MOLECULAR FIELD-BASED QSAR STUDIES AND DOCKING ANALYSIS OF MERCAPTOQUINAZOLINONE BENZENE SULFONAMIDE DERIVATIVES AGAINST hCA XII

P. Gopinath and M. K. Kathiravan
Department of Pharmaceutical Chemistry, SRMIST; Chennai, Tamil Nadu, India, 603203.

ABSTRACT
Selective targeting of the tumor-associated hCA XII isozyme is a promising strategy to obtain effective and safer agents in cancer therapy. A series of mercapto-quinazolinone benzene sulfonamide derivatives were subjected to molecular field analysis to derive 3D-QSAR models. Structural properties such as physicochemical, topological, electro-topological, and quantum-chemical descriptors were calculated using the Molecular Design Suite of V-life MDS 4.6 Software. The contour map generated from the SA-kNN model explains the significance of electrostatic and steric descriptors for hCA XII binding interaction. Molecular docking studies favored structural insights in association with QSAR. Most interacting residues Asn67, Gln92, Thr199 and His119 stabilized the compounds in the active pocket. The results suggest structural insights as well as highlight the key binding features of mercapto-quinazolinone benzene sulfonamide derivatives against hCA XII which can be utilized for the design and development of potent leads.

Keywords: Molecular Field Analysis, Simulated Annealing, Genetic Algorithm, Molecular Docking

INTRODUCTION
Human Carbonic anhydrase (hCA) iso-enzymes have an important role in the essential cellular process like respiration, pH regulation, tumorigenicity, electrolyte secretion, along with the bio-synthesis reactions which requires substrates like HCO$_3^-$ and CO$_2$ (example, ureagenesis, lipogenesis, and glucogenesis) and many similar processes. Metal center bound inhibitor participates in the process of shuttling a proton between the water molecule (bound with metal ion) and its environment. This results in an enhanced formation of metal hydroxide, a catalytic active species of an enzyme. Generally, an inhibitor is bound to the metal ion present in the enzyme active site as anions in the deprotonated state. However, an alternate inhibition mechanism under which an inhibitor was interacting with zinc coordinated hydroxide ion/water molecule, or in some cases do not interact with the zinc core at all, has been recently described. As presently known, there were fifteen different CA isozymes in humans that belong to the α-CA class; exhibiting variable tissue distribution, enzyme kinetics, subcellular locations and different expression levels.

hCA IX and XII with dimeric transmembrane were CA isoforms associated with the humans having an extra-cellular active site, which acts as the markers of a broad spectrum of different hypoxic tumor types. Both the hCA IX and XII regulate extra and intracellular pH & also are seen in the normal tissues. However, the expressions of hCA XII are not modulated with hypoxia, and were seen to increase in aggressively increasing tumors. For treating or for imaging of the tumors, the hCA IX & XII are the target drugs as these were over-expressed in many types of cancer and have an effect on the process which promotes cell metastasis, invasion, or proliferation. Selective targeting on the hCA isoymes (IX and XII) follows a promising method for obtaining safe and effective agents in cancer therapy. This is normally achieved by two approaches. The ring method includes direct connections with sulfonamide group; and in “tail approach”, the different groups of tails were appending to a zinc-binding bearing ring group that could facilitate interactions with specific residues in an active site and was the most prevalent approach. It was known that ureido benzene sulfonamide (SLC-0111), a selective CA IX and XII inhibitor was presently in Phase-II/b clinical trials for metastatic and solid tumor therapy.
coumarins and coumarins show the most CA inhibition isoform-selective profile out of all of the different CAIs discovered till today (like the phenols, polyamines, xanthates, di-thiocarbamates, sulfamides, sulfamates, and sulfonamides). In this context, Quinazolinone scaffold which was widely used in medicinal formulations due to its biological significance was considered for QSAR study. The literature review presents various substituted benzene sulfonamide derivatives showing effective inhibition activities against the hCA isoform subset with sub-nanomolar inhibition constants. However, these studies explored all possible substitutions on quinazolinone ring but lack theoretical studies based on steric, electrostatic information needed for potential lead generation. Molecular modeling approaches will definitely provide the structural insights needed for generating useful lead compounds.

In order to develop novel leads with better selectivity towards hCA XII, QSAR studies (Quantitative Structure-Activity Relationship) help to correlate models having their biological activity compared to chemical structures. In continuation to our published work on carbonic anhydrase inhibitors, we herein report the 3D-QSAR models reflecting the key structural features of mercapto-quinazolinone based benzene sulfonamides for human carbonic anhydrase XII inhibition. In this investigation, stochastic optimization algorithms such as simulated annealing as well as the genetic algorithm were used for descriptor optimization coupled with kNN–molecular field analysis.

