieved from the Uniporter protein database (http://www.uniprot.org) and utilized. The Cartesian coordinates for the docking boxes were set at 30 Å for each dimension (x, y, and z) [19]. Additionally, the coordinates for the box center were specified as -4.555, 60.241, and 109.222 for X, Y, and Z, respectively. At the end of calculation ten different modes were generated, and the optimal conformation of energy was selected for further analysis [20].

2.3. Descriptors calculation

In cheminformatics, a molecular descriptor is a quantitative measure derived from a systematic and standardized process that encoded various molecular structural information into a symbolic representation of a molecule [21]. These descriptors are instrumental in conveying a wide array of molecular properties and are essential for developing robust QSAR models. Quantitative descriptors are essential in characterizing various phenomena across multiple fields, including chemistry, biology, and materials science [22]. They provide measurable and objective data that can be analyzed statistically. In hybrid docking QSAR the molecular descriptors are calculated form optimal structures of interested chemicals after docking of them to target protein (Fig. 1) [23].

In this work the Dragon 7.0 software (https://vcclab.org/lab/edragon/) [24], was employed in conjunction with the ChemDes program (http://www.scbdd.com/chemdes/) and *PaDEll* to calculate the pool of descriptors based on the optimal three-dimensional structure of organic compounds obtained from docking studies. In order to do this the optimized docking structures of Oridonin derivatives, were converted from **.pdb* to**.hin* format. Then frozen structurers derived from docking studies were subjected to descriptor calculation. A comprehensive set of descriptors is generated, encompassing 1556 descriptors by *DRAGON*, 604 descriptors by Chemdes, and 409 descriptors from the *PaDEL* descriptor tool. These descriptors span a broad spectrum of molecular characteristics that can providing an extensive dataset for subsequent modeling efforts. Prior to variable selection, data were meticulously preprocessed to ensure quality and reliability. This step involved the elimination of constant variables, near-constant variables, and descriptors with zero values across all samples. To prevent multicollinearity issues, descriptors with Pearson correlation coefficients exceeding 0.90 were also removed. The remaining 504 descriptors are used in variable selection step.

2.4. Feature selection

Variable selection in QSAR is a critical step that aids in building robust and interpretable quantitative relationships between molecular structures and biological activities. Selecting the relevant variables is essential for constructing concise and interpretable models [25]. This process identifies key independent variables for predicting the dependent variable, which is particularly useful in scenarios with numerous predictors, emphasizing the most impactful variables for the response [23]. The forward addition method was employed using SPSS software (IBM SPSS Statistics 19) to identify the most significant descriptors. This iterative process continued until the addition of new descriptors no longer significantly improved the model's performance. As can be seen in Figure 2, adding more than four descriptors to model did not improve the model's predictive power significantly the breakpoint procedure. Therefore, four descriptors were selected to development of QSAR models [26].

2.5. Models development

QSAR models were developed using two primary methods; Multiple Linear Regression (MLR) and Support Vector Machine (SVM). The correlation coefficient (R) was subjected to statistical analysis to evaluate the performance of the goodness-of-fit for the model. Additionally, the determination coefficient (R^2) standard error (SE) and mean-square error (MSE) were computed for each model. These statically parameters are calculated from the following equations

$$\mathbf{R^2} = 1 - \sum_i \frac{(y_i - \hat{y}_i)^2}{(y_i - \bar{y}_i)^2} \quad (1)$$

$$\mathbf{MSE} = \frac{1}{n}\sum_{i=1}^{n}((y_i - \hat{y}_i)^2 \quad (2)$$

$$\mathbf{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}((y_i - \hat{y}_i)^2} \quad (3)$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and $\hat{\boldsymbol{y}}_{\boldsymbol{i}}$is the average value of *EC50*. The recommended values of these statistics to ensure the reliability of predictions from the QSAR model are; a correlation coefficient (R) ≥ 0.8, along with a coefficient of determination (R^2) ≥ 0.6 for in vivo data [27]. In order to developing these models selected molecular descriptors and *EC50* are considered as independent and dependent variables, respectively. Equations for predicting *EC50* values were derived from the training data using MLR and subsequently validated using the test data.

3. RESULTS AND DISCUSSION

3.1 Docking results

To simulate how Oridonin derivatives bind to the Akt protein's active site docking studies are performed (Fig. 3). Obtained results indicate that hydrogen bond interactions, both as acceptors and donors, emerged as the most lucrative ligand-protein interactions [28]. Following the docking analysis, compounds 14, 20, and 33 were identified as the top-performing compounds with binding constant values of -7.7, -7.4, and -7.4 μM^{-1}, respectively that indicate, their potential to consider as active compounds. Active compound 33, for instance, acts as a hydrogen bond acceptor from MCF-7 and forms two interactions with *MDA* and *MB-231* as a hydrogen bond donor. Interestingly, the representative compound 33 exhibits the highest potency as anticancer agent against *Leukemia MCF-7* cells with low micro molar to submicron molar *EC50* values, that indicating its potential to use for treatment of chemo resistant *Leukemia*. Conversely, compound 9, was categorized as inactive compound. Based on both docking results and the QSAR models results.

3.2. QSAR models

The dataset, consisting of 57 compounds, was divided into a training set of 47 compounds (80%) for model construction and a test set of 10 compounds (20%) for model validation, utilizing the Kennard-Stone algorithm. Initially, stepwise multiple linear regression (SW-MLR) was employed to develop the quantitative structure-activity relationships model using a distinct set of descriptors. The resulting correlation equation, which exhibits predictive capability for the training set, is as follows:

*EC50 =52.421- 18.327 * bcutv1+ 9.602 * GATSp4 - 3.523 * Smax8 +.008 * EstateVSA1+ 2.076 * Chiv4pc-0.003 * petitjeanShapeIndex.0 +9.602 * WHIM.4* **(eq. 4)**

n_{train} *=47;* R^2_{train} *= 0.72;* $RMSE_{train}$ *= 0.42;* n_{test}*=10;* R^2_{test} *=0.77;* $RMSE_{test}$ *= 0.82*

here, n_{train} is the number of compounds in the training set, R^2_{train} is the squared correlation coefficient for the training set, $RMSE_{train}$ is the root mean square error for the training set, n_{test} is the number of compounds in the test set, R^2_{test} is the squared correlation coefficient for the test set, $RMSE_{test}$ is the root mean square error for the test set. The details of statistical parameters of equation (4) are indicated in Table 3. The SVM model was implemented using *STATISTICA* software (Version 14.5.0.12). The values of SVM parameters are optimized by continuous changing of them and monitoring the error of model for training and test sets. Then the optimized SVM is used to period the values of *EC50* for training and test sets. The experimental and SVM predicted EC_{50} values and their corresponding residuals are shown in Table 4. The optimized SVMs parameters are, kernel functions RBF, C=100, γ=0.700, and ε=0.900, outperformed SVM (R^2=0.89, 0.82; SE =0.44, 0.79 for training and test sets, respectively, compared to R^2=, 0.62, 0.80, SE =1.44, 1.33 for MLR). Important statistical parameters for both SVM and MLR are shown in Table 5, which can be used to comparison the performance of these models. Compression between these parameters and those indicated in Table 5, reveals the superiority of SVM over MLR model. The SVM calculated values of EC50 for both training and test sets are plotted against their experimental values in Fig. 4(a), which reveals good correlation

between them and also their residuals in Fig. 4(b), random distribution of residuals around the zero line indicate that there is no any biases in developed SVM model

3.3. Interpretation of descriptor

In this study, the stepwise multiple linear regression (SW-MLR) method identified seven descriptors (Table 6) crucial for predicting EC_{50}, offering insights into potential drug discovery for *Leukemia* treatment. The first descriptor, is chiv4pc (mean simple molecular connectivity) which is a connectivity descriptor that encapsulates relevant information on a branch point, with an emphasizing on the adjacent branch, and specifies heteroatom and valence information. The biological effect of studied Akt inhibition can therefore be enhanced by raising this descriptor index.

