

Lipophilicity and Biological Activity Study of Several Caspofungin Antifungal Drugs Using QSAR and Monte Carlo Methods

Nikkar, Maria; Sayyadi Kordabadi, Robabeh^{*+}; Alizadeh, Omid; Ghasemi, Ghasem

Department of Chemistry and Chemical Engineering, Rasht Branch, Islamic Azad University, Rasht, I.R. IRAN

ABSTRACT: QSAR investigations of Caspofungin derivatives were conducted using Multiple Linear Regression (MLR), Artificial Neural Network (ANN), and Monte Carlo Methods. The obtained results were compared and GA-ANN and ICA-MLR combinations showed the best performance according to their correlation coefficient (R^2) and Root Mean Sum Square Errors (RMSE). The most important physicochemical and structural descriptors were presented and discussed. Monte Carlo method revealed that the presence of a double bond with branching, a six-member cycle, the absence of halogens, the presence of sp^2 carbon connected to branching, the presence of Nitrogen and Oxygen atoms, absence of Sulphur and Phosphorus are the most important molecular features. The best Caspofungin derivative was exposed to reaction with Cu, Zn, Fe using B3lyp/6-311g/lanl2dz to investigate the stability of the formed complexes, from which the Zn complex was perceived to be the most stable one. It was concluded that QSAR study and the Monte Carlo method can lead to a more comprehensive understanding of the relation between physicochemical, structural, or theoretical molecular descriptors of drugs to their biological activities and Lipophilicity.

KEYWORDS: Caspofungin drugs; QSAR; Genetic Algorithm; Monte Carlo method.

INTRODUCTION

Caspofungin is the first member of Echinocandins that has been approved as first-line therapy for invasive candidiasis and as salvage therapy for invasive aspergillosis [1]. Caspofungin inhibits the enzyme β (1, 3)-D-glucan synthase of the fungal cell wall in a non-competitive manner leading to inhibition of the synthesis of β (1, 3)-D-glucan. The β (1, 3)-D-glucan is a crucial element of the cell wall of many fungal species and forms a solid three-dimensional matrix that is involved in shape determination and mechanical strength of the cell wall [2-4].

The logarithm of the partition coefficient between n-octanol and water also referred to as logP, has been widely used in Quantitative Structure-Activity Relationship (QSAR) studies as a key parameter for characterizing lipophilicity. Lipophilicity is described by partition processes taking place between two phases: non-polar (organic) and polar (typically water). XLOGP3 is a new method for the fast calculation of logP [5, 6].

The partition coefficient is hydrophilic ("water-loving") or hydrophobic ("water-fearing") scale a

* To whom correspondence should be addressed.

+E-mail address: sayyadi@iaurasht.ac.ir; sayyadi_04@yahoo.com

1021-9986/2022/12/4056-4074

19/\$/6.09

chemical substance and it is useful in estimating the distribution of drugs within the body. Hydrophobic drugs with high octanol-water partition coefficients are mainly distributed to hydrophobic areas such as lipid bilayers of cells. Conversely, hydrophilic drugs (low octanol/water partition coefficients) are found primarily in aqueous regions such as blood serum [7, 8].

The biological activity of a chemical compound is directly related to its physical properties. Lipophilicity is the most useful parameter in the assessment of biological activity of substances and in the predicting their toxic activity [9,10].

Modelling and optimization approaches that relate the descriptors (constitutional, geometrical, topological, quantum chemical, etc.) to the biological activity of drugs are named QSAR [11, 12]. Since logP can be predicted computationally, it can be included in prediction models as a molecular descriptor. Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), Simulated Annealing algorithm (SA) [13], Genetic Algorithm (GA) [14], and Partial Least Squares (PLS), are the most common mathematical methods that have utilized to describe the quantitative relationship between the molecular descriptors of the drugs and their properties [15,16].

CORAL has been proposed as a competent software for the QSAR studies. It uses Monte Carlo method to find the most important Simplified Molecular Input-Line Entry System (SMILES)-based descriptors and calculate their correlation weights to predict an endpoint (e.g., $-\log(\text{IC}_{50})$). In simple words, SMILES are lines of symbols, representing the molecular structure [17].

The possibility of complex formation with important metals in the body, including copper, iron and zinc and the complex's quality is considered as one of the theoretical approaches to investigate the function of a drug in vivo. The formation of a strong complex with a metal may indicate a decrease in the concentration of that metal by entering the drug into the body, in which case it is recommended to present the complex form of the drug with that metal instead of its basic form [18].

Imperialist Competitive Algorithm (ICA) is a new population-based optimization algorithm that was proposed by *Atashpaz-Gargari* and *Lucas* in 2007 [19] and since then it was employed in solving a variety of optimization problems [20-22]. The algorithm, starts with an initial population and each member of the population is called a country. The members (countries) are two

types: imperialists and colonies. The most powerful countries are selected as imperialists and the rest as the colonies of these imperialists [23].

In the present work, MLR and ANN modelling tools coupled with SA, GA and ICA optimization techniques and the Monte Carlo method were used to find the best set of descriptors that correlate the half maximal inhibitory concentration of Caspofungin drugs. In addition, the formation of complexes of the most effective Caspofungin derivative with Cu, Fe and Zn metals was investigated.

THEORETICAL SECTION

Linear and non-linear methods

Geometry optimizations of Caspofungin compounds were carried out using the B3lyp/6-311g at the Gaussian 03W [24] and then Dragon program was used for calculation of 3226 molecular descriptors for each of the 34 compounds and then SPSS program [25] was used to reduce the number of descriptors [26] to 1259 with the dependent variables biological activity ($-\log(\text{IC}_{50})$) and lipophilicity (XLOGP), respectively. Then a stepwise multiple linear regression procedure, was employed to select the best descriptors of the 1259 descriptors. Low standard deviation, least numbers of independent variables, high ability for prediction [27], high correlation coefficient (R) and RMSE are characteristics of an ideal model, where the RMSE is defined as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - y_o)^2}{n}} \quad (1)$$

In Equation (1), y_i is the desired output, y_o is the predicted value by the model, and n is the number of molecules in the data set.

In GA-ANN and SA-ANN QSAR methods, 1259 descriptors were considered as possible input of the ANN and fed into the input layer of the ANNs. The neural networks were all three-layer and Levenberg-Marquart algorithm [21] was applied for training of the networks. Modelling and optimization calculations were carried out using Matlab. 7.12.

As another methodology, the 1259 SPSS [25] screened descriptors were used as the feed to a MLR-ICA approach as the population matrix in order to find the six best descriptors for the gas phase. The developed algorithm of this work is depicted in Fig. 1. The ICA procedure begins from random points (matrix indices of descriptors) called

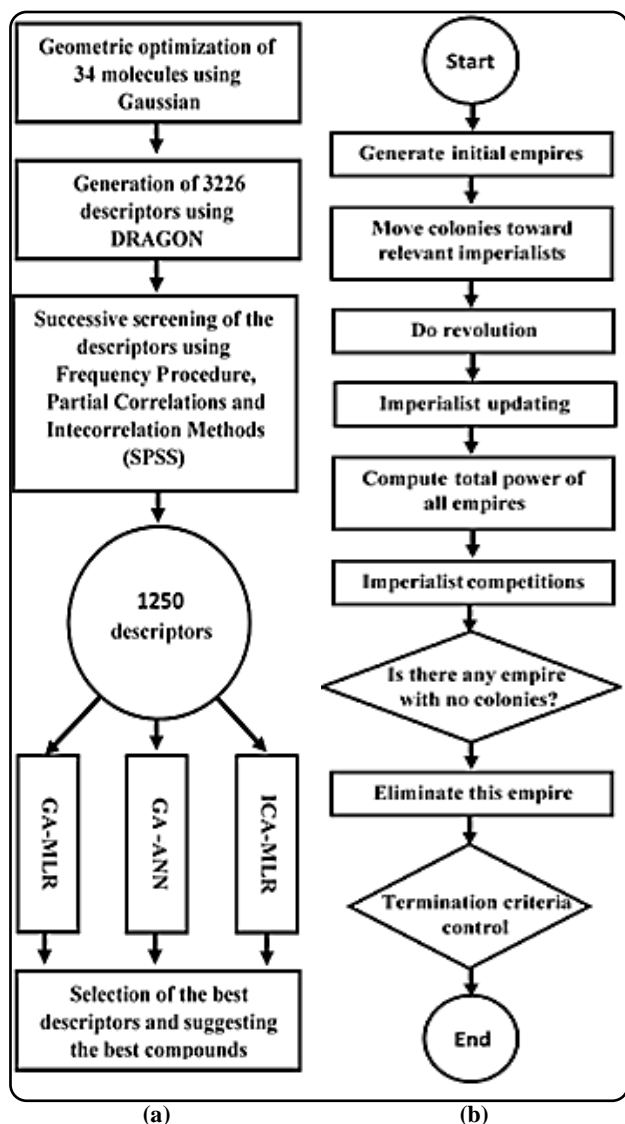


Fig. 1: Flowchart of the employed algorithms: (a) the general methodology (b) ICA algorithm.

the initial countries that are the counterpart of chromosomes in GA and it is a set of values of a candidate solution for the optimization problem [28]. The empires are sub-populations of the countries. Assimilation, which can be considered as a primitive form of Particle Swarm Optimization [29-31] moves all non-best countries (called colonies) in an empire toward the best country (called imperialist) in the same empire to find the colonies with the lowest error (RMSE of MLR-predicted $-\log(\text{IC}_{50})$ and XLOGP (lipophilicity) versus the empirical values). Different numbers of decision variables (nDes) and different numbers of empires (nEmp) were investigated to obtain the least RMSE and highest R^2 . The number of decision

variables (nDes) and a number of empires/ imperialists (nEmp) were considered 6 and 10, 20, 30, respectively.

In non-linear methods, 70%, 15%, and 15% of data sets were randomly chosen as training, validation, and test sets, respectively. The empirical values of $-\log(\text{IC}_{50})$ and lipophilicity ($\log P$) for these compounds were obtained from Pubchem [26]. The concept of IC_{50} is fundamental to pharmacology and it is the concentration of an inhibitor where the response (or binding) is reduced by half.

