Exploring 2D and 3D QSARs of benzimidazole derivatives as transient receptor potential melastatin 8 (TRPM8) antagonists using MLR and kNN-MFA methodology

7 AUTHORS, INCLUDING:

Mandar Avchar
R. C. Patel Institute of Pharmaceutical Educ…
2 PUBLICATIONS **23** CITATIONS

Dhumal Dinesh
Institute of Chemical Technology, Mumbai
14 PUBLICATIONS **16** CITATIONS

Savita Patil
R. C. Patel Institute of Pharmaceutical Educ…
16 PUBLICATIONS **65** CITATIONS

Shailesh Jain
Nirma University
21 PUBLICATIONS **112** CITATIONS

ORIGINAL ARTICLE

Exploring 2D and 3D QSARs of benzimidazole derivatives as transient receptor potential melastatin 8 (TRPM8) antagonists using MLR and kNN-MFA methodology

Kamlendra Singh Bhadoriya [a,*], **Narender K. Kumawat** [a],
Suvarna V. Bhavthankar [a], **Mandar H. Avchar** [a], **Dinesh M. Dhumal** [a],
Savita D. Patil [b], **Shailesh V. Jain** [c]

[a] *Drug Design and Development Department, R. C. Patel Institute of Pharmaceutical Education and Research, Shirpur, District Dhule, Maharashtra 425405, India*
[b] *H. R. Patel Institute of Pharmaceutical Education and Research, Shirpur, District Dhule, Maharashtra 425405, India*
[c] *Department of Pharmaceutical Chemistry, Institute of Pharmacy, Nirma University, S. G. Highway, Ahmedabad 382481, Gujarat, India*

Abstract TRPM8 is now best known as a cold- and menthol-activated channel implicated in thermosensation. TRPM8 is specifically expressed in a subset of pain- and temperature-sensing neuron. TRPM8 plays a major role in the sensation of cold and cooling substances. TRPM8 is a potential new target for the treatment of painful conditions. Thus, TRPM8 antagonists represent a new, novel and potentially useful treatment strategy to treat various disease states such as urological disorders, asthma, COPD, prostate and colon cancers, and painful conditions related to cold, such as cold allodynia and cold hyperalgesia. Better tools such as potent and specific TRPM8 antagonists are mandatory as high unmet medical need for such progress. To achieve this objective quantitative structure–activity relationship (QSAR) studies were carried out on a series of 25 benzimidazole-containing TRPM8 antagonists to investigate the structural requirements of their inhibitory activity against cTRPM8. The statistically significant best 2D-QSAR model having correlation coefficient $r^2 = 0.88$ and cross-validated squared correlation coefficient $q^2 = 0.64$ with external predictive ability of pred_$r^2 = 0.69$ was developed by SW-MLR. The physico-chemical descriptors such as polarizabilityAHP, kappa2, XcompDipole, +vePotentialSurfaceArea, XKMostHydrophilic were found to show a significant correlation with biological activity in benzimidazole derivatives. Molec-

ular field analysis was used to construct the best 3D-QSAR model using SW-kNN method, showing good correlative and predictive capabilities in terms of $q^2 = 0.81$ and pred_$r^2 = 0.55$. Developed kNN-MFA model highlighted the importance of shape of the molecules, i.e., steric & electrostatic descriptors at the grid points S_774 & E_1024 for TRPM8 receptor binding. These models (2D & 3D) were found to yield reliable clues for further optimization of benzimidazole derivatives in the data set. The information rendered by 2D- and 3D-QSAR models may lead to a better understanding of structural requirements of cTRPM8 antagonists and also can help in the design of novel potent cTRPM8 antagonists.

1. Introduction

Temperature detection is essential for mammalian homeostasis which dictates our responses to the environment. Several members of the transient receptor potential (TRP) superfamily of ion channels function as molecular transducers of thermal stimuli (Sabnis et al., 2008). TRPM8 is a cold & menthol sensing ion channel that acts by converting thermal & chemical stimuli into neuronal signals and sensations of cooling/cold. TRPM8 proteins have amino- and carboxyl-terminal intracellular domains separated by a domain containing six transmembrane helices (S1–S6) in which the last two helices (S5 and S6) flank a loop, called the pore loop, which determines ion selectivity (Tsavaler et al., 2001).

TRPM8 transcripts are expressed in a sub-population of cold responsive primary afferent sensory neurons within the dorsal root and trigeminal ganglia, as well as in normal prostate and testicular tissues, and in taste papillae (Sabnis et al., 2008). Later studies detected TRPM8 mRNA or protein (or both) in a subset of sensory neurons from dorsal root ganglia and trigeminal ganglia, in nodose ganglion cells innervating the upper gut, gastric fundus, vascular smooth muscle, liver and in bladder urothelium and different tissues of the male genital tract (Voets et al., 2007). Three novel natural odorants linalool, geraniol and hydroxycitronellal that activate TRPM8, together with the fact that TRPM8 is expressed in the trigeminus, which belongs to the sensory system of the olfactory epithelium, suggest that TRPM8 could be an important 'Chemosensory Trigeminal Nerve Receptor' (Behrendt et al., 2004).

Recent studies show that TRPM8 is expressed in non-neuronal tissues like visceral organs and the sensory pathways supplying them, particularly the urinary tract (Hayashi et al., 2009; Zhao et al., 2010). Clinical trials using herbal remedies containing peppermint attenuate symptoms of colonic hypersensitivity, which manifests clinically in patients with irritable bowel syndrome (IBS) as abdominal pain (Ford et al., 2008; Merat et al., 2010). Studies in the tongue and skin indicate a peripheral mechanism whereby TRPM8 activation evokes antinociception attributed to the desensitization of TRPV1-mediated responses (Klein et al., 2010). TRPM8 has the following functions: Innocuous cold perception, behavioral thermoregulation, cold-mediated analgesia; intense cold nociception (Flockerzi, 2007).

TRPM8 is a non-selective voltage-gated cation channel permeable to both monovalent and divalent cations (McKemy et al., 2002; Peier et al., 2002) activated upon membrane depolarization, comparable to classical voltage-gated K^+, Na^+ and Ca^{2+}, but with a much weaker voltage sensitivity (Voets et al., 2004). TRPM8 ion channel, upon activation allows entry of Na^+ & Ca^{2+} ions to the cells and leads to depolarization & generation of action potential which eventually leads to sensation of cold (Sherkheli et al., 2007, 2008). Rohacs and colleagues proposed that Ca^{2+} influx through TRPM8 leads to activation of Ca^{2+}-dependent phospholipase C (e.g. PLCδ1), leading to depletion of cellular PIP_2 levels and channel desensitization. They also provided evidence that positively charged residues in the TRP domain, which is located C-terminal of TM6, participate in the interaction of the channel with membrane-bound PIP_2 (Rohacs et al., 2005). Electrophysiological studies have demonstrated that TRPM8 is expressed on the plasma and the endoplasmic reticulum (ER) membranes of cells, where it facilitates the influx or release of Ca^{2+} from the extracellular sources and the ER stores, respectively (Mahieu et al., 2007).

Capsazepine, thio-BCTC [N-(4-tert butyl-phenyl)-4-(3-chloropyridin-2-yl) tetrahydro-pyrazine-1(2H)-(thio)carboxamide] and BCTC [N-(4-tert.butyl-phenyl)-4-(3-chloropyridin-2-yl) tetrahydropyrazine-1(2H)-carboxamide] were found to be antagonists of TRPM8. Studies in cell lines revealed that TRPM8 is required for prostate cancer cell survival (Behrendt et al., 2004). Prolonged activation of TRPM8 variant receptor leads to enhanced production of various proinflammatory cytokines, suggesting a role for this receptor in cellular inflammatory responses mediated by cold air/temperature. TRPM8 activation and its downstream response elements could provide potential new targets of drug development for cold-induced respiratory diseases such as asthma (Sabnis, 2007). A recent report provides intriguing evidence for a role of TRPM8 in bladder function. It was shown that filling of the bladder with cold water induces a detrusor reflex and voiding in patients which states a role for TRPM8 in bladder function (Tsukimi et al., 2005). Thus, novel TRPM 8 variant receptors represent an attractive novel strategy to treat various disorders such as asthma, COPD, cancer and so on.