The second descriptor, is smax8 (mean Maximum of E-State value of specified atom type), which is defined as the mean maximum E-State value of a specified atom type within a molecule. This means that for a given type of atom, it can be calculate the E-State values for all occurrences of that atom in the molecule to identify the maximum E-State value for each occurrence, and then compute the average of these maximum values. This measure can be useful in quantitative structure-activity relationship studies, where it helps in predicting the biological activity or other properties of chemical compounds based on their molecular structure.

The next descriptors is estateVSA1 (mean MOE-type descriptors using Estate indices and surface area contributions) which is a valuable descriptor in computational chemistry and combines the benefits of E-State indices with surface area considerations, providing a comprehensive view of molecular properties that can aid in various applications, including drug discovery and molecular design.

The forth descriptor is GATSp4 (Geary autocorrelation descriptors based on atomic polarizability) that can provides insights into how atomic polarizability is distributed throughout a molecule, which can influence its chemical behavior and interactions with other molecules and GATSp4 is a valuable tool in computational chemistry that leverages Geary autocorrelation and atomic polarizability to provide a deeper understanding of molecular properties. Applications of GATSp4 in predictive modelling and QSAR studies make it a significant descriptor for researchers in drug discovery and molecular design [29].

The next descriptor is WHIM4 (mean Weighted Holistic Invariant Molecular Descriptor) which is a molecular descriptor used in computational chemistry and cheminformatics to characterize the three-dimensional structure of molecules. It provides insights into molecular properties that are important for understanding chemical behavior and biological activity. WHIM 4 can be used to assess the similarity between different molecules, aiding in virtual screening and drug design by identifying potential candidates with similar properties.

The sixth descriptor is BCUT1 (mean stands for "Burden-Centric Unique Topological" descriptors). These descriptors focus on the topological features of a molecule, emphasizing the connectivity and arrangement of atoms'. These types of descriptors are valuable tools in cheminformatics for quantifying molecular properties and understanding the relationship between structure and activity. Their applications in molecular diversity analysis, QSAR modeling, and virtual screening make them essential for researchers in drug discovery and related fields .ref

The final descriptors is Petitjean Shape Index.0 (mean Petitjean Shape Index) (PSI) descriptor a specific molecular descriptor that is quantifies the shape of a molecule based on its three-dimensional structure. This descriptor is essential for understanding how the spatial arrangement of atoms influences the molecular properties and behaviors and is a powerful tool for quantifying the shape of molecules in cheminformatics.

3.4. Applicability domain analysis

The applicability domain (AD) of a QSAR model is critical for validating the model's predictions and ensuring their reliability. To define the AD, a Williams plot is employed, which displays standardized residuals versus leverage values (h). This visualization aids in identifying outliers and influential compounds and can providing insights into the robustness of the model. The leverage equation is calculated by: $h_i = x_i (X^T X)^{-1} x_i^T$, where x_i represents the descriptor vector for interested compound and X is the descriptor matrix derived from the training set. The warning leverage value (h^*) calculated is as follow:

$h^* = 3(d + 1)/n$ **(eq. 5)**

In this equation, d is the number of predictor variables, and n is the number of compounds in the training set. According to the above explanation Williams plots for both the MLR and SVM models were generated using a warning leverage value of h*= 0.45 that are shown in Figure 5. As can be seen in these figures, the majority of compounds fall within the applicability domain.

3.5. ADMET analysis

The pharmacokinetic parameters for the five identified hits were determined to fall within the acceptable range intended for human use which are shown in Table 7 and Fig 6. Highlighted by bold chemicals in are hits new candidate which indicating their potential as new drugs according to their pharmacokinetic and ADME results.

4. CONCLUSION

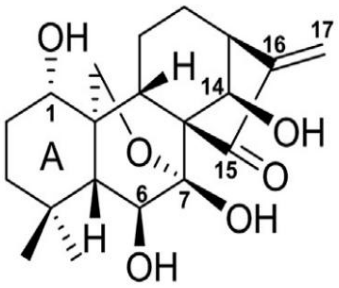
In this study some SVM and MLR models are developed based on molecular descriptors that are calculated from docking derived structures of interested Oridonin derivative's as B (AKT) kinase protein inhibitors. The models' predictive ability and robustness were assessed using various statistical parameters, such as RMSE and R^2. The MLR model yielded a clear output with an RMSE of 0.449. On the other hand, the SVM model**.** demonstrated more accurate predictions than the MLR model, with a $RMSE_{train}$ of 0.288 and a $RMSE_{test}$ of 0.790 for training and test sets respectively. These results indicate that the SVM method, when coupled with appropriate descriptors, can effectively predict the activity of new derivatives in the treatment of *Leukemia*. Analyzing of docking data and selected molecular descriptors indicate that steric and electronic interaction together with H-bond donner/acceptor ability of drugs candidate play important role on inhibitory activities (as EC50) of studied Oridonin derivatives. The visualization of the QSAR model and the docking mode into the target protein provided insights into the structure-activity relationship, offering explicit indications for designing improved Oridonin derivatives. Additionally, a virtual screening procedure was applied to a large commercial chemical database, resulting in 17 hits. These hits were further screened using the QSAR model for Akt inhibitory activity prediction, resulting in hits that were subsequently evaluated for their Absorption, Distribution, Metabolism, and Excretion (ADME) properties. The outcomes of this study provide valuable insights into the development of novel and potent Akt inhibitors, holding promise for the creation of new drugs for type 2 *Leukemia*.

REFERENCES

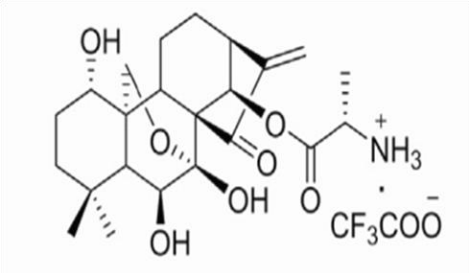
1. Anderson, et al. (1987). A line notation and computerized interpreter for chemical structures. *Report No. EPA/600/M-87/021*. U.S. Environmental Protection Agency, Environmental Research Laboratory-Duluth, Duluth, MN 55804.
2. Arber, D. A., et al. (2016). The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood, 127*(20), 2391-2405.
3. Bohanon, F. J., et al. (2015). Enhanced anti-fibrogenic effects of novel oridonin derivative CYD0692 in hepatic stellate cells. *Molecular and Cellular Biochemistry, 410*, 293-300.
4. Callaway, E., & Cyranoski, D. (2015). Anti-parasite drugs sweep Nobel Prize in medicine 2015. *Nature, 526*, 174-175.
5. Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis. (2020). *Applied Soft Computing Journal*.
6. Consonni, V., et al. (2002). Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 1. Theory of the novel 3D molecular descriptors. *Journal of Chemical Information and Computer Sciences, 42*(3), 682-692.
7. De Silva, C., et al. (2008). Inflammatory aortitis controlled by the Chinese herbal remedy Donglingcao Pian. *Rheumatology, 47*, 1257-1259.
8. Deininger, M., et al. (2017). Hematopoietic stem cell transplantation for chronic myeloid leukemia.
9. Deschler, B., & Lübbert, M. (2006). Acute myeloid leukemia: Epidemiology and etiology. *Cancer, 107*(9), 2099-2107.
10. Eberly, L. E. (2007). Multiple linear regression. In *Topics in Biostatistics* (pp. 165-187).
11. Helguera, M., et al. (2008). Applications of 2D descriptors in drug design: A DRAGON tale. *Current Topics in Medicinal Chemistry, 8*(18), 1628-1655.
12. Huang, M., et al. (2012). Terpenoids: Natural products for cancer therapy. *Expert Opinion on Investigational Drugs, 21*, 1008-1018.