Monte Carlo method

CORAL [32] software was used for the calculation of Descriptor Correlation Weight (DCW) of the 34 Caspofungin compounds with a hybrid optimization scheme including Hydrogen-Suppressed molecular Graph (HSG), Hydrogen-Filled Graphs (HFG), and SMILES representation of molecular structures. Modeling using CORAL software was carried out for thresholds of 1 up to 3 and 100 and 50 epochs with dependent variables biological activity ($-\log(\text{IC}_{50})$) and lipophilicity (XLOGP), respectively [33]. The SMILES-based and Graph-based optimal descriptors were used from the other work [34]. The hybrid objective function for finding the optimal descriptors is defined as:

$$\text{DCW}(T, \text{Nepoch})^{\text{Hybrid}} = \text{DCW}(T, \text{Nepoch})^{\text{SMILES}} + \text{DCW}(T, \text{Nepoch})^{\text{Graph}} \quad (2)$$

Formation of Complexes

Active sites of the best Caspofungin derivative including N site were exposed to reaction with Cu, Fe and Zn and geometrical optimization of the metal complexes was carried out with B3lyp/6-311g/lanl2DZ at the Gaussian 03W [24,25, 35].

RESULTS AND DISCUSSION

Linear and non-linear combination methods

All the optimized structures of the Caspofungin compounds are presented in Fig. 2.

The RMSE and the correlation coefficient (R^2) in MLR-PCR, MLR-PLS1 and MLR-MLR for the predicted activity were found to be [0.2881 0.8461], [0.2832 0.8512] and [0.2334 0.8989], for biological activity and [0.2605 0.9632], [0.2576 0.9641] and [0.1593 0.9862], for lipophilicity, respectively. Furthermore, the calculated parameters indicated that MLR-MLR method was better than all other employed linear methods (MLR-PLS1 and MLR-PCR).

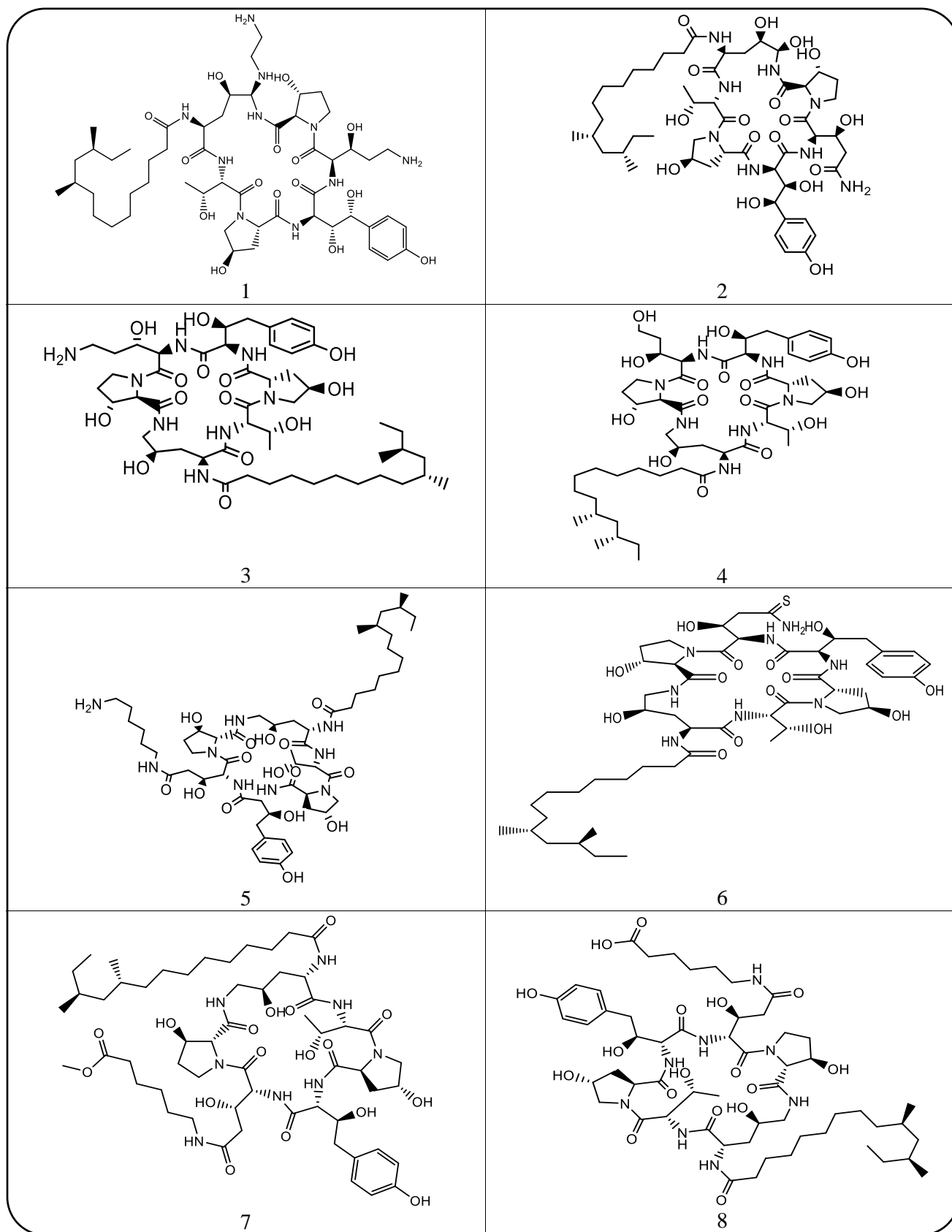
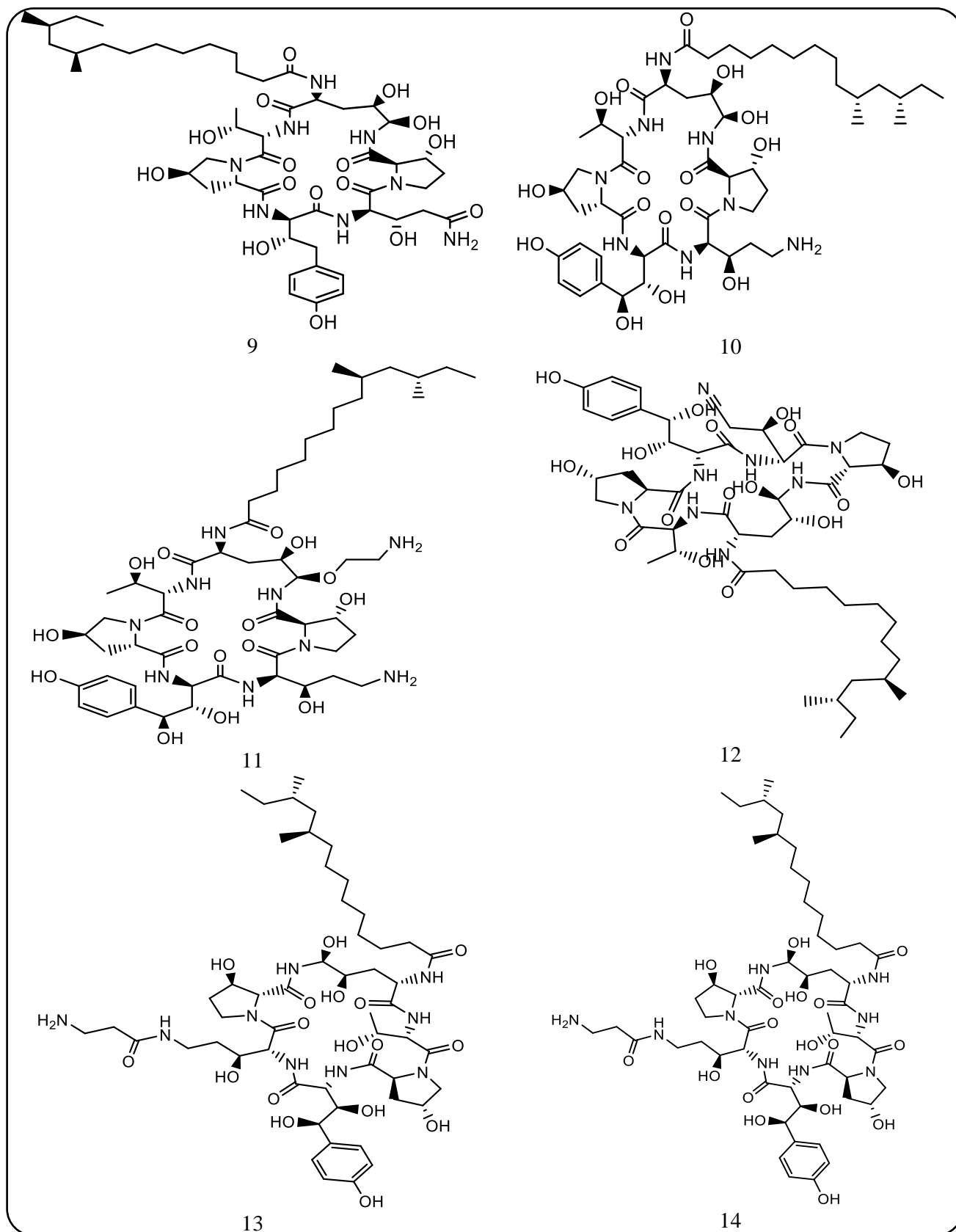
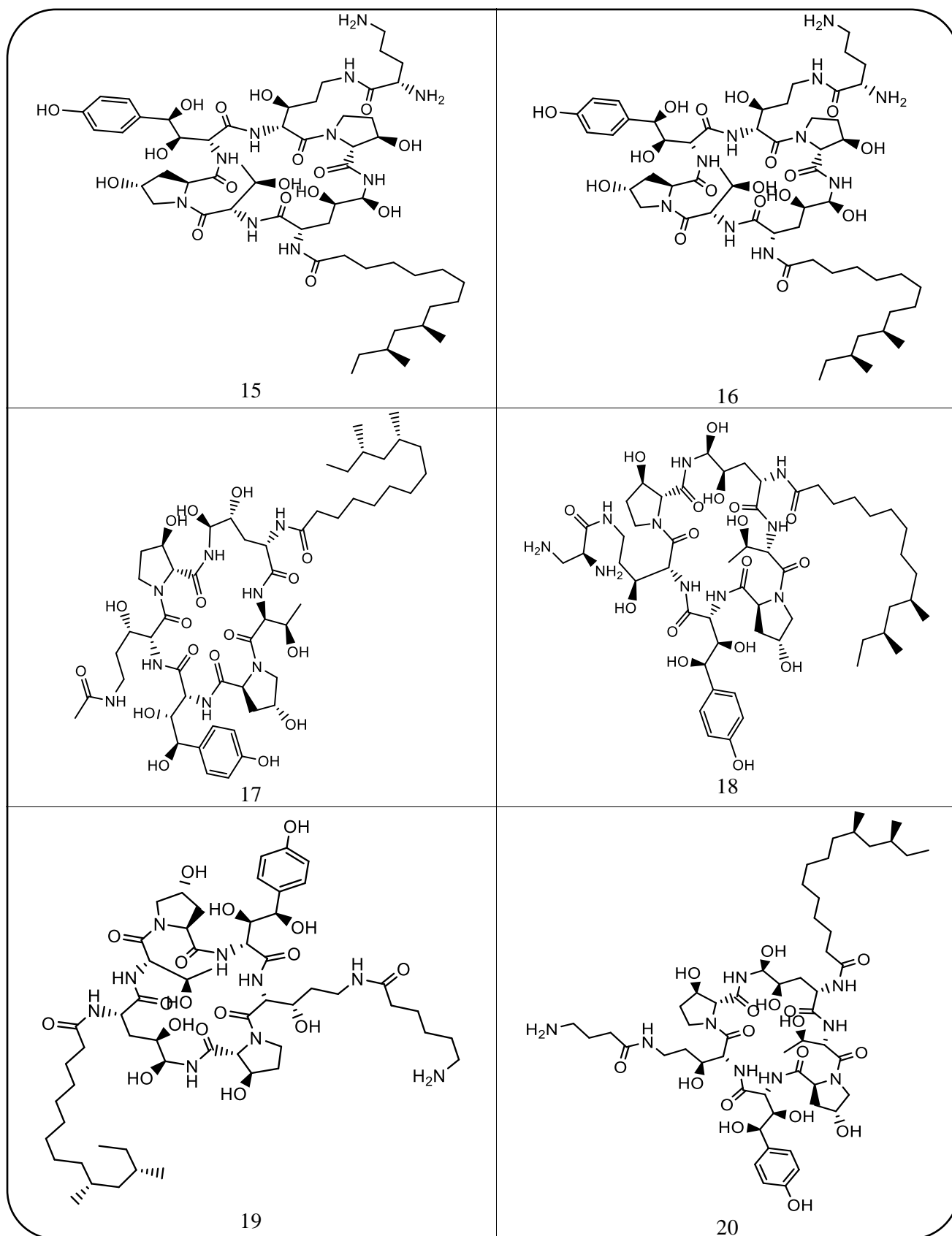