Molecular modeling study is an approach used to narrow down a library containing an extraordinarily high number of random molecules into a smaller list of the potentially effective inhibitors. The techniques of QSAR are valuable molecular modeling tools for drug design (Jain et al., 2012a,b). The quantitative structure–activity relationship (QSAR) approach became very useful and largely widespread for the prediction of biological activities, particularly in drug design. This approach is based on the assumption that the variations in the properties of the compounds can be correlated with changes in their molecular features (Jain et al., 2012c). The QSAR approach helps to correlate the specific biological activities or physical properties of a series of compounds with the measured or computed molecular properties of the compounds, in terms of

Table 1 Structures of dataset used for MLR and kNN-MFA QSAR analysis with corresponding experimental activities against cTRPM8 of benzimidazole derivatives.
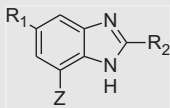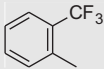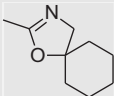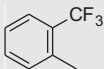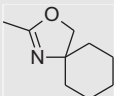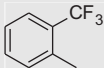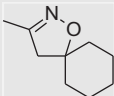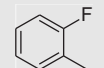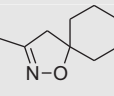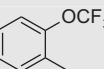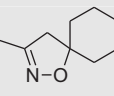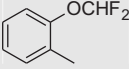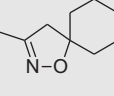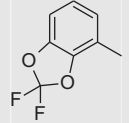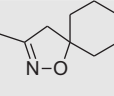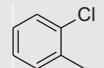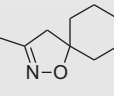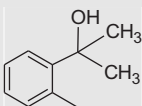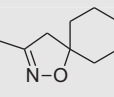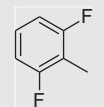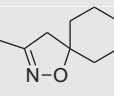

Compound	R$_1$	R$_2$	Z	cTRPM8 IC$_{50}$ (nM)	cTRPM8 pIC$_{50}$
1[b]			–H	61	7.21
2			–H	6.0	8.22
3			–H	3.1	8.51
4			–H	2.9	8.54
5			–H	1.1	8.96
6[b]			–H	2.5	8.6
7			–H	91	7.04
8			–H	2.4	8.62
9			–H	7.9	8.1
10[b]			–H	2.3	8.64

Table 1 *(continued)*
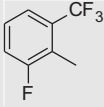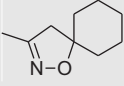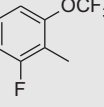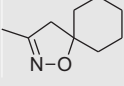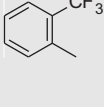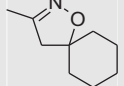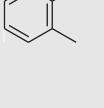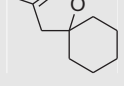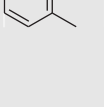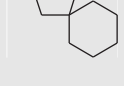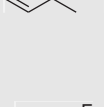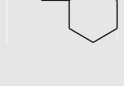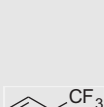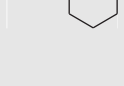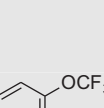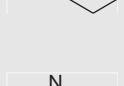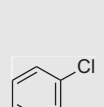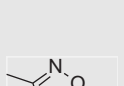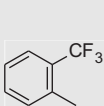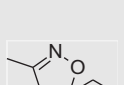
Compound	R_1	R_2	Z	cTRPM8 IC_{50} (nM)	cTRPM8 pIC_{50}
11[a, b]			–H	0.6	9.22
12[b]			–H	0.8	9.1
13[b]			–F	2.2	8.66
14[b]			–Cl	3.3	8.48
15[a]			–CN	0.7	9.15
16			–Br	2.7	8.57
17			–CF_3	1.7	8.77
18			–CF_3	0.8	9.1
19[a]			–CF_3	1.2	8.92
20[a]			–CF_3	1.0	9
21			–(CH_2)_3OH	0.2	9.7

Table 1 *(continued)*

Compound	R_1	R_2	Z	cTRPM8 IC_{50} (nM)	cTRPM8 pIC_{50}
22	(CF₃-substituted benzene)	(spiro isoxazoline, O)	–H	6.7	8.17
23	(CF₃-substituted benzene)	(spiro isoxazoline, CF₂)	–H	0.9	9.05
24	(CF₃-substituted benzene)	(spiro isoxazoline, O)	–CF_3	1.4	8.85
25	(CF₃-substituted benzene)	(spiro isoxazoline, O)	–Cl	4.0	8.4

[a] Indicates the compounds considered in the test set, rest of the compounds considered in the training set for 3D-QSAR study.
[b] Indicates the compounds considered in the test set, rest of the compounds considered in the training set for 2D-QSAR study.

descriptors (Hansch, 1969; Doweyko, 2008; Scior et al., 2009). QSAR methodologies save resources and expedite the process of the development of new molecules and drugs. Three dimensional quantitative structure–activity relationship (3D-QSAR) is widely used tool to identify the steric, electrostatic, and hydrophobic structural requirements of various drugs acting via receptor modulation for exerting biological activity. By the application of 3D-QSAR models, the number of compounds that need to be synthesized by a medicinal chemist can be reduced greatly. Thus, the time and cost of drug discovery and development can also be reduced (Bhadoriya et al., 2012a,b; Gaurav et al., 2010).

To the best of our knowledge, till date such QSAR study has not been reported on benzimidazole as cTRPM8 antagonists as discussed in this article. This inspired us to undertake this work. This study is aimed to elucidate the structural features of benzimidazole derivatives required for antagonism of cTRPM8 and to obtain predictive 2D- and 3D-QSAR models to guide the rational synthesis of novel antagonists. In this investigation, widely used technique, viz. stepwise (SW) has been applied for descriptor optimization, and MLR & kNN-MFA analyses have been applied for 2D- and 3D-QSAR models development. The generated models provide insight into the influence of various interactive fields on the activity and, thus, can help in designing and forecasting the inhibition activity of novel cTRPM8 antagonists.

2. Materials and methods

All molecular modeling studies (2D and 3D-QSAR) were carried out on a Windows XP workstation using the molecular modeling software package VLife Molecular Design Suite (VLife MDS) version 3.5 (VLife MDS 3.5, 2008).

2.1. Biological activity dataset for analysis

The 25 benzimidazole derivatives were selected for the dual 2D and 3D-QSAR model development. The inhibitory activities [IC_{50} (nM)] of benzimidazole derivatives against cTRPM8 have been studied by Parks et al. The in vitro inhibitory activity data against cTRPM8 of all 25 compounds were collected from the literature (Parks et al., 2011). For QSAR analysis, these biological activities [IC_{50} (nM)] were converted into pIC_{50} values which were used as the dependent variable in the model development. Table 1 shows the structure of 25 such compounds along with their biological activity values.

2.2. Computational details

The structures of all 25 compounds were drawn in Chem sketch version 12.0 (ACD/Chemsketch 12.0, 2009). All structures were cleaned and 3D optimized. Geometry optimization was carried out using Merck Molecular Force Field (MMFF) (Halgren, 1996a–d, 1999a,b) with distance dependent dielectric function and energy gradient of 0.001 kcal/mol Å.

2.3. Conformational analysis

Conformational analysis involves moving the atoms of a molecule in such a way that the total energy of the system is reduced based on an empirical representation of the interaction energy of the atoms of a molecule. The optimized compounds were subjected to conformational analysis and energy minimization using Monte Carlo conformational search with RMS gradient of 0.001 kcal/mol using a MMFF. Monte Carlo search method is a random search method for finding conformations of molecules (Metropolis et al., 1953).

It uses the Metropolis condition to accept or discard generated conformers. The conformers for all 25 optimized compounds were generated and selected the low energy conformer for each compound and used for further study. The basis of energy minimization is that the drug binds to effectors/receptors in the most stable form, i.e., the minimum energy form.

2.4. Molecular modeling for 2D-QSAR

2.4.1. Calculation of 2D Molecular descriptors

The purpose of molecular descriptor is to calculate the properties of molecules that serve as numerical characterizations of molecules in other calculations, such as QSAR, diversity analysis, and combinatorial library design. The energy-minimized geometry was used for the calculation of the various 2D molecular descriptors. A large number of theoretical 2D individual descriptors such as Mol. Wt., Volume, XlogP, smr, polarizabilityAHP; physicochemical such as kappa2, XcompDipole, +vePotentialSurfaceArea, XKMostHydrophilic, Polar Surface Area; alignment independent topological descriptors such as T_T_O_2, T_T_T_O, T_T_N_6, T_T_T_1 type have been computed for these geometrically optimized structures from the chemical structures of the compounds referred to above with a view to develop structure–activity relationship of benzimidazole derivatives against cTRPM8. A total of 938 molecular descriptors were calculated for our present investigation using VLife Sciences Molecular Design Suite and prior to the 2D-QSAR model development, the set of calculated molecular descriptors was reduced for further analysis. The preprocessing of the independent variables (i.e., 2D descriptors) was done by removing the invariable (constant column) which resulted in a total of 235 molecular descriptors to be used for QSAR analysis.