13. Jiang, Y. K. (2010). Molecular docking and 3D-QSAR studies on β-phenylalanine derivatives as dipeptidyl peptidase IV inhibitors. *Journal of Molecular Modeling, 16*(7), 1239-1249.
14. Koehn, F. E., & Carter, G. T. (2005). The evolving role of natural products in drug discovery. *Nature Reviews Drug Discovery, 4*, 206-220.
15. Koelmel, J. P., et al. (2019). Software tool for internal standard-based normalization of lipids, and effect of data-processing strategies on resulting values. *BMC Bioinformatics, 20*(1), 1-13.
16. Maitra, S., et al. (2015). CNN based common approach to handwritten character recognition of multiple scripts. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)* (pp. 1021–1025). https://doi.org/10.1109/ICDAR.2015.7333916.
17. Mezei, M. (2003). A new method for mapping macromolecular topography. *Journal of Molecular Graphics and Modelling, 21*(5), 463–472. https://doi.org/10.1016/s1093-3263(02)00203-6.
18. Miller, J., & Artemisinin, D. (2011). Discovery from the Chinese herbal garden. *Cell, 146*, 855-858.
19. Owona, B. A., & Schluesener, H. J. (2015). Molecular insight in the multifunctional effects of oridonin. *Drugs R&D, 15*, 233-244.
20. Particle swarm optimization for parameter determination and feature selection of support vector machines. (2008). *Expert Systems with Applications*.
21. Potemkin, V., & Grishina, M. (2008). Principles for 3D/4D QSAR classification of drugs. *Drug Discovery Today, 13*, 952–959. https://doi.org/10.1016/j.drudis.2008.07.006.
22. Roy, K., & Kar, S. (2015). How do we judge the predictive quality of classification and regression-based QSAR models? In *Frontiers in Computational Chemistry* (pp. 71-120). Bentham Science Publishers.
23. Sarumathi, T., et al. (2015). Statistica software: A state of the art review.
24. Surveillance, Epidemiology, and End Results (SEER) Program. (n.d.). Cancer Stat Facts: Leukemia. Available at: SEER.
25. Shalev-Shwartz, S., et al. (2016). Pegasos: Primal estimated sub-gradient solver for SVM. *Mathematical Programming, 127*(1), 3–30.
26. Vapnik, V., & Vladimir, N. (1997). The Support Vector method. In W. Gerstner, A. Germond, M. Hasler, & J. D. Nicoud (Eds.), *Artificial Neural Networks – ICANN'97* (Vol. 1327, pp. 261–271). Berlin, Heidelberg: Springer.
27. Xu, W., et al. (2006). Pharmacokinetic behaviors and oral bioavailability of oridonin in rat plasma. *Acta Pharmaceutica*, 1642-1646.
28. Yap, C. W. (2011). Pa DEL-descriptor: An open-source software to calculate molecular descriptors and fingerprints. *Journal of Computational Chemistry, 32*(7), 1466-1474.
29. Zhang, Y., et al. (2013). Novel nitrogen-enriched oridonin analogues with thiazole-fused A-ring: Protecting group-free synthesis, enhanced anticancer profile, and improved aqueous solubility. *Journal of Medicinal Chemistry, 56*, 5048-5058.

Table1. The structures of Oridonin derivatives which are used as data set* .

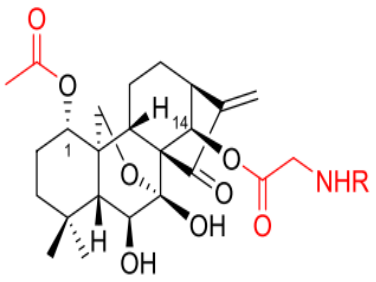


1.Oridonin

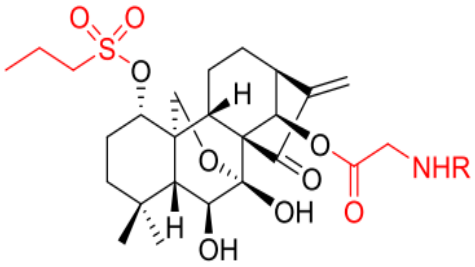


2. (HAO472)

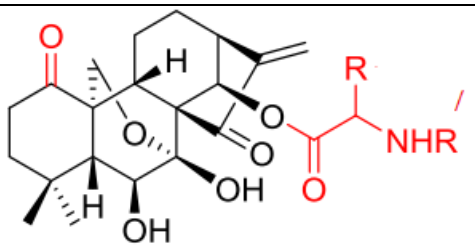
International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://www.theaspd.com/ijes.php



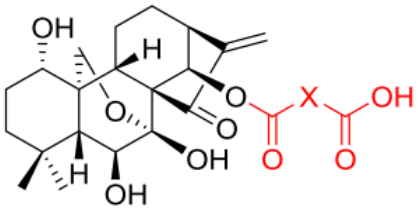
(3-5)



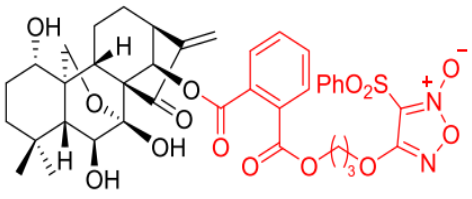
(6-8)



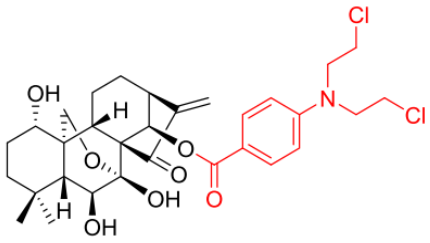
(9-11)



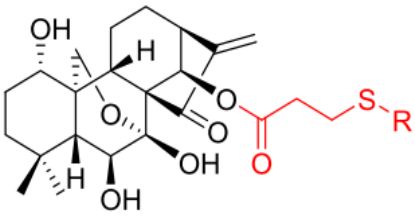
(12)



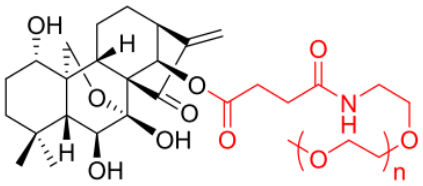
(13)



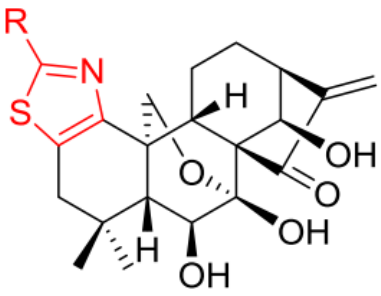
(14)



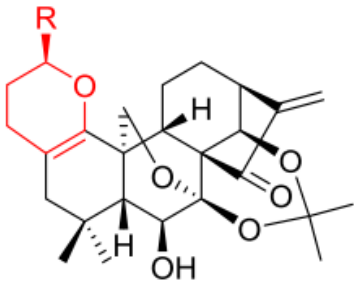
(15-16)



(17-20)

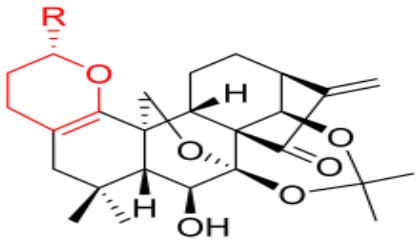


(21-31)

