Research Article

iMedPub Journals
www.imedpub.com

Archives of Medicine
ISSN 1989-5216

2020
Vol.12 No.5:30

DOI: 10.36648/1989-5216.12.5.331

An ANN-Based QSAR Model to Predict Anti-*Staphylococcus aureus* Activity of Oxadiazoles

Kamrani SME[1] and Hadizadeh F[2]*

1 Department of Medicinal Chemistry, School of Pharmacy, Mashhad University of Medical Sciences, Mashhad, Iran
2 Biotechnology Research Center, Pharmaceutical Technology Institute, Mashhad University of Medical Sciences, Mashhad, Iran

***Corresponding author:** Hadizadeh F

hadizadehf@mums.ac.ir; Smekamrani_pharmacy@gmail.com

Tel: 989139125565

Biotechnology Research Center, Pharmaceutical Technology Institute, Mashhad University of Medical Sciences, Mashhad, Iran.

Citation: Kamrani SME, Hadizadeh F (2020) An ANN-Based QSAR Model to Predict Anti-*Staphylococcus aureus* Activity of Oxadiazoles. Arch Med Vol.12 No.5: 30

Abstract

Background: Antibiotic-resistant bacteria are likely to be one of the most critical problems in the future; hence, more effort is needed to design and develop new types of antimicrobial agents. Quantitative Structure-Activity Relationship (QSAR) is a procedure which helps other researchers to design better chemical agents with more potent biological activity. QSAR can be defined as a quantitative relationship between the chemical structure and biological activity.

Results: In this study, a previously-synthesized oxadiazole library was used to build a QSAR model based on the Group Method of Data Handling (GMDH) method and Partial Least Squares (PLS) regression. Owing to their high correlation coefficients (R^2) for test and training data, both methods are sufficiently reliable. In this study, the active compounds of the library were used as a template to design new chemical compounds predicted to have a great anti-*Staphylococcus aureus* (*S. aureus*) activity according to PLS, GMDH, and docking methods. GMDH and PLS are highly flexible such that they can include other information like absorption, distribution, metabolism, and excretion (ADMT) and toxicity.

Conclusion: The selected methods can be used to handle huge amounts of data from a large library of chemical compounds and help research and development (R&D) process. Additionally, the designed model and the proposed compounds can help other researchers to find the best anti- *Staphylococcus aureus* chemical compounds.

Keywords: Oxadiazole; Quantitative structure-activity relationship; *Staphylococcus aureus*

Abbreviations: R&D: Research and Development; GMDH: Group Method of Data Handling; PLS: Partial Least Square; QSAR: Quantitative Structure Activity Relationship; ADME: Absorption, Distribution, Metabolism and Excretion; MRSA: Methicillin-Resistant *Staphylococcus aureus*; Pmic: -Log: Minimum Inhibitory Concentration

Received: August 09, 2020; **Accepted:** September 16, 2020; **Published:** September 23, 2020

Introduction

Antibiotic-resistant bacteria, whether gram-positive or gram-negative, are one of the most important health problems which are likely to turn into a critical threat in the future. This problem arises from the indiscriminate use of antibiotics. One of these resistant strains of bacteria is Methicillin-Resistant *S. aureus* (MRSA) [1]. Regarding the wide range of highly resistant strains observed in hospitals and communities, it is necessary to study and design new antibiotics to overcome the high rate of infectious diseases and their adverse consequences for the societies. One of these classes of chemical structures is 1,2,4-oxadiazole which has been extensively synthesized and investigated [2].

Developing Quantitative Structure-Activity Relationship (QSAR) models that delineate the relationships between the chemical descriptors and biological activities of the chemical compounds may have considerable advantages. These models can be developed based on artificial neural networks (e.g. MLP and RBF) or regressions (e.g. PLS, PCR, and MLR).

Chang and Mobashery synthesized a series of 1,2,4-oxadiazole and measured the plC_{50} of these compounds [3,4]. Leemans [5] used a library of 102 members of 1,2,4-oxadiazole to build a 3D-QSAR model with good predictive power (correlation coefficient for the training and test data sets were 0.88 and 0.61, respectively) [5] **(Figure 1)**. In the present study, Partial Least Squares (PLS) and Group Methods of Data Handling (GMDH) were used to develop a robust QSAR model for predicting the Pmic of oxadiazoles with high correlation coefficients. To this end, data mining was conducted to find appropriate descriptors having a better correlation with Pmic.

Ivahnenko (1971) presented a predictive model for identifying a nonlinear relationship between inputs and outputs **(Figure 2)**. Like other neural networks, the GMDH is also capable of predicting outcomes from the multivariate data and achieving the maximum accuracy and reliability by testing all possible structures of the polynomial regression models [6,7].

Partial Least Squares (PLS) regression is a popular method for soft modelling in QSAR. PLS is a method for constructing predictive models when there are multiple highly collinear factors. Today, PLS regression is one of the most widely used techniques in chemo metrics and related areas. It is also used in bio-informatics, sensometrics, neuroscience, and anthropology [8].

In this study, the Leemans' library was used to generate PLS- and GMDH-based QSAR model with a good correlation coefficient (R^2). These methods, which are defined as Matlab codes (.m), can be used by other drug designers to produce potent and non-toxic oxadiazoles with good pharmacokinetic characteristics, low toxicity, and few side effects.