2.4.2. Training and test set selection

In order to obtain a validated 2D-QSAR model for the purpose of meaningful prediction, biological activity dataset should be divided into the training and test sets. The sphere exclusion method was adopted for division of training and test sets comprising 18 and 7 compounds, respectively. This approach resulted in selection of compounds. 1, 6, 10, 11, 12, 13 and 14 as the test set for validating the quality of the models and the remaining 18 compounds as the training set for generating 2D-QSAR models.

2.4.3. Variable selection and model development

Variable selection is a key step in QSAR analysis. The reduced set of descriptors was then treated by forward stepwise variable selection for further reduction of non-significant descriptors and finally the optimum models with five significant descriptors were considered in our 2D-QSAR analysis. The variable/feature selection method can be used together with MLR regression analysis for constructing a 2D-QSAR model. The 2D-QSAR model was generated by using multiple linear regression (MLR) method by using VLife Molecular Design Suite (VLife MDS). MLR is the traditional and standard approach for multivariate data analysis. Multivariate analysis is the analysis of multidimensional data metrics by using statistical methods. It relates the dependent variable (biological activity) to a number of independent (predictor) variables (molecular descriptors) by using linear equations. This method

of regression estimates the values of the regression coefficients by applying least square curve fitting method.

2.4.4. Statistical parameters for 2D-QSAR models

The program computes the best model on the basis of coefficient of determination/squared correlation coefficient r^2; cross-validated correlation coefficient/cross-validated explained variance q^2, which is a relative measure of quality of fit; F-test (Fischer's value) which represents F-ratio between the variances of calculated and observed activity; pred_r^2, r^2 for external test set; DF, degree of freedom; Zscore, calculated by q^2 in the randomization test; best_rand_q^2, the highest q^2 value in the randomization test and alpha_rand_q^2, the statistical significance parameter obtained by randomization test. The coefficient of determination/squared correlation coefficient r^2 is a relative measure of fit by the regression equation. The coefficient of determination is the percent of the variation that can be explained by the regression equation. It represents the explained variance of the model, and is used as a measure of the goodness-of-fit of the model. The correlation coefficient values closer to 1.0 represent the better fit of the regression. The F-test reflects the ratio of the variance explained by the model and the variance due to the error in the regression. High values of the F-test indicate that the model is statistically significant. Cross-validated correlation coefficient/cross-validated explained variance q^2 is used as a measure of the internal performance, and sometime used to estimate predictivity. Validation parameter, pred_r^2 was calculated for evaluating the predictive capacity of the model. A value of pred_r^2 greater than 0.5 indicates the good predictive capacity of the QSAR model. However, a QSAR model is considered to be predictive, if the following conditions are satisfied: $r^2 > 0.6$, $q^2 > 0.6$ and pred_$r^2 > 0.5$ (Golbraikh and Tropsha, 2002). The low standard error of pred_r^2se, q^2se and r^2se shows absolute quality of fitness of the model. The generated QSAR model was validated for predictive ability inside the model by using cross validation (LOO) for q^2 and external validation, which is a more robust alternative method by dividing the data into training set & test set and calculating pred_r^2. The high pred_r^2 and low pred_r^2se show high predictive ability of the model.

2.5. Molecular modeling for 3D-QSAR

2.5.1. Molecular alignment

Proper alignment of molecules is the most critical input and a crucial step in the ligand based 3D-QSAR modeling method to obtain meaningful results (Tanneeru et al., 2011; Sharma and Kohli, 2012). Molecular alignment is useful for studying shape variation with respect to the base structure selected for alignment (VLife MDS 3.5, 2008). Energy-minimized and geometry-optimized structures of molecules were aligned by the template-based method (Ajmani et al., 2006), where a template structure is defined and used as a basis for alignment of a set of molecules. The template structure, i.e. 5-phenyl-1H-benzimidazole ring, was used for the alignment by considering the common elements of the series as shown in Fig. 1. The reference molecule is chosen in such a way that it is the most active among the series of molecules considered. The reference molecule is the molecule on which the other molecules of the align dataset get aligned based on the chosen template (VLife MDS
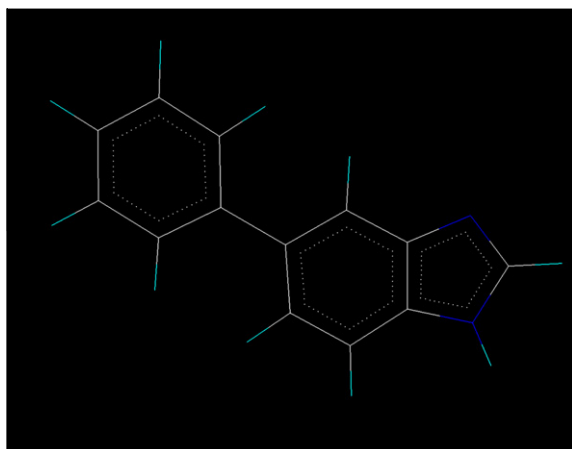
Figure 1 5-Phenyl-1*H*-benzimidazole ring as common template used for alignment of benzimidazole derivatives.
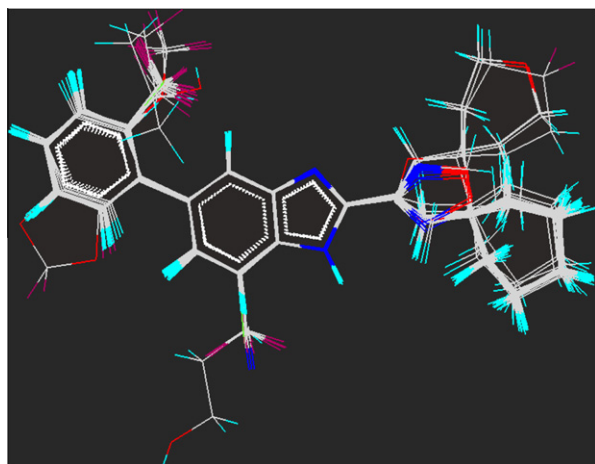

Figure 2 Stereo view of template based alignment of benzimidazole derivatives on the base template.

3.5, 2008). Compound 21 possessed very high inhibitory activity against cTRPM8 which made it is a valid lead molecule and, therefore, was chosen as a reference molecule. After optimizing, the template structure and the reference molecule were used to superimpose all molecules from the series using the template alignment method in VLife MDS 3.5 software (VLife MDS 3.5, 2008) to obtain optimal alignment between the molecular structures necessary for ligand–receptor interactions. This adjusts the geometry of the molecules such that their steric and electrostatic fields match the fields of the template molecule (Radhika et al., 2010). The superimposition of all molecules based on minimizing RMS deviation is shown in Fig. 2.

2.5.2. Molecular descriptors

MFA is a method for quantifying the interaction energy between a probe molecule and a set of aligned target molecules in a rectangular grid box and can be useful in establishing QSAR (Nakka and Guruprasad, 2012). This approach is effective for the analysis of data sets where activity information is available but the structure of the receptor site is unknown. It

attempts to postulate and represent the essential features of a receptor site from the aligned common features of the molecules that bind to it (Sahu et al., 2011). The aligned biologically active conformations of benzimidazole were used for the calculation of molecular fields. Molecular fields are the electrostatic, steric and hydrophobic interaction energies which are used to formulate a relationship among electrostatic, steric and hydrophobic properties together with the biological activities of compounds. Each conformation is taken in turn and the molecular fields around it are calculated. Molecular descriptors such as steric, electrostatic and hydrophobic fields have been calculated utilizing VLife MDS 3.5 software which allows the user to choose probe, grid size, and grid interval for the generation of descriptors (Ghosh and Bagchi, 2009). This is done by generating 3D rectangular grids around the molecule and calculating the interaction energy between the molecule and probe group placed at each grid point. Using Tripos force field (Clark et al., 1989), steric, electrostatic and hydrophobic fields were computed at each grid point considering Gasteiger-Marsili charges (Gasteiger and Marsili, 1980). Methyl probe of charge + 1 with 10.0 kcal/mol electrostatic and 30.0 kcal/mol steric cut-off was used for field generation. A value of 1.0 was assigned to the distance-dependent dielectric constant. A total of 3432 three dimensional descriptors were calculated using VLife MDS software. These included electrostatic, steric and hydrophobic field descriptors (1144 for each electrostatic, steric & hydrophobic) for all the compounds in separate columns. These interaction energy values at the grid point are considered for relationship generation using kNN method and utilized as descriptors to decide nearness between molecules (Ajmani et al., 2006).

