ORIGINAL RESEARCH PAPER
Volume - 10 | Issue - 08 | August - 2021 | PRINT ISSN No. 2277 - 8179 | DOI : 10.36106/ijsr
INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

# DEVELOPMENT AND VALIDATION OF QSAR MODEL FOR BIO - ACTIVITY PREDICTION OF THIAZOLIDINE- 4-ONE BEARING BENZIMIDAZOLE DERIVATIVES

Pharmacy

**Kadiyala Harshitha*** M.pharmacy, Pharmacology, post graduated from Andhra University. *Corresponding Author

**Pedapudi Sumeeth** M.Pharmacy, Post Graduated From Jawaharlal Nehru Technological University Kakinada.

## ABSTRACT

Plethora of strategies can be used in design of drugs. These include screening of natural products, screening synthetic compound from libraries, computer aided drug designing, 'me too' drugs, mimicking the natural ligand, drug repurposing and serendipity. QSAR approach attempts to identify and quantify the physicochemical properties of drug and see if any properties have an effect on drugs biological activity. The main aspects involved in QSAR approach to scientific research include: the concept of molecular descriptors and chemo metric tools. Present study is based on the assumption that the activity of a molecule is related to its structure so that similar molecules have similar activity so that QSAR modeling can be related to risk assessment. Basic purpose of this study is lead optimization seeking the most active compounds of a series, ultimately minimizing the expense, delay, and manpower required and is to see if the relationship is meaningful and quantify the goodness of the fit. Our study aimed at QSAR which is a very complex relationship between the structural properties of a drug and its biological behaviour. Here we took Benzimidazole since its derivatives are well known for anti-inflammatory activity and also recently have been discovered to have anticancer activity.30 QSAR models were successfully built. The good correlation between experimental and predicted pIC50 values for validation set compounds and cross validation proved the reliability of the QSAR model 10. The QSAR models revealed the importance of different physicochemical properties governing biological activity. Introduction of hydroxyl groups (VSA_POL) on the aromatic ring at R2 decreases activity and introduction of methoxy groups (fr-ether) on the aromatic ring at R2 increases activity.

## KEYWORDS

## INTRODUCTION

QSAR approach attempts to identify and quantify the physicochemical properties of drug and see if any properties have an effect on drugs biological activity. The main aspects involved in QSAR approach to scientific research include: the concept of molecular descriptors and chemo metric tools.[1] QSAR is a mathematical model of relationship between the structural properties of a drug and its biological behavior.

Analysis of QSAR include[1]
- PLS
- MLR
- ANN modeling.

QSAR is based on the assumption that the activity of a molecule is related to its structure so that similar molecules have similar activity. It is frequently used in aquatic toxicology, especially in chemical regulation and risk assessment.[2]

CLASSIC QSAR:
This is basically a lead optimization process seeking the most active compounds of a series, thus systematically minimizing the expense, delay, and manpower required.[2]

QSAR in simplest terms is a method of building computational models which attempt to find a statistically significant co relation between a range of compounds and their biological activity. In terms of drug design, structure refers to properties or descriptors of molecules,their substituents and their interaction energy field functions correspond to experimental biological end point like binding activity, toxicity, or rate constants, while chemometric methods include PLS, PCA, ANN etc.[2]

In the simplest situations, a range of compounds are synthesized in order to vary one physicochemical property (log1/c), a graph is then plotted against the biological activity on y – axis verses physicochemical parameter on x – axis. It is then necessary to draw the best possible line through the data points on the graph by a LINEAR REGRESSION ANALYSIS BY LEAST SQUARE ANALYSIS. The principle is that if we draw a line through a set of data points, most of the points will be scattered on the either side of the line. the one closest to the data points will be the best line.

The best line through the points will be the line where the total (sum of squares of verticals) is minimum. The equation for the straight line is $y = k_1 x + k_2$ where $k_1$ and $k_2$ are constants. By varying $k_1$ and $k_2$ we can obtain many equations until the best line is obtained.[3]

The next stage is to see if the relationship is meaningful and quantify the goodness of the fit. REGRESSION or CORELATION CO EFFICIENT is the measure of how well the equation explains the variation in activity observed in terms of physicochemical parameters present in the equation.[3]

## PHYSICOCHEMICAL PROPERTIES:

Many physical, structural and chemical properties have been studied by QSR approach, but most common are hydrophobic, electronic, and steric properties which can be quantified. Hydrophobic properties can be easily quantified. QSAR study considers how these properties show their effect on biological activity.[2]

## OTHER PHYSICOCHEMICAL PARAMETERS

The physicochemical properties also include
- Dipole moment
- Hydrogen bond
- Confirmational isomerism and
- Inter atomic distances.

Difficulties in quantifying these properties limit the use of these parameters, however several QSAR formulae have been developed based on highest occupied or the lowest unoccupied molecular orbitals of the test compounds. The calculation of these orbitals can be carried out using semi empirical quantum mechanical methods. Indicator variables for different substituents can also be used.[4]

## HANSCH EQUATION

The biological activity of most of the drugs is related to a combination of physicochemical properties. Simple equations are considered in this case involving only one parameter are relevant and only if the other parameters are kept constant. In reality, this is not easy to achieve equations which relate biological activity to different parameters. These equations are known as Hansch equations and they usually relate biological activity to the most commonly used physicochemical properties.[4]

## ADVANCES IN QSAR

QSARs relates physical and chemical properties of molecules to their biological activities by using easily calculable descriptors and simple statistical methods like Multiple Linear Regression (MLR) to build a model which describes both the activity of the data set and can predict activities for further sets of untested compounds. These types of descriptors often fail to take into account the three-dimensional nature of chemical structures which obviously play a part in ligand-receptor

binding, and hence activity. Steric, hydrophobic and electrostatic interactions are important to know if a molecule will interact optimally at its active site. It is logical for these interactions to be modelled to find the location in space around the molecule that are both acceptable and forbidden. The preceding QSAR methods usually do not take into account the 3D structure of the molecules or their targets such as enzymes and receptors. So, efforts have been made to explore structure-activity studies of ligands that take into account the known X-ray structures of proteins and enzymes, as well as the interaction of drugs with models of their receptors. Following are some of advanced approach to QSAR methodology.[1]