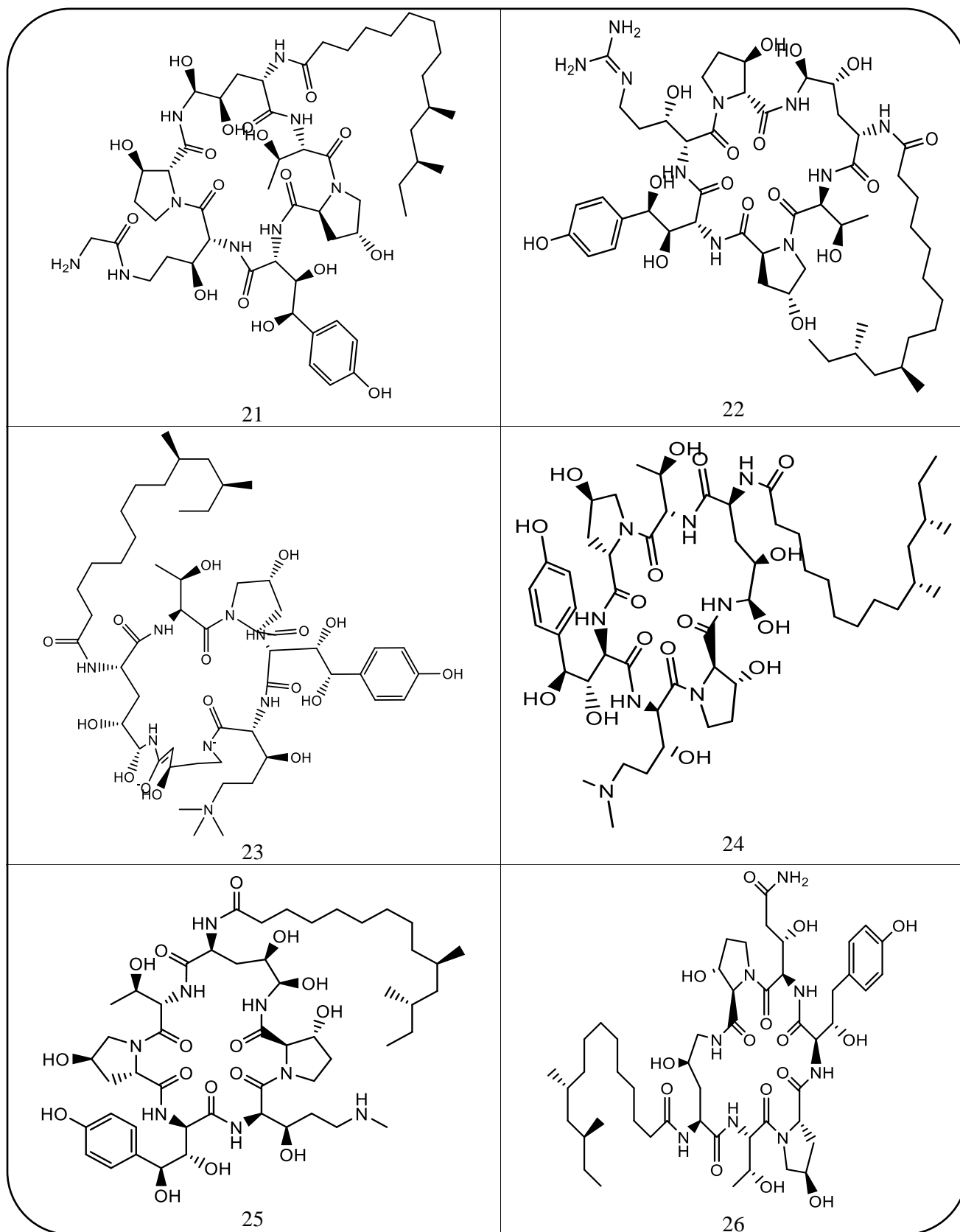
Fig. 2: Optimized structure of the Caspofungin derivatives used to build QSAR models with B3lyp/6-311g in gas phase.



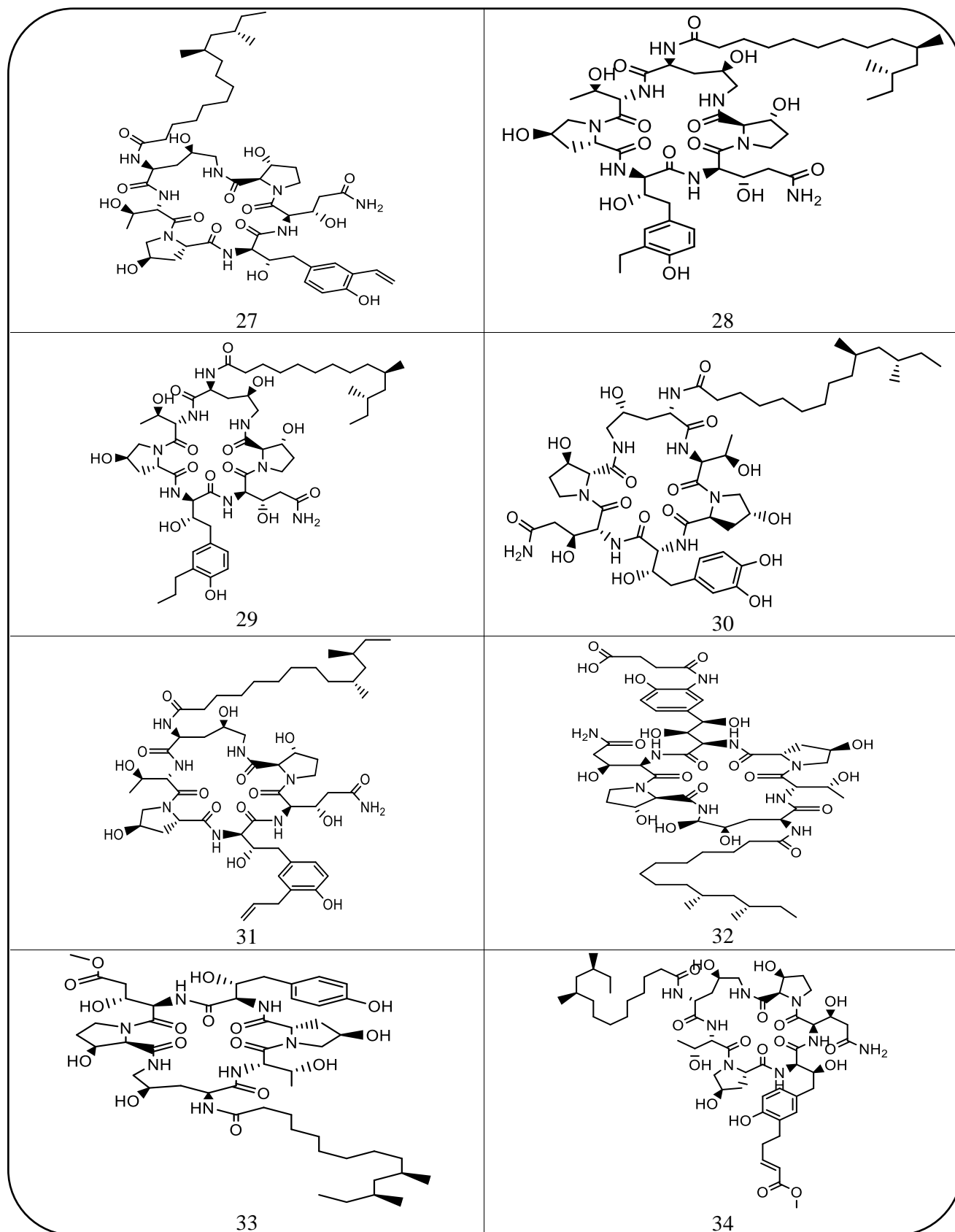
Continuation of the Fig. 2: Optimized structure of the Caspofungin derivatives used to build QSAR models with B3lyp/6-311g in gas phase.



Continuation of the Fig. 2: Optimized structure of the Caspofungin derivatives used to build QSAR models with B3lyp/6-311g in gas phase.



Continuation of the Fig. 2: Optimized structure of the Caspofungin derivatives used to build QSAR models with B3lyp/6-311g in gas phase.



Continuation of the Fig. 2: Optimized structure of the Caspofungin derivatives used to build QSAR models with B3lyp/6-311g in gas phase.