2.5.3. Division of a dataset into training and test sets

The sphere exclusion algorithm (Golbraikh and Tropsha, 2003) was adopted for division of training and test data sets. This algorithm allows construction of the data sets by using descriptor space occupied by the representative points, such that the test set compounds represent a range of biological activities similar to the training set; thus, the test set is truly a representative of the training set. In order to assess the similarity of the distribution pattern of the compounds in the generated sets, statistical parameters (with respect to the biological activity) i.e. mean, maximum, minimum and standard deviation were calculated for the training and test sets. For selection of training and test sets, we were ensured that the compounds have uniform spread (training and test) in terms of both activity and chemical space. This was achieved by setting aside four compounds as test set. Sphere exclusion approach resulted in the selection of compounds 11, 15, 19 and 20 as the test set for validating the quality of the models and the remaining 21 compounds as the training set for generating 3D-QSAR models. The test was used to ascertain the predictive power of the model.

2.5.4. Forward stepwise as feature (variable) selection method

Feature selection is a key step in QSAR analysis. Chance correlations and multi-collinearity are two major problems often encountered when attempting to find generalized QSAR models for use in drug design. An integral aspect of any model-building exercise is the selection of an appropriate set of features with low complexity and good predictive accuracy.

This process forms the basis of a technique known as feature selection or variable selection. Among several search algorithms, stepwise (SW), genetic algorithm (GA) and simulated annealing (SA) based feature selection procedures are most popular for building QSAR models and can explain the situation more effectively (Guyon and Elisseeff, 2003; Darlington, 1990; Zheng and Tropsha, 2000). In SW forward variable selection algorithm, the search procedure begins with developing a trial model step by step with a single independent variable and to each step, independent variables are added one at a time, examining the fit of the model by using the kNN-MFA procedure. Thus, the model is repeatedly altered from the previous one by adding or removing a predictor variable in accordance with the 'stepping criteria' (in this case, $F = 4$ for inclusion for the forward selection method). The method continues until there is no more significant variable remaining outside the model.

In the selected equations, the cross-correlation limit was set at 0.5, the number of variables at 10 and the term selection criteria at r^2. An F value was specified to evaluate the significance of a variable. The variance cut-off was set at 0.0, and scaling as none. Additionally kNN parameter setting was done in which the number of maximum neighbors and the number of minimum neighbors were set at 5 & 2 respectively and the prediction method was selected as the distance-based weighted average.

2.5.5. kNN-MFA methodology for building QSAR models

The kNN technique is a conceptually simple approach to pattern recognition problems (VLife MDS 3.5, 2008). The k-nearest neighbor (kNN) method is one of the simplest machine learning algorithms, most commonly used for classifying a new pattern (e.g. a molecule) (Verma et al., 2010). The kNN methodology relies on a simple distance learning approach whereby an unknown member is classified according to the majority of its k-nearest neighbors in the training set. The nearness is measured by an appropriate distance metric (e.g., a molecular similarity measure calculated using field interactions of molecular structures). The standard kNN method is implemented simply as follows: (1) Calculate the distances between an unknown object (u) & all the objects in the training set; (2) Select k objects from the training set most similar to object u, according to the calculated distances; and (3) Classify object u with the group to which the majority of the k objects belongs (Sharaf et al., 1986). An optimal k value is selected by optimization through the classification of a test set of samples or by leave-one-out cross-validation. The variables & optimal k values were chosen using stepwise variable selection method. kNN-MFA with stepwise (SW) variable selection method employs a stepwise variable selection procedure combined with kNN to optimize (i) The number of nearest neighbors (k) and (ii) The selection of variables from the original pool.

2.5.6. k-Nearest neighbor QSAR (kNN weighted average method)

The kNN method was also used to develop a QSAR model using continuous variable i.e. using activity as pIC_{50} values. In this case, by using a developed kNN QSAR model the activity of a molecule can be predicted using weighted average activity (Eq. (1)) of the k most similar molecules in the training set.

$$\hat{y}_i = \sum w_i y_i \tag{1}$$

where y_i and $\hat{y}_i$ are the actual and predicted activity of the ith molecule respectively, and w_i are weights calculated using (Eq. (2)).

$$w_i = \frac{\exp(-dj)}{\sum_{j=1}^{k} \exp(-dj)} \tag{2}$$

The similarities were evaluated as the inverse of Euclidean distances (d_j) between molecules (Eq. (3)) using only the subset of descriptors corresponding to the model. Where, k is number of nearest neighbors in the model.

$$d_{ij} = \left[\sum_{m=1}^{Vn} (X_{im} - X_{jm}) \right]^{1/2} \tag{3}$$

where X is the matrix of selected descriptors (Vn) for the kNN QSAR model.

2.5.7. Model validation and evaluation

This is done to test the internal stability and predictive ability of the QSAR models.

2.5.7.1. Internal and external validations. Internal validation was carried out using leave-one-out (q^2, LOO) method (Cramer et al., 1988). For calculating q^2, each molecule in the training set was eliminated once and the activity of the eliminated molecule was predicted by using the model developed by the remaining molecules. The cross-validated coefficient, q^2, was calculated using Eq. (4).

$$q^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{\text{mean}})^2} \tag{4}$$

where y_i and $\hat{y}_i$ are the actual and the predicted activity of the ith molecule in the training set, respectively, and y_{mean} is the average activity of all molecules in the training set.

However, a high q^2 value does not necessarily give a suitable representation of the real predictive power of the model for cTRPM8 antagonists. So, an external validation is also carried out in this study. The external predictive power of the model is assessed by predicting pIC_{50} value of four test set molecules, which are not included in the QSAR model development. The predictive ability of the selected model is also confirmed by pred_r^2.

For external validation, the activity of each molecule in the test set was predicted using the model developed by the training set. The pred_r^2 value is calculated as follows (Eq. (5))

$$\text{pred\_}r^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{\text{mean}})^2} \tag{5}$$

where y_i and $\hat{y}_i$ are the actual and the predicted activity of the ith molecule in the test set, respectively, and y_{mean} is the average activity of all molecules in the training set.

Both summations are over all molecules in the test set. Thus the pred_r^2 value is indicative of the predictive power of the current kNN-MFA model based on the external test set.

2.5.7.2. Randomization test. To evaluate the statistical significance of a QSAR model for an actual data set, one tail hypothesis testing was used (Zheng and Tropsha, 2000; Gilbert, 1976).

The robustness of the models for training sets was examined by comparing these models to those derived for random data sets. Random sets were generated by rearranging the activities of the molecules in the training set. The statistical model was derived using various randomly rearranged activities (random sets) with the selected descriptors and the corresponding q^2 were calculated. The significance of the models hence obtained was derived based on the calculated $Zscore$; Eq. (6).

A $Zscore$ value is calculated by the following formula:

$$Zscore = \frac{(h - \mu)}{\sigma} \tag{6}$$

where h is the q^2 value calculated for the actual data set, μ is the average q^2, and σ is its standard deviation calculated for various iterations using models build by different random data sets.

The probability (α) of significance of the randomization test is derived by using calculated $Zscore$ value as given in the literature (Shen et al., 2003).

2.5.7.3. Evaluation of the quantitative model. The developed 3D-QSAR model was evaluated using the following statistical measures: N, number of observations (molecules) in the training set; Number of nearest neighbors, number of k-nearest neighbor in the model; q^2, cross-validated r^2 (by leave one out) which is a relative measure of quality of fit; pred_r^2, r^2 for external test set; q^2se, standard error of cross-validation and pred_r^2se, standard error of external test set prediction. However, a QSAR model is considered to be predictive, if the following conditions are satisfied: $q^2 > 0.6$ and pred_$r^2 > 0.5$ (Golbraikh and Tropsha, 2002). The low standard error of pred_r^2se and q^2se shows absolute quality of fitness of the model. The high pred_r^2 and low pred_r^2se show high predictive ability of the model.

The q^2 and pred_r^2 values were used as deciding factors in selecting the optimal models.

3. Result and discussion

The 2D- and 3D-QSAR studies of 25 benzimidazole derivatives for inhibitory activity against cTRPM8 (Table 1) through MLR and kNN methodology, respectively, based on SW feature selection method were performed using VLife MDS 3.5 software (VLife MDS 3.5, 2008). Sphere exclusion method was selected for the division of whole data set into training and test sets (Table 1) for choosing uniformly distributed molecules in both sets. For selection of training and test sets, we were ensured that the molecules have uniform spread (training and test) in terms of both activity and chemical space. pIC$_{50}$ was selected as dependent variable and remaining all the variables were selected as independent variables. Selection of com-

pounds in the training set and test set is a key and important feature of any QSAR model. Therefore care was taken in such a way that biological activities of all compounds in test lie within the maximum and minimum value range of biological activities of the training set of compounds. A unicolumn statistics for training set and test set were generated to check correctness of selection criteria for training and test set compounds (Table 2).

The maximum and minimum values in training and test sets were compared in a way that:

1. The maximum value of pIC$_{50}$ of the test set should be less than or equal to the maximum value of pIC$_{50}$ of the training set.
2. The minimum value of pIC$_{50}$ of the test set should be higher than or equal to the minimum value of pIC$_{50}$ of the training set.

