ORIGINAL RESEARCH ARTICLES

2D QSAR MODEL BASED ON 1,2-DISUBSTITUTED BENZIMIDAZOLES IMPDH INHIBITORS

Maithili S. Pokle[a], Rashmi D. Singh[a] and Madhura P. Vaidya[a]*

(Received 04 August 2021) (Accepted 18 January 2022)

ABSTRACT

Quantitative structure activity relationship (QSAR) analysis of 1, 2-disubstituted benzimidazoles IMPDH inhibitors was studied for their antibacterial activity. The 2D QSAR model was developed using molecular suite (VLife MDS 4.3.1) on a set of 38 molecules. Multiple Linear Regression (MLR) was implemented for building a robust 2D QSAR model with various variable selection methods. The generated QSAR model emphasized that electronic, spatial, lipophilic and structural parameters play an important role in binding of benzimidazole derivatives to the receptor and thus in turn facilitates the further optimization of novel IMPDH inhibitors before synthesizing.

Keywords: Substituted benzimidazole, antibacterial, 2D QSAR, MLR

INTRODUCTION

A century ago, a cut could be deadly if it got infected but with the discovery of the first antibiotic penicillin, deaths due to infection went down rapidly. Since then, antibiotics have played a major role in increasing life expectancy and reducing morbidity and mortality. However, we are standing at the threshold of this golden epoch where the risk of running out of viable antibiotics is high. According to WHO and CDC, antibiotic resistance is one of the biggest threats to the health of human beings worldwide[1,2]. The need to design novel antibiotics to tackle this deficit is now more than ever to keep this constant war with micro-organisms at bay.

One such target is bacterial inosine-5'-monophosphate dehydrogenase (IMPDH) enzyme, which has recently gained a lot of attention due to reliance on this enzyme for the proliferation of microbes and the possibility of selective inhibition over host IMPDH. It catalyses the pivotal step in guanine nucleotide biosynthesis, the conversion of IMP to XMP and thus its inhibition can lead to anti-bacterial activity[3,4].

Umejiego et al., devised an HTS to identify inhibitors that target the NAD^+ site of *Cryptosporidium parvum* IMPDH (CpIMPDH) which is the divergent part of IMPDH[5]. In this screening study, benzimidazole analogue showed a lot of promise and thus was used as a starting point for further structure modifications for optimization of the activity[6]. The most active compounds owed their potency to the naphthyl group substituted in place of the aniline ring. The study of the structure of the CpIMPDH-IMP-C_{64} complex indicated the presence of a binding cavity near the aniline group and substitution with bulkier groups like naphthalene allows for better binding with the enzyme[7]. Replacing the methoxy group present on the para position of the phenyl ring with halogens like Cl, Br and F, increased the potency by 10 to 20-fold. In addition to this, the para position was found the best for halogen substitution to enhance the anti-bacterial activity[6].

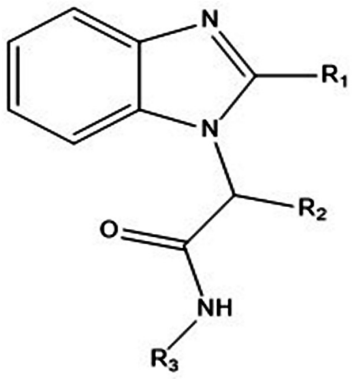
Biological systems are complex in nature and thus different models are used in drug research. One way to model quantitatively the bioactivity of a drug is by using QSAR[8]. It quantifies the increase or decrease in activity with the change in the structure and is a mathematical relationship linking chemical structure and pharmacological activity in a quantitative manner for a series of compounds. These models help predict the biological activity of compounds without synthesizing them in the wet-lab and thus expedite the process.

