International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 11-No.4 (2025) pp. 7270-7288
http://www.ijcesen.com
Research Article

ISSN: 2149-9144

Computational screening and qsar study of bastadins as acat1 inhibitors

Nabila Taib[1,2]*, Noureddine Tchouar[3]

[1]Department of Biotechnology, Faculty of Natural and Life Sciences, University of Science and Technology of Oran Mohamed BOUDIAF, P.O. Box 1503, El Mnaouer, 31000 Oran, Algeria
[2]Laboratory of Process and Environmental Engineering (LIPE), Faculty of Chemistry, University of Science and Technology of Oran Mohamed BOUDIAF, P.O. Box 1503, El Mnaouer, 31000 Oran, Algeria
*** Corresponding Author Email:** nabila.taib@univ-usto.dz- **ORCID:** 0009-0006-2667-7548

[3]Laboratory of Process and Environmental Engineering (LIPE), Faculty of Chemistry, University of Science and Technology of Oran Mohamed BOUDIAF, P.O. Box 1503, El Mnaouer, 31000 Oran, Algeria
Email: lamosi2002@yahoo.fr- **ORCID:** 0000-0003-4061-9273

Article Info:

DOI: 10.22399/ijcesen.3994
Received : 03 August 2025
Accepted : 29 September 2025

Keywords

Bastadins;
ACAT1 inhibitors;
QSAR;
ANN;
MLR;
MNLR.

Abstract:

In the search for new and effective anticancer agents, we performed a QSAR study on a series of sixteen bastadins to evaluate their potential as ACAT1 inhibitors and predict their antiangiogenic activity. Our goal was to establish a clear correlation between their biological responses and a set of molecular descriptors, applying principal component analysis (PCA), multiple linear regression (MLR), multiple nonlinear regression (MNLR), and an artificial neural network (ANN). Among the models generated, the best MLR and MNLR approaches achieved determination coefficients (R^2) of 0.71 and 0.91. To further assess their reliability, we performed an external validation on a test set of three compounds, confirmed their predictive accuracy, and yielded R^2 test values of 0.70 and 0.83, respectively. Furthermore, the ANN model, built with a 4-4-1 architecture, showed excellent performance, achieving a correlation coefficient of 0.96 with leave-one-out cross-validation coefficients (Q^2) of 0.79. These results indicate that the selected descriptors and calculated parameters are sufficient to reliably predict the biological activity of bastadins as ACAT1 inhibitors, providing a solid basis for the computer-aided design of novel anticancer agents.

1. Introduction

Cancer continues to be one of the leading causes of death worldwide and, in many regions, poses a major public health challenge [1]. Among the treatment options available, chemotherapy using potent chemical agents to destroy cancer cells remains a cornerstone, especially for metastatic or widespread tumors where surgery and radiotherapy are effective only for localized cancers. This central role of chemotherapy has drawn significant attention from researchers, driving ongoing efforts to discover and develop new, more effective anticancer drugs. Quantitative structure–activity relationship (QSAR) study have become indispensable in pharmaceutical research, especially for the early identification of promising drug candidates [2,3,4]. The classical QSAR approach assumes a clear and predictable relationship between a compound's molecular structure and its physical properties, chemical affinity, or biological activity. With suitable mathematical models and calculated molecular descriptors, it is possible to predict the biological activity of untested compounds. Once built and validated, a QSAR model can guide the design of new molecules by integrating various physical, chemical, and biological variables. Beyond drug design, QSAR methods also find applications in toxicology, food chemistry, and other scientific fields [5]. Macrocyclic compounds occupy a special place in medicinal chemistry due to

their unique structural features and high binding affinities, which often translate into potent and selective bioactivity. Marine-derived bastadins are brominated macrocyclic tyrosine-based metabolites that have shown a broad spectrum of biological activities, including antiangiogenic effects [6]. In particular, bastadins have emerged as promising inhibitors of Acyl-CoA: cholesterol acyltransferase 1 (ACAT1), an enzyme increasingly recognized for its role in tumor progression and chemoresistance [7,8]. In this study, as part of our search for innovative anticancer agents, we focused on sixteen bastadins macrocyclic compounds. Using a combination of SAR and QSAR approaches, including principal component analysis (PCA), multiple linear regression (MLR), multiple nonlinear regression (MNLR), and artificial neural network (ANN) modeling, we aimed to predict their antiangiogenic potential as ACAT1 inhibitors and elucidate the structural determinants governing their biological activity.

2. Computational Method

2.1. Data set collection and biological activities

The bastadins dataset used in this study was compiled from previously published research and includes both bastarane and isobastarane types. Sixteen selected compounds were assessed for their antiangiogenic potential by measuring their half-maximal inhibitory concentrations (IC_{50}) against human umbilical vein endothelial cells (HUVEC), with the results expressed in micromolar units [9,10].To minimize data skewness for the QSAR analysis, the IC_{50} values representing the concentration needed to achieve 50% inhibition of proliferative activity were converted to their logarithmic form (pIC_{50}). The substitution patterns and molecular structures of these ligands, along with their corresponding IC_{50} and pIC_{50} values, are summarized in Scheme 1 and Table 1, respectively.

2.2. Molecular descriptors

The molecular descriptors for the sixteen bastadins under investigation were predicted in silico using the HyperChem 8.0.6 [11] and Gaussian 09W software packages [12], along with the Molinspiration cheminformatics platform [13]. The studied molecules were first optimized using molecular mechanics with the MM+ force field. Their geometries were then fully refined using the semi-empirical PM3 method. Finally, the PM3-optimized structures of the sixteen compounds were employed to calculate various physicochemical descriptors through the QSAR module in HyperChem.These descriptors include molar refractivity (MR), surface area grid (S), hydration energy (HE), molar volume (V), molecular weight (MW), molar polarizability (Pol), and the partition coefficient octanol/water (LogP). LogP and MR were calculated using atomic contribution approaches, where the hydrophobicity constants for LogP and the atomic refractivity values for MR both empirically established were originally reported by Viswanadhan and colleagues [14]. Gavezotti's atomic radii, together with the grid-based volume calculation method from Bodor et al., form a well-established computational approach to estimate both solvent-accessible and van der Waals surface-bounded molecular volumes [15,16].The energy of hydration (EH) plays a crucial role in stabilizing molecular conformations. It is modeled based on the molecule’s accessible surface area in an aqueous environment, allowing estimation of its overall energy and associated thermodynamic effects [17]. Molecular polarizability is affected by the weight and surface of the molecule. According to recent research, molecular orientation and interactions are influenced by the variations in the polarizability of functional groups on polar molecules [18]. Various methods are employed to determine a system’s molecular weight (MW), including non-intact approaches that involve specialized solvent dissolution, as well as intact techniques like NMR and fluorescence measurements [20].

Furthermore, the Molinspiration program was used to calculate important molecular descriptors such as topological polar surface area (TPSA), hydrogen bond donors (HBD), hydrogen bond acceptors (HBA), and the number of rotatable bonds (NRB). These descriptors play a crucial role in drug-likeness, assessing molecular properties such as oral bioavailability, aqueous solubility, and membrane permeability [21].