**EXPERIMENTAL**

**Biological Activity Data Set for Analysis**

The 3D-QSAR studies were conducted using a VLifecMDS, Licensed software system version 4.6. 2020. The dataset containing mercapto-quinazolinone benzene sulfonamides as human carbonic anhydrase XII inhibitors reported by Adel S. El-Azab et al. and Murat Bozdag et al. was selected for the study and the data was listed in Table-1.

| Compound | R_1 | R_2 | R_3 | R_4 | Ki (nM) | pKi |
|----------|-----|-----|-----|-----|---------|-----|
| 1^c      | H   | H   | H   | H   | 0.59    | 9.2291 |
| 2^a      | CH_3| H   | H   | H   | 3.1     | 8.5086 |
| 3^c      | C_2H_5| H | H   | H   | 3.9     | 8.4089 |
| 4^a      | CH_2CN| H | H   | H   | 8.6     | 8.0655 |
| 5^a      | Bn  | H   | H   | H   | 38.4    | 7.4157 |
| 6^b      | 4-CN-Bn| H | H   | H   | 17.6    | 7.7545 |
| 7^a      | 4-F-Bn| H | H   | H   | 17.2    | 7.7645 |
| 8^b      | 4-CH_3-Bn| H | H   | H   | 28.2    | 7.5498 |
| 9^c      | 4-NO_2-Bn| H | H   | H   | 25.6    | 7.5918 |
| 10^a     | CH_2-CH_2-piperidin-N-yl | H | H   | H   | 20.2    | 7.6946 |
| 11^a     | CH_2COPh| H | H   | H   | 13      | 7.8861 |
| 12^a     | CH_2CO(4-Cl-Ph) | H | H   | H   | 10.8    | 7.9666 |
| 13^a     | CH_2CO(4-F-Ph) | H | H   | H   | 9.1     | 8.0409 |
| 14^a     | CH_2CONH_2| H | H   | H   | 2.4     | 8.6198 |
| 15^a     | CH_2CONHPh| H | H   | H   | 38.4    | 7.4157 |
| 16^a     | CH_2CONH(4-Cl-Ph) | H | H   | H   | 31.6    | 7.5003 |
MERCAPTOQUINAZOLINONE BENZENE SULFONAMIDE DERIVATIVES

P. Gopinath and M. K. Kathiravan

Computational Details

Molecule 3D geometry optimization and energy minimization were done by batch calculations in VLife MDS 4.6 software using a dielectric function (distance-dependent), MMFF: ‘Merck Molecular Force Field’ and the RMS: ‘Root Mean Square’ with iteration limits set to 10000, and gradient to 0.001Kcal/mol Å.

Conformation Generation of Molecules

For every molecule, multiple conformations were generated using systematic conformational search in which, rotatable bonds were identified. All atomic distances were computed, for every conformer generated by the rotation set. Systematic conformational analysis was a lengthy procedure due to the involved computations dealing with the following unbounded polynomial.

\[ N = \frac{360}{\Delta} \times n \]

Where

- \( N \) = Total conformations examined with an exhaustive search.
- \( \Delta \) = rotation increment in degrees also called the step size of rotation.
- \( n \) = the number of rotatable bonds

Each bond is a part of 4 connected atoms. The 1st and 4th atoms were required for measuring torsional angle in the conformation, with the middle two atoms forming a rotatable bond. A low-energy conformer in every compound is considered for studying 3D-QSAR.

Molecules Alignment

In kNN-MFA study, the molecules were aligned in a 3D space, with specific atoms super-imposed based on distances and positions. Positions of every atom are crucial in the kNN–MFA study for the reason that the calculation of the descriptors was done using a 3D-space grid. It is extremely important to properly align the structures, as well as the development of valid and effective 3DQSAR models. Geometry-optimized and energy-minimized molecule structures were aligned using the template-based technique where the structure template was used. The structure of the template, i.e. mercapto-quinazolinone ring attached to the N-ethyl bridge of benzene-sulfonamide, was used in alignments by considering all the common elements present in a series. The reference molecule was selected so that it would be highly active among all of the molecule series. Compound 31 with high hCA XII inhibitory activity was selected as the reference molecule. The reference molecule and the template structure, after optimizing were used for superimposing all molecules in a series, following the template alignment method in VLife MDS 4.6.
All of the molecules were super-imposed by minimizing the RMS deviation, furthermore, the molecules were used in building 3D-QSAR kNN–MFA models.

**Calculation of Molecular Field Descriptors for 3D-QSAR Analysis**

The over-laid molecule set after super-imposition was positioned in the center of the grid box or the lattice, which were used to calculate interacting energies between ligands and the different probe atoms placed along the intersection of the grid box or the lattice. The biologically aligned active conformations of mercapto-quinazolinone benzene sulfonamide derivatives were used to calculate the molecular field interaction energies (for hydrophobic, electrostatic and steric) that were computed on a lattice point of a grid using a charge +1 methyl probe and the Gasteiger–Marsili charges. To perform descriptor calculation 30.0 kcal/mol steric and 10.0 kcal/mol electrostatic cut-off was used. Descriptor was the term used for indicating the field values on lattice points and the values of interaction energy as the descriptors were used for generating relationships. The dielectric constant dependent on the distance was assigned a value 1.0. VLife MDS software was used to calculate a total of 6006 3-D descriptors. This includes 2002 descriptors for each hydrophobic, electrostatic, and steric field parameter for every compound in separate columns. To carry out the QSAR analysis, every invariable column was removed from the worksheet as it does not contribute to QSAR.

**Design of the Training and Test Sets using SEM (Sphere Exclusion Method)**

The 34 molecules data set was divided into: training, cross-validation test, and external-validation tests set by SEM algorithm with 2.23 as the dissimilarity value which helps in estimating the sphere exclusion radius. The SEM algorithm helps in constructing the training sets considering all descriptor space regions at the representation points.