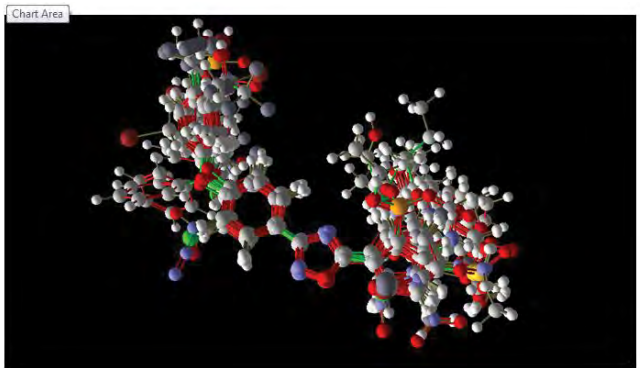


Figure 1 Aligned structure of compounds used to design QSAR model.

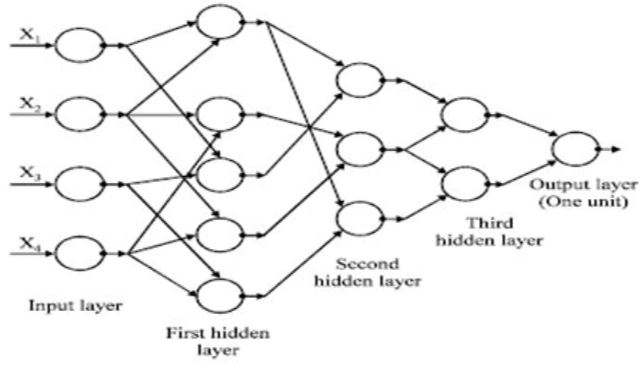


Figure 2 General structure of GMDH.

Material and Methods

Generation and selection of descriptors

The 3D structure of 102 compounds selected from Leemans' library [5] was drawn using ChemOffice 2012 package [9]. The complete dataset of the structures and biological activity is presented as a table in the supplementary materials.

Then, structure minimization was carried out using the Austin Model 1 (AM1) semi-empirical method in Gaussian 9 software. Descriptors of these compounds were calculated using various software including MOE [10], HyperChem, Gaussian 9 [11], and Dragon 5 [12]. A total of 3260 descriptors for each compound was calculated. Notably, all these descriptors are not necessarily required to produce an effective predictive model. Afterward, that dataset was imported into the SPSS modeler 18 software package [13]. Using feature selection module, about 300 descriptors were selected for building models. The outcome of SPSS modeler showed which variables should be excluded and which ones should be used to develop the model.

GMDH

In this study, the GMDH shell software was used to create a predictive model [14]. A total of 100 best descriptors (based on SPSS modeler output) was selected to build GMDH artificial neural network. The software randomly divided the compounds into two groups, namely training (n = 92) and test (n = 10) compounds. The results are shown in **Figure 3**. All the results and data sets are available in supplementary materials (https://github.com/smekamrani/ANN-Based-QSAR). A correlation coefficient (R^2) higher than 0.5 indicates the reliability of a model. According to the GMDH output, the correlation coefficients (R^2) were estimated to be about 0.80 and 0.76 for test and training data sets, respectively, indicating the good prediction ability of the model. Raw polynomial equation (the output of GMDH shell) which is implemented in Matlab (.m file) and the datasets used to build the model are available in supplementary materials.

PLS

In the present study, PLS was performed using crossval.m and pls.m that are available in MATLAB 2014 toolbox [15]. Crossval function using Leave-one-out method and NIPALS algorithm determined the number of principal components (PCs). After that, the PLS function, with various principal components, was used to measure the PLS coefficients. Then, the PLS predicted outcomes were obtained by multiplying the matrices of the coefficients in the test and training descriptors. Afterward, a code which automatically loads data sets was written, and the PLS was done (see supplementary materials for more details). **Figure 4** shows the graphs of the test and training data sets. Correlation coefficients (R^2) obtained in the training and test data sets were 0.88 and 0.76,