**Free Wilson Analysis**

The Free-Wilson approach is a structure-activity based methodology because it incorporates the contribution of various structural fragments to the overall biological activity. Indicator variables are used to denote the presence or absence of a particular structural feature. It is represented by the equation

$$BA = \Sigma a_i x_i + \mu$$

Where BA is the biological activity, $\mu$ is the overall activity, $a_i$ is the contribution of each structural feature, $x_i$ denotes the presence ($x_i = 1$) or absence ($x_i = 0$) of a particular structural fragment. This approach was easy to apply. However, it had its drawbacks, mostly centered on the large number of parameters and subsequent loss of the statistical degree of freedom. Fujita and Ban proposed a simplified approach that solely focused on the activity of group contribution.

$$\text{Log } A/A_0 = \Sigma G_i X_i$$

where A and $A_0$ represent the biological activity of the substituted and unsubstituted compounds respectively, while $G_i$ is the activity due to the substituent, $X_i$ has a value of 1 or 0 that corresponds to the presence or absence of that substituent. The delineation of these models led to an explosive development of QSAR analysis and related approaches.[4]

**APPLICATIONS:**

- Prediction of biological activity by rational means.
- Prediction could reduce the requirement of lengthy and painful animal testing.
- Promoting of greener chemistry by eliminating inactive molecules at the designing phase.
- Rationalization of the mechanism of action within a series of molecules.
- Biological activities of newer molecules can be used to predict QSAR equations.
- QSAR models have been used for risk management.[1]

**PARTIAL LEAST SQUARES**

- Partial least square is a popular method for soft modeling in industrial applications. Research in science and engineering often involves using controllable and easy to measure variables to explain, regulate, product the behavior of other variables (responses).[6]

- Partial least square is a method for constructing predictive models when the factors are many and highly collinear. PLS is not usually appropriate for screening out factors that have a negligible effect on the response but when prediction is the goal and there is no practical need to limit the number of measured factors, PLS can be a useful tool.[7]

- PLS was developed in the 1960's by Herman wold as an econometric technique, but some of its most avid proponents are chemical engineers and chemometricians. In addition to spectrometric calibration as discussed above, PLS has been applied to monitoring and controlling industrial processes; a large process can easily have hundreds of controllable variables and dozens of outputs.[6]

**CHEM DES**

Chem Des is an integrated platform for molecular descriptor computation

- ChemDes is a free web platform to calculate molecular descriptors, which provides more than 3679 molecular descriptors that are divided into 61 logical blocks.[8]
- In addition, it provides 59 types of molecular fingerprint systems for drug molecules. Molecular descriptors are experimentally measured or theoretically – derived properties of a molecule.[8]
- They are quantitative representations of physical, chemical or topological characteristics of molecules that summarize our knowledge and understanding of molecular structure and activity from different aspects.[8]

**BENZIMIDAZOLE**

Imidazole is the basic nucleus of the parent compound in the selected series. Imidazole or imidazoline is an azapyrrole, the nitrogen atoms are separated by one carbon atom. This compound was earlier called as glyoxalin and ammonia.[9]

The benzo derivative of imidazole is referred to as benzimidazole. Although benzimidazole is the common name of the parent compound of the series, other names such as 1, 3-benzodiazole are often used. An essential component of the of search for new leads in drug design program is the synthesis of molecules, which are novel but still resemble known biological actives by virtue of presence of some pharmacophoric groups. Benzimidazole derivatives are well known of their anti-inflammatory activity and recently have been discovered to have anti-cancer activity.[10]

Benzimidazole which is the core structure in various synthetic pharmaceuticals displaying a broad spectrum of biological activity such as anti-inflammatory activity, analgesic activity etc.

- When the coxibs were marketed, evidence for a new side effect appeared and rofecoxib were banned in 2004.[11]
- These compounds exhibit numerous biological properties such as anti- bacterial, anti-fungal, anti-tumor, anti-arrhythmic etc.[11]

**MATERIALS AND METHODS**

**Materials**

- Chem Draw software (version 8.0)

CHEM DRAW SOFTWARE: The Chem Draw platform is a digital drawing tool for chemists and biologists to create publication-ready, scientifically intelligent drawings for use in databases and publications.

It is a molecular editor first developed in 1985 by David A. Evans and Stewart Rubenstein. It is used for generating 2D molecular models. It helps in conversion of chemical structure to name, chemical name to structure, structure clean up.

**Chem Des:**

It is a free web platform for calculating molecular descriptors and fingerprints, which provides more than 3,679 molecular descriptors that are divided into 61 logical blocks. It provides 59 types of molecular fingerprint systems for drug molecules, including topological fingerprints, electro-topological state (E-state) fingerprints, MACCS keys, FP4 keys, atom pairs fingerprints, topological torsion fingerprints and Morgan/circular fingerprints.

Chem Des allows users to compute **3679** molecular descriptors from several open-source packages.

A QSAR study requires calculation of molecular descriptors; almost 200 physicochemical descriptors were generated by CHEM DES DESCRIPTORS.

- Chemopy Descriptors 1175
- CDK Descriptors
- RDsKit Descript0rs 196
- Pybel Descriptors **24**
- BlueDesc Descriptors **174**
- PaDEL Descriptors **1875**

**MOLECULAR DESCRIPTORS:**

Molecular descriptors are measured experimentally or derived theoretically for a molecule. To be more specific, they are quantitative representations of physical, chemical or topological characteristics of molecules that summarize our understanding of molecular structure and activity from different aspects. Molecular fingerprints are property profiles of a molecule, usually in forms of bit or count vectors with the vector elements indicating the existence or the frequencies of certain properties, respectively. Both molecular descriptors and fingerprints play an important role in QSAR/SAR analysis, similarity-based compound search, virtual molecule screening, target molecule ranking and the other drug discovery process.