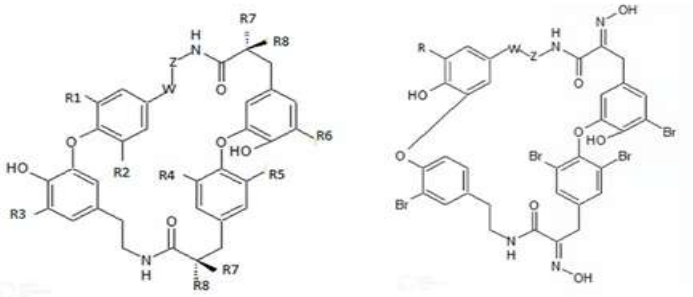

Figure 1. *2D structures of bastadins*

Table 1. *Dataset of bastadins and their corresponding experimental activities*

N°	*R*	*R 1*	*R 2*	*R3*	*R 4*	*R 5*	*R6*	*R7,R8*	*W—Z*	*Molecular formula*	*$IC_{50}/\mu M$*	*pIC_{50}*
1	-	Br	H	Br	Br	Br	Br	R7,R8= NOH		$C_{34}H_{25}Br_5N_4O_8$	0.053	7.276
2	-	Br	H	Br	Br	Br	Br	R7,R8= NOH		$C_{34}H_{27}Br_5N_4O_8$	0.062	7.208
3	-	Br	Br	Br	Br	Br	Br	R7,R8= NOH		$C_{34}H_{26}Br_6N_4O_8$	0.052	7.284
4	-	Br	H	Br	Br	H	Br	R7,R8= NOH		$C_{34}H_{26}Br_4N_4O_8$	0.054	7.268
*5**	-	Br	H	H	Br	Br	Br	R7,R8= NOH		$C_{34}H_{26}Br_4N_4O_8$	0.053	7.276
6	H	-	-	-	-	-	-	-		$C_{34}H_{28}Br_4N_4O_8$	0.52	6.284
7	Br	-	-	-	-	-	-	-		$C_{34}H_{27}Br_5N_4O_8$	0.12	6.921
8	-	Br	Br	Br	Br	Br	Br	R7=R8= H		$C_{34}H_{28}Br_6N_2O_6$	4.00	5.398
9	-	Br	Br	Br	Br	Br	Br	R7=H,R8=NH_2		$C_{34}H_{30}Br_6N_4O_6$	1.00	6.000
10	-	Br	Br	Br	Br	Br	Br	R7=NH_2,R8= H		$C_{34}H_{30}Br_6N_4O_6$	0.61	6.215
*11**	-	Br	H	Br	Br	Br	H	R7,R8= NOH		$C_{34}H_{28}Br_4N_4O_8$	0.16	6.796
*12**	-	H	H	Br	H	H	Br	R7,R8= NOH		$C_{34}H_{30}Br_2N_4O_8$	0.48	6.319
13	-	H	H	Br	H	H	H	R7,R8= NOH		$C_{34}H_{31}BrN_4O_8$	1.5	5.824
14	-	H	H	H	H	H	H	R7,R8= NOH		$C_{34}H_{32}Br_0N_4O_8$	2.6	5.585
15	-	Br	Br	NH_2	Br	Br	NH_2	R7,R8= NOH		$C_{34}H_{30}Br_4N_6O_8$	0.96	6.018

The Gaussian 09W software package was also applied to measure quantum chemical parameters. The minimized compound basis set to calculate the electronic descriptors, which are reported in Table 2. A key chemical descriptor affecting polarizability is polar surface area (PSA), which is utilized in molecular modeling to predict molecule transport characteristics such as intestine absorption or blood-brain barrier crossing [19].

Table 2. *Electronic descriptors included in the QSAR modeling*

Dipole moment D, Total Energy E, LUMO Energy, HOMO Energy, LUMO+1 Energy, HOMO-1 Energy, Hardness η, Electronegativity μ, Electrophilicity index ω,
$\mu = -(E_{HOMO}+E_{LUMO}) / 2$, $\eta = (E_{LUMO}- E_{HOMO}) / 2$, $\omega = \mu^2/ 2\eta$

2.3. Multi-Parameter Optimization (MPO) method

Multi-parameter optimization (MPO) is a drug development technique that considers numerous properties in order to predict drug-likeness and identify effective and safe medication candidates [26]. MPO methods vary from empirical rules to more complex computer procedures; among these rules, we carry out rules of thumb and calculated metrics [27.28.29].Rules of thumb are the most prevalent approach used to evaluate the quality of compounds in relation to variables other than potency, which provide suggestions for desired compound features. Several rules have been developed; the Lipinski's rule is the most widely utilized [27,28].The selected compounds were evaluated using Lipinski's rule, or "rule of five." This is a concept frequently used in drug-likeness, which is based on analysis of physicochemical properties to predict if a biologically active molecule is likely to have the chemical and physical qualities to be orally bioavailable. A compound is more likely to be active orally if it satisfies the following requirements:
According to Lipinski's rule of five, an orally a ctive medication can only violate one of these conditions [30.31].

(1) Molecular weight (MW) ≤500 Da
(2) Octanol water partition coefficient log*P* ≤5
(3) Hydrogen bond donors (HBD) ≤5
(4) H-bond acceptors (HBA) ≤10

Calculated metrics, on the other hand, seek to combine potency with other characteristics to create a single metric that can be tracked during optimization. The earliest and most often used measures are Ligand Efficiency (LE) and Lipophilicity Efficiency (LipE) [28], which are defined as follows:

$$LE = 1{,}4pIC50/NH$$

Where NH is the number of heavy atoms. So LE decreases with an increasing number of heavy atoms.Ligand Efficiency (LE) is a particularly essential statistic in fragment drug design because it prioritizes small compounds with lower potency over bigger, greater potency molecules [32,33].

$$LipE = pIC50 - logP$$

Lipophilicity efficiency (LipE) is a key factor in compound promiscuity, and optimized compounds should be more selective, targeting LipE values between 5 and 7 or even higher. Ideally, LipE should be between 5 and 9 or over 10 [34].

2.4. Statistical Methods
Principal Component Analysis, Multiple Linear Regression, Multiple Nonlinear Regression

The antiangiogenic activity of sixteen bastadins on human umbilical vein endothelial cells (HUVEC) was investigated to develop a QSAR model. To better understand the impact of these compounds on their inhibitory activity expressed as pIC50, QSAR statistical approaches can be used to examine molecular descriptors and build robust models to predict pIC50.

We utilized three strategies available in the

XLSTAT program version 2025: principal component analysis (PCA), multiple linear regression (MLR) and multiple nonlinear regression (MNLR). Molecular descriptors were initially submitted to PCA using the correlation matrix to reduce the number of computed descriptors and choose relevant ones to quantitatively represent response activity [35]. After selecting molecular descriptors, the MLR proposes a linear model to predict the inhibitory actions of new compounds. The MNLR approach is commonly used in QSAR modeling to enhance structure-activity relationships. We created an MNLR model using descriptors and training set molecules from MLR.For model construction, we randomly separated the data set into two sets: training and test. We generated the internal validation indices and external validation parameters to assess the dependability and quality of our created models; the model can be used as a tester to identify new ligands with improved anti-angiogenic activity.

Artificial Neural Networks (ANN) analysis

The artificial neural networks (ANN) method was used to create nonlinear models of structure-activity relationships (QSAR) for accurately characterizing the examined compounds based on a set of molecular descriptors derived from multiple linear regression (MLR) and their observed activity. The ANN analysis was performed using the Matlab R2025a program with the Levenberg-Marquardt algorithm [36]. There are three layers of neurons in it: the input layer, the hidden layer, and the output layer [37].The descriptors obtained in the multiple linear regression models constituted the input layer; the input layer included four artificial neurons with linear activation functions. The number of artificial neural networks in the hidden layer was varied through experimentation; the hidden layer was made up of 10 artificial neurons. A single neuron serves as the output layer for sigmoid function activation; the architecture of the used ANN models is described in Figure 1.

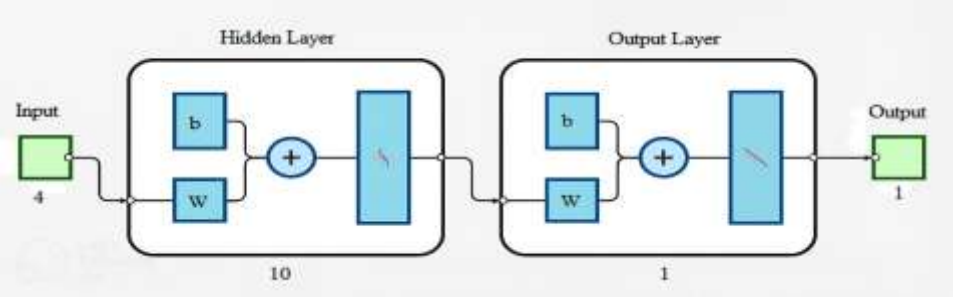

Figure 2. *The ANN architecture*

For ANN analysis, the data sets of molecules were randomly separated into three sets: training, validation, and testing.

Validation of model - Cross Validation-

Cross validation is the most frequent technique for validating the stability and predictive capability of the QSAR model; it includes splitting a dataset into subsets, training the model on part of them, and testing it on the rest. The Leave-One-Out (LOO) cross-validation (CV) is a specific instance in which k equals n (the total number of samples). Each data point is used once as a test, and the remainder is used for training. This validation procedure enables the estimation of the cross-correlation coefficient (Q^2), which serves as a crucial measure of ANN model quality. A Q^2 value greater than 0.5 is generally regarded as acceptable, whereas values above 0.9 indicate excellent predictive reliability [5]. In this study, the performance of the ANN model was validated using the Leave-One-Out (LOO) method implemented in MATLAB (version R2025a).

3. Results and discussion

3.1. Multi-Parameter Optimization (MPO) and drug-likeness evaluation

A key objective of this study was to explore the physicochemical properties of sixteen bastadins. For this purpose, we applied both empirical rules and metric-based approaches to evaluate their antiangiogenic activity (pIC50) against HUVEC [9,10], as summarized in

Table 1. The analyzed parameters included molecular weight (MW), octanol–water partition coefficient (LogP), hydrogen bond acceptors (HBA), hydrogen bond donors (HBD), ligand efficiency (LE), and lipophilic efficiency (LipE). The calculated results, obtained using HyperChem 8.0.6 [11] and the Molinspiration online platform [13], are presented in Table 3.