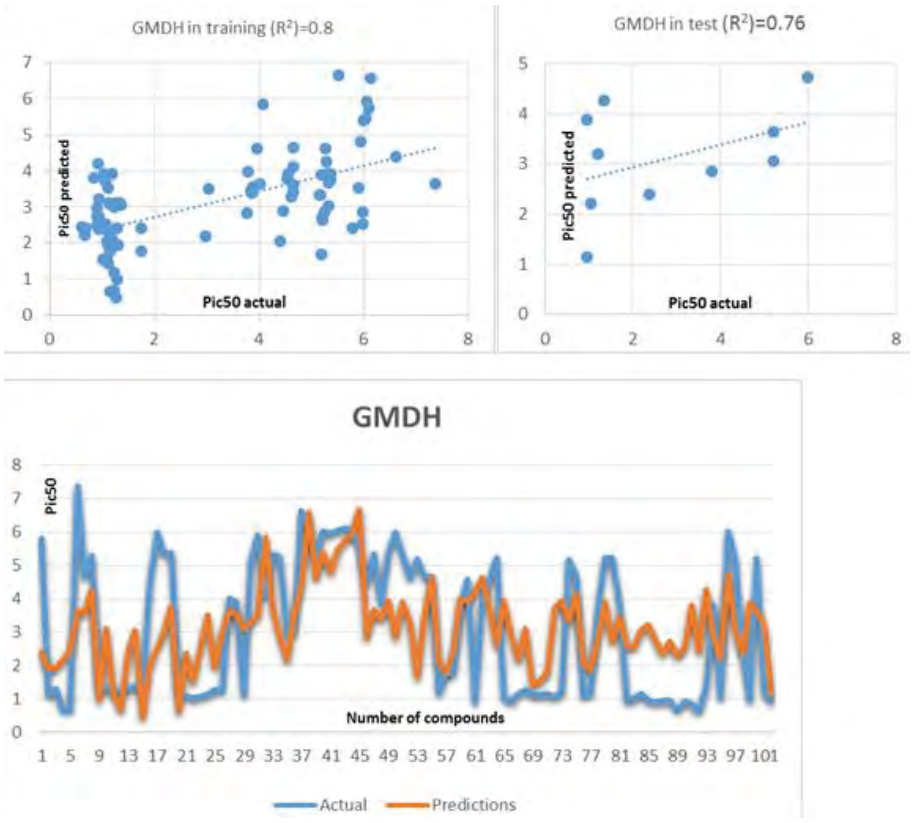


Figure 3 Results of GMDH model in both test and training data.