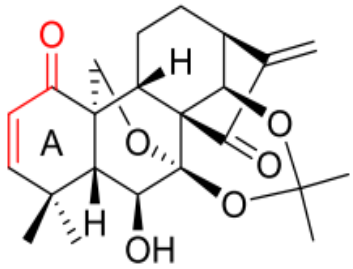


(32-36)

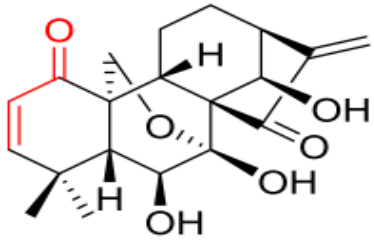
International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://www.theaspd.com/ijes.php



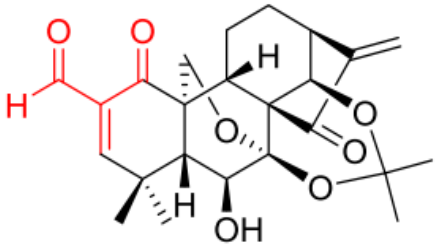
(37-41)



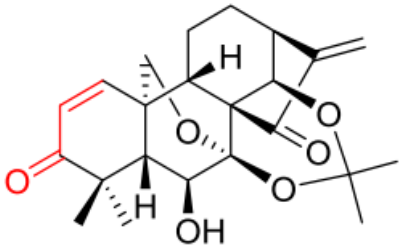
(42)



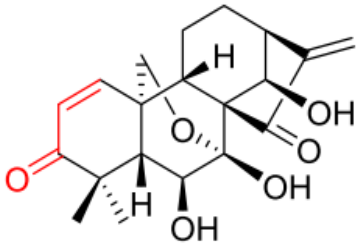
(43)



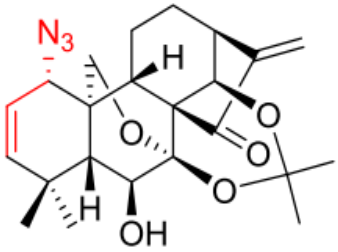
(44)



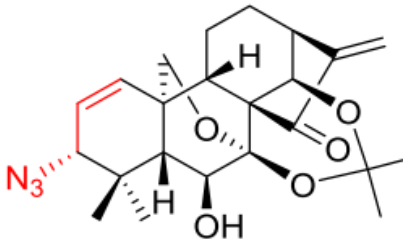
(45)



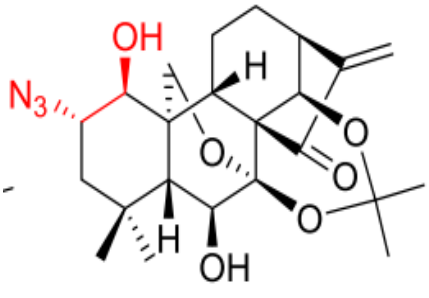
(46)



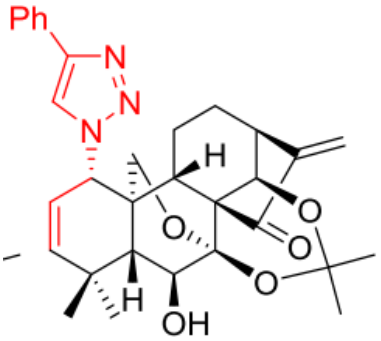
(47)



(48)

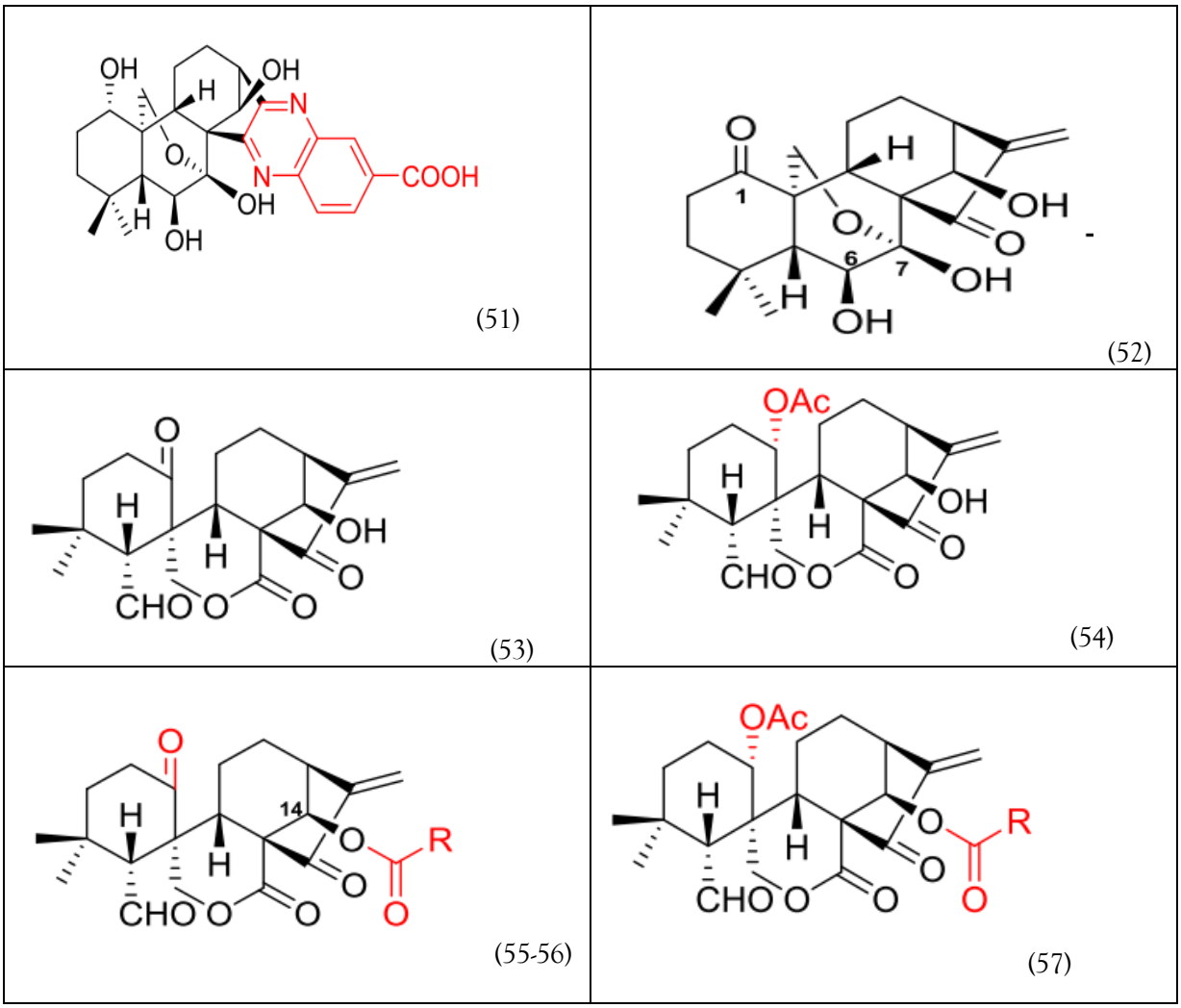


(49)



(50)

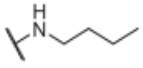
International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://www.theaspd.com/ijes.php

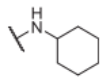
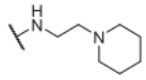
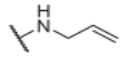
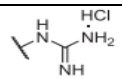
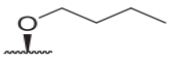
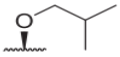
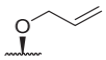
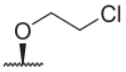
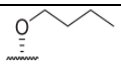


*R, R/ and X are shown in Table 2.