**Table: 1 List Of PYBEL Descriptors With Their Meanings**

| DESCRIPTOR NAMES | DESCRIPTION |
|---|---|
| HBA1 | Number of hydrogen bond acceptors 1 |
| HBA2 | Number of hydrogen bond acceptors 2 |
| HBD | Number of hydrogen bond donors |
| A – Bonds | Number of aromatic bonds |

| | |
|---|---|
| Atoms | Number of atoms |
| Bonds | Number of bonds |
| D – Bonds | Number of double bonds |
| N F | Number of fluorine atoms |

**Table: 2 List Of RDKIT Descriptors With Their Meanings**

| **Topological descriptors** | **Balaban's J index, BertzCT** | |
|---|---|---|
| Connectivity descriptors Chi indices | Chi0, Chi1, Chi0v, Chi1v, Chi2v…. Chi4n | 12, quantifies complexity of molecules |
| MOE-type descriptors | EState_VSA1….. EState_VSA11 | 11, determine surface area contributions of different groups in the structure. |
| MOE-type descriptors | PEOE_VSA1 - PEOE_VSA14 | 14, Charge descriptors |
| MOE-type descriptors | SMR_VSA1 - SMR_VSA10 | 10,Van Der Waals surface area in relation to various groups |
| MOE-type descriptors | SlogP_VSA1 - SlogP_VSA12 | 12, Van Der Waals surface area in relation to logP contribution from various groups |

• XL STAT SOFTWARE: It is an application of PLS (partial least square method) for developing a QSAR model. XLSTAT software(version 2018) .This is a data analysis software .This allows easy integration with Microsoft excel software .This solution allows data analysis, data mining, data testing, data modelling and data visualization. XLSTAT sensory has been designed to provide a detailed view and insight into the information related to customers and products. XLSTAT enables users to make forecasts using a user-friendly interface. XLSTAT biomed has been developed for bio and medical personnel to make correct and timely decisions based on research and data analysis.[29]

**Features Of XLSTAT:**

• Data sampling
• Distribution sampling
• Coding by ranks
• Histograms
• 2-sample comparison of variance tests

**HOW DOES PARTIAL LEAST SQUARES WORK?**

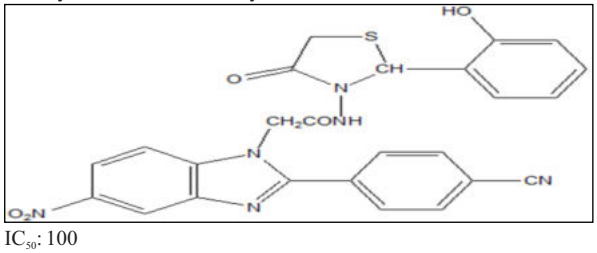
In principle, MLR can use many factors. However, when the number of factors get too large (eg: greater than the number of observations), you are likely to get a model that fits the sampled data perfectly but that will fail to predict the new data well. This phenomenon is called as over –fitting. In such cases although there are many manifest factors, there may be only few underlying factors that account for most of the variation in the response. The general idea of PLS is to try to extract these latent factors, accounting for as much of the manifest factor variation as possible while modelling responses well. For this reason, PLS also taken to mean "PROJECTION TO LATENT STRUCTURE. However, It should be noted that the term "LATENT" does not have the same technical meaning in the context of PLS as it does for other multivariate techniques. The overall goal is to use these factors to predict the responses in the population. This is achieved indirectly by latent variables T and U from sampled factors and responses respectively.[7]

**METHODOLOGY**

The QSAR study includes four phases:

**METHODS**

• A set of benzimidazole – 4 –one derivatives with anti – inflammatory activity and associated activity data were collected from literature.



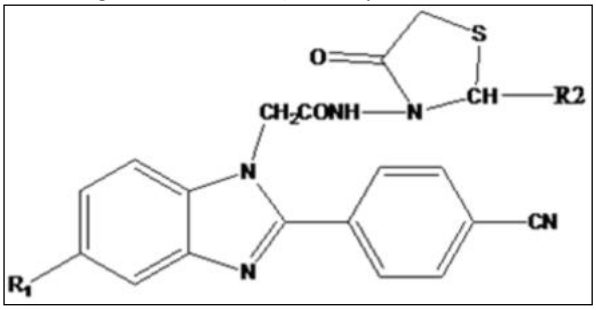

$IC_{50}$: 100
$pIC_{50}$: 4.605

• The biological activity data ($IC_{50}$ in n M) was converted into their molar units and further to negative log scale ($pIC_{50}$) and subsequently used as a dependent variable for QSAR analysis.

• Models of compounds were constructed using chem office software and converted to SMILS (simplified molecular – input line entry system).

**Phase: 1 Selection Of Drug Database**

• A set of benzimidazole - 4 - one derivatives with anti-inflammatory activity and associated activity data were collected from literature
• The biological activity data ($IC_{50}$ in µM) was converted into their molar units and further to negative log scale - $pIC_{50}$ and subsequently used as a dependent variable for QSAR analysis.

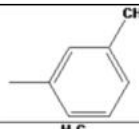
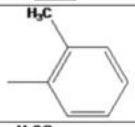
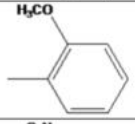
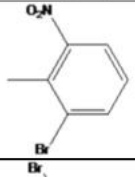
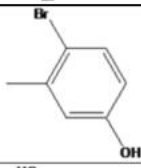
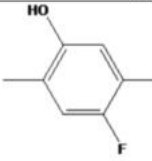
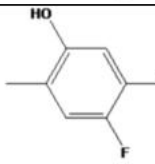
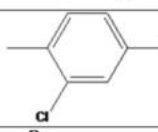
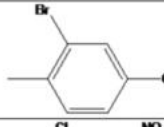
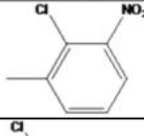
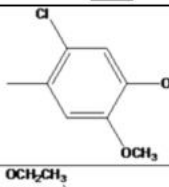
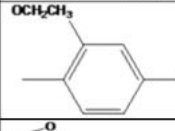
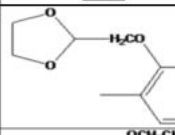
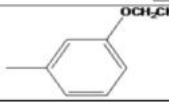