**Variable (Feature) Selection Methods**

The aim here is to calculate an optimal variable producing a significant 3D-QSAR model relating to the compound structures and binding affinities. Molecular field analysis is a 3D-QSAR method that relates the biological molecular activity with electrostatic and steric interactions as per the Coulomb and Lennard-Jones potentials, respectively. Multi-collinearity and chance correlations are the two problems that occur while attempting to find the general QSAR models used for designing drugs. In every model-building exercise, an integral part was to select an appropriate set of features having good predictive accuracy and low complexity. For this purpose, variable selection techniques like simulated annealing and genetic algorithm coupled with model building methods like kNN–molecular field analysis, multiple linear regression analysis were applied to establish the 3D-QSAR models. The kNN method was the simplest algorithm in machine learning applied to classify new patterns (like in this case a molecule). The k-nearest neighbor method works on the approach of simple learning distance, where the unknown number was classified as per the majority of their kNN in a training set. The kNN standard method was implemented as:

1. Estimate the distance between ‘u’ (an unknown object) and every other object in a training set;
2. From training sets, ‘k’ objects are selected, same as that of the object ‘u’, depending on the calculated distances,
3. Object ‘u’ was classified with groups to which maximum of the ‘k’ objects belong. The optimal ‘k’ value was selected by optimization by classifying the test set of the samples or by the validation method of the leave-one-out cross.

The variable ‘k’ values were selected by following the step-wise variable selection technique. It makes use of a step-wise method for selecting the variable combined with the kNN method for optimizing k (nearest neighbors) and variable selection out of the original pool. This step-to-step search method begins by forming a trial model using single independent variables and adding the independent variables, one step after another, examining a fit for the model on every step (with the procedure of weighted kNN cross-validation). This method is used till there were no more remaining significant models outside the model. Once both test and training sets were generated, the kNN method was applied to the generated descriptors on a grid. Hydrophobic, electrostatic and steric, energies were estimated on-grid lattice points.
using +1 charged methyl probe. These values of interaction energies were considered for generating
relations and were used as descriptors to decide the distance between the different molecules. Simulated
annealing follows the same physical annealing process, involving a system heated to a very high-
temperature value and cooling it to a pre-set room temperature. The system samples the possible
distributed configuration as per the Boltzmann distribution method so that at equilibrium, the lower
energy states were most populated. On every step, eqn.-1 is calculated.

\[ d = (V_{\text{new}})^2 - (V_{\text{old}})^2 \]  

(1)

\( V_{\text{new}} \) was accepted in case \( d > 0 \), or with \( -d/T \) probability \( \exp \); here, ‘T’ is the temperature control
parameter. Idea here was, to begin with, a higher ‘T’ value, to make sure all steps were accepted and
reduce ‘T’ gradually in the simulation process. Steps that help in improving the solutions were only
accepted during the simulation progress. An acceptance ratio was calculated for every ‘m’ step. When the
set convergence criterion, \( \beta \) was \( \geq \) acceptance ratio, the process was stopped, giving ‘V’ vector a minimal
cost.

In the present method regression model value ‘\( r^2 \)’, and minimum energy configuration was a better
solution with a subset of \( V \). Temperature was a control parameter and allows the higher energy
configurations (\( V_s \) with high cost), also this search method was likely to be trapped in the local minima.
For simulating the evolutions of the system, a metropolis method can also be applied. On every step, this
algorithm chooses \( V_{\text{new}} \) from \( V_{\text{old}} \). In every move, ‘V’ was changed by changing the descriptor(s) out of a
pool of other descriptors. Holland described a genetic algorithm, which was amongst the most commonly
used stochastic optimization methods mimicking a natural selection and evolution. This algorithm class
was inspired by the natural process of evolution in which, species with higher fitness under certain
conditions could survive and prevails to the next generation; an optimal species could be adapted by
mutation or crossover of the better ones. In this method, the chromosomes and their fitness in a species
demonstrated a set of molecular descriptors and cross-validation was also done for predicting the QSAR
model’s accuracy.

Internal and External Validations
The ‘\( q^2 \)’, LOO (leave-One-Out) methodology was followed to carry out internal validation. To calculate
the value of ‘\( q^2 \)’, every training set molecule was eliminated sequentially, the same descriptors were used
to refit the models, and this refit model was used to predict the biological activity of every eliminated
molecule. Following equation 2 was used to calculate the cross-validated coefficient: ‘\( q^2 \),

\[ (q)^2 = 1 - \frac{\sum (y_i - y^\wedge_i)^2}{\sum (y_i - y_{mean})^2} \]  

(2)

Here, ‘\( y_i \)’ was the actual and ‘\( y^\wedge_i \)’ was predicted \( i^{\text{th}} \) molecule activity in the training set. The average
activity of every molecule was depicted by ‘\( y_{mean} \)’ in the training set. Here the high value of ‘\( q^2 \)’ does not
give a suitable presentation of real predicted power for hCA XII inhibitor model. Hence, some external
validation was also conducted in this research. For this model, the external predictive power was
estimated by predicting the value of pKi for 4-different test molecule sets, which are not used in the 3D-
QSAR model. For a selected model their predictive ability was confirmed by ‘\( \text{pred}_r^2 \)’. The external
validations were done for the activity of every test set molecule predicted by the model developed from
the training sets. The value of ‘\( \text{pred}_r^2 \)’ was estimated using eqn.-3.