To further delve into the SAR relationship of these molecules and figure out which descriptors play a major

[a] Department of Pharmaceutical Chemistry, MET Institute of Pharmacy, Bandra (W), Mumbai- 400 050, Maharashtra, India
* For Correspondence: E-mail: madhurav_iop@met.edu

https://doi.org/10.53879/id.59.04.13117

Table I: Structure, experimental and predicted activity of quinazolines used in training and test set using MLR analysis



Compound No.	Compound	R1	R2	R3	IC50 (nm)	pIC50
1	C[a]	4-thiazolyl	-H	4-OMePh	1200 ± 200	5.9208
2	C39[a]	4-thiazolyl	-H	4-SMePh	120 ± 40	6.9208
3	C43[d]	4-thiazolyl	-H	4-i-PrPh	5000	5.3010
4	C9[b]	4-thiazolyl	-H	4-FPh	900 ±100	6.0457
5	C10[a]	4-thiazolyl	-H	4-ClPh	120 ± 40	6.9021
6	C11[a]	4-thiazolyl	-H	4-CF_3Ph	220 ± 40	6.6578
7	C58[b]	4-thiazolyl	-H	4-CNPh	370 ± 60	6.4317
8	C14[a]	4-thiazolyl	-H	4-BrPh	60 ± 30	7.221
9	C40[d]	4-thiazolyl	-H	4-SO_2MePh	5000	5.3010
10	C45[a]	4-thiazolyl	-H	4-OCF_3Ph	140 ± 50	6.853
11	C20[d]	4-thiazolyl	-H	2-ClPh	5000	5.3010
12	C48[a]	4-thiazolyl	-H	3-ClPh	490 ± 40	6.3098
13	C86[b]	4-thiazolyl	-H	3,4-DiClPh	30 ± 10	7.5228
14	C93[a]	4-thiazolyl	-H	3-CN, 4-ClPh	30 ± 10	7.5228
15	C90[c]	4-thiazolyl	-H	2-Naph	7± 4	8.1549
16	C28[d]	4-thiazolyl	-H	1-(4-Cl) Naph	5000	5.3010
17	C79[b]	4-thiazolyl	-CH_3	4-ClPh	60 ± 10	7.2218
18	C87[a]	4-thiazolyl	-CH_3	3,4-(OCH_2CH_3) Ph	240 ± 40	6.6197
19	C24[d]	4-thiazolyl	-CH (iPr)	4-ClPh	5000	5.3010
20	C18[d]	2-thiazolyl	-H	4-ClPh	5000	5.3010
21	C61[c]	2-thiazolyl	-H	4-ClPh	30 ± 10	7.5228
22	C64[a]	2-thiazolyl	-H	4-BrPh	28 ± 9	7.5228
23	C74[a]	2-thiazolyl	-CH_3	4-ClPh	23 ± 4	7.638
24	C84[a]	2-thiazolyl	-H	3,4-DiClPh	18 ± 5	7.7447
25	C97[a]	2-thiazolyl	-H	2-Naph	8 ± 3	8.0969
26	C67[a]	5-thiazolyl	-H	4-ClPh	35 ± 9	7.4559

27	C62[a]	2-thiophenyl	-H	4-ClPh	20 ± 20	7.698
28	C100[a]	1-methyl-4-pyrazolyl	-H	4-ClPh	35 ± 8	7.4559
29	C16[a]	2-pyridyl	-H	4-ClPh	43 ± 9	7.3665
30	C85[a]	2-pyridyl	-H	3,4-DiClPh	22 ± 5	7.657
31	C91[c]	2-pyridyl	-H	2-Naph	8 ± 3	8.0969
32	C92[a]	2-pyridyl	-H	3-CN, 4-ClPh	22 ± 10	7.657
33	C65[b]	2-pyrrolyl	-H	4-ClPh	80 ± 10	7.0969
34	C69[a]	2-oxazolyl	-H	4-ClPh	170 ± 10	6.769
35	C17[b]	Ph	-H	4-ClPh	210 ± 30	6.6777
36	C31[a]	4-ClPh	-H	4-ClPh	450 ± 20	6.346
37	C59[a]	4-FPh	-H	4-ClPh	870 ± 20	6.0404
38	C38[d]	3-thiazolyl	-H	4-OMePh	5000	5.3010

a = Training set *b = Test*

c = external *d = exempted from the study due to unusually high activity*

roles in determining the activity, we conducted computer aided 2D-QSAR study on a set of 38 benzimidazole analogues employing multiple linear regression (MLR). The models were evaluated for robustness using cross-validation and external test set prediction methods. The results obtained from these studies can be used for further drug design and optimization of IMPDH anti-microbial activity.