Table2. Structural details of Oridonine derivatives and their experimental EC_{50} (μM^{-1}) (numbers are identical as Table 1).

No.	R	R/	X	$EC_{50}(\mu M^{-1})$
1	-	-	-	6.60
2	HAO472	-	-	1.30
3	H	-	-	0.60
4	COCH=CHOOH	-	-	0.90
5	$CO(CH_2)_4COOH$	-	-	1.20

6	H	-	-	1.00
7	COCH=CHCOOH	-	-	0.20
8	$CO(CH_2)_4COOH$	-	-	2.00
9	H	$CO(CH_2)_4COOH$	-	1.00
10	CH_2Ph_1	$CO(CH_2)_4COOH$	-	0.80
11	CH_2Ph_1	COCH=CHOOH	-	3.40
12	-	-	$CH_2CH(C_{12}H_{25})$	0.44
13	-	-	-	2.20
14	-	-	-	1.38
15	$SCH_2CH(OH)CH_2OH$	-	-	0.14
16	H	-	-	0.24
17	-	-	PEG_{5KDa}-SA-ORI	1.74
18	-	-	PEG_{10KDa}-SA-ORI	5.69
19	-	-	PEG_{20KDa}-SA-ORI	1.22
20	-	-	PEG_{40KDa}-SA-ORI	9.56
21		-	-	1.39
22		-	-	1.48
23		-	-	1.27
24		-	-	1.22

25		-	-	1.74
26		-	-	1.40
27		-	-	1.31
28		-	-	0.87
29		-	-	0.39
30		-	-	4.76
31	-	-	-	8.11
32		-	-	2.64
33		-	-	1.11
34		-	-	0.21
35		-	-	0.22
36		-	-	0.35
37		-	-	0.33
38		-	-	0.52

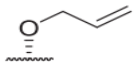
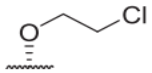
39		-	-	0.14
40		-	-	0.26
41		-	-	14.21
42	-	-	-	1.99
43	-	-	-	0.93
44	-	-	-	0.61
45	-	-	-	0.80
46	-	-	-	1.27
47	-	-	-	6.60
48	-	-	-	1.30
49	-	-	-	0.60
50	-	-	-	0.90
51	-	-	-	1.20
52	-	-	-	1.00
53	-	-	-	0.20
54	-	-	-	2.00
55	methyl	-	-	0.70
56	phenyl	-	-	2.40
57	methyl	-	-	6.60

Table 3. Statistical details of developed MLR model

Variable	**Unstandardized Coefficients**		**t**	**Sig.**
	B	**Std. Error**		
(Constant)	-19.16	4.04	-4.73	0.00
Chiv4pc	-19.16	4.04	-4.73	0.00
Smax8	2.39	0.33	7.11	0.00
EState_VSA1	-6.57	1.66	-3.95	0.00
GATSp4	-0.04	0.01	-2.54	0.01
WHIM.4	11.71	3.19	3.66	0.00
BCUT.1	9.97	2.89	3.44	0.00
petitjeanShapeIndex.0	-0.07	0.02	-2.67	0.01

able 4. The experimental and SVM predicted EC_{50} values and their residuals.

No.	**EC_{50}exp**	**EC_{50}pre**	**res**
1	6.60	5.80	0.70
2	1.30	0.6	0.6
3	0.60	0.63	-0.03
4	0.90	1.11	-0.21
5	1.20	1.56	-0.36
6	1.00	1.56	-0.56
7	0.20	0.76	-0.56
8	2.00	1.63	0.36
9	1.00	2.22	-1.22
10	0.80	1.44	-0.64
11	3.40	3.53	-0.13
12	0.44	-0.28	0.72
13	2.20	2.23	-0.03
14	1.38	1.98	-0.60
15	0.14	1.35	-1.21
16	0.24	-0.17	0.41
17	1.74	2.96	-1.22
18	5.69	4.84	0.84
19	1.22	2.33	-1.11
20	9.56	2.23	7.32
21	1.39	1.78	-0.39
22	1.48	2.07	-0.59
23	1.27	1.19	0.07
24	1.22	1.24	-0.02
25	1.74	1.69	0.04
26	1.40	2.28	-0.88
27	1.31	2.37	-1.06

28	0.87	0.70	0.16
29	0.39	1.20	-0.81
30	4.76	4.28	0.47
31	8.11	7.54	0.56
32	2.64	3.50	-0.86
33	1.11	0.52	0.58
34	0.21	1.49	-1.28
35	0.22	-0.61	0.83
36	0.35	-0.01	0.36
37	0.33	1.31	-0.98
38	0.52	1.01	-0.49
39	0.14	0.92	-0.78
40	0.26	1.34	-1.08
41	14.21	13.00	1.20
42	1.99	2.28	-0.29
43	0.93	2.22	-1.29
44	0.61	1.42	-0.81
45	0.80	0.94	-0.14
46	1.27	0.93	0.33
47	6.60	5.84	0.75
48	1.30	0.66	0.63
49	0.60	0.63	-0.03
50	0.90	1.11	-0.21
51	1.20	1.56	-0.36
52	1.00	1.56	-0.56
53	0.20	0.76	-0.56
54	2.00	1.63	0.36
55	0.700	-1.01	1.71
56	2.400	2.62	-0.22
57	6.600	5.84	0.75

Table 5. Statistical parameters of SVM and MLR model.

Parameter	**Set**	**MLR**	**SVM**
R^2	training	0.62	0.89
R^2	test	0.80	0.82
MES	training	1.36	1.46
MES	test	1.85	3.86
SE	training	1.44	0.44
SE	test	1.33	0.79

Table 6: A summary of the descriptors utilized in model construction.

No	Symbol	Class	Meaning
1	Chiv4pc	Simple molecular connectivity	Simple molecular connectivity
2	Smax8	Electro topological State Indices	Maximum of E-State value of specified atom type
3	EState_VSA1	MOE-type descriptors	MOE-type descriptors using Estate indices and surface area contributions
4	GATSp4	Geary autocorrelation descriptors	Geary autocorrelation-lag4/weighted by atomic polarizabilities
5	WHIM.4	WHIM descriptors	Unweighted WHIM descriptors
6	BCUT.1	Burden descriptors (64)	Burden descriptors based on atomic mass
7	Petitjean Shape Index.0	Petitjean based on topology	Petitjean Index based on molecular geometrical distance matrix

Table 7. The pharmacokinetic parameters of identified hits (Heavy atoms, Aromatic heavy atoms, Fraction Csp3, Rotatable bonds, H-bond acceptors, H-bond donors, MR, TPSA and XLOGP3*.