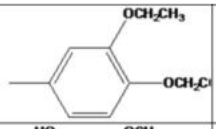
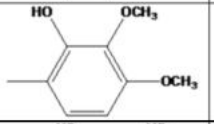
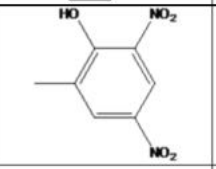
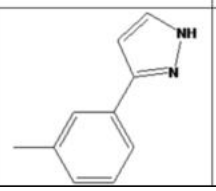

Structure

**Table: 3 Structure Table**

| Compound code | R1 | R2 | IC50 values | Smiles |
|---|---|---|---|---|
| b(1) | Cl | OCH3 | 5.12 | Clc1cc2nc(-c3ccc(C#N)cc3)n(CC(=O)N3C(c4ccc(C)cc4)SCC3=O)=O)c2cc1.[O]C |
| b(2) | Cl | Br | 10.84 | Clc1cc2nc(-c3ccc(C#N)cc3)n(CC(=O)N3C(c4ccc(Br)c4)SCC3=O)=O)c2cc1 |
| b(3) | NO2 | Cl | 8.67 | O=C1CSC(c2ccc(Cl)cc2)N1NC(=O)Cn1c(-c2ccc(C#N)cc2)nc2c1ccc([N+]([O-])=O)c2)=O |
| b(4) | NO2 | Br | 11.79 | O=C1CSC(c2ccc(C)cc2)N1NC(=O)Cn1c(-c2ccc(C#N)cc2)nc2c1ccc([N+]([O-])=O)c2)=O.[Br] |
| b(5) | NO2 | HO | 100 | O=C1CSC(c2ccccc2O)N1NC(=O)Cn1c(-c2ccc(C#N)cc2)nc2c1ccc([N+]([O-])=O)c2)=O |
| b(6) | Cl | Cl | 8.65 | Clc1cc2nc(-c3ccc(C#N)cc3)n(CC(=O)N3C(c4cc(Cl)c(C)cc4)SCC3=O)=O)c2cc1.[Cl] |
| b(7) | NO2 | HO | 17.65 | O=C1CSC(c2ccc(O)ccc2O)N1NC(=O)Cn1c(-c2ccc(C#N)cc2)nc2c1ccc([N+]([O-])=O)c2)=O |
| b(8) | NO2 | Cl | 100 | O=C1CSC(c2ccccc2Cl)N1NC(=O)Cn1c(-c2ccc(C#N)cc2)nc2c1ccc([N+]([O-])=O)c2)=O |
| b(9) | NO2 | Cl | 8.06 | O=C1CSC(c2ccc(Cl)cc2)N1NC(=O)Cn1c(-c2ccc(C#N)cc2)nc2c1ccc([N+]([O-])=O)c2)=O |
| b(10) | NO2 | Br | 10.42 | O=C1CSC(c2ccc(Br)c2)N1NC(=O)Cn1c(-c2ccc(C#N)cc2)nc2c1ccc([N+]([O-])=O)c2)=O |
| b(11) | NO2 | OCH3 | 4.93 | O=C1CSC(c2ccc(C)cc2)N1NC(=O)Cn1c(-c2ccc(C#N)cc2)nc2c1ccc([N+]([O-])=O)c2)=O.[O]C |

PRINT ISSN No. 2277 - 8179 | DOI : 10.36106/ijsr

| | | | | |
|---|---|---|---|---|
| b(12) | $NO_2$ |  | | O=C1CSC(c2ccccc(CBr)c2)N1NC(Cn1c(-c2ccc(C#N)cc2)nc2c1ccc([N+]([O-])=O)c2)=O |
| b(13) | $NO_2$ |  | | O=C1CSC(c2ccccc2C)N1NC(Cn1c(-c2ccc(C#N)cc2)nc2c1ccc([N+]([O-])=O)c2)=O |
| b(14) | Cl |  | 8.39 | O=C1CSC(c2ccccc2C)N1NC(Cn1c(-c2ccc(C#N)cc2)nc2c1ccc([N+]([O-])=O)c2)=O |
| b(15) | Cl |  | 100 | Clc1cc2nc(-c3ccc(C#N)cc3)n(CC(NN3C(c4c(Br)cccc4[N+]([O-])=O)SCC3=O)=O)c2cc1 |
| b(16) | Cl |  | 23.6 | Clc1cc2nc(-c3ccc(C#N)cc3)n(CC(NN3C(c4cc(O)ccc4Br)SCC3=O)=O)c2cc1 |
| b(17) | Cl |  | 11.33 | Clc1cc2nc(-c3ccc(C#N)cc3)n(CC(NN3C(c4cc(F)c(F)cc4O)SCC3=O)=O)c2cc1 |
| b(18) | $NO_2$ |  | 10.87 | O=C1CSC(c2cc(F)c(F)cc2O)N1N\C(Cn1c(-c2ccc(C#N)cc2)nc2c1ccc([N+]([O-])=O)c2)=O |
| b(19) | Cl |  | 4.82 | Clc1cc2nc(-c3ccc(C#N)cc3)n(CC(NN3C(c4c(Cl)cc(Cl)cc4)SCC3=O)=O)c2cc1 |
| b(20) | Cl |  | 11.77 | Clc1cc2nc(-c3ccc(C#N)cc3)n(CC(NN3C(c4ccc(C)cc4Br)SCC3=O)=O)c2cc1.F[C](F)F |
| b(21) | Cl |  | 18.99 | Clc1cc2nc(-c3ccc(C#N)cc3)n(CC(NN3C(c4c(Br)cccc[N+]([O-])=O)c4Cl)SCC3=O)=O)c2cc1 |
| b(22) | Cl |  | 4.89 | Clc1cc2nc(-c3ccc(C#N)cc3)n(CC(NN3C(c4cc(OC)c(OC)cc4Cl)SCC3=O)=O)c2cc1 |
| b(23) | Cl |  | | Clc1cc2nc(-c3ccc(C#N)cc3)n(CC(NN3C(c4ccc(OC)cc4OCC)SCC3=O)=O)c2cc1 |
| b(24) | $NO_2$ |  | | O=C1CSC(c2ccccc2C)N1CC(Nn1c(-c2ccc(C#N)cc2)nc2c1ccc([N+]([O-])=O)c2)=O.[C]C1OCCO1.O |
| b(25) | Cl |  | | Clc1cc2nc(-c3ccc(C#N)cc3)n(CC(NN3C(c4c(Br)ccc(OCCCBr)c4)SCC3=O)=O)c2cc1 |