MATERIALS AND METHODS

Selection of molecules

Data set of 38 1,2-disubstituted benzimidazole derivatives (Table I) were selected from a published article to use for our 2D- QSAR study[6]. 10 molecules having exceptionally poor activity were removed from the dataset. The reported anti-bacterial activities were converted from IC_{50} value to pIC_{50} to get a linear relationship. To evaluate different models, data was divided into training and test set using square exclusion, random selection and manual selection methods.

Molecular modelling

All computerized experiments were performed using the Dell laptop having the Intel dual-core processor and the Windows XP operating system using the software molecular design suite, namely the VLife MDS[9]. The structures were drawn and converted to smiles format using Zinc15 database website[10] and then converted to mol2 format using VLife MDS software. These molecules were then subjected to energy minimization using the Merck molecular force field and force field charges followed by the Austin model 1 with 10000 as the maximum number of cycles, 0.01 as convergence criteria (the root mean square gradient) and 1.0 as constant (medium dielectric constant which is one in a vacuum) in dielectric properties. The values of the default of 30.0 and 10.0 kcal mol^{-1} were used as electrostatic and steric energy cut offs.

2D-QSAR ANALYSIS

Calculation of descriptors[11]

Approximately 1000 descriptors were calculated after energy optimization of the data set molecules. Various types of physicochemical descriptors such as–Individual descriptors (molecular weight, volume, H-acceptor and donor count, X logP, SMR, polarizability AHC, etc.), retention index that is Chi, atomic valence conductivity index that is ChiV, path count, Chi chain, ChiV chain,chain path count, cluster, cluster path, Kappa element count, element count, Dipole moment, electrostatic descriptors, distance based topological descriptors (Connectivity Index, Wiener Index, Radius of gyration, Moment of Inertia, etc.), Estate numbers (SddC count, SsNH3 count, SsCH3 count etc.), Estate contributions (SsCH3E-index, StCHE-index, SssNH2E-index, SsOHE-index etc.), Information theory based (Id Average, Ipc, Id, Idw Average, etc.), Semi empirical (LUMO Energy, QM DipoleY, etc.), Hydrophobicity XlogpA, Hydrophobicity XlogpK, Hydrophobicity Slogp A, Hydrophobicity SlogpK and Polar Surface Area descriptors were calculated. Around 700 Alignment Independent (AI) descriptors like

T_2_O_7, T_C_O_1, T_N_N_5, etc., and Atom type count descriptors were also calculated. The descriptors that were constant for all the molecules were removed as they did not contribute to the QSAR.

Generation of training and test sets and evaluation

To evaluate different models, data was divided into training and test sets using square exclusion, random selection and manual selection method. The training set was used to develop the QSAR model and the test set was used to assess the QSAR model's predictive power[12].

Various methods like sphere exclusion method, random selection method and manual data selection method were used to divide the data set into training and test sets[13].

Generation of 2D-QSAR models based on the data set

2D QSAR model was constructed using multiple linear regression (MLR) method for a series of 1,2-disubstituted benzimidazole derivatives as antimicrobial agents using software QSAR pro (VLife Sciences). MLR develops a mathematical linear relationship between activity (the dependant Y variable) and the descriptors (the independent X variable) such as[14] –

Activity = *f* (physicochemical properties or theoretical descriptors) + error

(Here, the error means either a model bias or observational variability.)

MLR extends the regression analysis to include more than one independent variable[15].

It is based on the least square method which is used to calculate regression coefficient (r^2). The model is fitted such that the sum of the square of the difference of observed and predicted value is minimized. The values of r^2 should be less than 1 and more than 0.7 for a good correlation between activity and descriptor.

RESULTS AND DISCUSSION

Multitude of 2D-QSAR models were produced using the MLR analysis in tandem with stepwise forward-backward, forward, backward, genetic algorithm and simulated annealing variable selection method. Methods like sphere exclusion, random and manual selection were used multiple times to separate the data set into test and training set. The test and training set combinations were accepted if they followed unicolumn statistics which states that the maximum of test is less than the maximum of training set and minimum of test is greater than that of the training set. This ensures that the test is a representative of the training set. The ratio of test set to training set was set at 80:20. A training set of 22 molecules and test set of 6 molecules was used for generation of QSAR model. External Validation Test set of 4 molecules was used for validation.