Code	MW	#Heavy atoms	#Aromatic heavy atoms	Fraction Csp3	#Rotatable bonds	#H-bond acceptors	#H-bond donors	MR	TPSA	XLOGP3
a	625.68	43	0	0.72	11	12	4	148.97	211.21	0.77
b	607.65	44	6	0.55	9	10	4	154	176.53	1.44
c	607.65	44	6	0.55	9	10	4	154	176.53	1.44
d	591.65	42	0	0.77	12	11	4	145.49	185.76	0.72
e	655.75	45	0	0.81	14	12	4	159.06	211.21	1.14

- Code of chemicals are identical as Fig. 6.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://www.theaspd.com/ijes.php

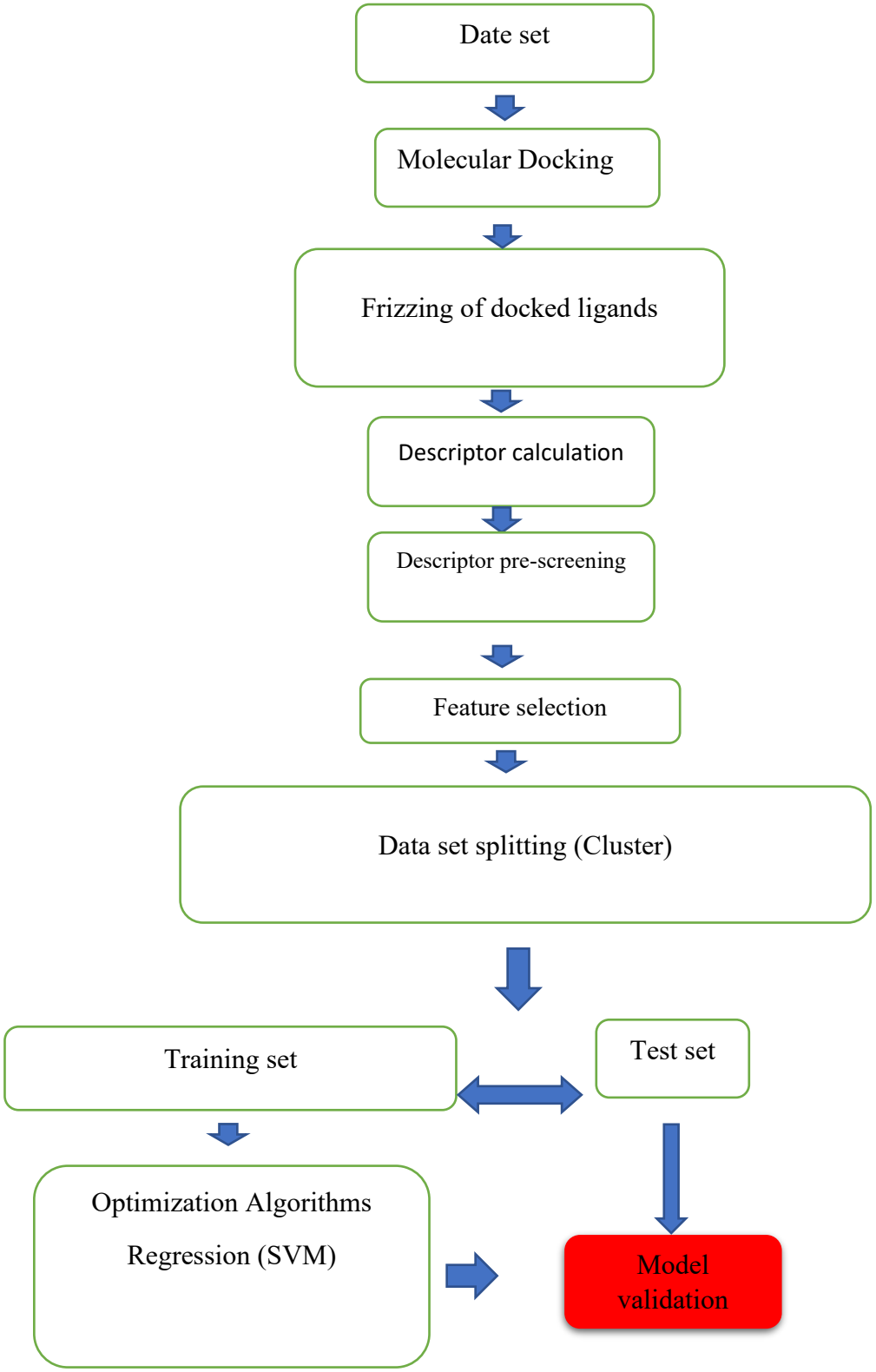


Fig. 1. QSAR workflow for modeling Akt1 inhibitors.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://www.theaspd.com/ijes.php

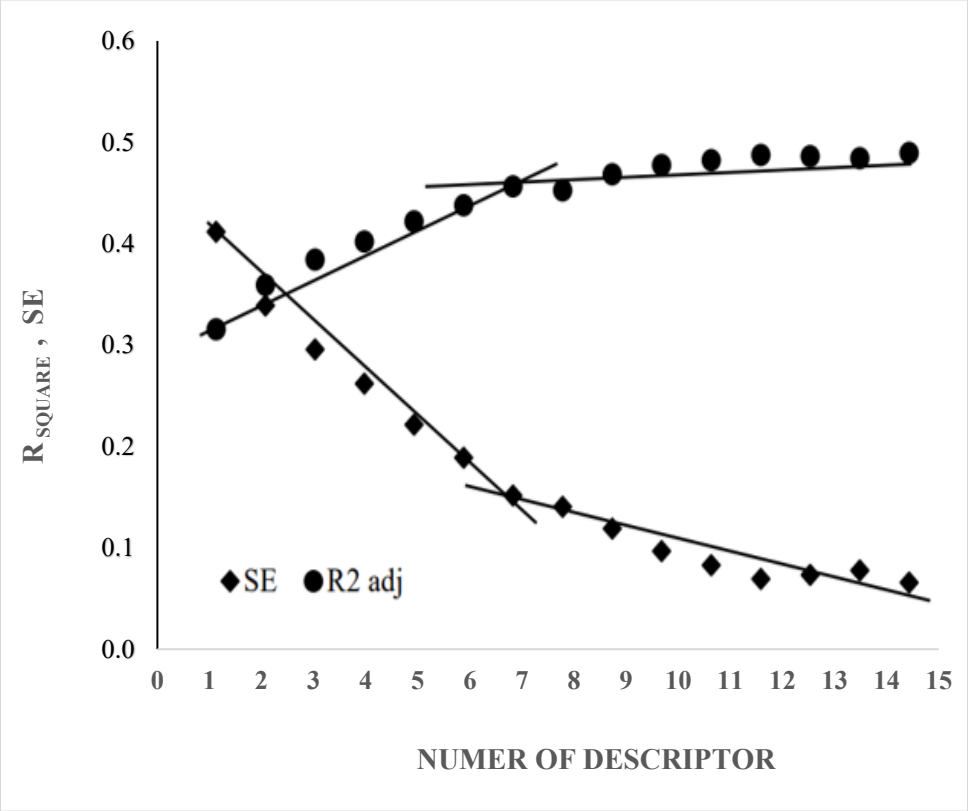
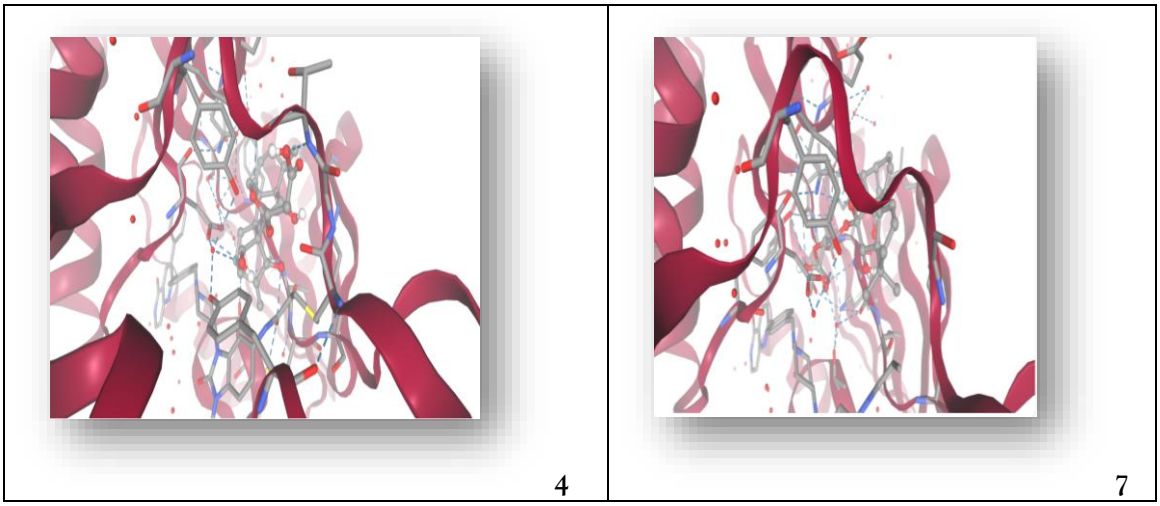


Fig. 2. Variations in the correlation coefficient (R) and standard error (SE) relative to descriptor count.