Table 1: Test and valid series in QSAR models*

QSAR Models	Biological activity		lipophilicity	
	Test series	valid series	Test series	valid series
MLR-SA	9,26,29,30,34	4,10,14,15,22	3,5,20,23,32	2,7,14,20,23
SA-ANN	3,5,10,19,25	2,4,12,24,28	1,9,10,12,13	4,19,10,13,12
MLR-GA	4,6,22,24,25	5,11,15,23,26	3,5,20,23,32	2,12,15,19,27
GA-ANN	4,9,19,25,34	2,7,14,20,23	1,9,10,12,13	4,19,23,26,32

The naming are according to Fig. 2.

The selection procedure of the most effective descriptors by employing SA-MLR, SA-ANN, MLR-GA and ANN-GA resulted [RMSE R²] parameters for the predicted biological activity as [0.3244 0.9107], [0.2144 0.9198], [0.19743 0.9278] and [0.18323 0.9504], while these parameters for the predicted lipophilicity were obtained to be [0.32435 0.9436], [0.38993 0.9180], [0.12655 0.9922] and [0.19464 0.9943], respectively for the mentioned approaches. Table 1 shows test and valid series in nonlinear models in compounds in predicted biological activity and predicted lipophilicity, respectively.

The obtained results demonstrated that the GA-ANN method led to better results predictive ability than the other QSAR models. Therefore, the selected descriptors using GA-ANN are discussed here.

The most effective descriptors turned out to be ESpm11x, RDF135e, Du, JGI3, DP05 and ESpm04x for the biological activity and ESpm07d, G(N...N), ALOGP for the lipophilicity. ESpm11x (Spectral moment 11 from edge adj. matrix Weighted by edge degrees/ Edge adjacency indices) and ESpm04x (Spectral moment 04 from edge adj. matrix Weighted by Edge degrees/ Edge adjacency indices) and ESpm07d (Spectral moment 08 from edge adj. matrix weighted by dipole moments) are edge adjacency indices and these descriptors were used in molecular graphs. Molecules as weighted graphs were used for calculation of the novel index, in which the elements of edges set were substituted by the bond orders between connected atoms in the molecule [27]. RDF135e (Radial distribution function-13.5 / Weighted by atomic sanderson electronegativities) is RDF descriptor and these descriptors are independent on the number of atoms, i.e; the size of a molecule, it is unique regarding the three-dimensional arrangement of the atoms, and it is invariant against translational and rotational entire molecule [30]. Du (D total accessibility index / unweighted) is WHIM

descriptor that was built in such a way as to capture the relevant molecular 3D information regarding molecular size, shape, symmetry, and atom distribution with respect to invariant reference frames [27]. JGI3 (mean topological charge index of order 3) is topological charge indices and DP05 (molecular profile no. 5) is Randic molecular profiles. The Topological Charge Indices were proposed to evaluate the charge transfer between pairs of atoms and therefore, the global charge transfer in the molecule [27]. The Randic molecular profile DP_k is derived from the distance distribution moments of the geometric matrix G as the average row sum of its entries raised to the kth power and normalized by the factor k! [27]. G(N...N) (sum of geometrical distances between N...N) is a geometrical descriptor, that incorporate information on the magnitude of the displacement between the polarized field (center of charge) and the molecular centroid (center of mass) [27]. ALOGP (Ghose-Crippen octanol-water partition coeff. (logP) is the Molecular properties descriptor that is one method for the calculation of the n-octanol/water partition coefficient based on similarities in the structure or properties of chemical compounds [27].

The predicted values of -logIC₅₀ and (XLOGP) using the GA-ANN are plotted against the observed values in Figure 3, which indicates a very strong agreement.

The correlation coefficient (R²) of the predicted set in GA-ANN model is acceptable in comparison to previous works [28,29].

A sensitivity analysis was carried out to investigate the effects of ESpm11x, RDF135e, Du, JGI3, DP05, and ESpm04x descriptors on the negative logarithm half maximal inhibitory concentration.

The obtained results demonstrated that [range Optimum value/range] parameters for the ESpm11x, RDF135e, Du, JGI3, DP05, and ESpm04x descriptor as [17.884-17.649 maximum], [140.034-24.877 maximum],

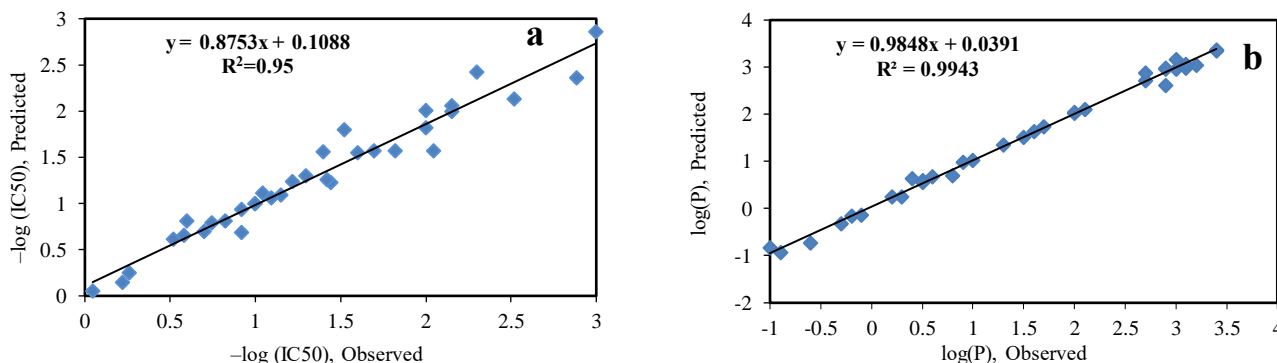


Fig. 3: Predicted versus observed values of a) $-\log(\text{IC}_{50})$ and b) lipophilicity (XLOGP) of Caspofungin complexes using GA-ANN approach.

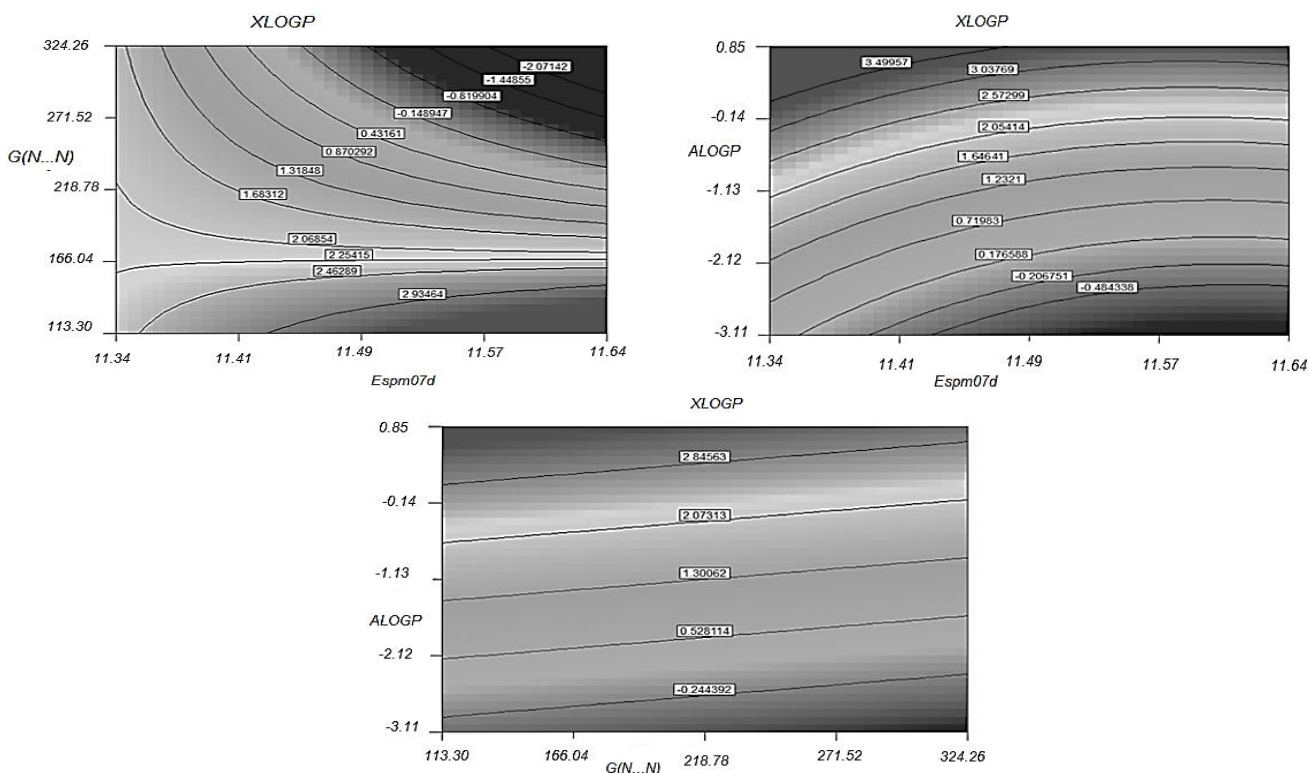


Fig. 4: The Contour plots of the ESpm07d, $G(N\dots N)$ and ALOGP descriptors versus XLOGP.

[0.512-0.379 maximum], [0.056-0.052 maximum], [12.583-10.209 0-12.583], [8.344-8.153 maximum], respectively for the predicted biological activity in GA-ANN method.

Therefore, for designing new drugs, ESpm11x, RDF135e, Du, JGI3 and ESpm04x descriptors in predicted biological activity are recommended to be at their maximum value and DP05 descriptor in the range of 0 to 12.583.

The contour plots of the ESpm07d, $G(N\dots N)$ and ALOGP descriptors versus XLOGP were plotted using the Design-Expert 7.00 Trial in GA-ANN method [32] (Figure 4).

Figure 4 shows the interaction effects of the ESpm08d and $G(N\dots N)$ descriptors on XLOGP. A large and negative

interaction effect is observed, so that increasing ESpm08d at low $G(N\dots N)$ values increases XLOGP.

While, at high $G(N\dots N)$ values, increasing ESpm07d decreases XLOGP. At the same time, the $G(N\dots N)$ descriptor, lonely has a decreasing effect on XLOGP throughout the studied conditions: as $G(N\dots N)$ increases, XLOGP decreases, and this decline is more severe at higher values of ESpm07d.

With growing ALOGP, the amount of XLOGP increases and changing of ESpm07d in the studied range, does not have a significant effect on this trend. In addition, ESpm08d has a declining effect on XLOGP in the shown range.