This observation showed that the test set was interpolative and derived within the minimum–maximum range of the training set.

Statistically significant models were selected, considering the term selection criterion as r^2, q^2 and pred_r^2. The models were validated by both internal and external validation procedures. Best 2D- and 3D-QSAR models are chosen for discussion.

3.1. 2D-QSAR

The various 2D-QSAR models were developed using multiple linear regression (MLR) method. 2D-QSAR equations were selected by optimizing the statistical results generated along with variation of the descriptors in these models. The fitness/pattern plots were also generated for evaluating the dependence of the biological activity on various different types of the descriptors. The frequency of use of a particular descriptor in the population of equations indicated the relevant contributions of the descriptors.

The best regression equation obtained is represented in Eq. (7):

$$\begin{aligned}
pIC_{50} = {} & 5.40081 - 0.0509337 \, polarizabilityAHP \\
& + 0.757011 kappa2 - 0.200476 XcompDipole \\
& - 0.0100592 + vePotentialSurfaceArea \\
& + 2.11017 XKMostHydrophilic \tag{7}
\end{aligned}$$

The equation explains 88% ($r^2 = 0.88$) of the total variance in the training set. It also has an internal (q^2) and external (pred_r^2) predictive ability of ~64% and ~69% respectively. The F-test = 17.57 shows the statistical significance of

Table 2 Unicolumn statistics of the training and test sets for inhibitory activity against cTRPM8.

Data set	Column name	Average	Max.	Min.	SD	Sum
2D-QSAR						
Training	pIC$_{50}$	8.6483	9.7000	7.0400	0.5674	155.6700
Test	pIC$_{50}$	8.5586	9.2200	7.2100	0.6550	59.9100
3D-QSAR						
Training	pIC$_{50}$	8.5374	9.6990	7.0410	0.5931	179.2864
Test	pIC$_{50}$	9.0744	9.2218	8.9208	0.1383	36.2976

Max., maximum; Min., minimum; SD, standard deviation.

99.99% of the model which means that probability of failure of the model is 1 in 10,000. In addition, the randomization test shows confidence of ~99.9% that the generated model is not random and hence it is chosen as the QSAR model (Table 3).

The descriptors selected for 2D-QSAR modeling inhibitory activity against cTRPM8 of benzimidazole derivatives are summarized in Table 4 and the correlation matrix between the physico-chemical descriptors influencing the inhibitory activity against cTRPM8 is presented in Table 5.

The fitness plot of observed versus predicted activity provides an idea about how well the model was trained and how well it predicts the activity of the external test set. From the fitness plot it can be seen that the model is able to predict the activity of training set quite well (all points are close to regression line) as well as external test set up to 69% (all points are close to regression line) providing confidence in predictive ability of the model (Fig. 3). The predicted activities & residuals of both train-

Table 3 Statistical results of 2D-QSAR equation generated by MLR method and 3D-QSAR model generated by forward stepwise variable selection kNN-MFA method for benzimidazole derivatives.

S. No.	Statistical parameter	Results	
		2D-QSAR	3D-QSAR
1	r^2	0.88	–
2	r^2se	0.23	–
3	q^2	0.64	0.81
4	q^2se	0.41	0.26
5	pred_r^2	0.69	0.55
6	pred_r^2se	0.37	0.43
7	F-test	17.57	–
8	N	18	21
9	Nearest neighbor	–	5
10	Degree of freedom	12	18
11	Zscore_q^2	3.00144	–
12	best_rand_q^2	0.47898	–
13	alpha_rand_q^2	0.01000	–
14	Contributing descriptors	1. Polarizability AHP	1. E_1024 (-0.3652, -0.3433)
		2. kappa2	2. S_774 (-0.1842, -0.1634)
		3. XcompDipole	
		4. +vePotential Surface Area	
		5. XKMost Hydrophilic	

Table 4 Selected physico-chemical descriptors used in 2D-QSAR model with values.

Compound	pIC$_{50}$	Polarizability AHP	kappa2	XcompDipole	+vePotential Surface Area	XKMost Hydrophilic
1	7.21	41.039	8.1648	−0.569695	247.668274	0.313501
2	8.22	41.039	8.1648	−0.569456	268.623474	0.518432
3	8.51	41.039	8.1648	−3.406923	290.065186	0.510254
4	8.54	39.386	7.438017	−4.336603	265.042358	0.45733
5	8.96	41.676	8.741253	−3.268443	254.273575	0.539697
6	8.6	42.965	8.859375	−2.379686	275.87738	0.507704
7	7.04	41.63	7.796982	−0.4448	300.308105	0.491227
8	8.62	41.405	7.438017	−3.266764	215.876434	0.501297
9	8.1	45.619	8.1648	−2.982738	250.004211	0.468389
10	8.64	39.295	7.679584	−3.186778	254.287094	0.466665
11	9.22	40.948	8.408284	−3.277809	262.351349	0.524292
12	9.1	41.585	8.981844	−0.533051	222.778351	0.555079
13	8.66	40.948	8.408284	−3.464247	290.227386	0.530688
14	8.48	42.967	8.408284	−4.729369	287.389801	0.594294
15	9.15	43.025	8.981844	−5.593012	314.157806	0.467935
16	8.57	43.665	8.408284	−3.465478	283.43042	0.617796
17	8.77	40.948	8.408284	−5.661723	307.554199	0.556983
18	9.1	42.601	9.141498	−5.951039	321.074005	0.607054
19	8.92	43.238	9.707555	1.997233	262.491089	0.641335
20	9	42.967	8.408284	−4.956248	290.760132	0.598902
21	9.7	47.181	10.16595	−3.760355	300.417053	0.469463
22	8.17	39.841	8.1648	−1.984283	215.499512	0.262306
23	9.05	40.857	8.340496	−3.813835	241.814011	0.53442
24	8.85	41.403	9.141498	−4.099267	281.497498	0.35867
25	8.4	41.769	8.408284	−1.156562	216.434906	0.344089

Table 5 Correlation matrix for physico-chemical descriptors influencing the inhibitory activity against cTRPM8 (2D-QSAR model).

	Polarizability AHP	kappa2	XcompDipole	+vePotential surface area	XKMost Hydrophilic	Score
Polarizability AHP	1	0.61941	−0.024815	0.298363	0.260522	5
kappa2	0.61941	1	0.007689	0.398268	0.172653	5
XcompDipole	−0.024815	0.007689	1	−0.442572	−0.071619	5
+vePotential Surface Area	0.298363	0.398268	−0.442572	1	0.470572	5
XKMost Hydrophilic	0.260522	0.172653	−0.071619	0.470572	1	5
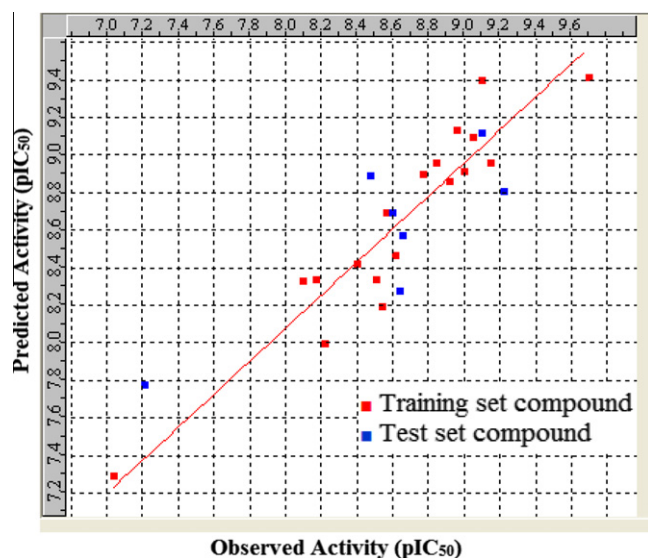
Figure 3 Fitness plot of observed activity (pIC$_{50}$) versus predicted activity (pIC$_{50}$) for training set & test set compounds according to 2D-QSAR MLR model.

ing & test set compounds by 2D-QSAR model are shown in Table 6.