All compounds were examined for drug-likeness [38]. To meet Lipinski's rule and exhibit suitable oral bioavailability, a compound must maintain an appropriate balance between its aqueous solubility and its ability to passively diffuse across biological barriers [39].

Lipinski's rules are grounded in well-established physical and chemical principles. Hydrogen bonding enhances a molecule's solubility in water but must be broken for it to penetrate the lipid bilayer. As the number of hydrogen bonds increases, transferring the molecule from the aqueous environment to the lipid layer becomes more difficult, thereby reducing its capacity for passive diffusion [40]. According to Table 3, all of the tested derivatives except for molecules 15 and 16 meet Lipinski's rules (3) and (4). This compliance indicates that these compounds generally possess lower polarity, which tends to favor their ability to cross biological membranes, ultimately enhancing their potential for good oral absorption.

Compounds with a lower molecular weight tend to display better oral bioavailability. Smaller molecules dissolve more easily in aqueous environments and can pass through biological membranes more efficiently, which are necessary processes for absorption in the gastrointestinal tract. In contrast, as molecular weight increases, the energy required to generate larger solvent holes rises, reducing solubility and absorption [28]. Table 3 indicates that all the molecules have a molecular weight exceeding 500 Da. According to Lipinski's rules (1), this may limit their oral absorption because larger molecules often dissolve less in water and pass less easily through cell membranes. As a result, these compounds might have low oral bioavailability and may be better suited for alternative administration routes, such as intravenous or intramuscular delivery, unless to have chemical modification to improve their delivery.

Moreover, Doak et al. demonstrated that the 500 Da molecular weight threshold in Lipinski's Rule of Five is not a strict barrier to the development of oral drugs. Their analysis of approved medicines and clinical candidates revealed many orally active compounds with molecular weights well above this limit classified as belonging to the beyond Rule of Five (bRo5) category. Such molecules often feature notable structural and physicochemical adaptations that enable them to bypass the limitations imposed by their larger size [41].

In our study, the sixteen molecules have a macrocyclic structure and high molecular weight, which may give them certain advantages for drug behavior and membrane crossing. Large ring structures can fold in ways that hide polar groups, making it easier for them to pass through membranes even if they are heavy. Because macrocycles are more rigid, they can move across lipid membranes more easily and are less likely to be broken down by metabolic enzymes. These features can help macrocyclic drugs reach the bloodstream successfully, even when they do not fit the usual Rule of Five limits [42]. Log P represents a compound's distribution between a hydrophobic medium (n-octanol) and an aqueous phase (water). It is calculated by taking the ratio of the drug's concentration in n-octanol to its concentration in water. This metric is frequently used to measure oral solubility and forecast absorption capacity.

A rise in logP generally leads to lower water solubility, which can diminish the extent of absorption. Negative Log P values (LogP < 0) show that the molecule is extremely hydrophilic, resulting in high solubility in aqueous environments, increased gastric tolerance, and efficient renal excretion. On the other hand, positive LogP values (LogP > 0) indicate that a compound is mostly lipophilic, which helps it cross membranes, bind to plasma proteins, and be metabolized. However, this often reduces water solubility and can affect stomach tolerance [43]

In this study, all measured Log P values are in the negative range, indicating strong solubility in water and enhanced gastric tolerance. The highest value within the series is observed for compound 3, with a logP of (-0.55).

The count of hydrogen bond acceptors (O and N atoms) and donors (NH and OH groups) is important in medication design because it

influences a compound's capacity to absorb and penetrate biological membranes [44].
For the substances studied, these values stayed within Lipinski's limits, below 10 and 5, respectively. Compounds that break more than one of these rules often have lower bioavailability and are less likely to show good drug-like properties. [45,46]

With the exception of molecules 15 and 16, all the compounds examined possess five hydrogen bond donors, while molecule 8 contains four. A reduced number of hydrogen bond donors often increases a compound's lipophilicity, which improves its ability to pass cell membranes and enter the cell interior [39].
The results indicate that all molecules follow Lipinski's recommendations, meaning that they are unlikely to cause oral bioavailability concerns in theory. However, molecules 15 and 16 diverge from these criteria, implying that they may experience limitations in oral bioavailability.

Lipophilic Efficiency (LipE) and Ligand Efficiency (LE) are important metrics in drug design because they help determine the balance between potency and physicochemical qualities. Evaluating these measures helps discover compounds that combine effective biological activity with desirable drug-like features, leading to the selection of promising candidates for further development [47.48].Table 3 shows that all compounds have LipE values between 6 and 10, indicating a good balance between potency and lipophilicity and suggesting effective optimization.
Among all the analyzed compounds, compound 5 presented the maximum LE value of 0.204, identifying it as the top-performing molecule.

3.2. Structure Activity Relationship for bastadins

For the sixteen bastadins, other physicochemical properties were examined, including molar refractivity (MR), hydration energy (HE), molar volume (V), surface area grid (S), and polarizability (Pol) using HyperChem 8.0.6. Other parameters, such as the number of rotatable bonds (NRB) and topological polar surface area (TPSA), were calculated with the Molinspiration online database. The findings are presented in Table 4

In addition, we examined seven quantum properties of the bastadins, as summarized in Table 5. These include the dipole moment (D), total energy (E), energies of the frontier orbitals (ELUMO, EHOMO, ELUMO+1, and EHOMO–1), and the electrophilicity index (ω), all calculated using Gaussian 09

In molecular modeling, polarizability, molar refractivity, and molecular volume are closely related. Polarizability reflects how easily a molecule's electron cloud can be distorted by an external electric field, and it typically increases with molecular size and weight [49].
Molar refractivity, which is derived from the Lorentz–Lorenz relationship, incorporates both the refractive index and molecular volume, thereby serving as an indirect indicator of steric effects and the space occupied by atoms or functional groups [50]. Because molar refractivity is mathematically proportional to polarizability, compounds with greater electron density and larger volumes tend to exhibit higher values of both descriptors [51].
In this sense, molecular volume is a primary determinant influencing both molar refractivity and polarizability, making these parameters valuable in quantitative structure–activity relationship (QSAR) studies where they help to rationalize drug absorption, permeability, and receptor–ligand recognition [52].
An increase in the molecular weight and overall size of bastadins leads to higher values of both molar refractivity and polarizability (Table 4). This result is consistent with the Lorentz–Lorenz equation [53], which links these properties to molecular volume [54]. Within the studied series, compound 16 stands out by exhibiting the highest values of polarizability (87.00 Å³), molar refractivity (242.87 Å³), and molecular volume (2053.40 Å³). On the other hand, compound 14 represents the smallest structure, displaying comparatively lower polarizability (65.55 Å³), reduced molar refractivity (184.44 Å³), and a smaller volume (1588.73 Å³).
In QSAR investigations, the surface area grid is a mathematical tool that divides the region around a molecule into a grid and measures the solvent-accessible surface area of the molecule. This metric is important because molecular surface exposure is strongly connected to solubility, membrane permeability, and, ultimately, bioavailability [55].

Surface area grid values in the range of 819.28 to 1043.66 Å² show that these compounds have an extended solvent-accessible surface, consistent with relatively high molecular size and greater potential for identifying biological targets [28]. Such large surface areas help molecules interact better with their targets, but they may also reduce oral bioavailability to some extent, since the molecules do not pass easily through membranes.

The calculated surface area follows the same increasing pattern as polarizability, molar refractivity, and volume in these molecules. Compound 16 shows the maximum surface value (1043.66 Å²).