Table 2: Statistical parameters of ICA-MLR models with different nEmp (Max.It=500).

nVar- nEmp	Predicted (Biological Activity)		Predicted (lipophilicity)	
	R ²	RMSE	R ²	RMSE
6-10	0.9209	0.2064		
6-20	0.9441	0.1736		
6-30	0.9348	0.1874		
6-40	0.9384	0.1822		
3-10			0.9896	0.1387
3-20			0.9913	0.1265
3-30			0.9913	0.1265
3-40			0.9913	0.1265

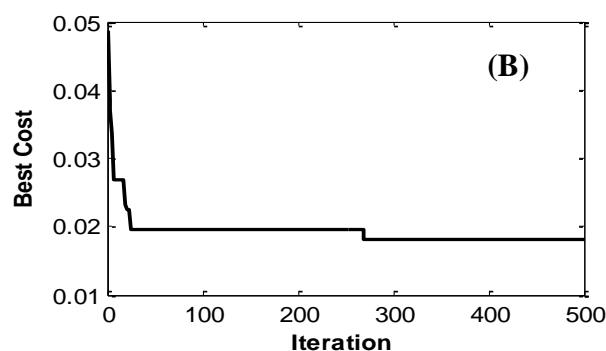
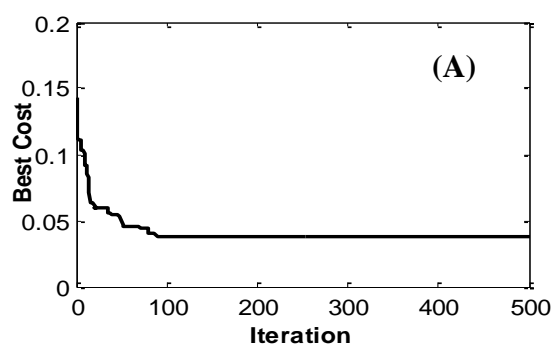


Fig. 5: Plot between best cost values versus the variation of Iteration A) predicted biological activity B) lipophilicity (XLOGP) predicted.

In other words, the interaction effects of the Espm08d and AIOGP descriptors in the studied conditions were negligible.

The two descriptors G(N...N) and ALOGP have no significant interaction effects. Throughout the study range, ALOGP has an increasing effect and G(N...N) has a decreasing effect on XLOGP (i.e., XLOGP raises with growing of ALOGP and reducing of G(N...N)).

Molecular descriptors generation with MLR-ICA approach

As a first trial, 500 number of iterations were done to find the most powerful empires and, subsequently, the best descriptors. A plot of the best cost values versus the number of iterations is represented in Fig. 5. It implies that there is no variation in the best cost (MSE) after about 300 iterations. However, in order to ensure that the best descriptors are captured, the number of iterations for the rest of computations was set to 500 in the $-\log IC_{50}$ and lipophilicity (XLOGP) calculations, respectively.

In order to choose the most suitable number of empires, the model was run using different numbers of empires and the results are demonstrated in Table 2. According to this table, the optimum number of empires was chosen as 20 in the $-\log IC_{50}$ and lipophilicity (XLOGP) calculations, respectively.

Plots of the predicted versus empirical values of $-\log IC_{50}$ and XLOGP are depicted in Figure 6. The figure implies that the developed model possesses a high correlation coefficient, indicating that the experimental and predicted values are well correlated.

The best-selected descriptors using MLR-ICA Method with nDes=6 and nEmp=20 are CIC2 (Complementary information content index (neighborhood symmetry of 2-order)/ Information indices), DISPp (dCoMMAX value/ Weighted by atomic polarizabilities), ESpm05x (spectral moment 05 from edge adj. matrix weighted by edge degrees/edge adjacency indices), EEig14r (eigenvalue n.14 from edge adj. matrix weighted by resonance integrals/edge adjacency indices), F05[C-N] (frequency

Table 3: Optimum value/range descriptors in ICA-MLR method in gas phase

Descriptor (lipophilicity)	range	Optimum value/range
CIC2	2.065-2.509	minimum
DISPp	0.144-0.487	minimum
Espm05x	9.449-9.65	minimum
EEig14r	3.024-3.131	minimum
F05[C-N]	25-30	25-29.5
RDF015m	52.218-60.802	minimum

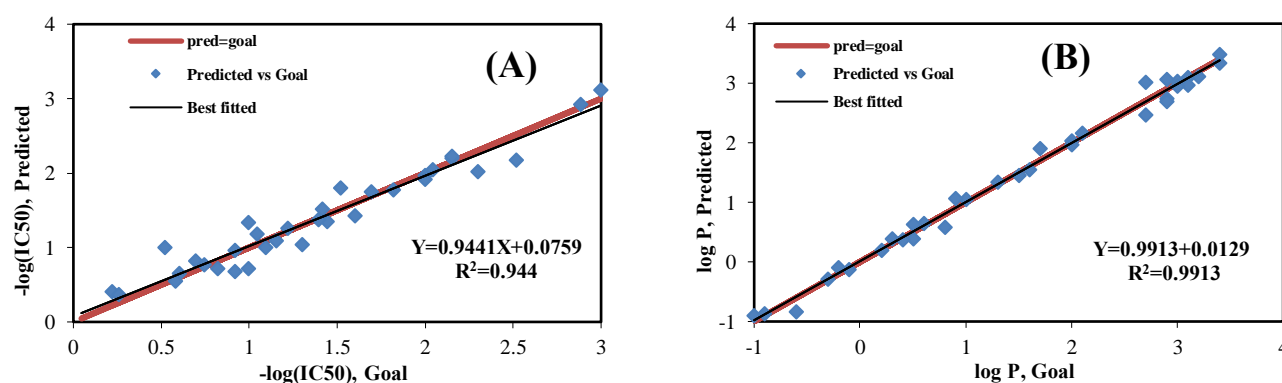


Fig. 6: A) Plot between predicted values versus Goal with $nVar=6$ and $nEmp=20$, for biological activity. B) Plot between predicted values versus Goal of XLOGP with $nVar=3$ and $nEmp=20$.

of C - N at topological distance 5/2D frequency fingerprints), RDF015m (radial Distribution Function – 1.5/ weighted by atomic masses/RDF descriptors [27]) with the dependent variables biological activity ($-\log(\text{IC}_{50})$).

CIC2 is information indices. The total information content (I) is obtained by multiplying the mean information content by the number of elements [31].

Optimum value/range descriptors in ICA-MLR method are depicted in table 3 with biological activity ($-\log(\text{IC}_{50})$) values. According to this table, for designing new drugs, CIC2, DISPp, ESPm05x, EEig14r descriptors are recommended to be at their minimum value, while a range of values between 25 and 29.5 is the best for F05[C-N].

The best selected descriptors using MLR-ICA Method in lipophilicity (XLOGP) predicted with $nDes=3$ and $nEmp=20$ are EEig01x (Eigenvalue 01 from edge adj. matrix weighted by edge degrees/ Edge adjacency indices), EEig12x (Eigenvalue 12 from edge adj. matrix weighted by edge degrees/ Edge adjacency indices) and

ALOGP (Ghose-Crippen octanol-water partition coeff. ($\log P$)/ Molecular properties).

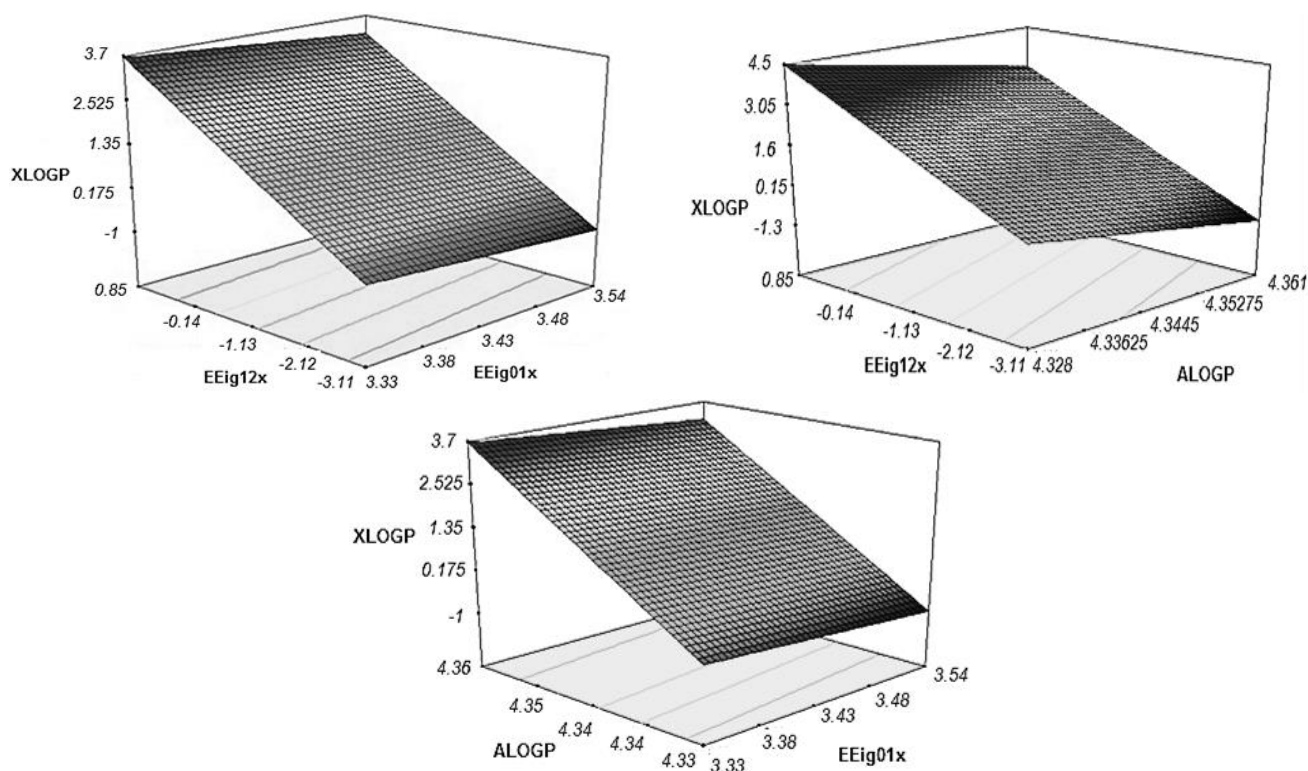
The two-dimensional plots of the XLOGP versus EEig01x, EEig12x and ALOGP descriptors are plotted using the Design-Expert 7.00 Trial [29] (Fig. 7). It was shown that the XLOGP value grows with EEig12x and ALOGP descriptors. While it reduces with EEig01 descriptor.