The best model is selected based on multiple factors like the values of correlation coefficient (r^2), cross-validation correlation coefficient (q^2) and predicted correlation coefficient for the external test set (pred_r^2) which represents the predictive power of the model and F ratio. The F ratio compares the variance in the data explained by the model to the variance attributed to error. High F value indicates that the model is statistically significant. The plot of observed versus predicted activities for the compounds is represented in Fig. 1.

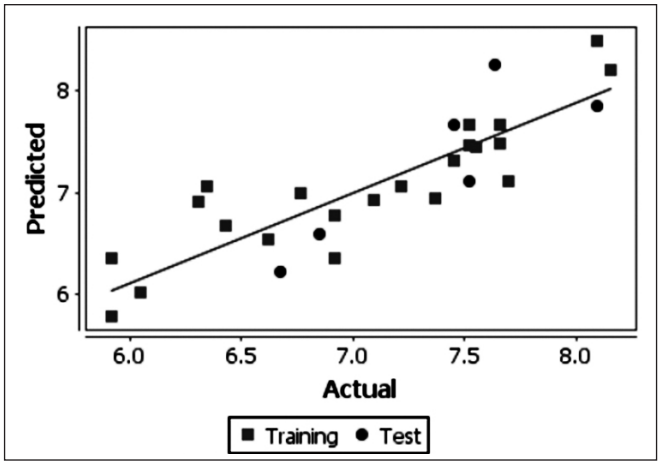


Fig. 1: The plot of observed versus predicted activities for the compounds

The best QSAR equation predicted the training data with a r^2 of 0.7591 together with q^2 estimating to 0.5381.

pIC_{50} = 0.6923 (± 0.1189) T_2_2_0 + 3.5778 (±0.8926) Chi V3 cluster- 0.47879 (±0.1179) T _2_2_5+ 0.2165 (± 0.0900) Dipole moment+ 0.0135 (± 0.0059) Quadrupole 2 + 0.0047

N = 22, D_f = 16, r^2 = 0.7591, q^2 = 0.5381, F = 10.0839, pred_r^2 = 0.5305, r^2 se =0.3832, q^2se = 0.5306

Interpretation of model

The statistically significant penta-parametric model with multiple regression methods with the coefficient of

determination (r^2) = 0.7591 can explain 75.91% of the variance in the observed activity values. The low standard error of r^2_se = 0.3832 demonstrates the accuracy of the model. The acceptable cross-validated squared correlation coefficient of this model was 0.5381, which shows a reasonable internal prediction power of this model. Another parameter for the predictivity of the test set compound is high pred_r^2 = 0.5302 which is showing the average external predictive power of the model.

Quadrupole2 signifies the magnitude of the second tensor of quadrupole moments. Dipole moment (DM) is an electronic descriptor, calculated from partial charges on the molecule. It indicates the strength and orientation behaviour of a molecule in an electrostatic field. The descriptor reflects the overall polarity of the molecule. Both the magnitude and the components (X, Y, Z) of the dipole moment are calculated. It is estimated by utilizing partial atomic charges and atomic coordinates. Overall charge distribution and the polarity of a molecule correlate positively to the biological activity. The positive correlation of DM signifies that an increase in the polarity of the molecules will lead to an increase in the activity. It also illustrates the non–covalent, electronic interactions between the receptor and inhibitor molecules. Chi V3 indices reflect the atom identities, bonding environments, and the number of bonding hydrogens. Moreover, the size, branching, unsaturation, cyclic and chemical nature of various chemical species are determined by molecule connectivity. They are the third-order cluster valence connectivity index (3) which encodes steric information and structural complexity, such as degree of substitution, length, and heteroatom content of substituted rings. It is highly sensitive to changes in branching and its value rapidly increases with the degree of branching. Therefore, an increase in the size, structural complexity, ring substitution pattern and degree of branching on the basic skeleton will increase the antagonist activity. T_2_2_0 is the count of the number of double-bonded atoms separated from any other bond by 0 bonds in a molecule. It shows a positive correlation with activity. An increase in double bond increases the activity.T_2_2_5 shows the number of double-bonded atoms separated from any other double-bonded atom any 5 bonds in a molecule. A negative correlation with this activity indicates the importance of more double bond.