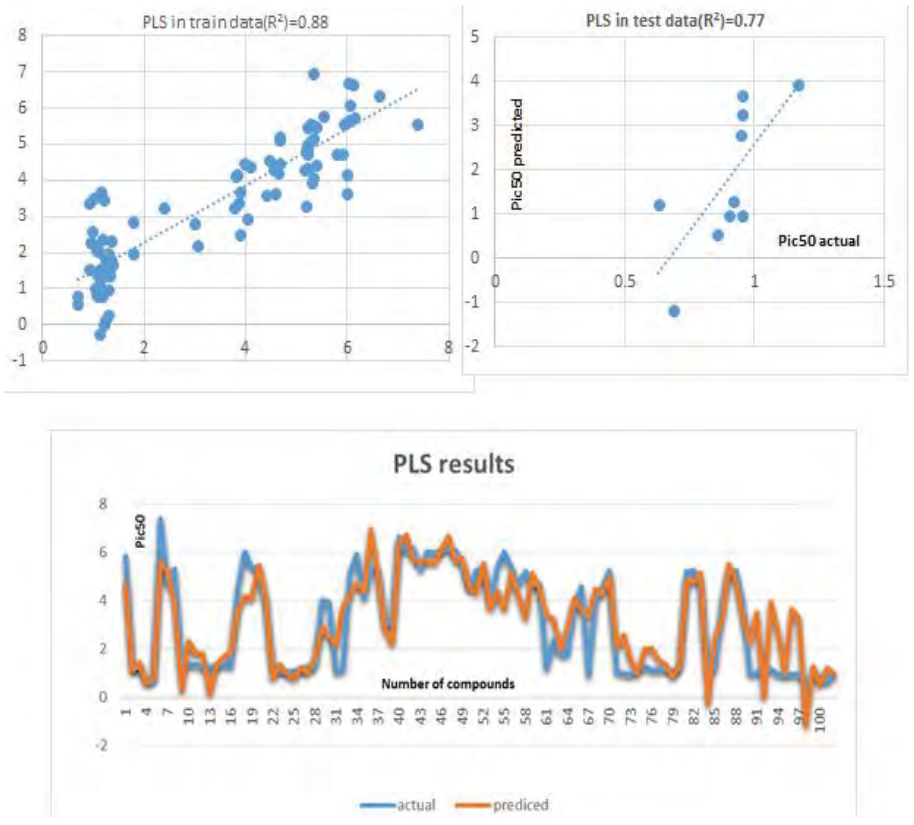


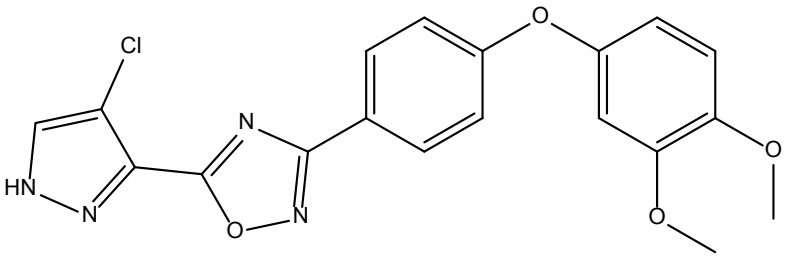
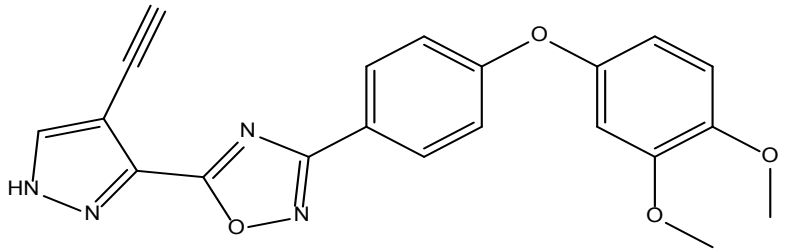
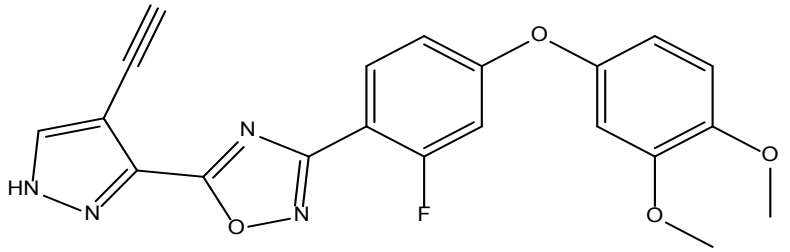
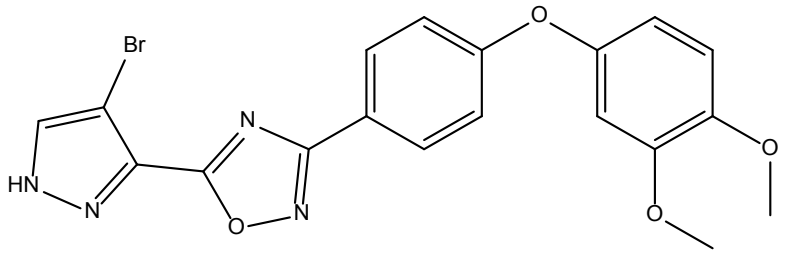
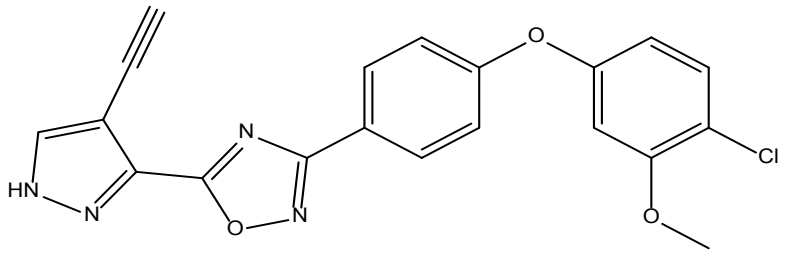
Figure 4 Results of PLS model in both test and training data.

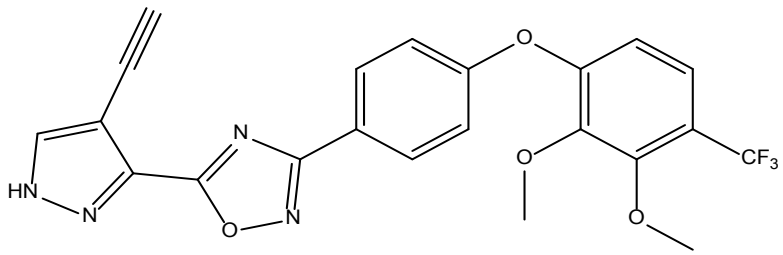
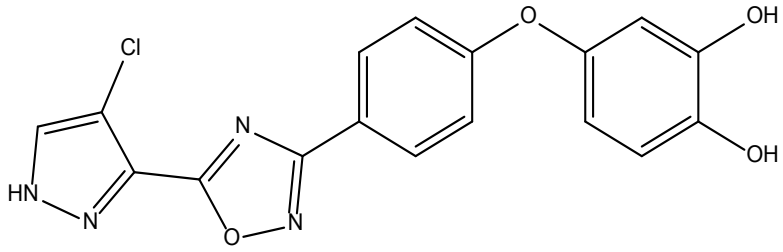
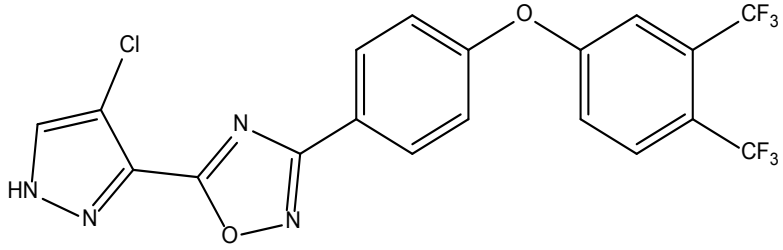
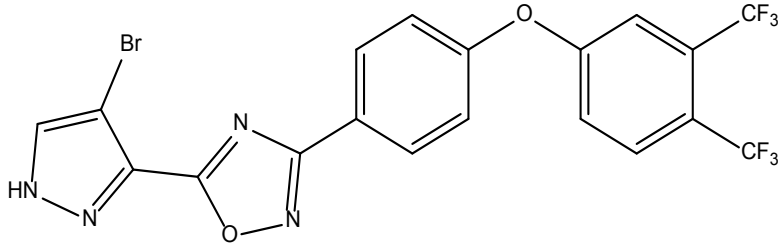
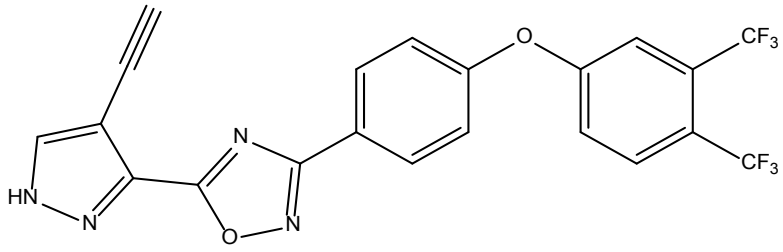
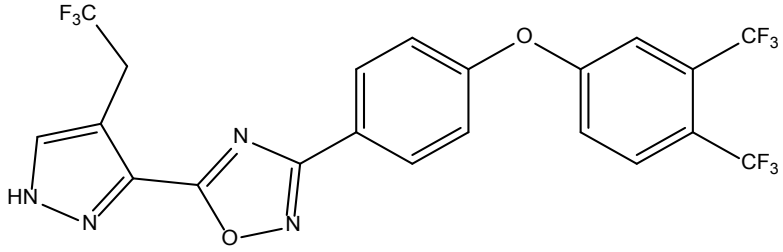
respectively. In-depth information about descriptors and Matlab codes (.m file) are presented in the supplementary materials. The complete data sets of the structures and biological activities are also available in the supporting information.