Result of the Monte Carlo method

The statistical parameters of the models obtained using molecular graphs (HSG) and SMILES are shown in Table 4. The performance of the models was compared with each other by the criterion of the predictability in test set (R_m^2) which should be larger than 0.5, correlation coefficient (R^2) in each set, cross-validated correlation coefficient (Q^2) [36]. The results showed that for of the three splits, the threshold of one gives the best results. The results with thresholds of 2, 1 for the probe 2, 1 including the function of $-\log(\text{IC}_{50})$ and lipophilicity (XLOGP) in terms of DCW and the model fit results are presented in Table 4.

Table 4: The best split models in Monte Carlo Method

Split 1: (T=2, prob=2 in Biological Activity)	
$-\log\text{IC}_{50} = -6.4654180 (\pm 0.2053103) + 0.1431490 (\pm 0.0042867) * \text{DCW} (2,100)$	
n=14, $R^2= 0.8970$, $Q^2=0.8590$, $s=0.256$ (training set)	
n=10, $R^2= 0.9446$, $Q^2=0.9279$, $s=0.487$ (calibration set)	
n=10, $R^2= 0.9667$, $Q^2=0.9499$, $s=0.259$ (test set), $R^2_{\text{m TEST}}= 0.961$	
Split 1: (T=1, prob=1 in lipophilicity)	
$\text{XLOGP} = 3.9123046 (\pm 0.0189362) + 0.0690579 (\pm 0.0004887) * \text{DCW} (1,50)$	
n=14, $R^2= 0.9940$, $Q^2= 0.9917$, $s=0.101$ (training set)	
n=10, $R^2= 0.9996$, $Q^2= 0.9993$, $s=0.191$ (calibration set)	
n=10, $R^2= 0.9838$, $Q^2= 0.9757$, $s=0.247$ (test set), $R^2_{\text{m TEST}}= 0.9818$	

Fig. 7: The two-dimensional plots of the XLOGP versus $ES_{\text{pm}08d}$, Hy and $ALOGP$ descriptors.

The variation of correlation coefficient (test set) with respect to threshold and the number of epochs are plotted in figure 8. This figure confirms that [2 80] and [1 25] are the most appropriate values for threshold and number of epochs including the functions of $-\log\text{IC}_{50}$ and lipophilicity (XLOGP), respectively.

Molecular features are given according to their

correlation weights in table 5. Molecular feature with negative correlation weights are omitted due to their inverse effect on the $-\log\text{IC}_{50}$ value. The higher the correlation weight of a molecular feature, the lower the value of IC_{50} , therefore, the features are more significant. Definitions of the molecular features are given by Kumar and Chauhan [36].

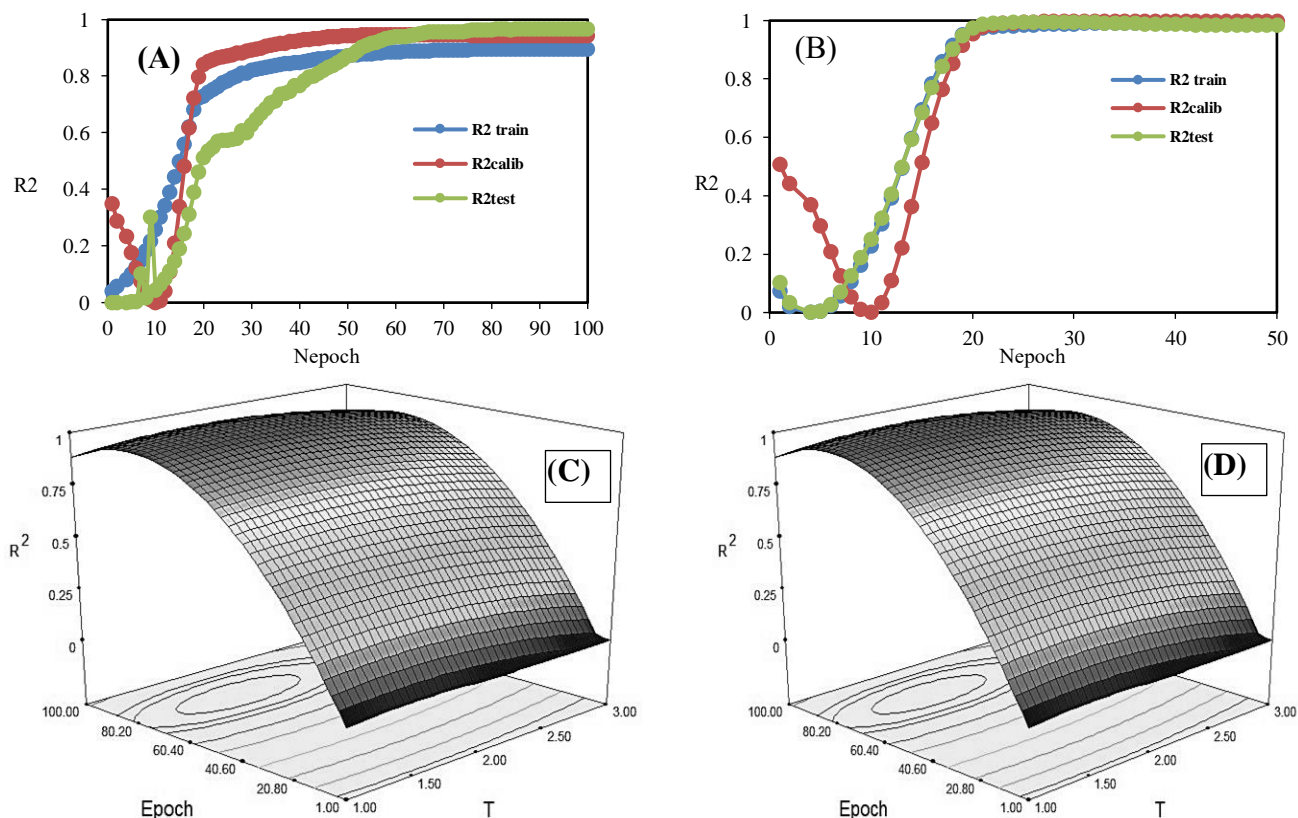


Fig. 8: The variation of correlation coefficient for test set by threshold and number of epochs. (A) in predicted biological activity, (B) in predicted lipophilicity Effects of the number of epochs. (C) in predicted biological activity, (D) in predicted lipophilicity 3-D surface plot of R^2 according to the threshold and the number of epochs.

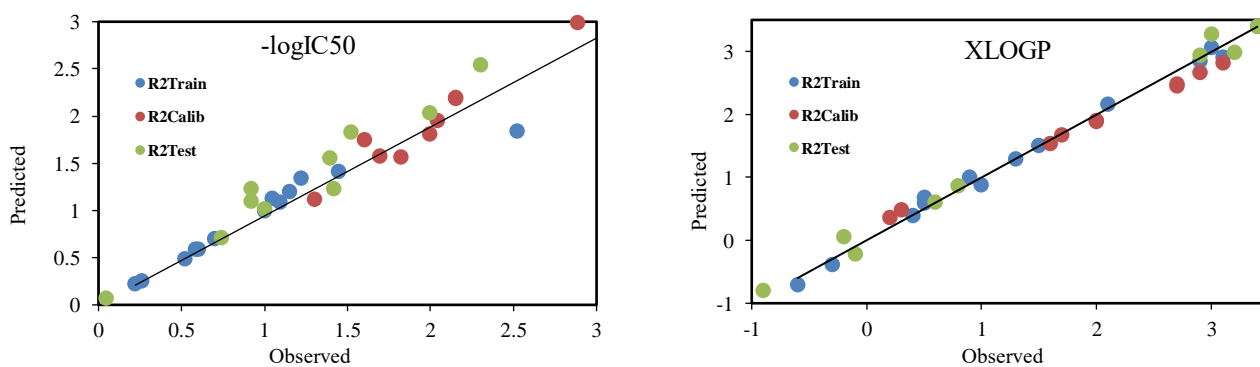


Fig. 9: Correlation between experimental and predicted $-\log IC_{50}$ (A) and lipophilicity (B).

The experimental and calculated activities ($-\log IC_{50}$) and lipophilicity (XLOGP) for the sequence of compounds are plotted against each other in figure 9. A good correlation between the calculated and empirical values of $-\log IC_{50}$ and lipophilicity (XLOGP) can be observed in this figures that approves the appropriateness of the developed model.

According to table 5 with the dependent variables biological activity ($-\log(IC_{50})$), in molecules 2-5

Presence of cyclic rings (1.....) and in molecule 14 Presence of double bond in the cyclic rings (=...1.....) and in molecules 13, 14 (C3.....0., C4.....0.) three and four member cycle with Oxygen and molecule 19 absence of halogens and in molecule 30 presence of Nitrogen and Oxygen, but absence of Sulphur and Phosphorus, Presence of Nitrogen connected to ring in molecule 21 (N...2.....), three and two of sp² carbon in molecule 26, 24

Table 5: SMILES attributes with positive correlation weights for the best split

SMILES attributes (in Biological Activity)	CWs	SMILES attributes (in lipophilicity)	CWs	SMILES attributes (in lipophilicity)	CWs
1.....	4.07724	1.....	3.09748	(.....)	-1.06429
2.....	5.71551	=...(.....)	2.07757	C.....(.....)	-1.47728
3.....	4.90776	=...1.....	3.11165	N.....	-2.55480
=...1.....	4.74122	C...C.....	3.24872	N...C.....	-2.12365
C3.....0...	4.99824	C6.....1...	3.90976	NNC-C...312.	-2.64102
C6.....1.	4.64154	EC0-C...1.	2.83369	NOSP11000000	-2.52392
HALO00000000	7.82137	HALO00000000	2.95020	EC0-N...1....	-2.14048
N...2.....	4.86327	N...2.....	3.29612	NNC-N...110	-2.38960
NOSP11000000	5.92507	NNC-C...110	5.05800	C.....	-0.37527
NNC-C...321.	6.87278	BOND10000000	2.86656	EC0-C...3...	-0.41136
NNC-C...330	4.67748	N...4.....	3.29235	O.....(.....)	-1.11744
C4.....0.	4.72196	N...=.....	2.61706	O.....	-1.77481

(NNC-C...321., NNC-C...330) increase the $-\log IC_{50}$ values and reduce the half maximal inhibitory concentration (IC_{50} value) and these are the most important molecular features.

Also, with the dependent variables lipophilicity (XLOGP), Presence of cyclic rings, presence of double bond with branching and cyclic ring, Presence of two sp^2 carbon connected, one six-member cycle, number of carbon connected to cyclic ring, absence of halogens, Presence of double bond, presence of four N atom increase the XLOGP values in Hydrophobic drugs with high octanol-water partition coefficients.