Table 3. *Pharmacological activities and properties involved in MPO method for bastadins*

compound	MW (uma)	Log P	HBA	HBD	Lipinski score of 4	pIC_{50}	LE	LipE
1	1017.11	-0.94	10	5	3	7.276	0.200	8.216
2	1019.13	-0.61	10	5	3	7.208	0.198	7.818
3	1098.03	-0.55	10	5	3	7.284	0.196	7.834
4	938.22	-0.99	10	5	3	7.268	0.203	8.258
5*	938.22	-0.99	10	5	3	7.276	0.204	8.266
6	940.23	-0.66	10	5	3	6.284	0.176	6.944
7	1019.13	-0.61	10	5	3	6.921	0.190	7.531
8	1040.03	-2.03	6	4	3	5.398	0.157	7.428
9	1070.06	-3.87	8	5	3	6.000	0.168	9.870
10	1070.06	-3.87	8	5	3	6.215	0.174	10.085
11*	940.23	-0.66	10	5	3	6.796	0.190	7.456
12*	782.44	-0.76	10	5	3	6.319	0.184	7.079
13	703.55	-0.81	10	5	3	5.824	0.173	6.634
14	624.65	-0.86	10	5	3	5.585	0.170	6.445
15	970.26	-4.09	10	8	2	6.018	0.162	10.108
16	1054.34	-3.35	12	12	1	5.236	0.126	8.586

Table 4. *Values of physicochemical descriptors used in the regression analysis*

Nº	LOG P	MW (UMA)	MR (Å³)	HE (KCAL/MOL)	V (Å³)	S (Å²)	POL (Å³)	NRB	TPSA (Å²)
1	-0.94	1017.11	222.24	-30.17	1887.95	1008.74	78.49	0	182.30
2	-0.61	1019.13	222.11	-30.22	1906.02	1017.01	78.68	0	182.30
3	-0.55	1098.03	229.65	-22.15	1891.29	941.18	81.31	0	182.30
4	-0.99	938.22	214.70	-32.10	1840.64	982.68	75.87	0	182.30
5*	-0.99	938.22	214.70	-31.10	1819.51	957.27	75.87	0	182.30
6	-0.66	940.23	214.58	-26.28	1819.05	948.84	76.06	0	182.30
7	-0.61	1019.13	222.11	-25.62	1875.49	982.58	78.68	0	182.30

8	-2.03	1040.03	221.09	-15.21	1845.73	967.20	78.17	0	117.12
9	-3.87	1070.06	227.39	-21.43	1873.17	952.38	80.87	0	169.17
10	-3.87	1070.06	227.39	-20.45	1884.90	957.09	80.87	0	169.17
11*	-0.66	940.23	214.58	-29.02	1796.10	905.34	76.06	0	182.30
12*	-0.76	782.44	199.51	-32.25	1731.05	898.70	70.81	0	182.30
13	-0.81	703.55	191.98	-32.69	1621.29	830.58	68.18	0	182.30
14	-0.86	624.65	184.44	-30.51	1588.73	819.28	65.55	0	182.30
15	-4.09	970.26	221.68	-37.54	1873.57	942.44	78.76	0	234.35
16	-3.35	1054.34	242.87	-45.89	2053.40	1043.66	87.00	4	306.10

Table 5. *Values of electronic descriptors used in the regression analysis*

N°	***D (Deb)***	***E (eV)***	E_{LUMO} ***(eV)***	E_{HOMO} ***(eV)***	E_{LUMO+1} ***(eV)***	E_{HOMO-1} ***(eV)***	***ω (eV)***
1	9.50	-408241.92	-1.88	-6.11	-1.66	-6.18	3.77
2	11.76	-408275.50	-2.01	-6.05	-1.21	-6.27	4.02
3	8.12	-478305.12	-1.82	-6.23	-1.43	-6.44	3.67
4	8.27	-338212.40	-1.64	-6.13	-1.62	-6.30	3.36
5*	11.53	-338212.17	-1.82	-6.05	-1.58	-6.11	3.66
6	10.18	-338245.67	-1.99	-6.01	-1.33	-6.11	3.98
7	10.91	-408275.42	-2.01	-6.00	-1.40	-6.34	4.02
8	11.07	-471265.81	-1.23	-6.04	-1.17	-6.50	2.74
9	6.81	-474278.83	-1.22	-6.14	-1.07	-6.26	2.75
10	8.71	-474278.54	-1.29	-6.06	-1.24	-6.47	2.83
11*	4.65	-338246.20	-1.99	-5.83	-1.44	-6.35	3.98
12*	6.80	-198187.02	-1.85	-5.93	-1.46	-6.34	3.71
13	1.46	-128157.39	-1.43	-5.79	-1.39	-6.01	2.99
14	8.41	-58127.76	-1.59	-5.84	-1.28	-6.10	3.25
15	3.12	-341259.79	-1.82	-5.24	-1.37	-5.56	3.64
16	4.74	-349360.68	-1.96	-5.36	-1.43	-5.97	3.94

Hydration energy plays a crucial role in controlling the stability of various molecular conformations in aqueous environments [56]. The most important hydration energy in the absolute value is that of compound 16 (45.89 kcal/mol). Regarding compound 8, it shows the minimum absolute value (15.21 kcal/mol) (Table 4).

The variations in hydration energy were directly related to the number of polar groups present in each compound. Compound 16 carries 12 hydrogen bond donors and 12 acceptors, which greatly enhances its capacity to establish interactions with water, resulting in higher hydration energy. On the other hand, compound 8, limited to 4 donors and 6 acceptors, forms fewer hydrogen bonds with the solvent and therefore displays a reduced hydration energy.

In medicinal chemistry, the number of rotatable bonds (NRB) is considered a key molecular feature describing conformational flexibility. Molecules with many freely rotating single bonds can easily adopt different shapes in solution. Although such flexibility can improve the ability to fit into biological targets, it usually has a negative impact on oral drug performance, as highly flexible compounds tend to cross membranes less efficiently and may be metabolized more rapidly. For this reason, NRB is frequently applied in drug-likeness criteria, where compounds with fewer than 10 rotatable bonds are generally associated with favorable oral absorption profiles [28].

Topological Polar Surface Area (TPSA) is an important descriptor that reflects the contribution of polar atoms, mainly oxygen and nitrogen with their attached hydrogens, to the molecular surface. It is widely applied in drug design because it strongly correlates with pharmacokinetic properties such as intestinal absorption, Caco-2 permeability, and blood–brain barrier penetration. In general, compounds with TPSA values below 140 Å² tend to show better oral bioavailability, while those with values under 90 Å² are more likely to cross the blood–brain barrier [57]. Compounds with TPSA values of 140 Å² or more are expected to exhibit poor intestinal absorption [58].

TPSA of bastadins was found in the range of 117.12 to 306.10 Å². All compounds present TPSA values above 140 Å², except compound 8, which has TPSA under 140 Å² (Table 4).

The evaluation of rotatable bonds revealed that all compounds were rigid (0 NRB), except compound 16, which has 4 NRB. The high rigidity observed in most molecules may favor stronger binding affinity and stability within the active site.

All compounds show high TPSA values together with no rotatable bonds. This combination makes them rigid and strongly polar, which reduces their ability to cross membranes and the blood–brain barrier. However, this property can also be beneficial since it increases selectivity, prevents central nervous system penetration, and favors peripheral activity.

The analysis of electronic descriptors shows that molecule 3 is the most stable, with the lowest total energy value of -478305.12 eV in comparison with compound 14 (-58127.76 eV). This stability might be linked to its inhibitory behavior. Compound 2 exhibits the highest dipole moment (11.76 D), which can be attributed to resonance, where electron donation from the nucleus is directed toward the electron-attracting groups; the results are shown in Table 5.

To provide a more detailed description of the electronic properties, we also investigated frontier molecular orbitals. In particular, we examined the HOMO (highest occupied molecular orbital) and the LUMO (lowest unoccupied molecular orbital), together with the neighboring orbitals HOMO–1 and LUMO+1.

Studying these orbitals is important because the HOMO reflects the electron-donating ability of a molecule, while the LUMO indicates its capacity to accept electrons. The additional orbitals (HOMO–1 and LUMO+1) give complementary insights into electronic transitions and molecular reactivity, thus allowing a more comprehensive interpretation of the compound's behavior [59].

In this study, the electrophilicity index (ω) was calculated to assess the electron-accepting capacity of the compounds. Higher ω values indicate stronger electrophilic character, while lower values reflect weaker electron-accepting ability [60].

Among the studied molecules, compounds 2 and 7 displayed the highest electrophilicity index (4.02 eV), highlighting their strong electron-accepting ability compared to the other compounds, whereas compound 8 showed a lower ω (2.74 eV), reflecting reduced electrophilic power and a lower tendency to accept electrons (Table 5).

3.3. Quantitative Structure-Activity Relationships Studies (QSAR) of bastadins

In this part, we aimed to build a reliable QSAR model to analyze and predict the biological activity (pIC_{50}) of bastadins. This allowed us to examine how physicochemical descriptors are related to biological response and to highlight the molecular features that influence activity. The application of mathematical and statistical approaches to experimental findings represents a valuable strategy for identifying novel compounds with strong inhibitory potential [5]. Our approach begins with Principal Component Analysis (PCA), which serves to reduce redundancy by selecting the most

relevant descriptors from a set of correlated variables. It is followed by a stepwise regression procedure, where descriptors without statistical significance are excluded, retaining only those contributing to a robust MLR model. (including the critical probability: p-value < 0.05 for all descriptors and the complete model).

Principal Component Analysis (PCA)

Among the 15 calculated molecular descriptors (Tables 4 and 5) that quantitatively characterize the 16 compounds, only those showing a strong correlation with the biological activity and a weak inter-correlation were selected to build a reliable and efficient multiple linear regression (MLR) model for predicting the inhibitory activity of bastadins. To achieve this, a principal component analysis (PCA) was carried out on all descriptors using Pearson's correlation matrix with the aid of XLSTAT software (Table 6).