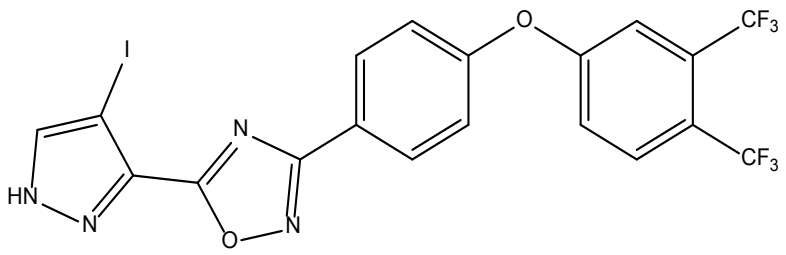
Docking and designing of the new compounds

Leemans [5] created a 3D-QSAR map (CoMFA) which suggests where negative or positive groups should be included in the lead compound to increase the biological activity. The present study used the most potent compounds of Leemans' library (i.e., 40,42,47 and 46) to design the new compounds predicted to be sufficiently potent by PLS and GMDH (**Table 1** and **Figure 5**). In addition, the compounds were docked to investigate their binding mode using MOE 2015 (https://www.chemcomp.com). The X-ray

Table 1 Statistical data of each model.

Name	Structure	PLS Prediction Pmic	GMDH Prediction Pmic
1D		13.8	2.7
2D		8.0	1.7
3D		7.4	1.75
4D		8.2	3.5
5D		10.5	3.1

6D		12.3	3.98
7D		6.9	3.96
8D		12.5	11.57
9D		12.6	12.63
10D		14.2	10.26
11D		21.6	11.38

12D		15.7	13.54

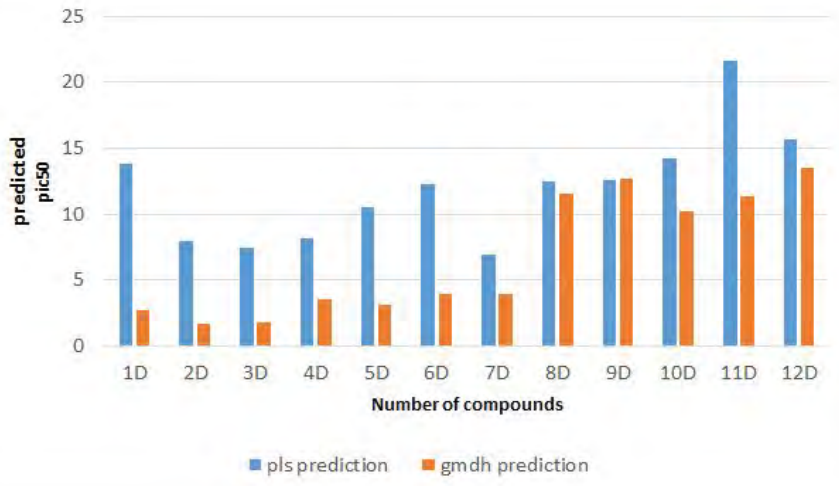


Figure 5 Predicted value of designed structure by each model (GMDH, PLS).

crystal structure of the *S. aureus* Sortase A and co-crystallized LPATG (PDB code: 2KID) [16] was downloaded from www.rcsb.org. The catalytic pocket of the enzyme composed of His^{120}, Cys^{184}, and Tyr^{194} is crucial to inhibit the enzyme activity and antibiotic potency [16,17]. Sortase A Transpeptidase is a therapeutically important membrane-bound enzyme in Gram-positive bacteria. It catalyzes the cell wall anchoring domain of important virulence factors to the surface of the organism. London DG scoring function was used to dock the active compounds of the Leemans' library and that of the selected library.

Quantum mechanical studies

Salter [18] in a study showed the correlation between the computational studies and the experimental data. He compared ΔH of naphthalene and azulene in the quantum mechanical and experimental studies and concluded that they are nearly equal. The present study aims to carry out a roughly similar work to attain thermochemical information of the designed compounds. The compound showed in **Figure 6** was used as the lead compound (i.e., the designed compound without any chemical substitution), and the Density Functional Theory (DFT) was done within the basis set of B3LYP/6-31G (D) for all compounds to predict ΔH and ΔG in the event that the respective substitution is applied in the hypothetical lead compound.