International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://www.theaspd.com/ijes.php

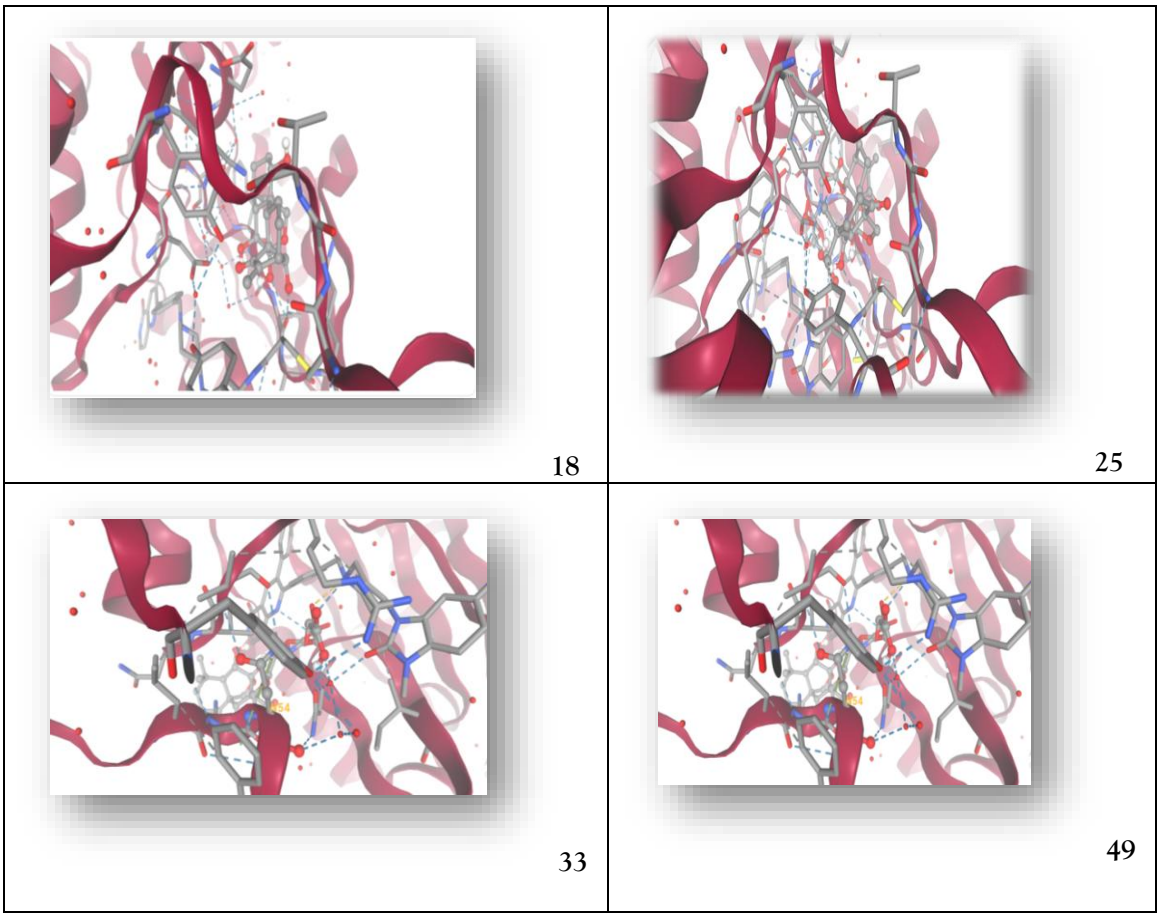
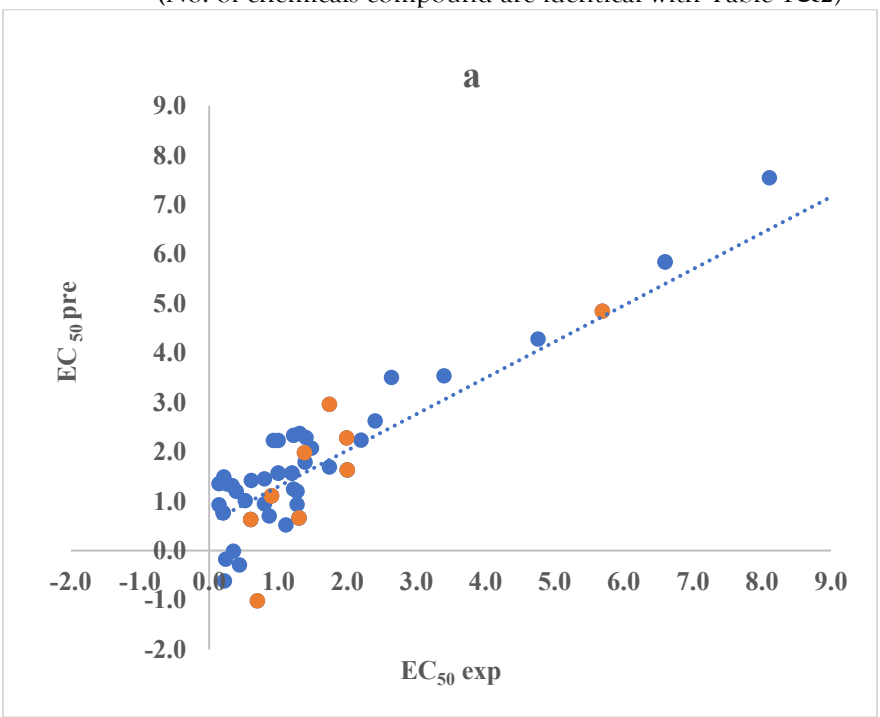


Fig. 3. Docking result between Akt and some Oridonin derivatives. (No. of chemicals compound are identical with Table 1&2)



International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://www.theaspd.com/ijes.php

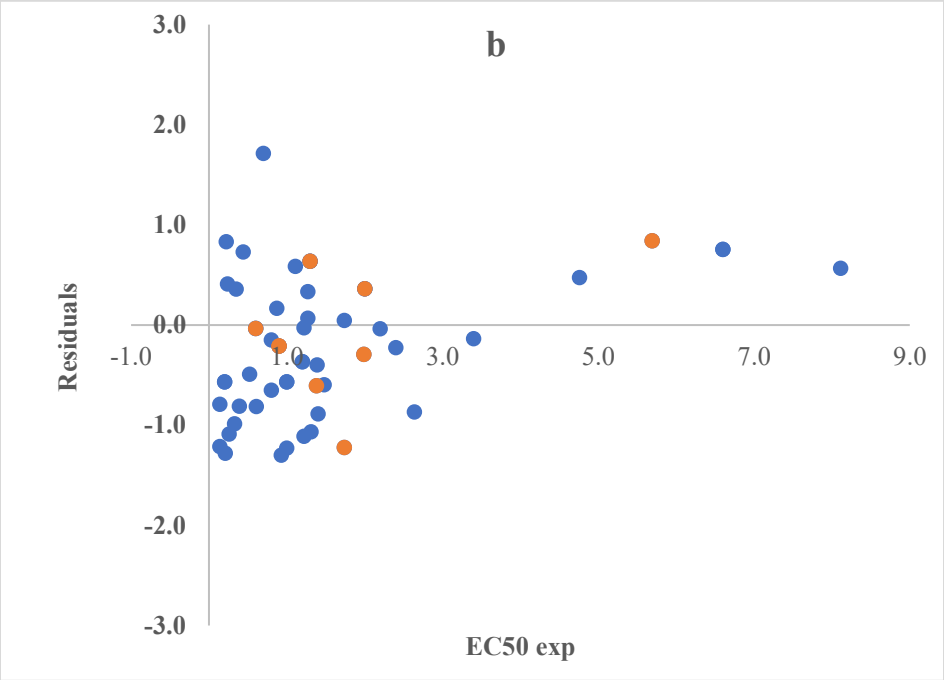


Fig. 4. The plot of SVM predicted against the experimental values of EC_{50} (a) and residuals (b).