\[ \text{Pred}_r^2 = 1 - \frac{\sum (y_i - y^\wedge_i)^2}{\sum (y_i - y_{mean})^2} \]  

(3)

Where ‘\( y_i \)’ was the actual and ‘\( y^\wedge_i \)’ was predicted \( i^{\text{th}} \) molecule activity in the test set. An average activity
of every molecule is depicted by ‘\( y_{mean} \)’ in a training set. The value of ‘\( \text{pred}_r^2 \)’ indicates the predictive
power of the system as per the external test set.
Molecular Docking Protocol
The optimized lowest energy conformations of the best dataset compounds (compound 1, 21, 23, 28, 30 and 31) generated from Avogadro tool were subjected to molecular docking studies to study the residue interactions, H-bonds and binding energy scores using Autodock V4.2.6. The original configuration of hCA XII, pdb id: 6EBE was derived from the protein database (www.rcsb.org). Using the Modeller V9.23 program missing residues were fixed along with the addition of hydrogen atoms and removal of existing ligands.

Protein was prepared for docking by adding geometric polar hydrogen and the pdbqt file was allocated to Kollman's united atomic charges. Histidine protonation residue states were examined when ND1 was assigned to zinc-bound histidines and NE2 to the residual histidines. Ligand file formation was done by adding polar hydrogens and applying gasteiger charges. Torsions were detected in the ligands and a file of pdbqt was produced. The Autogrid option makes it possible to choose an active site and to set the grid size to 60*60*60 points having 0.375 Å spacing and calculate the energetic map with a distance-dependent dielectric constant function. Autodock bound parameters file was subjected to ensure correct radius, well-depth, and charges associated with the metalloprotein. The grid box comprises the enzyme's active binding site and provides enough space to rotate and translate the ligand. For ligand conformation poses, orientations inside the active site hCA XII were employed in the Lamarck genetic algorithm. The following are the optimized parameters; Maximum energy assessments are raised to 25,000,000 per run in the population, and the gene mutation rate was 0.02. For the performance of the docking, all other parameters are set to the default and metal parameters were added. The search algorithm analysis generates ligand pose at CA XII binding site, taking into concern the roto-translational and internal degrees of freedom of the ligand. Best poses were identified by means of binding affinity and pose retrieval was done using Discovery studio visualizer non-bonding interaction predictions and bond distances. LigRMSD tool generated RMSD values between the reference and the predicted structures were used to confirm whether the docking simulation has predicted a close-match docked pose or not. The co-crystallized ligand has been redocked to 6EBE for the identification of docking parameters that are useful for docking compound design.

RESULTS AND DISCUSSION
The dataset compounds were geometry optimized using the MMFF force field applying distance-dependent dielectric constant and Gasteiger-Marsili charges. The compounds with hCA XII inhibitory activity (Table-1) were subjected to a genetic algorithm and simulated annealing feature selection methods using VLife MDS 4.6 software. The training and test set were segregated (Table-1) by the sphere exclusion algorithm technique and further, the models were validated by both internal (cross-validation) and external validation procedures.

The UniColumn statistics of training and test sets reveal the acceptable selection of compounds into the partition sets. The max and min values in the test and training sets were compared as follows:

1. The max value of pKi of the test set should be ≤ the max value of pKi of the training set.
2. The min value of pKi of the test set should be ≥ the min value of pKi of the training set.

After attaining the proper alignment, the test set looked interpolative (7.5498 to 9.2676) and was derived within the min-max range of the training set (7.3449 to 9.3279). The mean and standard deviation of pKi values of sets of training and test provides insights into the relative difference of mean and point density distribution (along with mean). The mean of the test set (8.5934) was higher than the training set (8.2604) indicating the presence of relatively more active molecules as compared to the inactive ones. Also, the similar standard deviation in both sets indicates that the spread in both the sets with their respective mean was comparable. Statistically significant best 3D-QSAR models were selected, considering the term selection criterion as ‘q^2', ‘pred_r^2', and the standard error values.

Interpretation of 3D QSAR models
It was known that multiple linear regression makes strong assumptions about the form of f(X). In a linear relationship between X and Y, if the true relationship was far from linear, then the resulting model will provide a poor fit to the data, and any conclusions drawn from it will be suspected. On the other hand,
kNN (k-Nearest Neighbor) does not assume an explicit form for f(X), providing a more flexible approach. In this present investigation, two widely used techniques simulated annealing (SA) and genetic algorithm (GA) was applied for 3D descriptor optimization. GA deals with parameters such as population size, mutation rate, and cut length whereas SA involves temperature, cooling rate, and absolute temperature. GA provides quality solutions if a large enough population was set but a large population greatly increases its run time. In SA, the cooling rate that was very close to one should be set for optimal solutions but this increases the number of iterations the algorithm will perform.