Descriptors exhibiting strong intercorrelation ($|r| > 0.90$, based on Pearson's correlation matrix) were excluded from the dataset. Similarly, descriptors with negligible influence on the biological response, defined by correlation coefficients with inhibitory activity ($|r| \leq 0.1$), were also removed. Finally, from the remaining set, only those descriptors exhibiting the weakest interdependence with one another were retained for model construction (61.62).

The correlation coefficient between Pol and inhibitory activity was very low (r = 0.076), indicating that polarizability has no significant impact on the activity and was therefore excluded from the model.

LUMO and ω are perfectly correlated (r (LUMO, ω) = 0.998),

. E and MW are highly correlated (r (E, MW) = 0.975).

The correlation matrix revealed an almost perfect linear relationship between LUMO and ω (r = 0.998) as well as between E and MW (r = 0.975), indicating strong redundancy among these descriptors; therefore, one variable from each highly correlated pair was removed to minimize multicollinearity in the model.

. MW, MR, V and S are highly correlated (r (MW, MR) = 0,953; r (MW, V) = 0,905); r (MW, S) =0,825; r (MR, V) = 0.977; r (MR, S) = 0.862; r (S, V) = 0.932),

Descriptors MW, MR, V, and S exhibited very strong inter-correlations and were therefore all excluded from the dataset in order to prevent multicollinearity and improve the robustness of the model.

. TPSA and HE are correlated r (TPSA, HE) = 0.874),

A strong correlation was observed between descriptors TPSA and HE (r = 0.874). To prevent multicollinearity, one of the correlated descriptors was discarded from the dataset.

According to the analysis results, the selected descriptors were retained for building the multiple linear regression model: E, ELUMO, EHOMO, ELUMO+1, EHOMO-1, D, LogP and TPSA.

Table 6. *Correlation matrix (Pearson (n)) between different obtained descriptors*

Variables	**pIC50**	**LUMO**	**HOMO**	**Log p**	**E**	**MW**	**D**	**ω**	**LUMO+1**	**S**	**MR**	**HE**	**Pol**	**TPSA**	**V**	**HOMO-1**
pIC50	**1**															
LUMO	-0.458	**1**														
HOMO	**-0.558**	-0.226	**1**													
Log p	**0.528**	-0.471	-0.454	**1**												
E	-0.303	-0.083	0.347	0.341	**1**											
MW	0.262	-0.040	-0.192	-0.392	**-0.975**	**1**										

D	0.425	-0.106	**-0.651**	0.329	-0.405	0.317	**1**									
ω	0.455	**-0.998**	0.208	0.452	0.032	0.090	0.131	**1**								
LUMO+1	**-0.541**	**0.518**	-0.018	-0.430	-0.192	0.123	0.051	-0.484	**1**							
S	0.284	-0.297	-0.054	-0.242	**-0.743**	**0.825**	0.421	0.339	-0.130	**1**						
MR	0.115	-0.127	0.043	-0.489	**-0.864**	**0.953**	0.162	0.175	0.067	**0.862**	**1**					
HE	0.115	**0.527**	**-0.732**	0.064	-0.482	0.290	0.493	**-0.501**	0.475	-0.093	0.027	**1**				
Pol	0.076	**-0.123**	0.078	**-0.514**	**-0.850**	**0.938**	0.130	0.170	0.064	**0.851**	**0.999**	-0.001	1			
TPSA	-0.224	-0.478	**0.749**	-0.330	0.160	0.061	-0.472	0.473	-0.271	0.291	0.347	**-0.874**	0.340	**1**		
V	0.151	-0.253	0.097	-0.431	**-0.796**	**0.905**	0.204	0.298	-0.025	**0.932**	**0.977**	-0.108	**0.976**	0.429	**1**	
HOMO-1	-0.281	-0.244	**0.796**	-0.318	0.401	-0.278	**-0.519**	0.212	-0.155	-0.100	-0.100	**-0.719**	**-0.100**	**0.638**	-0.040	**1**

Multiple Linear Regressions (MLR)

The QSAR models were generated using the experimental antiangiogenic activity values (pIC50) of the 16 bastadins combined with the chemical descriptors selected by the PCA approach. The data set was randomly divided into two sets: a training set (thirteen compounds) and a testing set (three compounds: 5, 11, and 12)

Multiple Linear Regression (MLR) is one of the most widely used techniques in QSAR studies. The performance of an MLR model is generally evaluated using key statistical parameters, including the correlation coefficient (R), the coefficient of determination (R^2), the mean squared error (MSE), the mean calibration error (MCE), the Fisher's statistic (F), and the significance level (p-value). The following equation describes the best obtained linear QSAR model using the multiple linear regression (MLR) method:

$$\mathbf{pIC50} = 2.55 - 1.89\ 10^{-6}\ \mathbf{E} - 2.11\ \mathbf{E_{LUMO+1}} - 1.32\ \mathbf{E_{LUMO}} - 1.04\ 10^{-2}\ \mathbf{TPSA}$$

$$N = 16,\ R = 0.84,\ R^2 = 0.71,\ MSE = 0.158,\ MCE = 0.257,\ F = 4.709,\ P = 0.03$$

The multiple linear regression model showed a good overall fit, with a correlation coefficient of R = 0.84 and a coefficient of determination of $R^2 = 0.71$. This indicates that about 71% of the variance in the biological activity is explained by the selected descriptors. The prediction errors were also acceptable, with a mean squared error (MSE) of 0.158 and a mean calibration error (MCE) of 0.257, highlighting a good consistency between fitted and predicted activities.

The statistical significance of the model was supported by the Fisher test (F=4.709, p=0.03), showing that the regression is significant at the 95% confidence level and the relationship between the selected descriptors and the biological activity is not due to chance.

The negative coefficients of ELUMO+1, ELUMO, TPSA, and total energy indicate that increases in these properties are associated with a decrease in activity, with ELUMO+1 and ELUMO exerting the most significant

influence. The correlations of predicted and observed pIC50 values are illustrated in Figure 2.

Taken together, these results demonstrate that the proposed model is both statistically valid and reasonably predictive, providing a reliable basis for the analysis of structure–activity relationships in this series of compounds and guiding the design of new compounds in this series.

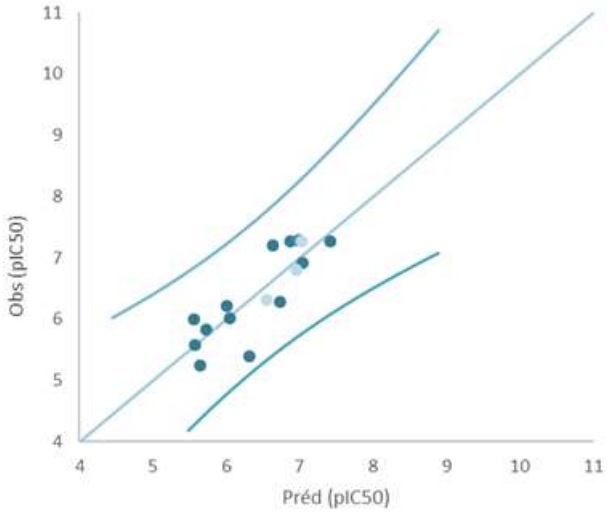

Figure 3. *Correlations of observed and predicted pIC50 values using MLR*

The potential multicollinearity among the four selected descriptors was assessed by computing their Variance Inflation Factors (VIFs), as presented in Table 7. The VIF for each descriptor is calculated using the formula $VIF=1/(1-R^2)$, where R represents the correlation coefficient between the descriptor under consideration and all other descriptors in the mode1.

Values of VIF exceeding 10 indicate strong multicollinearity, meaning that the variance of the estimated coefficient may be inflated and the descriptor may appear more influential than it actually is. Conversely, a VIF close to 1 suggests minimal correlation with other variables, reflecting a more stable and robust model [5].

Table 7: *The variance inflation factors (VIF) of descriptors in QSAR model*

	E	E_{LUMO+1}	E_{LUMO}	TPSA
Tolerance	0.941	0.736	0.596	0.694
VIF	1.063	1.358	1.676	1.440

The Variance Inflation Factors for the descriptors ranged from 1.063 to 1.676, indicating low multicollinearity and confirming that the model is robust, with the coefficients not significantly influenced by correlations among the variables.

The descriptors identified in the MLR model were subsequently employed as input variables for conducting multiple nonlinear regression (MNLR) analysis.