CONCLUSION

The quantitative structure-activity relationship (QSAR) model gave a good correlation between physicochemical descriptors and biological activity and was found to be statistically significant. The model shows that-an increase in the polarity or polar surface area of the molecules will lead to an increase in the activity. An increase in the size, structural complexity, ring substitution pattern and degree of branching on the basic skeleton will increase the antagonist activity. The biological activity is sensitive to changes in branching and its value rapidly increases with the degree of branching. An increase in double bond increases the activity and the proximity of these double bonds should be less. An increase in the electronegativity of the molecule will cause an increase in the activity.

An overall view of the generated QSAR model emphasizes that electronic, spatial, lipophilic and structural parameters play an important role in the bonding of the benzimidazole derivatives to its receptor.

ACKNOWLEDGEMENT

The authors wish to express gratitude to VLife Science Technologies Pvt. Ltd., for providing the software for the study.

REFERENCES

1. CDC. Antibiotic Resistance Threats in the United States, 2019, Atlanta, GA: U.S. Department of Health and Human Services, CDC.
2. WHO. Global Action Plan on Antimicrobial Resistance,2015, WHO Press, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland.
3. Jackson R.C., Weber G. and Morris H.P.: IMP dehydrogenase, an enzyme linked with proliferation and malignancy, **Nature**, 1975, 256,331-333.
4. Gilbert H.J., Lowe C.R. and Drabble W.T.: Inosine 5'-monophosphate dehydrogenase of *Escherichia coli*. Purification by affinity chromatography, subunit structure and inhibition by guanosine 5'-monophosphate, **Biochem. J.**, 1979, 183, 481-494.
5. Umejiego N.N., Gollapalli D., Sharling L., Volftsun A., Lu J., BenjaminN.N., Stroupe A.H., Riera T.V., Striepen B. and Hedstrom L: Targeting a prokaryotic protein in a eukaryotic pathogen: identification of lead compounds against cryptosporidiosis, **Chem. Bio.**, 2008,15(1), 70–77.
6. Kirubakaran S., Gorla S.K., Sharling L., Zhang M., Liu X., Ray S.S., MacPherson I.S., Striepen B., Hedstrom L. and Cuny G.D.: Structure-activity relationship study of selective benzimidazole-based inhibitors of *Cryptosporidium parvum* IMPDH, **Bioorg. Med. Chem. Lett.**, 2012, 22, 1985-1988.
7. Macpherson I. S., Kirubakaran S. and Gorla S. K.: The structural basis of Cryptosporidium specific IMP dehydrogenase inhibitor selectivity, **J. Am. Chem. Soc.**, 2010, 132(4), 1230–1231.
8. Trinajstić N., Nikolić S.and Carter S.: QSAR: Theory and Application. Kemija u industriji, **J. Chemists Chem. Eng.**, 1989, 38, 469-484.

9. VLife MDS: Molecular Design Suite, Vlife Sciences Technologies Pvt. Ltd., Pune, India, 2010. (www.vlifesciences.com)
10. Sterling, T. and Irwin, J.J.: Zinc 15- Ligand Discovery for everyone, **J. Chem. Inf. Model.**, 2015, 55(11), 2324-2337.
11. "Molecular descriptor definitions in VLife QSAR." VLife adding value to life, Nova Lead Pharma, 2014.(Access date- 10/1/2020) (https://www.vlifesciences.com/support/QSAR_Descriptor_Defination s_faqs_Answer.php).
12. Afshar M. and Sallantin J.: Selecting quantitative structure–activity relationship models, in: Computer-Assisted Drug Design: in **Compr. Medi. Chemi. II**, Elsevier, 2007,4.30, 74-774.
13. Golbraikh A. and Tropsha A.: Predictive QSAR Modeling based on Diversity Sampling of Experimental Datasets for the Training and Test Set Selection, **J. Comp. Mol. Des.**, 2002, 16(5-6), 347-369.
14. Roy K., Das R.N. and Kar S.A.: Primer on QSAR/QSPR Modelling: Fundamental Concepts, Springer international publishing AG Switzerland, 2015.
15. Kulkarni V.M. and Bothra K.G.: Drug Design, Nirali Prakashan, Pune2007, pp. 187-235.