The present paper was an attempt in the direction seeking the development of better QSAR models by different feature selection methods for mercapto-quinoxalinone benzene sulfonamide derivatives. The model statistics for the training set were determined by calculating $r^2$ and $q^2$, whereas external validation was determined by predictive $r^2$ (pred $r^2$) for the test set compounds. 3D QSAR models developed by different feature selection and model building methods were evaluated statistically and the results were presented in Table-2. The observed and predicted activity values of the training and test set compounds selected for different 3D QSAR models were presented in supplementary Table-1.

| S. No. | pKi | SA-kNN Model | GA-kNN Model | SA-MLR Model | GA-MLR Model |
|--------|-----|---------------|--------------|--------------|--------------|
|        |     | Predicted | Residual | Predicted | Residual | Predicted | Residual | Predicted | Residual | Predicted | Residual |
| 1      | 9.229 | 9.012 | 0.217 | 9.237 | -0.008 | 9.075 | 0.154 | 9.019 | 0.21 |
| 2      | 8.509 | 8.601 | -0.092 | 8.772 | -0.263 | 8.824 | -0.315 | 8.729 | -0.22 |
| 3      | 8.409 | 8.286 | 0.123 | 8.579 | -0.17 | 8.486 | -0.077 | 8.46 | -0.051 |
| 4      | 8.066 | 8.564 | -0.498 | 8.885 | -0.819 | 8.528 | -0.462 | 8.646 | -0.58 |
| 5      | 7.416 | 7.632 | -0.216 | 7.842 | -0.426 | 7.66 | -0.244 | 7.735 | -0.319 |
| 6      | 7.755 | 7.607 | 0.148 | 7.713 | 0.042 | 7.81 | -0.055 | 7.691 | 0.064 |
| 7      | 7.764 | 7.782 | -0.018 | 7.826 | -0.062 | 7.732 | 0.032 | 7.628 | 0.136 |
| 8      | 7.55 | 7.577 | -0.027 | 7.756 | -0.206 | 7.594 | -0.044 | 7.575 | -0.025 |
| 9      | 7.592 | 7.893 | -0.301 | 7.771 | -0.179 | 7.832 | -0.24 | 7.628 | -0.036 |
| 10     | 7.695 | 7.915 | -0.22 | 8.345 | -0.65 | 7.751 | -0.056 | 8.049 | -0.354 |
| 11     | 7.886 | 7.559 | 0.327 | 7.718 | 0.168 | 7.638 | 0.248 | 8.028 | -0.142 |
| 12     | 7.967 | 7.771 | 0.196 | 7.683 | 0.284 | 7.627 | 0.34 | 7.599 | 0.368 |
| 13     | 8.041 | 7.733 | 0.308 | 7.774 | 0.267 | 7.965 | 0.076 | 7.78 | 0.261 |
| 14     | 8.62 | 8.287 | 0.333 | 8.834 | -0.214 | 8.665 | -0.045 | 9.026 | -0.406 |
| 15     | 7.416 | 7.733 | -0.317 | 7.77 | -0.354 | 7.523 | -0.107 | 7.422 | -0.006 |
| 16     | 7.5 | 8.004 | -0.504 | 7.753 | -0.253 | 7.525 | -0.252 | 7.488 | 0.012 |
| 17     | 8.06 | 7.706 | 0.354 | 7.713 | 0.347 | 7.577 | 0.303 | 7.738 | 0.322 |
| 18     | 7.345 | 7.829 | -0.484 | 7.826 | -0.481 | 7.519 | -0.174 | 7.92 | -0.575 |
| 19     | 7.648 | 7.912 | -0.264 | 7.795 | -0.147 | 7.674 | -0.026 | 7.637 | 0.011 |
| 20     | 7.772 | 7.615 | 0.157 | 7.796 | -0.024 | 7.861 | -0.089 | 7.878 | -0.106 |
| 21     | 9.208 | 9.016 | 0.192 | 9.047 | 0.161 | 9.177 | 0.031 | 8.955 | 0.253 |
| 22     | 9.194 | 9.023 | 0.171 | 9.049 | 0.145 | 8.848 | 0.346 | 8.907 | 0.287 |
| 23     | 9.268 | 9.027 | 0.241 | 8.57 | 0.698 | 9.062 | 0.206 | 8.771 | 0.497 |
| 24     | 9.181 | 9.018 | 0.163 | 8.577 | 0.604 | 8.517 | 0.664 | 8.65 | 0.531 |
| 25     | 8.839 | 9.201 | -0.362 | 8.632 | 0.207 | 8.876 | -0.037 | 8.676 | 0.163 |
| 26     | 9.119 | 9.023 | 0.096 | 8.565 | 0.554 | 9.058 | 0.061 | 8.806 | 0.313 |
| 27     | 8.257 | 9.016 | -0.759 | 8.861 | -0.604 | 8.705 | -0.448 | 8.869 | -0.612 |
| 28     | 9.215 | 8.858 | 0.357 | 8.638 | 0.577 | 8.812 | 0.403 | 8.642 | 0.573 |
| 29     | 9.137 | 8.851 | 0.286 | 8.573 | 0.564 | 8.766 | 0.371 | 8.509 | 0.628 |
| 30     | 9.268 | 9.016 | 0.252 | 9.031 | 0.237 | 9.114 | 0.154 | 8.997 | 0.271 |
| 31     | 9.328 | 9.172 | 0.156 | 9.019 | 0.309 | 8.895 | 0.433 | 8.732 | 0.596 |
| 32     | 8.241 | 8.384 | -0.143 | 9.028 | -0.787 | 8.037 | 0.204 | 8.61 | -0.369 |
| 33     | 9.187 | 9.268 | -0.081 | 8.838 | 0.349 | 9.399 | -0.212 | 8.635 | 0.552 |
| 34     | 8.148 | 8.451 | -0.303 | 8.818 | -0.67 | 8.214 | -0.066 | 8.317 | -0.169 |
Table-2: Comparison of the 3D QSAR models for mercapto-quinazolinone based benzenesulfonamides as human carbonic anhydrase XII inhibitors.