| | | | |
|---|---|---|---|
| b(26) | Cl |  | Clc1cc2nc(-c3ccc(C#N)cc3)n(CC(NN3C(c4ccc(OCC#C)c(OCC)c4)SCC3=O)=O)c2cc1 |
| b(27) | Cl |  | Clc1cc2nc(-c3ccc(C#N)cc3)n(NC(CN3C(c4ccc(OC)c(C)c4O)SCC3=O)=O)c2cc1.[O]C |
| b(28) | Cl |  | Clc1cc2nc(-c3ccc(C#N)cc3)n(NC(CN3C(c4ccc([N+]([O-])=O)cc[N+]([O-])=O)c4O)SCC3=O)=O)c2cc1 |
| b(29) | | | Clc1ccc2c(cc(-c3ccc(C#N)cc3)n2CC(NN2C(c3cc(C(C)(C)C)cc(C(C)(C)C)c3O)SCC2=O)=O)c1 |
| b(30) | Cl |  | Clc1cc2nc(-c3ccc(C#N)cc3)n(CC(NN3C(c4ccc(OC)cc4OCC)SCC3=O)=O)c2cc1 |

**Phase 2: Construction Of QSAR Molecular Models**

QSAR models of compounds were constructed using Chem office software and converted to SMILES (simplified molecular-input line-entry system).

**Phase 3 Generation Of PYBEL Descriptors**

- Select Webserver option on http://www.scbdd.com/chemdes/
- Select PYBEL Descriptor calculator from menu.
- Submit smiles of the structure under study.
- Download the data.

**Generation Of RD KIT Descriptors**

- Select Webserver option on http://www.scbdd.com/chemdes/
- Select RDKIT Descriptor calculator from menu.
- Submit smiles of the structure under study.
- Download the data. A total of 90 2D descriptors and 106 1D descriptors were generated. Some of the 2D descriptors were selected for QSAR model building.

**Generation Of CHEMOPY Descriptors**

- Select Webserver option on http://www.scbdd.com/chemdes/
- Select CHEMOPY Descriptor calculator from menu.
- Submit smiles of the structure under study.
- Download the data.

Some of the 2D descriptors were selected for QSAR model building.

**Phase 4: QSAR Model Development**

- Selection of explanatory variables (molecular descriptors) (quantitative)
- Division of active set and validation set of molecules.
- Using PLS-R to build QSAR models, their validation and selection of best model based on $Q^2$ values.

**RESULTS**

In the present work, efforts have been made to find out the structural requirements for optimum COX1 inhibitory activity of Thiazolidine-4-one bearing benzimidazole derivatives.

Quantitative models were developed taking into consideration the structural contribution from various substituents using partial least squares regression method.

In the first step three different models were built using three, four and five parameters respectively to select the optimum number of explanatory variables.

For a data set containing not more than 50 compounds not more than five explanatory variables can be used to build a relevant PLS model.

The data of the predicted $IC_{50}$ values and the corresponding residuals have been represented in the following table.

**Table: 4a Model With Five Parameters**

| Pic 50 | Pred(Pic 50) | Residual |
|---|---|---|
| 5.290 | 5.244 | 0.046 |
| 4.960 | 4.917 | 0.043 |
| 5.060 | 4.770 | 0.290 |
| 4.920 | 4.715 | 0.205 |
| 4.000 | 4.662 | -0.662 |
| 5.060 | 5.352 | -0.292 |
| 4.000 | 4.854 | -0.854 |
| 5.300 | 4.828 | 0.472 |
| 5.076 | 4.715 | 0.361 |
| 4.000 | 4.424 | -0.424 |
| 4.620 | 4.333 | 0.287 |
| 4.940 | 4.847 | 0.093 |
| 4.960 | 4.973 | -0.013 |
| 5.310 | 5.063 | 0.247 |
| 4.920 | 4.917 | 0.003 |
| 4.720 | 4.654 | 0.066 |
| 5.310 | 5.178 | 0.132 |
| 4.720 | 4.871 | -0.151 |
| 5.090 | 4.816 | 0.274 |
| 4.980 | 4.427 | 0.553 |

Fig: 5 Variable importance of the descriptive variables
$Q^2$: -0.017

**Table:4b Model With Four Parameters**

| Pic 50 | Pred(Pic 50) | Residual |
|---|---|---|
| 5.290 | 5.244 | 0.046 |
| 4.960 | 4.917 | 0.043 |
| 5.060 | 4.770 | 0.290 |
| 4.920 | 4.715 | 0.205 |
| 4.000 | 4.662 | -0.662 |
| 5.060 | 5.352 | -0.292 |
| 4.000 | 4.854 | -0.854 |
| 5.300 | 4.828 | 0.472 |
| 5.076 | 4.715 | 0.361 |
| 4.000 | 4.424 | -0.424 |
| 4.620 | 4.333 | 0.287 |
| 4.940 | 4.847 | 0.093 |
| 4.960 | 4.973 | -0.013 |
| 5.310 | 5.063 | 0.247 |
| 4.920 | 4.917 | 0.003 |
| 4.720 | 4.654 | 0.066 |
| 5.310 | 5.178 | 0.132 |
| 4.720 | 4.871 | -0.151 |
| 5.090 | 4.816 | 0.274 |
| 4.980 | 4.427 | 0.553 |