3.1.1. Interpretation of 2D-QSAR MLR model

It is apparent from Equation (7) and Fig. 4 that the descriptor kappa2 plays a pivotal role in determining activity. This descriptor (kappa2) signifies second kappa shape index. This is a positively contributing descriptor toward inhibitory activity against cTRPM8 and its contribution is approx 34% (Fig. 4). The positive coefficient of kappa2 (33.70%) showed that increase in the values of this descriptor is beneficial for the inhibitory activity against cTRPM8 (like in compounds 5, 12, 15, 18, 19, 21 and 24). The next important descriptor which influences the activity is XcompDipole that signifies the x component of the dipole moment (external coordinates). This descriptor is inversely proportional to the activity (−25.91%). The negative coefficient of XcompDipole (−25.91%) showed that increase in the values of this descriptor is detrimental for the inhibitory activity against cTRPM8 (like in compounds 1, 2, 7, 15, 17, 18, 20, 21 and 24). The negative coefficient (−21.17%) of +vePotentialSurfaceArea (signifies total van der Waals surface area with positive electrostatic potential of the molecule) showed that increase in the values of this descriptor is detrimental for the inhibitory activity against cTRPM8 (like in compounds 5, 7, 8, 11, 12 and 23). As a positive contributing descriptor, XKMostHydrophilic (most

hydrophilic value on the vdW surface) is also a physico-chemical descriptor influencing activity variation and is directly proportional to activity. The positive coefficient of XKMostHydrophilic (13.09%) showed that increase in the values of this descriptor is beneficial for the inhibitory activity against cTRPM8 (like in compounds 1, 5, 12, 13, 16, 17, 18, 19, 20, 22 and 23). The last descriptor is polarizabilityAHP which is an individual descriptor that evaluates molecular polarizability using atomic hybrid polarizability (AHP) and negatively contributes to the biological activity (−6.12%). The negative coefficient of polarizabilityAHP showed that increase in the values of this descriptor is detrimental for the inhibitory activity against cTRPM8 (like in compounds 5, 8, 9, 10, 11, 12, 13, 14, 17, 23 and 24).

3.2. 3D-QSAR

Three dimensional quantitative structure–activity relationship studies of benzimidazole derivatives with reported activities against cTRPM8 were performed using stepwise variable selection method for developing kNN-MFA models based on steric, electrostatic & hydrophobic fields. Several 3D-QSAR models were generated using stepwise variable selection method resulted several statistically significant models, of which the corresponding best model is reported herein.

The model selection criterion is the value of q^2, the internal predictive ability of the model and that of pred_r^2, the ability of the model to predict the activity of external test set. For inhibitory activity against cTRPM8, selected the template-based 3D-QSAR model with 21 training set compounds was found to be statistically most significant, especially with respect to the internal predictive ability ($q^2 = 0.81$) of the model. As the cross-validated correlation coefficient (q^2) is used as a measure of reliability of prediction, the correlation coefficient suggests that our model is reliable and accurate. The LOO cross-validated analysis of the best model gave rise to a q^2 value of 0.81 suggesting that the model is a useful tool for predicting inhibitory activity against cTRPM8. A data set of four compounds was selected as the test set from the original data of 25 compounds for the validation experiments. The predictive ability of this forward stepwise variable selection kNN-MFA model was evaluated by predicting the biological activities of the test set compounds. Residual values obtained by subtraction of predicted activities from biological activities were found near to zero. Therefore it was concluded that the resultant QSAR model has good predictive ability. The value of pred_r^2 was obtained for the test set and gave better results, with a value of 0.55, which means 55% predictive power for the external test set. Thus, our model displays good predictivity in regular cross-validation (Table 3). The actual, predicted activities by 3D-QSAR model & residuals of both training &

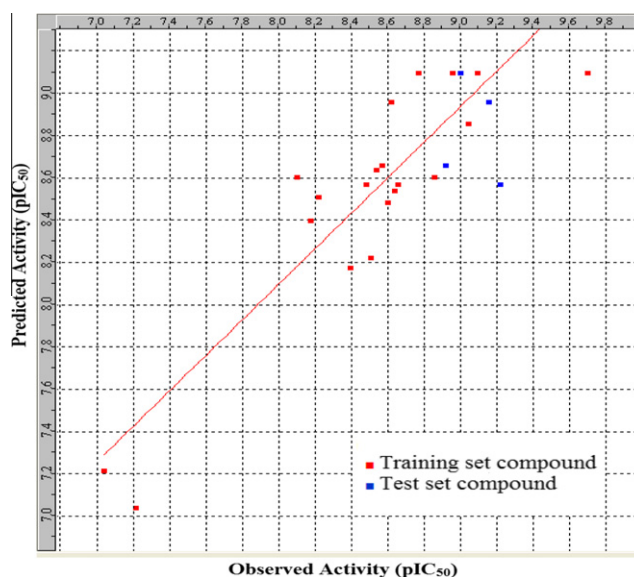**Table 6** Comparative observed and predicted activities of benzimidazole derivatives by 2D and 3D-QSAR models.

Compound	Observed activity (pIC_{50})	2D-QSAR model		3D-QSAR model	
		Predicted activity	Residual	Predicted activity	Residual
1	7.21	7.7758	−0.5658	7.04096	0.16904
2	8.22	7.99739	0.22261	8.50864	−0.28864
3	8.51	8.3333	0.1767	8.22185	0.28815
4	8.54	8.19372	0.34628	8.63827	−0.09827
5	8.96	9.13163	−0.17163	9.09691	−0.13691
6	8.6	8.6924	−0.0924	8.48149	0.11851
7	7.04	7.28773	−0.24773	7.21467	−0.17467
8	8.62	8.46375	0.15625	8.95861	−0.33861
9	8.1	8.32962	−0.22962	8.60206	−0.50206
10	8.64	8.2786	0.3614	8.5376	0.1024
11	9.22	8.80477	0.41523	8.56864	0.65136
12	9.1	9.11929	−0.01929	9.09691	0.00309
13	8.66	8.57523	0.08477	8.56864	0.09136
14	8.48	8.88879	−0.40879	8.56864	−0.08864
15	9.15	8.95726	0.19274	8.95861	0.19139
16	8.57	8.68928	−0.11928	8.65758	−0.08758
17	8.77	8.89697	−0.12697	9.09691	−0.32691
18	9.1	9.39548	−0.29548	9.09691	0.00309
19	8.92	8.85975	0.06025	8.65758	0.26242
20	9	8.91009	0.08991	9.09691	−0.09691
21	9.7	9.416	0.284	9.09691	0.60309
22	8.17	8.33597	−0.16597	8.39794	−0.22794
23	9.05	9.09351	−0.04351	8.85387	0.19613
24	8.85	8.95924	−0.10924	8.60206	0.24794
25	8.4	8.41931	−0.01931	8.17393	0.22607
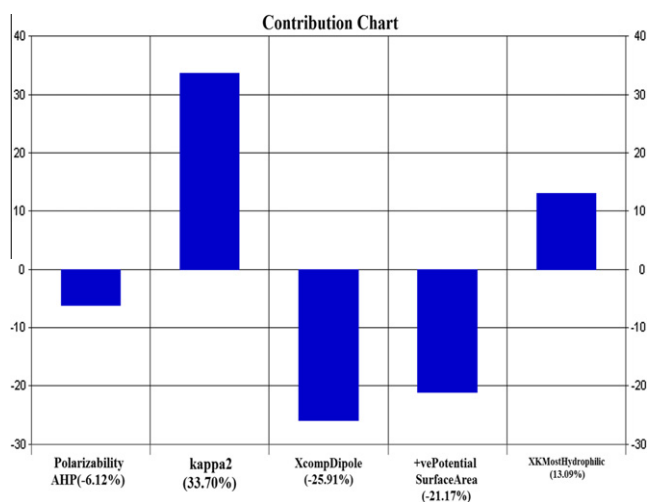
Figure 4 Contribution of descriptors for biological activity developed using MLR equation.

Figure 5 Fitness plot of observed activity (pIC_{50}) versus predicted activity (pIC_{50}) for training set & test set compounds according to 3D-QSAR SW-kNN-MFA model.

test set compounds are given in Table 6. The fitness plot of observed versus predicted activity of both training & test set compounds helped in cross-validation of kNN-MFA 3D-QSAR model and are depicted in Fig. 5.

The descriptors selected for 3D-QSAR modeling inhibitory activity against cTRPM8 of benzimidazole derivatives are summarized in Table 7 and the correlation matrix between the electrostatic and steric descriptors influencing the inhibitory activity against cTRPM8 is presented in Table 8.

The plot of contributions of steric and electrostatic field interactions (Fig. 6) indicates relative regions of the local fields

(steric and electrostatic fields) around the aligned molecules. Green and blue balls represent steric and electrostatic field effects, respectively.

In 3D-QSAR studies, 3D data points generated around benzimidazole pharmacophore were used to optimize the electrostatic & steric requirements of the benzimidazole nucleus

Table 7 Descriptors used in 3D-QSAR model with values.