Furthermore, Presence of branching, Presence of sp^2 carbon connected to branching, Presence of Nitrogen atom and connected to Carbon atom, Number of sp^2 carbon, presence of Nitrogen and Oxygen absence of Sulphur and Phosphorus, Presence of Oxygen and oxygen with branching decrease XLOGP values in hydrophilic drugs.

Structures of Cu, Fe, Zn complexes with octreotide No. 1

Among the 34 selected derivatives, Compound No. 1 was chosen as the best potential drug for the rest investigation using Complex formation study because it has the highest $-\log IC_{50}$ value.

Nitrogen atoms in compound No. 1 of the Caspofungin derivatives as metal-binding sites. It can be observed that the predicted values are in good accordance with their

experimental values [33, 34]. Optimized bond length of Caspofungin complexes with Cu, Fe, and Zn demonstrated that Cu-N, Zn-N, and Fe-N bond lengths as [1.943, 1.998], [1.975, 1.901] and [1.852, 1.891, 1.886], respectively. The structure-optimized Caspofungin complexes are presented in Figure 10.

The energy of complex structure formation for Cu, Zn, Fe complexes for Caspofungin No. 1 were obtained to be -5351.35759 Hf, -5489.5997 Hf, -4974.6827 Hf, respectively.

The Copper(II) and Zinc(II) complexes of antifungal drugs with nitrogen-donor ligands have shown remarkable antifungal and antibacterial activities and could represent novel agents for potential clinical use[44,45]. It is proposed that chelation reduces the polarity of the metal ion and could enhance its lipophilic character, which subsequently favors the permeation through the lipid layers of the cell membrane and blocks the metal binding sites on enzymes of microorganisms [18,37].

Iron plays a role in the development of resistance to antifungal drug therapies and Combined treatment of antifungals with compounds targeting iron assimilation is a promising approach to combat opportunistic fungal infections, particularly mucormycosis [38].

It was shown that Zn complex of Caspofungin No. 1 is more stable than that of Fe and Cu complexes. Therefore, when the Caspofungin drug is injected into the body,

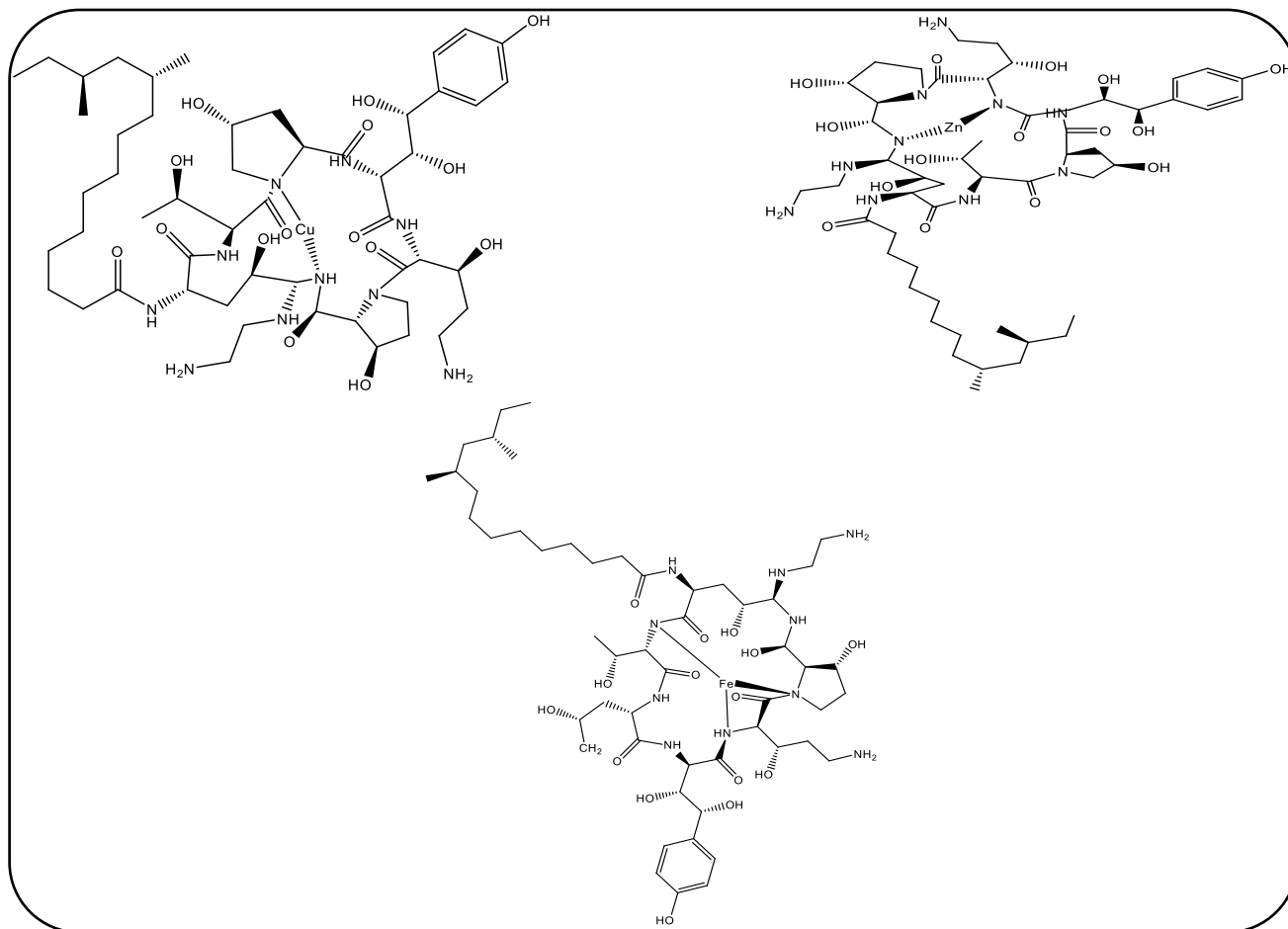


Fig. 10: Optimized structure of the Caspofungin complexes with B3lyp/6-311g/lan12dz.

the amount of Zn in the body is expected to decrease and reduced Zn absorption in the body and probably skin problems. As a result, Caspofungin drug might be better to enter the body in a complex form with Zn. In this case, Nitrogen site are involved with Zn. Therefore, the complex does not form with other metals, which can reduce the side effects of the Caspofungin drug.

It is worthy to note that Zn is an essential element for living cells and its complexes have been identified as anticancer species [39]. The Zn(II) ions play important roles in many cellular processes ,including gene expression [40], apoptosis [41], enzyme regulation [42] and neurotransmission [43].

CONCLUSIONS

In this study, five different approaches were employed to study the structure-activity relationships of 34 Caspofungin derivatives. The GA-ANN and MLR-ICA showed the best performance among the considered approaches.

In biological activity study, the results proved that ESpm11x, RDF135e, Du, JGI3, DP05 and ESpm04x descriptors in the GA-ANN method were the most significant descriptors. In lipophilicity study, the best selected descriptors are EEig12x, EEig01x and ALOGP descriptors in MLR-ICA method.

In Monte Carlo method, the structural descriptors include the Presence of cyclic rings, the Presence of double bonds in the cyclic rings, three and four member cycles with Oxygen, the absence of halogens, the presence of Nitrogen but the absence of Sulphur, and Phosphorus, six-member cycle, number of carbon connected to the cyclic ring, Presence of sp^2 carbon connected to branching are the most important molecular features.

It was concluded that the simultaneous use of Monte Carlo and linear and non-linear methods gives deeper and more comprehensive knowledge about the effects of molecular and structural descriptors on the activity and lipophilicity of drugs and provides better insights to design

new drugs. In addition, the study of complex formation of the best Caspofungin derivative with the important metals in the live body showed that it is recommendable to use this drug in its complex form with Zn instead of its molecular form.

Acknowledgement

The authors gratefully acknowledge the support provided by the Islamic Azad University of Rasht.

Received: Dec. 12, 2021 ; Accepted: Mar. 14, 2022

References

- [1] Cornely O.A., Bassetti M., Calandra T., Garbino J., Kullberg B.J., Lortholary O., W Meersseman, Akova M., Arendrup M.C., Arikan-Akdagli S., Bille J., Castagnola E., Cuenca-Estrella M., Donnelly J.P., Groll A.H., Herbrecht R., Hope W.W., Jensen H.E., Lass-Flörl C., Petrikos G., Richardson M.D., Roilides E., Verweij P.E., Viscoli C., Ullmann A.J., [ESCMID Guideline for the Diagnosis and Management of Candida Diseases 2012: Nonneutropenic Adult Patients](#), *Clin. Microbiol. Infect.*, **18**: 19-37 (2012).
- [2] Chen Y.C., Kuo S.F., Chen, F.J., Lee C.H., [Antifungal Susceptibility of Candida Species Isolated from Patients with candidemia in southern Taiwan, 2007–2012: Impact of New Antifungal Breakpoints](#), *Mycoses*, **60**: 89-95 (2017).
- [3] Latg J-P., [The Cell Wall: A Carbohydrate Armour for the Fungal Cell](#), *Mol. Microbiol.*, **66**: 279-290 (2007).
- [4] Letscher-Bru V., Herbrecht R., [Caspofungin: the First Representative of a New Antifungal Class](#), *J. Antimicrob. Chemother.*, **51**: 513-521 (2003).
- [5] Wang R., Fu Y., Lai L., [A New Atom-Additive Method for Calculating Partition Coefficients](#), *J. Chem. Inf. Comput. Sci.*, **37**: 615-621 (1997).
- [6] Wang R., Gao Y., Lai L., [Calculating Partition Coefficient by Atom-Additive Method](#), *Drug Discov. Des.*, **19**: 47-66 (2000).
- [7] Sangster J., ["Octanol–Water Partition Coefficients: Fundamentals and Physical Chemistry"](#), Wiley Series in Solution Chemistry, Chichester: John Wiley & Sons Ltd, 184 (1997).
- [8] Shargel L., Susanna WP., Yu AB., ["Applied Biopharmaceutics & Pharmacokinetics \(6th ed.\)"](#), McGraw-Hill Medical, New York, (2012).
- [9] Moriguchi I., Hirono S., Liu Q., Nagakome I., Matsushita Y., [Comparison of Reliability of log P Values for Drugs Calculated by Several Methods](#), *Chem. Pharm. Bull.*, **42**: 976 (1994).
- [10] Guziałowska-Tic1 J., [The use of QSAR Methods for Determination of N-Octanol/Water Partition Coefficient using the Example of Hydroxyester HE-1](#), E3S Web of Conferences 19, EEMS, (2017).
- [11] Sahebamee H., Yaghmaei P., Abdolmaleki P., Foroumadi A.R., [Quantitative Structure-Activity Relationships Study of Carbonic Anhydrase Inhibitors Using Logistic Regression Model](#), *Iran. J. Chem. Chem. Eng. (IJCCE)*, **32**: 19-29 (2013).
- [12] Ugochukwu Ibeji C., Oguejiofo U., Chukwuma Chime C., Godfrey Akpomie K., Onyinye Anarado C.J., Abiola Odewole O., Grishina M., Potemkin V., [Dehydroacetic Acid-Phenylhydrazone as a Potential Inhibitor for Wild-Type HIV-1 Protease: Structural, DFT, Molecular Dynamics, 3D QSAR and ADMET Characteristics](#), *Iran. J. Chem. Chem. Eng. (IJCCE)*, **40**: 215-230 (2021).
- [13] Černý V., [Thermodynamical Approach to the Traveling Salesman Problem: An Efficient Simulation Algorithm.](#), *J. Optimiz. Theory. App.*, **45**: 41-51 (1985).
- [14] Schmitt L.M., [Theory of genetic Algorithms](#). *Theor. Comput. Sci.*, **259**: 1-61 (2001).
- [15] Bertsimas D., Tsitsiklis J., [Simulated Annealing.](#), *Statist. Sci.*, **8**: 10-15 (1993).
- [16] Meanwell N.A., Wallace O.B., Fang H., Wang H., Deshpande M., Wang T., Yin Z., Zadjura L., Tweedie D.L., Yeola S., Zhao F., Ranadive S., Robinson BA., Gong Y.F., Wang H.G., Spicer T.P., Blair W.S., Shi P.Y., Colonna R.J., Lin P.F., [Inhibitors of HIV-1 Attachment. Part 2: An Initial Survey of Indole Substitution Patterns](#). *Bioorg. Med. Chem. Lett.*, **19**: 1977-1981 (2009).
- [17] Toropova A.P., Toropov A.A., Benfenati E., Gini G., Leszczynska D., Leszczynski J., [CORAL: Quantitative Structure–Activity Relationship Models for Estimating Toxicity of Organic Compounds in Rats](#), *J. Comput. Chem.*, **32**: 2727-2733 (2011).