Fig: 6 Variable importance of the descriptive variables
$Q^2$: 0.11

**Table: 4c Model With Three Parameters**

| Pic 50 | Pred(Pic 50) | Residual |
|---|---|---|
| 5.060 | 4.915 | 0.145 |
| 4.920 | 4.863 | 0.057 |
| 4.000 | 4.850 | -0.150 |
| 4.720 | 4.799 | -0.079 |
| 4.000 | 4.136 | -0.136 |
| 5.090 | 4.928 | 0.072 |
| 4.980 | 4.859 | 0.121 |
| 5.300 | 5.247 | 0.053 |
| 5.076 | 4.863 | 0.213 |
| 4.000 | 4.139 | -0.139 |
| 4.620 | 4.716 | -0.096 |
| 4.940 | 4.834 | 0.106 |
| 4.960 | 4.922 | 0.038 |
| 5.310 | 5.253 | 0.057 |
| 4.920 | 4.824 | 0.096 |
| 4.720 | 4.735 | -0.015 |
| 5.310 | 5.219 | 0.091 |
| 5.290 | 5.148 | 0.142 |
| 4.960 | 4.824 | 0.136 |
| 5.060 | 4.863 | 0.197 |

Fig: 7 Variable importance of the descriptive variables
$Q^2$: 0.265

Triparametric models showed better fit with better $Q^2$ value, hence 25 triparametric models were built taking into consideration various RD-KIT descriptors that were generated.

**Table4d: Model 1**

| $pIC_{50}$ | $Pred(pIC_{50})$ | Residual |
|---|---|---|
| 5.290 | 5.222 | 0.068 |
| 4.960 | 5.158 | -0.198 |
| 4.920 | 5.085 | -0.165 |
| 5.014 | 5.021 | -0.007 |
| 5.060 | 5.151 | -0.091 |
| 5.014 | 5.021 | -0.007 |
| 5.090 | 5.021 | 0.069 |
| 4.980 | 5.021 | -0.041 |
| 5.300 | 5.085 | 0.215 |
| 5.014 | 5.027 | -0.013 |
| 5.014 | 5.092 | -0.078 |
| 5.076 | 5.092 | -0.016 |
| 4.940 | 5.016 | -0.076 |
| 4.960 | 4.879 | 0.081 |
| 5.310 | 5.087 | 0.223 |
| 4.920 | 5.008 | -0.088 |
| 4.720 | 4.879 | -0.159 |
| 5.310 | 5.027 | 0.283 |
| 5.060 | 5.021 | 0.039 |
| 4.720 | 4.950 | -0.230 |
| 5.014 | 4.950 | 0.064 |
| 4.620 | 5.087 | -0.467 |

**Table4e: Model 2**

| $pIC_{50}$ | $Pred(pIC_{50})$ | Residual |
|---|---|---|
| 5.290 | 5.202 | 0.088 |
| 4.960 | 5.165 | -0.205 |
| 5.060 | 4.982 | 0.078 |
| 4.920 | 5.019 | -0.099 |
| 5.014 | 4.982 | 0.032 |
| 4.720 | 4.919 | -0.199 |
| 5.014 | 4.982 | 0.032 |
| 5.090 | 4.982 | 0.108 |
| 4.980 | 4.982 | -0.002 |
| 5.014 | 4.924 | 0.090 |
| 5.014 | 5.045 | -0.031 |
| 5.076 | 5.045 | 0.031 |
| 5.014 | 4.919 | 0.095 |
| 4.940 | 5.039 | -0.099 |
| 4.960 | 4.856 | 0.104 |
| 5.310 | 5.102 | 0.208 |
| 4.920 | 5.014 | -0.094 |
| 4.720 | 4.856 | -0.136 |
| 5.060 | 5.139 | -0.079 |
| 5.300 | 5.019 | 0.281 |
| 4.620 | 5.102 | -0.482 |
| 5.310 | 4.924 | 0.386 |

**Table 4f:model3**

| $pIC_{50}$ | $Pred(pIC_{50})$ | Residual |
|---|---|---|
| 4.960 | 4.942 | 0.018 |
| 5.060 | 4.956 | 0.104 |
| 4.920 | 5.144 | -0.224 |
| 5.014 | 4.956 | 0.058 |
| 5.060 | 5.088 | -0.028 |
| 4.720 | 4.914 | -0.194 |
| 5.014 | 4.956 | 0.058 |
| 5.090 | 4.956 | 0.134 |
| 4.980 | 4.956 | 0.024 |
| 5.300 | 5.144 | 0.156 |
| 5.014 | 5.054 | -0.041 |
| 5.014 | 4.998 | 0.016 |
| 5.076 | 4.998 | 0.078 |
| 5.014 | 4.914 | 0.100 |

| | | |
|---|---|---|
| 4.620 | 4.900 | -0.280 |
| 4.940 | 4.857 | 0.083 |
| 4.960 | 4.871 | 0.089 |
| 4.720 | 4.871 | -0.151 |
| 5.290 | 5.130 | 0.160 |
| 5.310 | 4.900 | 0.410 |
| 4.920 | 5.003 | -0.083 |
| 5.310 | 5.054 | 0.256 |

**Table 4g: Model4**

| $pIC_{50}$ | $Pred(pIC_{50})$ | Residual |
|---|---|---|
| 5.290 | 5.177 | 0.113 |
| 4.960 | 5.002 | -0.042 |
| 5.060 | 4.963 | 0.097 |
| 4.920 | 5.137 | -0.217 |
| 5.014 | 4.963 | 0.051 |
| 5.060 | 5.142 | -0.082 |
| 4.720 | 4.928 | -0.208 |
| 5.014 | 4.963 | 0.051 |
| 5.090 | 4.963 | 0.127 |
| 4.980 | 4.963 | 0.017 |
| 5.300 | 5.137 | 0.163 |
| 5.014 | 4.992 | 0.022 |
| 5.014 | 4.997 | 0.017 |
| 5.076 | 4.997 | 0.079 |
| 5.014 | 4.928 | 0.086 |
| 4.620 | 4.968 | -0.348 |
| 4.940 | 4.933 | 0.007 |
| 4.960 | 4.894 | 0.066 |
| 5.310 | 4.968 | 0.342 |
| 4.920 | 5.074 | -0.154 |
| 4.720 | 4.894 | -0.174 |
| 5.310 | 4.992 | 0.318 |