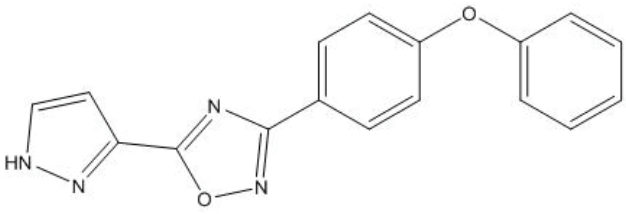


Figure 6 Lead compound (our designed compound without any Chemical substitute.

Results

This study showed that both the GMDH neural network and PLS are reliable tools to predict Pmic (the statistical information of each model is presented in **Table 2**). Previous studies that developed a 3D-QSAR model using this library accurately predicted Pmic and suggested which part of the molecule needs a substitution, whether of the negative or positive group. Both machine learning models, i.e., PLS and GMDH, are considerably flexible such that various data from different sources can be included, and large data sets can be accurately handled. Indeed, it is possible to training a QSAR model based on data associated with pharmacokinetic characteristics and toxicity. Therefore, compiling data from various sources would enable practical training and development of effective models and provide in-depth information helping to identify compounds that could be evaluated in clinical studies and supplied to the market.

Table 2 Structure of designed compound.

Compound	Sum of electronic and thermal Enthalpies	Sum of electronic and thermal Free Energies	From	To	ΔH (k.cal/mol)	ΔG (k.cal/mol)
lead	-642344.3	-642369.4				
1D	-1074347.3	-1074384.9	Lead	1D	-432003	-432016
2D	-833716.06	-833753.71	Lead	2D	-191372	-191385
3D	-896197.23	-896241.15	Lead	3D	-253853	-253872
4D	-2401047.7	-2401085.4	Lead	4D	-1758704	-1758716
5D	-1050501.9	-1050533.3	Lead	5D	-408158	-408164
6D	-1045419.1	-1045463.0	Lead	6D	-403075	-403094
7D	-1025282.3	-1025313.6	Lead	7D	-382938	-382945
8D	-1353884.2	-1353928.1	Lead	8D	-711540	-711559
9D	-2680371.3	-2680415.2	Lead	9D	-2038027	-2038046
10D	-1113246.6	-1113290.5	Lead	10D	-470903	-470922
11D	-1301631.4	-1301669.0	Lead	11D	-659287	-659300
12D	-5375809.1	-5375846.7	Lead	12D	-4733465	-4733478