Multiples nonlinear regression (MNLR)

In addition to MLR, we employed a nonlinear regression approach to refine the structure–activity relationship and quantitatively assess the influence of substituents. The analysis was carried out on the data matrix derived from the descriptors previously selected by MLR for the 16 studied compounds. Model quality was evaluated based on statistical parameters such as the correlation coefficient (R), the coefficient of determination (R^2), and the mean squared error (MSE), which guided the selection of the most reliable regression model. The QSAR model established through the multiple linear regression method can be expressed as follows:

$$\mathbf{pIC50} = -0{,}16 - 7{,}93\,\mathbf{E_{LUMO}} - 6{,}29\,10^{-7}\,\mathbf{E} + 9{,}06\,\mathbf{E_{LUMO+1}} + 0{,}04\,\mathbf{TPSA} - 2{,}05\,\mathbf{E_{LUMO}^2} +$$

$$3{,}88\ 10^{-12}\ \mathbf{E^2} + 3{,}73\ \mathbf{E_{LUMO+1}}^{\mathbf{2}} - 1{,}11\ 10^{-4}\ \mathbf{TPSA^2}$$

N = 16, R = 0.95, R2 = 0.91, MSE = 0.048

The MNLR model demonstrated a high level of reliability, as indicated by the statistical parameters. The correlation coefficient (R = 0.95) reflects an excellent linear relationship between the predicted and experimental activities. The coefficient of determination (R^2 = 0.91) shows that more than 91% of the variance in biological activity is explained by the selected descriptors, highlighting the robustness of the model. The mean squared error (MSE = 0.048) further supports the accuracy of the predictions, as lower error values indicate better model performance. Collectively, these results suggest that the developed MNLR model is both statistically significant and predictive, making it suitable for structure–activity relationship studies. The correlations of predicted and observed pIC50 values are illustrated in Figure 3.

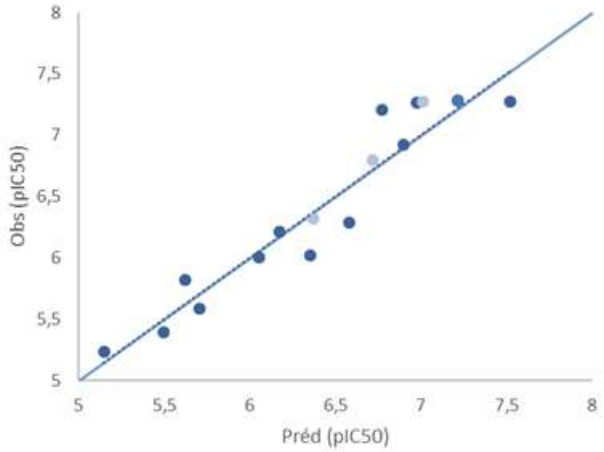

Figure 4. *Correlations of observed and predicted pIC50 values using MNLR*.

External validation

To evaluate the external validity of the constructed QSAR models, it is essential to test them on molecules that were not included in the training phase. In this study, the models generated from the dataset of 13 bastadins were subsequently applied to predict the biological activity of 3 test compounds. A comparison between the predicted pIC50 values and the experimentally measured ones indicates that the models provide reliable estimations for these test compounds (Table 10).

Table 8*: Comparative performance of the models generated through MLR and RNLM approaches*

	Training set			Test set		
	R	R^2	MSE	R	R^2	MSE
MLR	0.84	0.71	0.158	0.83	0.70	0.047
MNLR	0.95	0.91	0.048	0.91	0.83	0.026

The assessment of model performance revealed high predictive accuracy, with the test set coefficient of determination attaining 0.70 for the MLR model and 0.83 for the MNLR model. These values demonstrate that both statistical approaches provide acceptable reliability for the estimation of inhibitory activity based on the chosen descriptors, with MNLR showing slightly stronger predictive ability (Table 8).

Artificial Neural networks (ANN)

At this stage, a feed-forward artificial neural network was constructed, composed of four layers. The hidden layer used a sigmoid activation function, whereas the output layer applied a linear activation function. The adopted architecture corresponded to a 4-4-1 structure Figure 4).

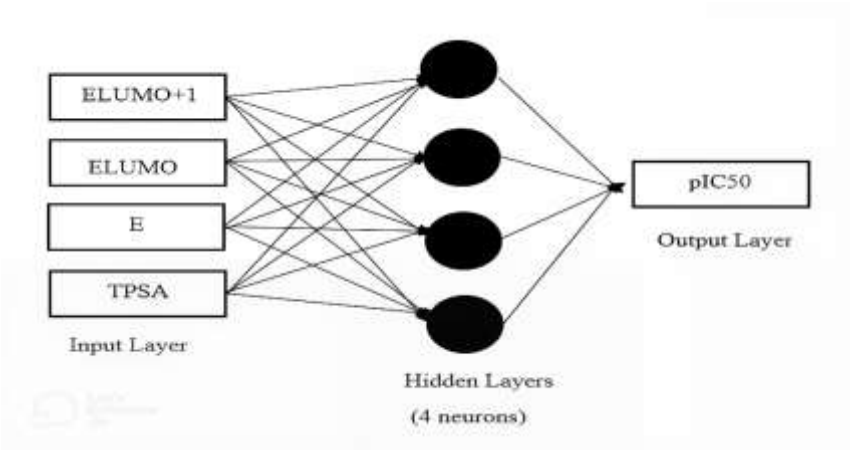

Figure 5. *Configuration [4-4-1] of the developed model and the four-layer artificial neural network's construction*

To evaluate the predictive performance of the ANN model, the dataset was randomly divided into three subsets. Approximately 70% of the compounds were assigned to the training set to optimize the network parameters, while 15% were used as a validation set to monitor the learning process and prevent overfitting. The remaining 15% was kept aside as an independent test set to assess the model's generalization ability. The strong agreement between experimental and predicted responses is evidenced by high correlation coefficients (R and R^2) together with very low mean squared error (MSE) values across training, validation, and test phases (Table 9). The relationship between the activities predicted by the artificial neural network and the experimental values is presented in Figure 5.

Table 9. *Statistical indicators (R^2, R, and MSE) describing the performance of the ANN model across the training, validation, and test phases.*

Model	Samples	MSE	R	R2
Training	12	0.0209	0.981	0.962
Validation	2	0.0302	1.000	1.000
Test	2	0.0031	1.000	1.000

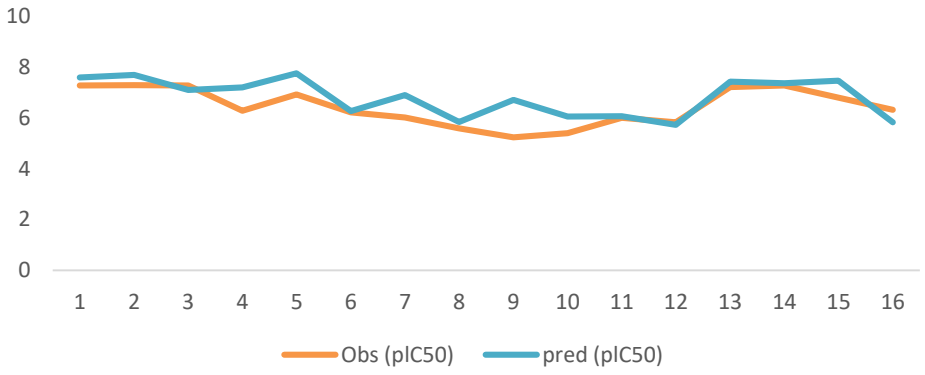

Figure 6. *Comparison of Predicted and Observed pIC50 Values calculated using ANN model*

The graph shows that predicted and observed pIC50 values are very close, meaning the ANN model can reliably estimate compound activity. Small differences are normal, the high R^2 and low error confirm strong accuracy. Overall, the model is effective for predicting pIC50 and can help prioritize promising drug candidates in discovery. In this section, the artificial neural network (ANN) model was subjected to validation through a cross-validation strategy, specifically the leave-one-out (LOO) procedure, carried out with MATLAB software.

In this method, one compound is removed from the dataset at a time, and its activity is predicted using the model built from the remaining compounds. The process is repeated until all molecules have been tested once, which allows an independent evaluation of the model's robustness and predictive ability [63].

The statistical parameters obtained from this validation serve as indicators of the predictive reliability and overall performance of the developed ANN model:

$Q^2_{LOO} = 0.79$ $MSE_{LOO} = 0.117$

The predictive ability was quantified by the cross-validated coefficient of determination **(Q^2_{LOO} = 0.79)** and the mean squared error obtained after LOO **(MSE_{LOO} = 0.117).** These results indicate that the model possesses satisfactory internal robustness and predictive reliability.

The comparison between the activities predicted through ANN leave-one-out cross-validation and the corresponding experimental values is illustrated in Figure 6.