**Table4h: Model 5**

| $pIC_{50}$ | $Pred(pIC_{50})$ | Residual |
|---|---|---|
| 5.290 | 5.177 | 0.113 |
| 4.960 | 5.002 | -0.042 |
| 5.060 | 4.963 | 0.097 |
| 4.920 | 5.137 | -0.217 |
| 5.014 | 4.963 | 0.051 |
| 5.060 | 5.142 | -0.082 |
| 4.720 | 4.928 | -0.208 |
| 5.014 | 4.963 | 0.051 |
| 5.090 | 4.963 | 0.127 |
| 4.980 | 4.963 | 0.017 |
| 5.300 | 5.137 | 0.163 |
| 5.014 | 4.992 | 0.022 |
| 5.014 | 4.997 | 0.017 |
| 5.076 | 4.997 | 0.079 |
| 5.014 | 4.928 | 0.086 |
| 4.620 | 4.968 | -0.348 |
| 4.940 | 4.933 | 0.007 |
| 4.960 | 4.894 | 0.066 |
| 5.310 | 4.968 | 0.342 |
| 4.920 | 5.074 | -0.154 |
| 4.720 | 4.894 | -0.174 |
| 5.310 | 4.992 | 0.318 |

**Table4i: Model 6**

| $pIC_{50}$ | $Pred(pIC_{50})$ | Residual |
|---|---|---|
| 5.290 | 5.105 | 0.185 |
| 5.060 | 5.013 | 0.047 |
| 4.920 | 5.013 | -0.093 |
| 5.060 | 5.105 | -0.045 |
| 4.720 | 4.875 | -0.155 |
| 5.014 | 5.013 | 0.001 |
| 5.090 | 5.013 | 0.077 |
| 4.980 | 5.013 | -0.033 |
| 5.300 | 5.013 | 0.287 |
| 5.014 | 5.013 | 0.001 |
| 5.014 | 5.013 | 0.001 |
| 4.620 | 5.036 | -0.416 |
| 4.940 | 4.894 | 0.046 |
| 4.960 | 4.802 | 0.158 |
| 5.310 | 5.105 | 0.205 |
| 4.920 | 4.892 | 0.028 |
| 4.720 | 5.013 | -0.293 |
| 5.310 | 5.315 | -0.005 |
| 4.960 | 5.105 | -0.145 |
| 5.014 | 4.944 | 0.070 |
| 5.014 | 5.013 | 0.001 |
| 5.076 | 5.013 | 0.063 |

**Table4j: Model 7**

| $pIC_{50}$ | $Pred(pIC_{50})$ | Residual |
|---|---|---|
| 4.960 | 5.088 | -0.128 |
| 5.060 | 5.026 | 0.034 |
| 5.014 | 4.941 | 0.073 |
| 5.060 | 5.088 | -0.028 |
| 4.720 | 4.856 | -0.136 |
| 5.014 | 5.026 | -0.012 |
| 5.090 | 5.026 | 0.064 |
| 4.980 | 5.026 | -0.046 |
| 5.300 | 5.026 | 0.274 |
| 5.014 | 5.026 | -0.012 |
| 5.014 | 5.026 | -0.012 |
| 5.014 | 5.026 | -0.012 |
| 4.620 | 5.003 | -0.383 |
| 4.940 | 4.907 | 0.033 |
| 4.960 | 4.845 | 0.115 |
| 5.310 | 5.088 | 0.222 |
| 4.920 | 4.944 | -0.024 |
| 5.310 | 5.332 | -0.022 |
| 5.290 | 5.088 | 0.202 |
| 4.920 | 5.026 | -0.106 |
| 5.076 | 5.026 | 0.050 |
| 4.720 | 5.026 | -0.306 |

**Table4k: Model 8**

| $pIC_{50}$ | $Pred(pIC_{50})$ | Residual |
|---|---|---|
| 5.290 | 5.111 | 0.179 |
| 4.960 | 5.111 | -0.151 |
| 5.060 | 5.028 | 0.032 |
| 4.920 | 5.028 | -0.108 |
| 5.014 | 4.942 | 0.071 |
| 5.060 | 5.111 | -0.051 |
| 4.720 | 4.857 | -0.137 |
| 5.014 | 5.028 | -0.014 |
| 4.980 | 5.028 | -0.048 |
| 5.300 | 5.028 | 0.272 |
| 5.014 | 5.028 | -0.014 |
| 5.076 | 5.028 | 0.048 |
| 5.014 | 5.028 | -0.014 |
| 4.620 | 5.025 | -0.405 |
| 4.960 | 4.815 | 0.145 |
| 5.310 | 5.111 | 0.199 |
| 4.920 | 4.919 | 0.001 |
| 5.310 | 5.317 | -0.007 |
| 5.090 | 5.028 | 0.062 |
| 5.014 | 5.028 | -0.014 |
| 4.940 | 4.897 | 0.043 |
| 4.720 | 5.028 | -0.308 |

**Table4l: Model 9**

| $pIC_{50}$ | $Pred(pIC_{50})$ | Residual |
|---|---|---|
| 5.290 | 5.100 | 0.190 |
| 4.960 | 5.100 | -0.140 |
| 5.060 | 5.015 | 0.045 |
| 4.920 | 5.015 | -0.095 |
| 5.014 | 4.943 | 0.071 |
| 4.720 | 4.871 | -0.151 |
| 5.090 | 5.015 | 0.075 |
| 4.980 | 5.015 | -0.035 |
| 5.300 | 5.015 | 0.285 |
| 5.014 | 5.015 | -0.002 |
| 5.076 | 5.015 | 0.061 |
| 5.014 | 5.015 | -0.002 |

| | | |
|---|---|---|
| 4.620 | 5.027 | -0.407 |
| 4.940 | 4.897 | 0.043 |
| 4.960 | 4.813 | 0.147 |
| 5.310 | 5.100 | 0.210 |
| 4.720 | 5.015 | -0.295 |
| 5.310 | 5.307 | 0.003 |
| 5.060 | 5.100 | -0.040 |
| 5.014 | 5.015 | -0.002 |
| 5.014 | 5.015 | -0.002 |
| 4.920 | 4.905 | 0.015 |