Compound	pIC$_{50}$	E_1024	S_774
1	7.21	0.2339	−0.208062
2	8.22	0.16111	−0.171235
3	8.51	0.191864	−0.142129
4	8.54	−0.377701	−0.173954
5	8.96	−0.364871	−0.174306
6	8.6	−0.413408	−0.154159
7	7.04	0.26824	−0.141407
8	8.62	−0.365198	−0.163383
9	8.1	−0.446816	−0.17988
10	8.64	−0.377819	−0.173457
11	9.22	−0.414777	−0.178667
12	9.1	−0.348411	−0.18209
13	8.66	−0.391136	−0.179788
14	8.48	−0.392356	−0.175788
15	9.15	−0.366994	−0.180889
16	8.57	−0.39333	−0.17858
17	8.77	−0.343279	−0.166872
18	9.1	−0.357838	−0.184159
19	8.92	−0.386703	−0.190586
20	9	−0.320989	−0.175561
21	9.7	−0.327277	−0.167751
22	8.17	−0.42539	−0.239621
23	9.05	−0.660477	0.664457
24	8.85	−0.386963	−0.032763
25	8.4	−0.412358	−0.241024

Table 8 Correlation matrix for electrostatic and steric descriptors influencing the inhibitory activity against cTRPM8 (3D-QSAR model).

	E_1024	S_774	Score
E_1024	1	−0.311113	2
S_774	−0.311113	1	2
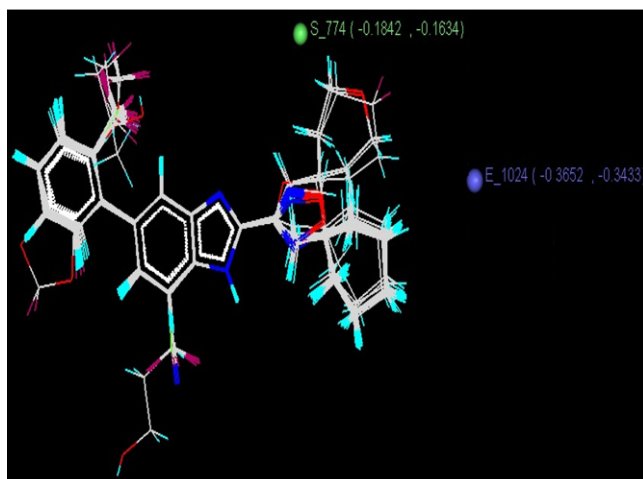
Figure 6 Developed 3D-QSAR SW-kNN-MFA model field plot for steric and electrostatic interactions of all 25 compounds.

for inhibitory activity against cTRPM8. The range of property values in the generated data points helped for the design of new chemical entities. These ranges were based on the varia-

tion of the field values at the chosen points using the most active molecule & its nearest neighbor set. The points generated in SW-kNN-MFA 3D-QSAR model are S_774 (−0.1842, −0.1634) and E_1024 (−0.3652, −0.3433) that is steric & electrostatic interaction fields at lattice points 774 & 1024 respectively (Fig. 6). These points suggested the significance and requirement of steric & electrostatic properties as mentioned in the ranges in parenthesis for SAR & maximum biological activities of benzimidazole analogs.

From 3D-QSAR model, it is observed that negative values in steric field descriptor indicated the requirement of negative steric potential for enhancing the biological activity of benzimidazole analogs. Therefore less steric substituents were preferred at the position of generated data points S_774 (−0.1842, −0.1634) around benzimidazole pharmacophore. Similarly the negative values of electrostatic descriptors suggested the requirement of electronegative groups like NO$_2$, SO$_2$R, CN, SO$_2$Ar, COOH, F, Cl, Br, I, COOR, OR, OH etc. at the position of generated data point E_1024 (−0.3652, −0.3433) around benzimidazole pharmacophore for maximum activity.

Two data points generated at the position of R$_2$ around benzimidazole nucleus were steric point S_774 (−0.1842, −0.1634) and electrostatic point E_1024 (−0.3652, −0.3433). These points may show requirement of less steric substituents containing electronegative groups.

4. Conclusion

The present study elucidates key structural requirements for the inhibitory activity against cTRPM8 utilizing 2D and 3D-QSAR models. This work reveals how the inhibitory activities against cTRPM8 of various benzimidazole derivatives may be treated statistically to uncover the molecular characteristics which are essential for high activity. The 2D- and 3D-QSAR studies have been carried out on a series of benzimidazole derivatives against cTRPM8. Both 2D- and 3D-QSAR models are statistically significant. The 2D-QSAR model obtained by SW-MLR method indicates that increase in kappa2 and XKMostHydrophilic or decrease in XcompDipole, +vePotentialSurfaceArea and polarizabilityAHP values of benzimidazole derivatives leads to increases in potency of molecules (pIC$_{50}$). The 3D results reveal that less bulky substituents with electronegative groups at R$_2$ position of benzimidazole ring are required for potent cTRPM8 antagonists. A combination of the above models is useful in understanding the structural requirements for design of novel, potent and selective cTRPM8 antagonists. Reliability of the models was confirmed by several statistical analyses and thus, the proposed models, due to the good predictive ability, offer a useful alternative for determining inhibitory activity against cTRPM8 of newly designed molecules. These efforts will guide synthetic medicinal chemists to design and synthesize new compounds with an increased biological activity in comparison to the reported compounds.

References

ACD/Chemsketch 12.0, 2009.

Ajmani, S., Jadhav, K., Kulkarni, S.A., 2006. Three-dimensional QSAR using the k-nearest neighbor method and its interpretation. J. Chem. Inf. Model. 46, 24–31.

Behrendt, H.-J., Germann, T., Gillen, C., Hatt, H., Jostock, R., 2004. Characterization of the mouse cold-menthol receptor TRPM8 and vanilloid receptor type-1 VR1 using a fluorometric imaging plate reader (FLIPR) assay. Br. J. Pharmacol. 141, 737–745.

Bhadoriya, K.S., Jain, S.V., Bari, S.B., Chavhan, M.L., Vispute, K.R., 2012a. 3D-QSAR study of indol-2-yl ethanones derivatives as novel indoleamine 2,3-dioxygenase (IDO) inhibitors. J. Chem. 9, 1753–1759. http://dx.doi.org/10.1155/2012/368617.

Bhadoriya, K.S., Sharma, M.C., Jain, S.V., Raut, G.S., Rananaware, J.R., 2012b. Three-dimensional quantitative structure–activity relationship (3D-QSAR) analysis and molecular docking-based combined in silico rational approach to design potent and novel TRPV1 antagonists. Med. Chem. Res.. http://dx.doi.org/10.1007/s00044-012-0226-4.

Clark, M., Cramer III, R.D., Van, O.N., 1989. Validation of the general purpose Tripose 5.2 force field. J. Comput. Chem. 10, 982–1012.

Cramer III, R.D., Patterson, D.E., Bunce, J.D., 1988. Comparative molecular field analysis (CoMFA) 1. Effect of shape on binding of steroids to carrier proteins. J. Am. Chem. Soc. 110, 5959–5967.

Darlington, R.B., 1990. Regression and Linear Models. McGraw-Hill, New York.

Doweyko, A.M., 2008. QSAR: dead or alive? J. Comput. Aided Mol. Des. 22, 81–89.

Flockerzi, V., 2007. An introduction on TRP channels. HEP 179, 1–19.

Ford, A.C., Talley, N.J., Spiegel, B.M.R., Foxx-Orenstein, A.E., Schiller, L., Quigley, E.M.M., Moayyedi, P., 2008. Effect of fibre, antispasmodics, and peppermint oil in the treatment of irritable bowel syndrome: systematic review and meta-analysis. BMJ 337, a2313. http://dx.doi.org/10.1136/bmj.a2313.

Gasteiger, J., Marsili, M., 1980. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. Tetrahedron 36, 3219–3228.

Gaurav, A., Yadav, M.R., Giridhar, R., Gautam, V., Singh, R., 2010. 3D-QSAR studies of 4-quinolone derivatives as high-affinity ligands at the benzodiazepine site of brain $GABA_A$ receptors. Med. Chem. Res.. http://dx.doi.org/10.1007/s00044-010-9306-5.

Ghosh, P., Bagchi, M.C., 2009. QSAR modeling for quinoxaline derivatives using genetic algorithm and simulated annealing based feature selection. Curr. Med. Chem. 16, 4032–4048.

Gilbert, N., 1976. Statistics. W.B. Saunders Co., Philadelphia, PA.

Golbraikh, A., Tropsha, A., 2002. Beware of q^2! J. Mol. Graph. Model. 20, 269–276.

Golbraikh, A., Tropsha, A., 2003. QSAR modeling using chirality descriptors derived from molecular topology. J. Chem. Inf. Comput. Sci. 43, 144–154.

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182.

Halgren, T.A., 1996a. Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. J. Comp. Chem. 17, 520–552.

Halgren, T.A., 1996b. Merck molecular force field. III. Molecular geometries and vibrational frequencies. J. Comp. Chem. 17, 553–586.

Halgren, T.A., 1996c. Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data and empirical rules. J. Comp. Chem. 17, 616–641.