- [18] Schatzschneider U., "Antimicrobial activity of organometal compounds: Past, Present, and Future Prospects", In "Advances in Bioorganometallic Chemistry", Elsevier, Amsterdam, 173–192 (2019).
- [19] Atashpaz-Gargari E., Lucas C., "Imperialist Competitive Algorithm: An Algorithm for Optimization Inspired by Imperialistic Competition", In *IEEE Congress on Evolutionary Computation*, Singapore, (2007).
- [20] Hosseini S., Al Khaled A., A Survey on the Imperialist Competitive Algorithm Metaheuristic: Implementation in Engineering Domain and Directions for Future Research. *Applied Soft Comput.*, **24**: 1078-1094 (2014).
- [21] Bigdeli N., Afshar K., Gazafroudi A.S., Ramandi M.Y., A Comparative Study of Optimal Hybrid Methods for Wind Power Prediction in Wind farm of Alberta, Canada, *Ren. Sustain. Energy. Rev.*, **27**: 20-29 (2013).
- [22] Aliniya Z., Mirroshandel S.A., A Novel Combinatorial Merge-Split Approach for Automatic Clustering using Imperialist Competitive Algorithm. *Expert. Syst. Appl.*, **117**: 243-266 (2019).
- [23] Shokrollahpour E., Zandieh M., Dorri B., A Novel Imperialist Competitive Algorithm for Bi-Criteria Scheduling of the Assembly Flowshop Problem, *Int. J. Prod. Res.*, **49**: 3087–3103 (2011).
- [24] Guha R., Serra J.R., Jurs P.C., Generation of QSAR Sets with A Self-Organizing Map, *J. Mol. Graph. Model.*, **23**: 1-14 (2004).
- [25] Frisch M.J., Trucks G.W., Schlegel H.B., Scuseria G.E., Robb M.A., Cheeseman J.R., Scalmani G., Barone V., Mennucci B., Petersson G.A., Nakatsuji H., Caricato M., Li X., Hratchian H.P., Izmaylov A.F., Bloino J., Zheng G., Sonnenberg J.L., Hada M., Ehara M., Toyota K., Fukuda R., Hasegawa J., Ishida M., Nakajima T., Honda Y., Kitao O., Nakai H., Vreven T., Montgomery J.A., Peralta J.E., Ogliaro F., Bearpark M., Heyd J. J., Brothers E., Kudin K.N., Staroverov V.N., Kobayashi R., Normand J., Raghavachari K., Rendell A., Burant J.C., Iyengar S.S., Tomasi J., Cossi M., Rega N., Millam J.M., Klene M., Knox J.E., Cross J.B., Bakken V., Adamo C., Jaramillo J., Gomperts R., Stratmann R.E., Yazyev O., Austin A.J., Cammi R., Pomelli C., Ochterski J.W., Martin R.L., Morokuma K., Zakrzewski V.G., Voth G.A., Salvador P., Dannenberg J., Dapprich S., Daniels A.D., Farkas Ö., Foresman J.B., Ortiz J.V., Cioslowski J., Fox D.J., <https://gaussian.com/glossary/g09>, *Gaussian 09* (Gaussian, Inc., Wallingford CT, 2009).
- [26] <https://pubchem.ncbi.nlm.nih.gov>.
- [27] Todeschini R., Consonni V., "Handbook of Molecular Descriptors", Wiley-VCH, (2000).
- [28] Sayyadi kord Abadi R., Alizadehdakhel A., Tajadodi Paskiabei S., A DFT and QSAR Study of Several Sulfonamide Derivatives in Gas and Solvent, *J. Korean Chem. Soc.*, **60**: 225 (2016).
- [29] Sayyadi Kord Abadi R., Alizadehdakhel A., Dorani Shiraz S., Ab Initio and QSAR Study of Several Etoposides as Anticancer Drugs: Solvent Effect, *Russ. J. Physic. Chem. B*, **11**: 307-317 (2017).
- [30] Toropova, A.P., Toropov A.A., CORAL Software: Prediction of Carcinogenicity of Drugs by Means of the Monte Carlo Method, *Eur. J. Pharm. Sci.*, **52**: 21-25 (2014).
- [31] Schuur J.H., Selzer, P., Gasteiger J., The Coding of the Three-Dimensional Structure of Molecules by Molecular Transforms and Its Application to Structure-Spectra Correlations and Studies of Biological Activity, *J. Chem. Inform. Comput. Sci.*, **36**: 334-344 (1996).
- [32] <http://www.insilico.eu/coral>
- [33] Fermi G., Perutz M.F., Shaanan B., Fourme R., The Crystal Structure of Human Deoxyhaemoglobin at 1.74 Å Resolution, *J. Mol. Biol.*, **175**: 159-174 (1984).
- [34] Stevanović N-L., Aleksic I., Kljun J., Bogojevic S.S., Veselinovic A., Nikodinovic-Runic J., Turel I., Djuran M.I., Glišić B.Đ., Copper(II) and Zinc(II) Complexes with the Clinically Used Fluconazole: Comparison of Antifungal Activity and Therapeutic Potential., *Pharmaceuticals*, **14**: 24-44 (2021).
- [35] Seif N., Farhadi A., Badri R., Kiasat A.R., An Experimental and Theoretical Study on Bicyclo-3,4-Dihydropyrimidinone Derivative: Synthesis and DFT Calculation, *Iran. J. Chem. Chem. Eng. (IJCCE)*, **39**: 21-33 (2020).
- [36] Kumar A., Chauhan S., Monte Carlo Method based QSAR Modelling of Natural Lipase Inhibitors using Hybrid Optimal Descriptors, *SAR. QSAR. Environ. Res.*, **28**: 179-197 (2017).
- [37] Sharma R.K., Katiyar D., Recent Advances in the Development of Coumarin Derivatives as Antifungal Agents, In *Recent Trends in Human and Animal Mycology*, 235–263 (2019).
- [38] Stanford F.A., Voigt K., Iron Assimilation during Emerging Infections Caused by Opportunistic Fungi with emphasis on Mucorales and the Development of Antifungal Resistance, *Genes*, **11**: 1296 (2020).

- [39] Vallee B.L., Falchu K.H., [The Biochemical Basis of Zinc Physiology](#), *Physiol. Rev.*, **73**: 79–118 (2020).
- [40] Falchuk K.H. [The Molecular Basis for the Role of Zinc in Developmental Biology](#), *Mol. Cell. Biochem.*, **188**: 41–48 (1998).
- [41] Zalewski, P.D., Forbes, I. J., Betts W.H., [Correlation of Apoptosis with Change in Intracellular Labile Zn\(II\) using Zinquin \[\(2-methyl-8-p-toluenesulphonamido-6-quinolyloxy\)Acetic Acid\], A New Specific Fluorescent Probe for Zn\(II\)](#). *Biochem. J.* **296**, 403-8 (1993).
- [42] Maret W., Jacob C., Vallee B.L., Ficher E., [Inhibitory Sites in Enzymes: Zinc Removal Andreactivation by Thionein](#), *Proc. Natl. Acad. Sci.*, **96**, 1936 –1940 (1999).
- [43] Cuajungco M.P, Lees G.J, [Zinc Metabolism in the Brain: Relevance to Human Neurodegenerative Disorders](#), *Neurobiol. Dis.*, **4**: 137-69 (1997).
- [44] Andrejević T.P., War zajtis B., Glišić B.D., Vojnovic S., Mojicevic M., Stevanović N.L., Nikodinovic-Runic J., Rychlewska U., Djuran M. I., [Zinc\(II\) Complexes with Aromatic Nitrogen-Containing Heterocycles as Antifungal Agents: Synergistic Activity with Clinically used Drug Nystatin](#), *J. Inorg. Biochem.*, **208**: 111089 (2020).
- [45] de Azevedo França J.A., Granado R., de Macedo Silva S.T., Santos-Silva G.D., Scapin S., Borba-Santos L.P., Rozental S., de Souza W., Martins-Duarte E.S., Barrias E., Rodrigues J.C.F., Navarro M., [Synthesis and Biological Activity of Novel Zinc-Itraconazole Complexes In Protozoan Parasites and Sporothrix spp](#), *Antimicrob. Agents Chemother.*, **64**: e01980-19 (2020).