| Statistical Parameter       | 3D QSAR results         |
|-----------------------------|-------------------------|
|                             | SA-kNN                  | GA-kNN                  |
| \(N_{\text{training}}\)    | 24                      | 24                      |
| Degree of freedom           | 19                      | 21                      |
| Nearest neighbor            | 2                       | 5                       |
| \(R^2\)                    | -                       | -                       |
| \(Q^2\)                    | 0.7540                  | 0.5937                  |
| F-test                     | -                       | -                       |
| \(R^2\)-se                 | -                       | -                       |
| \(Q^2\)-se                 | 0.3275                  | 0.4208                  |
| Pred \(R^2\)               | 0.9337                  | 0.6525                  |
| Pred \(R^2\)-se            | 0.2419                  | 0.5540                  |
| Ext Val \(R^2\)            | 0.9245                  | 0.8285                  |
| Ext Val \(R^2\)-se         | 0.2324                  | 0.3502                  |

**Contributing descriptors**

| Descriptor(Grid points)     | SA-MLR                  | GA-MLR                  |
|-----------------------------|-------------------------|-------------------------|
| \(S_{513}\) (-0.0366,-0.0349), \(E_{797}\) (2.2558, 2.3718), \(E_{459}\) (0.0023, 0.2310), \(S_{265}\) (-0.0003, -0.0003), \(S_{1150}\) (-0.0558, -0.0555) | | |
| Descriptor(Grid points)     | H_{110} (0.1192, 0.1267), S_{770} (-0.0053, -0.0044), S_{1655} (-0.0552, -0.0544) | | |

| \(N_{\text{training}}\)    | 24                      | 24                      |
| Degree of freedom           | 19                      | 21                      |
| Nearest neighbor            | -                       | -                       |
| \(R^2\)                    | 0.8539                  | 0.6840                  |
| \(Q^2\)                    | 0.7621                  | 0.5947                  |
| F-test                     | 22.2063                 | 15.1554                 |
| \(R^2\)-se                 | 0.2837                  | 0.3967                  |
| \(Q^2\)-se                 | 0.3619                  | 0.4493                  |
| Pred \(R^2\)               | 0.8161                  | 0.7563                  |
| Pred \(R^2\)-se            | 0.4030                  | 0.4639                  |
| Ext Val \(R^2\)            | 0.9576                  | 0.9321                  |
| Ext Val \(R^2\)-se         | 0.1740                  | 0.2204                  |

**Contributing descriptors**

| Descriptor(coefficient)     | S_{1213} \((-0.020±0.0000), \(H_{1711}\) \((-3.3433±0.5990), \(E_{1807}\) \((-1.2703±0.1716), \(E_{389}\) (2.0939±0.4129), \(S_{523}\) (18.6999±3.1413)\) | H_{110} \((-7.5111±0.1426), S_{246}\) \((199.1570±81.5787), S_{1655}\) \((4.8423±0.2307)\) |
| Constant:9.2697            | | Constant:10.6313 |

From statistical results in terms of cross-validated coefficient (\(q^2\)) and standard error values, it can be seen that the SA-MLR model has performed better compared to its counterpart with a genetic algorithm-based model. On the other hand, the SA-kNN model has performed better compared to its counterpart GA-kNN. Standard error values were found to be high for GA-MLR model compared to its counterpart SA-MLR model and the relatively same phenomenon was observed in the case of GA-kNN and SA-kNN models. Overall, the SA-kNN model has shown better statistical values as well as predictions of hCA XII inhibition activity. The value of pred-\(r^2\) for the test set was found to be 0.9337, which means 93 % predictive power for the external test set with a standard error value of 0.2419. The contribution plot arising out of 3D-QSAR studies provides some useful insights for a better understanding of the structural features of these compounds responsible for producing significant hCA XII inhibitors (Fig.-1). The descriptors \(S_{513}\), \(E_{797}\), \(E_{459}\), \(S_{265}\), and \(S_{1150}\) were the steric and electrostatic field energy of interactions between methyl probe and compounds at their corresponding spatial grid points of 513, 797,
459, 265, and 1150. These points suggested the significance and requirement of steric and electrostatic properties in the ranges in parenthesis (as shown in Table-2) for SAR and maximum biological activities of mercapto-quinazolinone based benzene sulfonamide analogs.