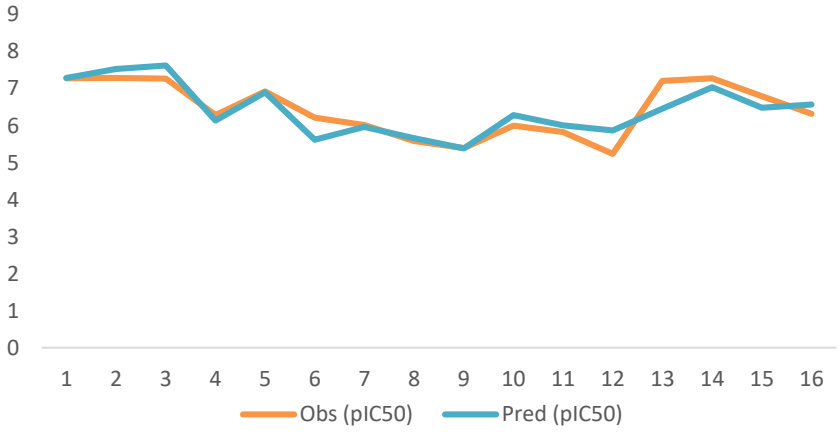

Figure 7. *Comparison of Predicted and Observed pIC50 Values calculated using ANN LOO-CV model*

Table 10: *Observed values and calculated of pI50 according to different methods.*

MOL	PIC50 (OBS)	MLR	MNLR	ANN	CV
		pIC50 (pred)	pIC50 (pred)	pIC50 (pred)	pIC50 (pred)
1	7.276	7.410	7.520	7.583	7.290
2	7.208	6.633	6.771	7.420	6.460
3	7.284	6.979	7.216	7.689	7.532
4	7.268	6.877	6.976	7.097	7.627
5*	7.276	7.030	7.010	7.362	7.032
6	6.284	6.727	6.580	7.202	6.139
7	6.921	7.033	6.900	7.753	6.894
8	5.398	6.316	5.498	6.045	5.387
9	6.000	5.558	6.054	6.060	6.285
10	6.215	6.008	6.174	6.266	5.620
11*	6.796	6.959	6.720	7.463	6.480
12*	6.319	6.552	6.368	5.824	6.570
13	5.824	5.719	5.621	5.721	6.012
14	5.585	5.566	5.704	5.8421	5.660
15	6.018	6.054	6.356	6.900	5.968
16	5.236	5.636	5.148	6.706	5.865

The graph shows a good agreement between the observed and predicted values under the ANN LOO validation. The two lines follow a very similar pattern, which means the model is able to capture the data behavior well. The small differences between them are normal and expected in predictive modeling. Overall, the close alignment confirms that the ANN model is reliable and effective for estimating pIC50 values, making it a useful tool in drug discovery for identifying promising compounds.Table 10 reports the experimental results with the predicted values obtained from MLR, MNLR, ANN, and cross-validation models.

4. Conclusion

In this work, Computational screening and SAR/QSAR analyses were performed to evaluate both the qualitative and quantitative influence of molecular structures on the anticancer activity of the compounds. Our findings indicate that this series of molecules complies with Lipinski's rules. Key physicochemical and electronic parameters, including Topological Polar Surface Area (TPSA), total energy (E), and Lowest Unoccupied Molecular Orbital (ELUMO), together with the neighboring orbital (ELUMO+1), were successfully employed to model the antiangiogenic activity of bastadins.

The different modeling approaches, MLR, MNLR, and ANN, were used to develop QSAR models. Among the developed models, the ANN showed the best predictive performance. The robustness and predictive ability of the models were further validated using independent training sets, test sets, and leave-one-out (LOO) cross-validation. Moreover, the predictions for the test compounds 5, 11, and 12 were in excellent agreement with the experimental results, confirming the reliability of the developed models.ve rall, the strong predictive performance, minimal residual errors, and successful cross-validation confirm that the proposed QSAR models are both reliable and effective for predicting the anticancer activity of this series of molecules. These results highlight the potential of bastadins as promising ACAT1 inhibitors for future computer-aided drug design and optimization aimed at developing innovative anticancer therapies.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Gibbs, J. B. (2000). Mechanism-based target identification and drug discovery in cancer research. Science, 287(5460), 1969–1973.

[2] Myint, K. Z., & Xie, X. Q. (2010). Recent advances in computational fragment-based drug design. International Journal of Molecular Sciences, 11(10), 3846–3866.

[3] Perkins, R., Fang, H., Tong, W., & Welsh, W. J. (2003). Quantitative structure–activity relationship methods: Perspectives on drug discovery and toxicology. Environmental Toxicology and Chemistry, 22(8), 1666–1679.

[4] Salum, L. B., & Andricopulo, A. D. (2009). Fragment-based QSAR: Perspectives in drug design. Molecular Diversity, 13(3), 277–285.

[5] El Mchichi, L., Aouidate, A., Chokrafi, F. Z., Ghaleb, A., Lakhlifi, T., Khalil, F., & Bouachrine, M. (2018). Prediction of biological activity of pyrazolo [3,4-b] quinolinyl acetamide by QSAR results. Rhazes: Green and Applied Chemistry, 3, 79–93.

[6] Ebada, S. S., & Proksch, P. (2012). The chemistry of marine sponges. In E. Fattorusso, W. Gerwick, & O. Taglialatela-Scafati (Eds.), Handbook of Marine Natural Products (pp. 191–293). Springer.

[7] Eguchi, K., Kato, H., Fujiwara, Y., Losung, F., Mangindaan, R. E. P., de Voogd, N. J., Takeya, M., & Tsukamoto, S. (2015). Bastadins, brominated-tyrosine derivatives, suppress accumulation of cholesterol ester in macrophages. Bioorganic & Medicinal Chemistry Letters, 25(22), 5389–5392.

[8] Sun, T., & Xiao, X. (2024). Targeting ACAT1 in cancer: From threat to treatment. Frontiers in Oncology, 14, 1395192.

[9]. Aoki, S., Cho, S.-H., Hiramatsu, A., Kotoku, N., & Kobayashi, M. (2006). Bastadins, cyclic tetramers of brominated-tyrosine derivatives, selectively inhibit the proliferation of endothelial cells. Journal of Natural Medicines, 60(3), 231–235.

[10]. Kotoku, N., Hiramatsu, A., Tsujita, H., Hirakawa, Y., Sanagawa, M., Aoki, S., & Kobayashi, M. (2008). Structure–activity relationship study of bastadin 6, an anti-angiogenic brominated-tyrosine derived metabolite from marine sponge. Archiv der Pharmazie: Chemistry in Life Sciences, 341(9), 568–577.

[11]. Hypercube, Inc. (2008). HyperChem (Molecular Modeling System) [Computer software]. Gainesville, FL: Hypercube, Inc.

[12]. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Petersson, G. A., Nakatsuji, H., Li, X., Caricato, M., Marenich, A. V., Bloino, J., Janesko, B. G., Gomperts, R., Mennucci, B., Hratchian, H. P., Ortiz, J. V., Izmaylov, A. F., … Fox, D. J. (2009). Gaussian 09 (Revision A.02) [Computer software]. Wallingford, CT: Gaussian, Inc.

[13]. Molinspiration Cheminformatics. (2019). Molinspiration Property Calculator. https://www.molinspiration.com/docu/mipc/index.html

[14]. Viswanadhan, V. N., Ghose, A. K., Revankar, G. R., & Robins, R. K. (1989). Prediction of hydrophobic (LogP) parameters from atomic contributions. Journal of Chemical Information and Computer Sciences, 29(3), 163–172.

[15]. Gavezzotti, A. (1983). The calculation of molecular volumes and the use of volume analysis in the investigation of structured media and of solid-state organic reactivity. Journal of the American Chemical Society, 105(16), 5220–5225.

[16]. Bodor, N., Gabanyi, Z., & Wong, C. K. (1989). A new method for the estimation of partition coefficient. Journal of the American Chemical Society, 111(11), 3783–3786.

[17]. Ooi, T., Obatake, M., Nemethy, G., & Scheraga, H. A. (1987). Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. Proceedings of the National Academy of Sciences of the United States of America, 84(10), 3086–3090.

[18]. Tanaka, M., Sugimoto, R., & Nakamura, N. (2025). Spontaneous orientation polarization driven by designing molecular asymmetry. Communications Materials, 6, 92.

[19]. Prasanna, S., & Doerksen, R. J. (2009). Topological polar surface area: A useful descriptor in 2D-QSAR. Current Medicinal Chemistry, 16(1), 21–41.

[20]. Tabani, H., Fernando, L. D., Heiss, C., & Azadi, P. (2025). A review of advanced strategies for molecular weight determination of insoluble polysaccharides: Developments and future trends. International Journal of Biological Macromolecules, 320(Pt 4), 146047.