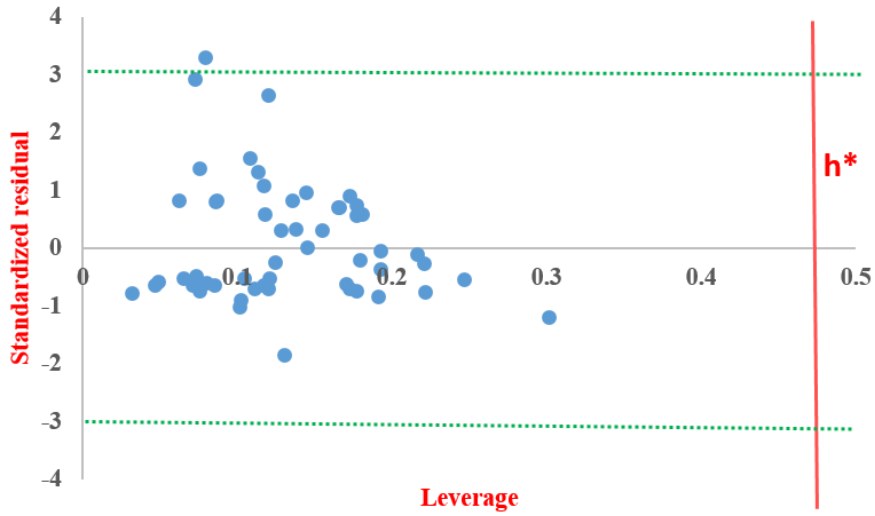
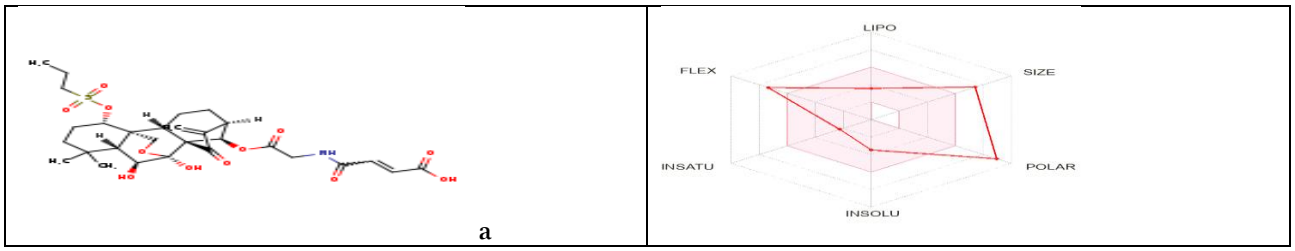


Fig. 5. The results of applicability domain analysis (Williams plot).



International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://www.theaspd.com/ijes.php

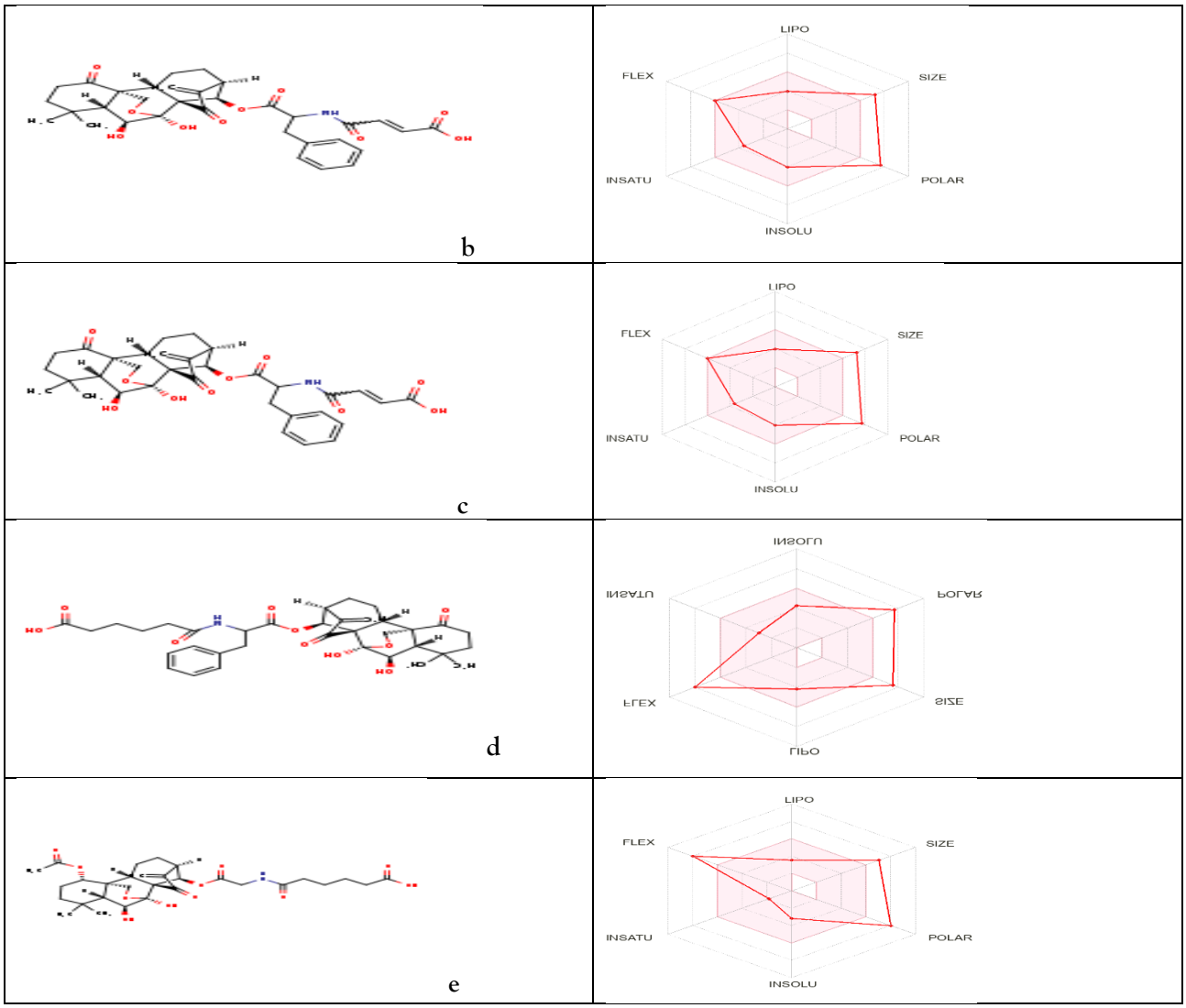


Fig. 6. Results of ADME analysis for Hit drug candidates.