![Contribution Plot for Steric and Electrostatic Interactions from SA-kNN MFA Model](image)

Fig.-1: The Contribution Plot for Steric and Electrostatic Interactions from SA-kNN MFA Model

From the 3D-QSAR SA-kNN model, it is observed that electrostatic descriptor E_797 with positive value is near to mercapto group indicating the electropositive groups were favorable. Compounds 2 and 3 with higher activity having electropositive substitution (CH₃, C₂H₅) on the mercapto group strongly support the above statement. The electropositive groups like propyl, isopropyl, butyl, and n-butane at the generated data point position E_797 around mercapto moiety will maximize the hCA XII inhibitory activity. Similar observations were seen with E_459 having positive values at the 6th position of the mercapto-quinazolinone ring. Compound 29 having nearby E_459 and E_797 descriptors show better activity compared to compound 34.

The presence of steric descriptor S_513 with a negative value at the 8th position of the mercapto-quinazoline ring indicates less steric or less bulky substituents will be favorable for human carbonic anhydrase XII inhibition activity. The above results are in close agreement with the experimental observations where compounds 24, 28, and 29 show high activity values. Therefore, steric substituents such as F, Cl, Br, and unsubstituted or less bulky groups like H, CH₃ were preferred at the 8th position of generated data point S_513. Similarly, S_1150 with a negative value near to benzene sulfonamide group indicates less steric or less bulky substituents will be favorable. S_265 with a negative value near to mercapto group indicates the necessity of less steric or less bulky groups. The simulated annealing-kNN model field plot (Fig.-1) along with the corresponding steric field range show that the ranges are more toward the negative side, meaning, decreasing the steric bulk of the substituent group is favorable at the respective substitution site. The molecular descriptor values of the dataset compounds obtained from the SA-kNN MFA model are given in supplementary Table-2.

**Molecular Docking Analysis**

The best active compounds of the dataset were subjected to molecular docking studies for binding affinity and H-bond interactions prediction. Compound 1 displayed favorable hydrogen bonds with the residues Asn67, Gln92, His119 and Thr199. Further Pi-Pi stacking with Leu198, Thr200, His94 and Ala65 have
added stability to the molecule in the binding pocket. The best energy conformation of compound 1 interacting with the hCA XII enzyme (pdb id: 6ebe) residues was displayed in Fig.-2 (Top left image). Apart from the nitrogen of sulfonamide bonding with Zn301 in all docked compounds, other residue interactions along with bond distances were presented in Table-3. The best confirmation from Autodock resulted in a good binding affinity of compound 1 with 6ebe as -8.6 K Cal / mol.

Table-3: Molecular Interactions and Binding Energy Values of Docked Complexes

| Compound | Binding Energy (Kcal/mol) | Hydrogen Bond (Distance Å) | Other Residue Interactions (Hydrophobic pi-pi stacking) |
|----------|--------------------------|-----------------------------|------------------------------------------------------|
| 1        | -8.6                     | Asn67 (2.86) with carbonyl oxygen of quinazolinone ring, Gln92 (1.96) with carbonyl oxygen of quinazolinone ring, His119 (2.07) with oxygen of sulfonamide group, Thr199 (1.95) with oxygen of sulfonamide group, Thr199 (2.29) with nitrogen of sulfonamide group | Leu198, Thr200 His94, Ala65, Val121 |
| 21       | -8.8                     | Asn67 (3.01) with carbonyl oxygen of quinazolinone ring, Gln92 (1.81) with carbonyl oxygen of quinazolinone ring, His119 (2.06) with | Phe95, Leu198, Thr200, His94, Ala65 |

Supplementary Table-2: Descriptor Values utilized in SA-kNN MFA 3D-QSAR Model

| Compound | pKi | S_1150 | S_265 | E_797 | E_459 | S_513 |
|----------|-----|--------|-------|-------|-------|-------|
| 1        | 9.229 | -0.034 | 1.976 | 0.156 | 0      | -0.056 |
| 2        | 8.509 | -0.051 | 4.148 | 0.145 | 0      | -0.056 |
| 3        | 8.409 | -0.061 | 4.888 | 0.141 | 0      | -0.056 |
| 4        | 8.066 | -0.065 | 4.907 | 0.302 | 0      | -0.056 |
| 5        | 7.416 | -0.074 | -0.6  | 0.152 | -0.001 | -0.056 |
| 6        | 7.755 | -0.083 | 1.505 | 0.228 | -0.001 | -0.056 |
| 7        | 7.764 | -0.081 | 2.058 | 0.226 | -0.001 | -0.056 |
| 8        | 7.55  | -0.07  | -0.039 | 0.143 | -0.001 | -0.056 |
| 9        | 7.592 | -0.085 | 2.673 | 0.25  | -0.001 | -0.056 |
| 10       | 7.695 | -0.047 | 4.958 | 0.116 | -0.001 | -0.057 |
| 11       | 7.886 | -0.069 | -2.084 | 0.424 | 0      | -0.063 |
| 12       | 7.967 | -0.045 | -1.931 | 0.2   | -0.001 | -0.058 |
| 13       | 8.041 | -0.048 | -2.517 | 0.219 | -0.001 | -0.058 |
| 14       | 8.62  | -0.084 | 4.796 | 0.222 | -0.001 | -0.056 |
| 15       | 7.416 | -0.04  | 0.057 | 0.182 | -0.001 | -0.058 |
| 16       | 7.5   | -0.048 | -1.931 | 0.221 | -0.001 | -0.057 |
| 17       | 8.06  | -0.097 | 6.232 | 0.264 | -0.001 | -0.056 |
| 18       | 7.345 | -0.074 | -1.65 | 0.404 | -0.001 | -0.062 |
| 19       | 7.648 | -0.098 | 5.67  | 0.252 | -0.001 | -0.056 |
| 20       | 7.772 | -0.066 | -0.747 | 0.486 | 0      | -0.067 |
| 21       | 9.208 | -0.035 | 2.372 | 0.002 | 0      | -0.056 |
| 22       | 9.194 | -0.037 | 2.256 | 0.231 | 0      | -0.056 |
| 23       | 9.268 | -0.035 | 2.201 | 0.073 | 0      | -0.056 |
| 24       | 9.181 | -0.043 | 1.359 | 0.297 | 0      | -0.056 |
| 25       | 8.839 | -0.035 | 2.103 | 0.196 | 0      | -0.056 |
| 26       | 9.119 | -0.034 | 2.146 | 0.098 | 0      | -0.055 |
| 27       | 8.257 | -0.036 | 2.001 | 0.252 | 0      | -0.06 |
| 28       | 9.215 | -0.062 | 2.057 | 0.141 | 0      | -0.056 |
| 29       | 9.137 | -0.062 | 2.083 | 0.134 | 0      | -0.056 |
| 30       | 9.268 | -0.04  | 0.712 | 0.433 | 0      | -0.052 |
| 31       | 9.328 | -0.055 | 2.499 | -0.143 | 0      | -0.06 |
| 32       | 8.241 | -0.134 | 3.192 | 0.27  | 0      | -0.056 |
| 33       | 9.187 | -0.05  | 2.739 | -0.607 | 0      | -0.056 |
| 34       | 8.148 | -0.072 | 1.86  | 0.032 | -0.001 | -0.057 |
### Table: Interactions of Mercaptoquinazolinone Benzene Sulfonamide Derivatives with 6ebe Protein