[21]. Schneider, G., Baringhaus, K. H., & Kubinyi, H. (2008). Molecular design: Concepts and applications (277 p.). Weinheim, Germany: Wiley-VCH.

[22]. Cramer, C. J. (2004). Essentials of computational chemistry: Theories and models (2e éd., 624 p.). Chichester, UK: Wiley.

[23]. Lee, C., Yang, W., & Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Physical Review B, 37(2), 785–789.

[24]. Becke, A. D. (1993). Density-functional thermochemistry. III. The role of exact exchange. The Journal of Chemical Physics, 98(7), 5648–5652.

[25]. Hariharan, P. C., & Pople, J. A. (1973). The influence of polarization functions on molecular orbital hydrogenation energies. Theoretica Chimica Acta, 28(3), 213–222.

[26]. Segall, D. M. (2012). [Review of drug design]. Current Pharmaceutical Design, 18(9), 1292–1310.

[27]. Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2012). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews, 64(Suppl.), 4–17.

[28]. Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., & Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. Journal of Medicinal Chemistry, 45(12), 2615–2623.

[29]. Zerroug, A., Belaidi, S., Benbrahim, I., Sinha, L., & Chtita, S. (2019). Molecular docking, 3D-QSAR, and ADMET studies on new pyridazine derivatives as anti-cancer agents. Journal of King Saud University – Science, 31(4), 595–601.

[30]. Benet, L. Z., Hosey, C. M., Ursu, O., & Oprea, T. I. (2016). BDDCS, the Rule of 5 and drugability. Advanced Drug Delivery Reviews, 101, 89–98.

[31]. Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews, 46(1–3), 3–26.

[32]. Carr, R. A., Congreve, M., Murray, C. W., & Rees, D. C. (2005). Fragment-based lead discovery: Introduction and overview. Drug Discovery Today, 10(14), 987–992.

[33]. Mortenson, P. N., & Murray, C. W. (2011). Assessing the lipophilicity of fragments and early hits. Journal of Computer-Aided Molecular Design, 25(8), 663–667.

[34]. Schultes, S., Graaf, C., Haaksma, E., Iwan, J. P., & Kramer, O. (2010). Ligand efficiency as a guide in fragment hit selection and optimization. Drug Discovery Today: Technologies, 7(3), e157–e162.

[35]. Koubi, Y., Moukhliss, Y., Sbai, A., El Masaoudy, Y., Maghat, H., El-Mernissi, R., Ajana, M. A., Bouachrine, M., & Lakhlifi, T. (2019). Antimicrobial evaluation against Escherichia coli (MTCC 1652) by using 1,4-disubstituted 1,2,3-

triazole and derivatives: QSAR study. Arabian Journal of Chemical and Environmental Research, 6(1), 57–69.

[36]. Renckens, C. N. M. (2007). Alternatieve klinisch-chemische laboratoria: Leveranciers van onjuiste of niet-bestaande diagnosen. Nederlands Tijdschrift voor Geneeskunde, 151(51), 2816–2819.

[37]. He, G., Feng, L., & Chen, H. (2012). Safety management research based on system engineering theory. In Proceedings of the International Symposium on Safety Science and Engineering in China (pp. 204–209). Procedia Engineering, 43.

[38]. Aurélien, R. (2006). Détermination et caractérisation des cibles de DSP1 chez Drosophila melanogaster (Doctoral dissertation, Université d'Orléans).

[39]. Rouane, A., Tchouar, N., Belaidi, S., Kerassa, A., & Lanez, T. (2023). Chemical structure, substitution effect, and drug-likeness applied to quercetin and its derivatives. Journal of Fundamental and Applied Sciences, 15(3), 1–12.

[40]. Fadel, F. Z., Tchouar, N., Belaidi, S., Soualmia, F., Oukil, O., & Ouadah, K. (2021). Computational screening and QSAR study on a series of theophylline derivatives as ALDH1A1 inhibitors. Journal of Fundamental and Applied Sciences, 13(2), 1–12.

[41]. Doak, B. C., Over, B., Giordanetto, F., & Kihlberg, J. (2014). Oral druggable space beyond the rule of 5: Insights from drugs and clinical candidates. Chemistry & Biology, 21(9), 1115–1142.

[42]. Garcia Jimenez, D., Poongavanam, V., & Kihlberg, J. (2023). Macrocycles in drug discovery: Learning from the past for the future. Journal of Medicinal Chemistry, 66(8), 5602–5617.

[43]. Schultes, S., Graaf, C., Haaksma, E., Iwan, J. P., & Kramer, O. (2010). Ligand efficiency as a guide in fragment hit selection and optimization. Drug Discovery Today: Technologies, 7(3), e157–e162.

[44]. Bemis, G. W., & Murcko, M. A. (1996). The properties of known drugs. 1. Molecular frameworks. Journal of Medicinal Chemistry, 39(15), 2887–2893.

[45]. Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews, 23(1–3), 3–25.

[46]. Singh, T., Adekoya, O. A., & Jayaram, B. (2015). Protein–ligand interaction fingerprints: Generating insights into binding affinities and mechanisms. Molecular BioSystems, 11(4), 1041–1051.

[47]. Johnson, T. W., Gallego, R. A., & Edwards, M. P. (2018). Lipophilic efficiency as an important metric in drug design. Journal of Medicinal Chemistry, 61(15), 6401–6420.

[48]. Reynolds, C. H. (2015). Ligand efficiency metrics: Why all the fuss? Future Medicinal Chemistry, 7(11), 1363–1365.

[49]. Cramer, C. J. (2013). Essentials of computational chemistry: Theories and models (2nd ed., 624 p.). Chichester, UK; Hoboken, NJ: John Wiley & Sons.

[50]. Hinchliffe, A. (2003). Molecular modelling for beginners (2nd ed., 410 p.). Chichester, UK: John Wiley & Sons.

[51]. Leach, A. R. (2001). Molecular modelling: Principles and applications (2nd ed., 744 p.). Harlow, England: Pearson Education/Prentice Hall.

[52]. Atkins, P., & de Paula, J. (2014). Physical chemistry (10th ed., 972 p.). Oxford, UK: Oxford University Press.

[53]. Kerassa, A., Belaidi, S., Harkati, D., Lanez, T., Prasad, O., & Sinha, L. (2016). Theoretical studies on the electronic and structural properties of bioactive compounds. Reviews in Theoretical Science, 4(1), 85–96.

[54]. Almi, Z., Belaidi, S., & Segueni, L. (2015). Computational investigation of physicochemical descriptors in drug design. Reviews in Theoretical Science, 3(3), 264–272.

[55]. Lee, B., & Richards, F. M. (1971). The interpretation of protein structures: Estimation of static accessibility. Journal of Molecular Biology, 55(3), 379–400.

[56]. Andrasi, M., Buglyo, P., Zekany, L., & Gaspar, A. (2007). Determination of stability constants of metal complexes in solution by potentiometric titration. Journal of Pharmaceutical and Biomedical Analysis, 43(3), 1040–1046.

[57]. Ertl, P., Rohde, B., & Selzer, P. (2000). Fast calculation of molecular polar surface area as a sum of fragment-based contributions. Journal of Medicinal Chemistry, 43(20), 3714–3717.

[58]. Viswanadhan, V. N., Ghose, A. K., Revankar, G. R., & Robins, R. K. (1989). Prediction of hydrophobic (log P) parameters from atomic contributions. Journal of Chemical Information and Computer Sciences, 29(3), 163–172

[59]. Fukui, K. (1982). Role of frontier orbitals in chemical reactions. Science, 218(4574), 747–754.

[60]. Parr, R. G., & Pearson, R. G. (1983). Absolute hardness: Companion parameter to absolute electronegativity. Journal of the American Chemical Society, 105(26), 7512–7516.

[61]. Tabti, K., Sbai, A., Maghat, H., Lakhlifi, T., & Bouachrine, M. (2024). Computational assessment of the reactivity and pharmaceutical potential of novel triazole derivatives: An approach combining DFT calculations, molecular dynamics simulations, and molecular docking. Arabian Journal of Chemistry, 17(1), 105376.

[62]. Vidal, R., Ma, Y., & Sastry, S. S. (2016). Principal component analysis. In Generalized principal component analysis (pp. 25–62). New York, NY: Springer.

[63]. Larif, M., Chtita, S., Adad, A., Hmamouchi, R., Bouachrine, M., & Lakhlifi, T. (2014). Predicting biological activity of chalcone (1,3-diphenyl-2-propen-1-one) derivatives: Cytotoxicity against HT-29 human colon adenocarcinoma cell lines by DFT-QSAR models. Journal of Computational Methods in Molecular Design, 4(4), 121–130.