**Table4m: Model 10**

| $pIC_{50}$ | Pred($pIC_{50}$) | Residual |
|---|---|---|
| 5.290 | 5.254 | 0.036 |
| 4.960 | 5.024 | -0.064 |
| 5.060 | 5.073 | -0.013 |
| 4.920 | 4.973 | -0.053 |
| 5.014 | 4.974 | 0.04 |
| 5.060 | 5.100 | -0.040 |
| 4.720 | 4.685 | 0.035 |
| 5.014 | 5.073 | -0.059 |
| 5.090 | 5.073 | 0.017 |
| 4.980 | 5.073 | -0.093 |
| 5.300 | 5.327 | -0.027 |
| 5.014 | 5.066 | -0.052 |
| 4 | 3.971 | 0.029 |
| 5.076 | 5.066 | 0.010 |
| 5.076 | 5.066 | 0.010 |
| 5.014 | 4.999 | 0.015 |
| 4.620 | 4.645 | -0.025 |
| 4.940 | 4.878 | 0.062 |
| 4.960 | 4.918 | 0.042 |
| 5.310 | 5.267 | 0.043 |
| 5.310 | 5.307 | 0.003 |
| 4.720 | 4.731 | -0.011 |

**Table4n: Correlation Matrix:**

| Variables | Q_VSA_HYD | VSA_POL | fr-ether | $pIC_{50}$ |
|---|---|---|---|---|
| Q_VSA_HYD | 1.000 | 0.973 | 0.872 | 0.932 |
| VSA_POL | 0.973 | 1.000 | 0.705 | 0.833 |
| fr-ether | 0.772 | 0.705 | 1.000 | 0.774 |
| $pIC_{50}$ | 0.876 | 0.921 | 0.865 | 1.000 |

Fig: 4-1 Correlation between $pIC_{50}$ and predicted $pIC_{50}$ values via internal validation
$Q^2 = 0.73$

Fig: 4-2 Correlation between $pIC_{50}$ and predicted $pIC_{50}$ values via external validation

$N_{training} = 17, N_{test} = 3, r^2 = 0.977, q^2 = 0.73$
The equation for MODEL 10 is as follows:
$pIC_{50} = 5.836440596386 7 - 7.122756825000 95E-02 * Q\_VSA\_HYD + 6.399570201120 77E-02$ *VSA_POL+5.63190331147629E-03*fr-ether

## DISCUSSION

The derived QSAR MODEL 10 showed significant correlation between the biological activity and molecular descriptors. The value of cross validated correlation coefficient ($q^2$ =0.73) using leave one out method suggested good internal predictivity(73%) of the equation. This model shows external predictivity(97%) with $r^2$ (0.977). The variable importance of various descriptors in the prediction of biological activity has been shown in Fig. .It indicates that MODEL 10 can be successfully applied to predict the COX inhibitory activity of these classes of molecules.

The descriptor Q_VSA_HYD describes the total hydrophobic Van der waals surface area. The increase in total hydrophobic surface area is found to contribute to an increase in biological activity. The replacement of the more lipophilic chloro substituent by the less lipophilic nitro substituent leads to a decrease of COX-1 inhibitory activity in the case of compounds 3 and 21 respectively.

The descriptor VSA_POL approximates the sum of Van der waals surface area of polar atoms (both HBD AND HBA such as –OH). Replacement of the dihydroxy derivative with halogens and methoxy groups has resulted in an increase of COX-1 inhibitory activity in the case of compounds 18 and 22 respectively.

The descriptor fr-ether describes the number of ether oxygens present in a molecule. Increase in the number of ether oxygens results in an increase of COX-1 inhibitory activity in the case of compounds 11 and 22.

## CONCLUSION

- Thirty QSAR models were successfully built.
- The good correlation between experimental and predicted $pIC_{50}$ values for validation set compounds and cross validation proved the reliability of the QSAR model 10.
- The QSAR models revealed the importance of different physicochemical properties governing biological activity.
- Introduction of hydroxyl groups (VSA_POL) on the aromatic ring at R2 decreases activity and introduction of methoxy groups (fr-ether) on the aromatic ring at R2 increases activity.

## REFERENCES


1. Wilson and Gisvold's Textbook of Organic Medicinal and Pharmaceutical Chemistry-11th edition.
2. Foye's Principals of Medicinal Chemistry ,6th edition David A .WILLIAMS.
3. Graham L. Patrick- An Introduction to Medicinal Chemistry,3rd edition.
4. Nantasenamat C,Isarankura – NA - Ayudhya C, Naenna T(2009)A Practical Overview of Quantitative Structure Activity Relationship.Excli J. 8:-74-88.
5. Nantasenamat C , Isarankura-Na-Ayudhya C , Prachayasittikul V (July 2010) Advances in computational method to predict the biological activities of compounds.Expert Opinion of drug discovery . 5(7):633-54.
6. An Introduction to partial least square regression Randall D. Tobias,SAS Institute Inc.,Cary,NC.
7. Afaf, H. H. Fahmy , S.H. Abdelwal Molecules, 5 (2000), p. 1429.
8. http://www.scbdd.com/chemdes/
9. Amari et al,.2002M. Amari , M. Fodili , B. Nedjar Kolli Reactivity studies on 4-aminopyrones: Access to benzimidazoles and benzimidazolone derivatives
10. Ansari and Lal, 2009 a K.F. Ansari, C. Lal Synthesis and evaluation of some new benzimidazole derivatives as potential antimicrobial agents .Eur. J. Med . Chem. , 44(2009),p. 2294.
11. Ansari and Lal , 2009 b K. F . Ansari, C. Lal SynthesisPhysicochemical properties and antimicrobial activity of some new benzimidazole derivatives .
12. SharmM, C.* and Kohili D.V, Indian Drug Manufacturers AssocIation Indian Drug Scientific and research publication Journal Vol 54 Oct 2017 issue no 10.
13. Mukherjee pulok k *, Bahadur shiv, Harwansh ranjit k , Chaudhary sushil k from Indian Drug Manufacturers Association Indian Drug Scientific and research publication Journal Vol 54 Nov 2017.