Halgren, T.A., 1996d. Merck molecular force field: I. Basis, form, scope, parameterization and performance of MMFF94. J. Comp. Chem. 17, 490–519.

Halgren, T.A., 1999a. MMFF VI. MMFF94s option for energy minimization studies. J. Comp. Chem. 20, 720–729.

Halgren, T.A., 1999b. MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. J. Comp. Chem. 20, 730–748.

Hansch, C., 1969. A quantitative approach to biochemical structure–activity relationships. Acc. Chem. Res. 2, 232–239.

Hayashi, T., Kondo, T., Ishimatsu, M., Yamada, S., Nakamura, K., Matsuoka, K., Akasu, T., 2009. Expression of the TRPM8-immunoreactivity in dorsal root ganglion neurons innervating the rat urinary bladder. Neurosci. Res. 65, 245–251.

Jain, S.V., Bhadoriya, K.S., Bari, S.B., 2012a. QSAR and flexible docking studies of some aldose reductase inhibitors obtained from natural origin. Med. Chem. Res. 21, 1665–1676. http://dx.doi.org/10.1007/s00044-011-9681-6.

Jain, S.V., Bhadoriya, K.S., Bari, S.B., Sahu, N.K., Ghate, M., 2012b. Discovery of potent anticonvulsant ligands as dual NMDA and AMPA receptors antagonists by molecular modelling studies. Med. Chem. Res. 21, 3465–3484. http://dx.doi.org/10.1007/s00044-011-9889-5.

Jain, S.V., Ghate, M., Bhadoriya, K.S., Bari, S.B., Chaudhari, A., Borse, J.S., 2012c. 2D, 3D-QSAR and docking studies of 1,2,3-thiadiazole thioacetanilides analogues as potent HIV-1 non-nucleoside reverse transcriptase inhibitors. Org. Med. Chem. Lett. 2, 22. http://dx.doi.org/10.1186/2191-2858-2-22.

Klein, A.H., Sawyer, C.M., Carstens, M.I., Tsagareli, M.G., Tsiklauri, N., Carstens, E., 2010. Topical application of L-menthol induces heat analgesia, mechanical allodynia, and a biphasic effect on cold sensitivity in rats. Behav. Brain Res. 212, 179–186.

Mahieu, F., Owsianik, G., Verbert, L., Janssens, A., De Smedt, H., Nilius, B., Voets, T., 2007. TRPM8-independent Menthol-induced Ca^{2+} release from endoplasmic reticulum and golgi. J. Biol. Chem. 282, 3325–3336.

McKemy, D.D., Neuhausser, W.M., Julius, D., 2002. Identification of a cold receptor reveals a general role for TRP channels in thermosensation. Nature 416, 52–58.

Merat, S., Khalili, S., Mostajabi, P., Ghorbani, A., Ansari, R., Malekzadeh, R., 2010. The effect of enteric-coated, delayed-release peppermint oil on irritable bowel syndrome. Dig. Dis. Sci. 55, 1385–1390.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21, 1087–1092.

Nakka, S., Guruprasad, L., 2012. The imidazolidone analogs as phospholipase D1 inhibitors: analysis of the three-dimensional quantitative structure–activity relationship. Med. Chem. Res. 21, 2517–2525. http://dx.doi.org/10.1007/s00044-011-9773-3.

Parks, D.J., Parsons, W.H., Colburn, R.W., Meegalla, S.K., Ballentine, S.K., Illig, C.R., Qin, N., Liu, Y., Hutchinson, T.L., Lubin, M.L., Stone Jr., D.J., Baker, J.F., Schneider, C.R., Ma, J., Damiano, B.P., Flores, C.M., Player, M.R., 2011. Design and optimization of benzimidazole-containing transient receptor potential melastatin 8 (TRPM8) antagonists. J. Med. Chem. 54, 233–247.

Peier, A.M., Moqrich, A., Hergarden, A.C., Reeve, A.J., Andersson, D.A., Story, G.M., Earley, T.J., Dragoni, I., McIntyre, P., Bevan, S., Patapoutian, A., 2002. A TRP channel that senses cold stimuli and menthol. Cell 108, 705–715.

Radhika, V., Kanth, S.S., Vijjulatha, M., 2010. CoMFA and CoMSIA studies on inhibitors of HIV-1 integrase-bicyclic pyrimidinones. J. Chem. 7 (S1), S75–S84.

Rohacs, T., Lopes, C.M.B., Michailidis, I., Logothetis, D.E., 2005. PI(4,5)P$_2$ regulates the activation and desensitization of TRPM8 channels through the TRP domain. Nat. Neurosci. 8, 626–634.

Sabnis, A.S., 2007. Expression and characterization of TRPM 8 receptors in lung epithelial cells. University of Utah Spencer S. Eccles Health Sciences Library, pp. 1–159.

Sabnis, A.S., Shadid, M., Yost, G.S., Reilly, C.A., 2008. Human lung epithelial cells express a functional cold-sensing TRPM8 variant. Am. J. Respir. Cell Mol. Biol. 39, 466–474.

Sahu, N.K., Bari, S.B., Kohli, D.V., 2011. Molecular modeling studies of some substituted chalcone derivatives as cysteine protease inhibitors. Med. Chem. Res.. http://dx.doi.org/10.1007/s00044-011-9900-1.

Scior, T., Medina-Franco, J.L., Do, Q.T., Martinez-Mayorga, K., Yunes Rojas, J.A., Bernard, P., 2009. How to recognize and workaround pitfalls in QSAR studies: a critical review. Curr. Med. Chem. 16, 4297–4313.

Sharaf, M.A., Illman, D.L., Kowalski, B.R., 1986. Chemometrics, Chemical Analysis Series. Wiley, New York.

Sharma, M.C., Kohli, D.V., 2012. A comprehensive structure–activity analysis 2,3,5-trisubstituted 4,5-dihydro-4-oxo-3H-imidazo [4,5-c] pyridine derivatives as angiotensin II receptor antagonists: using 2D- and 3D-QSAR approach. Med. Chem. Res.. http://dx.doi.org/10.1007/s00044-012-0040-z.

Shen, M., Xiao, Y., Golbraikh, A., Gombar, V.K., Tropsha, A., 2003. An in silico screen for human S9 metabolic turnover using k-nearest neighbor QSPR method. J. Med. Chem. 46, 3013–3020.

Sherkheli, M.A. et al., 2007. Selective TRPM8 agonists: a novel group of neuropathic analgesics. FEBS J. 274 (s1), 232. http://dx.doi.org/10.1111/j.0014-2956.2007.05861_4.x.

Sherkheli, M.A., Gisselmann, G., Vogt-Eisele, A.K., Doerner, J.F., Hatt, H., 2008. Menthol derivative WS-12 selectively activates transient receptor potential melastatin-8 (TRPM8) ion channels. Pak. J. Pharm. Sci. 21, 370–378.

Tanneeru, K., Reddy, B.M., Guruprasad, L., 2011. Three-dimensional quantitative structure–activity relationship (3D-QSAR) analysis and molecular docking of ATP-competitive triazine analogs of human mTOR inhibitors. Med. Chem. Res.. http://dx.doi.org/10.1007/s00044-011-9629-x.

Tsavaler, L., Shapero, M.H., Morkowski, S., Laus, R., 2001. Trp-p8, a novel prostate-specific gene, is up-regulated in prostate cancer and other malignancies and shares high homology with transient receptor potential calcium channel proteins. Cancer Res. 61, 3760–3769.

Tsukimi, Y., Mizuyachi, K., Yamasaki, T., Niki, T., Hayashi, F., 2005. Cold response of the bladder in guinea pig: involvement of transient receptor potential channel, TRPM8. Urology 65, 406–410.

Verma, J., Khedkar, V.M., Coutinho, E.C., 2010. 3D-QSAR in drug design – a review. Curr. Top. Med. Chem. 10, 95–115.

VLife MDS 3.5, 2008. Molecular Design Suite. Vlife Sciences Technologies Pvt. Ltd., Pune, India. <http://www.vlifesciences.com>.

Voets, T., Droogmans, G., Wissenbach, U., Janssens, A., Flockerzi, V., Nilius, B., 2004. The principle of temperature-dependent gating in cold- and heat-sensitive TRP channels. Nature 430, 748–754.

Voets, T., Owsianik, G., Nilius, B., 2007. TRPM8. HEP 179, 329–344.

Zhao, H., Sprunger, L.K., Simasko, S.M., 2010. Expression of transient receptor potential channels and two-pore potassium channels in subtypes of vagal afferent neurons in rat. Am. J. Physiol. Gastrointest. Liver Physiol. 298, G212–221.

Zheng, W., Tropsha, A., 2000. Novel variable selection quantitative structure–property relationship approach based on the k-nearest neighbor principle. J. Chem. Inf. Comput. Sci. 40, 185–194.