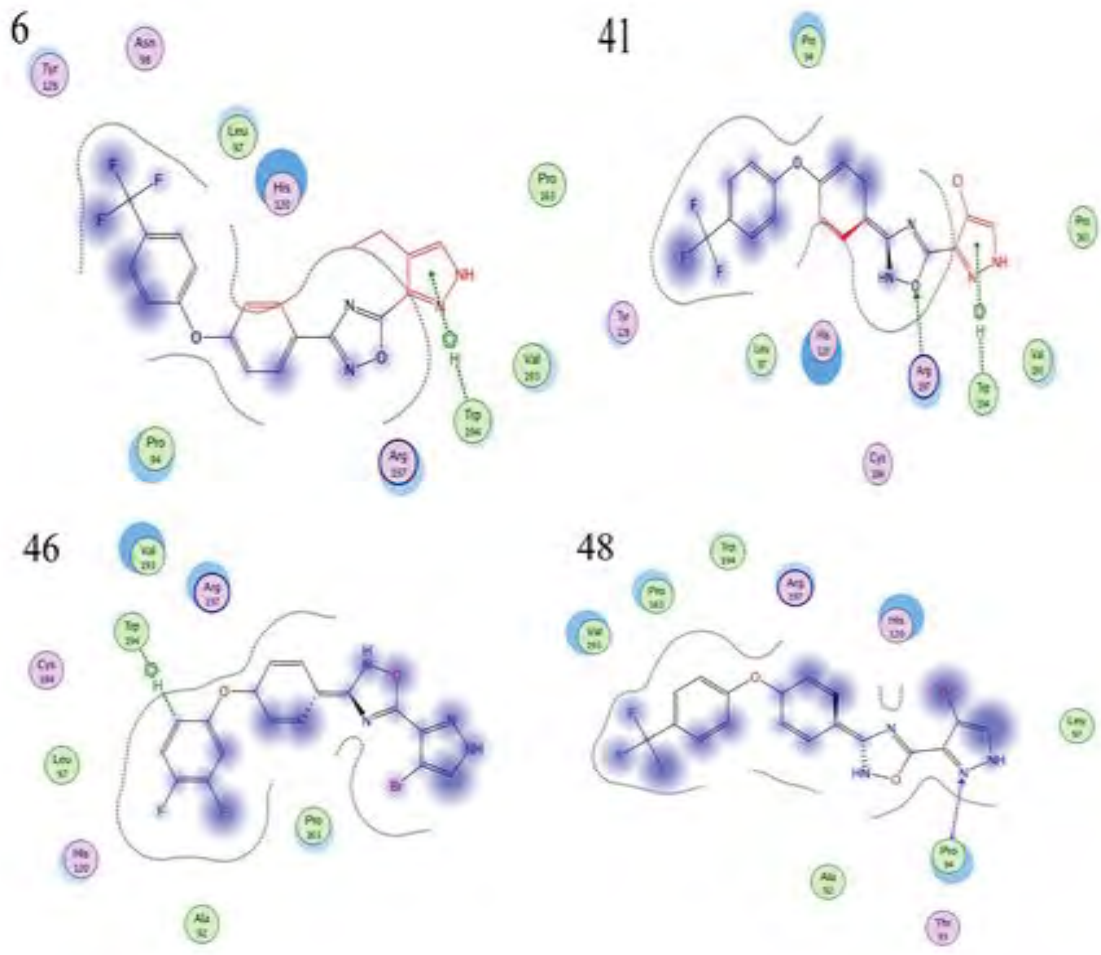


Figure 7 Interaction of most active compounds of Leemans's library in active site of enzyme. They have Aren cation interaction with Trp 194 mostly and Arg 197.

Docking studies showed that the most potent compounds of Leemans' library [5] have some interaction with some residues (Trp 194) in the active site of the enzyme. Some interactions with residues in the active-site pocket of the enzyme could also be seen in the designed compounds (residues Cys-184 and Arg-197). These results are shown in **Figures 7 and 8.** The designed compounds were screened by both models, i.e., GMDH and PLS, and it is predicted that they will have good biological activities.

Table 3 showed all the thermochemical information attained using the DFT calculation. In summary, it was predicted that substituting from the lead compound to 4D, 9D, and 12D compounds can result in more negative ΔH and ΔG. Indeed, there will be fewer problems for the synthesis of these compounds, and they will be more stable as well.

Discussion

This study designed a model with comparatively better correlation coefficients than those proposed in previous studies [5], especially on the test dataset. Note worthily, a drug designer should produce not only a potent compound but also a drug with considerably low toxicity and good pharmacokinetic characteristics. Kinetics and toxicity data can be substantially helpful to make effective models. Narrowing the research scope can also be effective in reaching a clinically practical compound. The GMDH as an algorithm for complex systems can be helpful in designing new chemical compounds. Nowadays, the cost of drug design and development is considerably high, and R&D researcher cannot only rely on traditional trials and error procedures. Computational drug design is a solution to reduce the cost of the R&D process.

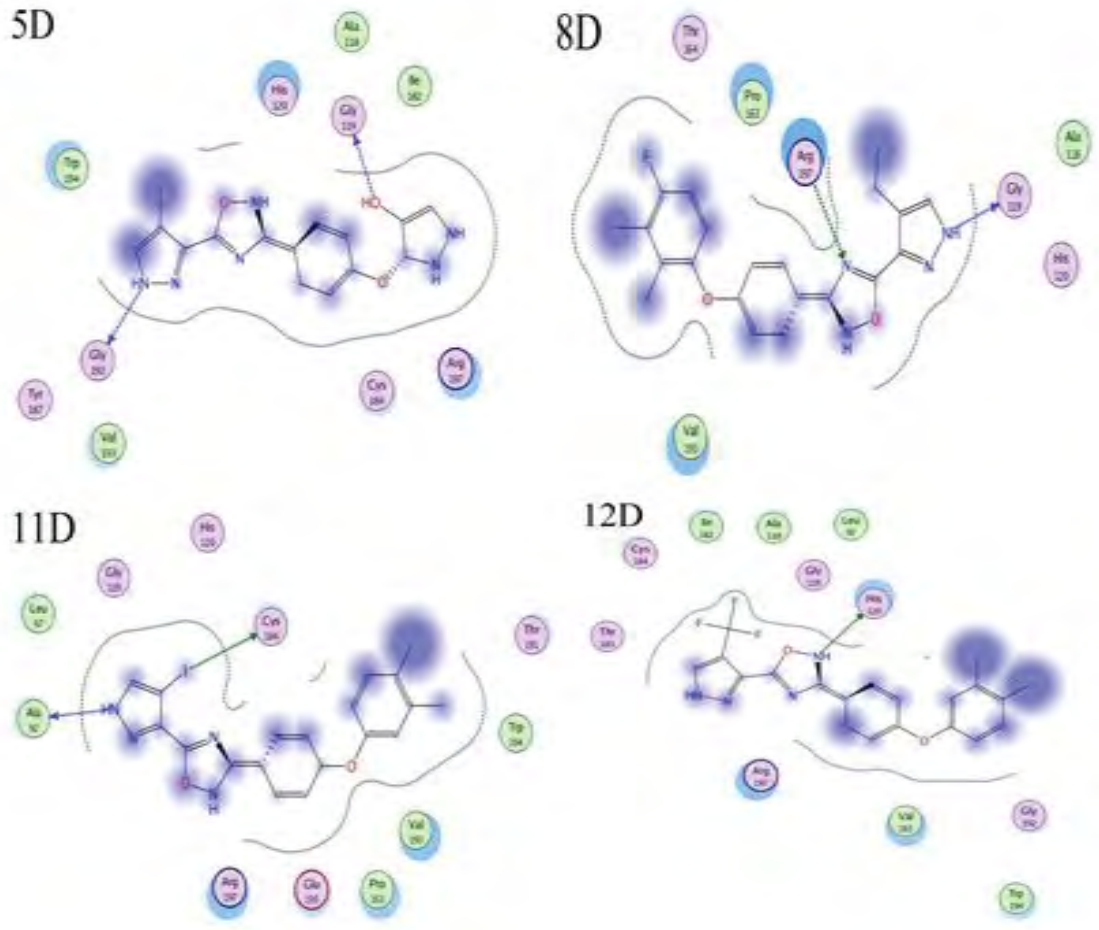


Figure 8 Interaction of some designed compounds in active site of enzyme. They have Aren cation interaction with Trp 194. Interaction with Arg 197 and Cys 184 may be a reason for high activity of compounds. Because they are located in catalytic site of enzyme.

Table 3 Thermochemical information from DFT calculation.

Method	Correlation coefficient (R^2) in training data	Correlation coefficient (R^2) in test data	Mean error in training data	Mean error in test data
PLS	0.88	0.66	0.70	1.28
GMDH	0.80	0.72	1.4	1.57

Conclusion

We can conclude that based on acceptable correlation coefficients (R^2), we can rely on PLS, and GMDH to design a practical compound based on previous libraries of compounds. The authors hope this study can be a step to guide other researchers toward the identification of effective anti-bacterial compounds with low toxicity and more efficacy.

Acknowledgement

We acknowledge Mashhad University of medical science for granting this project.

Conflict of Interest

This research funded by Mashhad University of medical science (number 25665). The author states that there is no conflict of interests.

References

1 Velázquez-Meza ME (2005) *Staphylococcus aureus* methicillin-resistant: Emergence and dissemination. Salud Pública de México 47: 381-387.

2 Boucher HW, Corey GR (2008) Epidemiology of Methicillin-Resistant *Staphylococcus aureus*. Clin Infect Dis 46: S344-S349.

3 Ding D, Boudreau AM, Leemans E, Spink E, Yamaguchi T, et al. (2015) Exploration of the structure–activity relationship of 1, 2, 4-oxadiazole antibiotics. Bioorg Med Chem Lett 25: 4854-4857.

4 Spink E, Ding D, Peng Z, Boudreau MA, Leemans E, et al. (2015) Structure–activity relationship for the oxadiazole class of antibiotics. J Med Chem 58: 1380-1389.

5 Leemans E, Mahasenan KV, Kumarasiri M, Spink E, Ding D, et al. (2016) Three-dimensional QSAR analysis and design of new 1, 2, 4-oxadiazole antibacterials. Bioorg Med Chem Lett 26: 1011-1015.

6 Rostami A, Hemmati-Sarapardeh A, Karkevandi-Talkhooncheh A, Husein MM, Shahaboddin M, et al. (2019) Modeling heat capacity of ionic liquids using group method of data handling: A hybrid and structure-based approach. I J Heat Mass Transfer 129: 7-17.

7 Menad NA, Noureddine Z, Hemmati-Sarapardeh A, Shamshirband S, Amir M, et al. (2019) Modeling temperature dependency of oil-water relative permeability in thermal enhanced oil recovery processes using group method of data handling and gene expression programming. Engineering Applications of Computational Fluid Mechanics 13: 724-743.

8 Barclay D, Higgins C, Thompson R (1995) The Partial Least Squares (pls) Approach to Casual Modeling: Personal Computer Adoption Ans Use as an Illustration.

9 Milne GW (2010) Software Review. Chem Bio Draw 12.0. ACS Publications.

10 Vilar S, Cozza G, Moro S (2008) Medicinal chemistry and the molecular operating environment (MOE): Application of QSAR and molecular docking to drug discovery. Current topics in medicinal chemistry 8: 1555-1572.

11 Frisch MJ (2009) Gaussian 09, Revision A. 1. Gaussian Inc. Wallingford CT, USA, 27: 34.

12 Todeschini R, Consonni V (2009) Molecular descriptors for chemoinformatics: volume I: alphabetical listing/volume II: appendices. John Wiley & Sons.

13 Seppelt R, Voinov A, Lange S (2012) Tools for environmental data mining and intelligent decision support. Iemss Org.

14 Koshulko O, Koshulko G (2011) Validation strategy selection in combinatorial and multilayered iterative GMDH algorithms. In: Proceedings of 4th international workshop on inductive modelling, Kyiv, Ukraine. Citeseer.

15 Guide (1998) The mathworks. Inc., Natick, MA, USA. 5: 333.

16 Suree N, Liew CK, Villareal VA, Thieu W, Fadeev EA, et al. (2009) The structure of the *Staphylococcus aureus* sortase-substrate complex reveals how the universally conserved LPXTG sorting signal is recognized. Journal of Biological Chemistry J Biol Chem 284: 24465-2477 .

17 Zhang Y, Bao J, Deng XX, He W, Fan JJ, et al. (2016) Synthesis, biological evaluation and molecular docking of 2-phenyl-benzo [d] oxazole-7-carboxamide derivatives as potential *Staphylococcus aureus* Sortase A inhibitors. Bioorganic & medicinal chemistry letters 26: 4081-4085.

18 Salter C, Foresman JB (1998) Naphthalene and Azulene I: semimicro bomb calorimetry and quantum mechanical calculations. J Chem Educ. 75: 1341.