| Substitution | Interaction Details | Protein Residues |
|--------------|---------------------|------------------|
| 23           | His119 (2.9) with oxygen of sulfonamide group, His119 (2.16) with nitrogen of sulfonamide group, Pro202 (2.29) with nitrogen of quinazolinone ring | Trp209, Phe131, Leu198, Trp5, His64 |
| 28           | Gln92 (1.79) with carbonyl oxygen of quinazolinone ring, His119 (2.12) with oxygen of sulfonamide group, His119 (2.39) with nitrogen of sulfonamide group, Thr199 (2.09) with oxygen of sulfonamide group, Thr199 (2.76) with nitrogen of sulfonamide group | Leu198, Ile91, His94, Val121 |
| 30           | Gln69 (3.62) with oxygen of methoxy group, His119 (2.14) with oxygen of sulfonamide group | Leu198, Leu60, His94, Trp5, Val121 |
| 31           | Asn67 (2.98) with carbonyl oxygen of quinazolinone ring, Gln92 (1.70) with carbonyl oxygen of quinazolinone ring, His119 (1.96) with oxygen of sulfonamide group, His119 (2.75) with nitrogen of sulfonamide group, Thr199 (2.21) with nitrogen of sulfonamide group, Thr199 (2.80) with oxygen of sulfonamide group, Leu198 (3.06) with phenyl ring, Asn244 (3.31) with methyl of methoxy group | Leu198, Thr200, His94, Ala65, Tyr7, His96, Val121 |
| Co-crys ligand | Thr200 (1.96) with oxygen of sulfonamide group, Thr199 (2.14) with oxygen of sulfonamide group, Asn62 (2.37) with oxygen of nitro group, His94 (3.61) with oxygen of nitro group, Asn67 (1.90) with oxygen of nitro group, Asn67 (2.31) with hydroxyl group attached to benzene sulfonamide, Glu69 (3.06) with fluorine atom, Ile91 (3.51) with fluorine atom | Phe131, Leu198 both interact with rings by Pi stacking |

Substitutions on mercaptoquinazolinone ring displayed different conformations and interactions which were in conjunction with QSAR model predictions. H-bond interactions were predicted in compound 21 with the residues Asn67, Gln92, His119, His94, His96 and Thr199 of 6ebe protein (Fig.-2 Top right image). Also, Pi-Pi interactions were observed with Phe95, Leu198, Thr200, His94 and Ala65. His94, His96 and Phe95 interacted with fluorine substitution at R2 position which should be noted complementary to predictions from QSAR analysis describing electropositive groups favourable at the position as shown by model descriptor E_459. In this context compound, 21 with a binding affinity of -8.8 Kcal / mol can be improved with electropositive substitution at R2 position.

This electron-withdrawing group at one end of the molecule may be responsible for zinc301 ‘bonding distance’ with sulfonamide nitrogen, which was found to be 2.31 Å as shown in Fig.-3 (Top right image). On the other hand, compound 31, the best active compound of the dataset with pKi value of 9.3279 has shown a bonding distance of 2.73 Å between zn301 and sulfonamide nitrogen (Fig.-3 Bottom left image) due to the presence of hydrogen atoms at R2 position. Moreover, a number of hydrogen bonds present in the compound31-6ebe complex also matter with favorable interactions and binding energy. H-bond interactions were predicted in compound 31 with the 6ebe residues Asn67, Gln92, His119, Leu198, Asn244 and Thr199 (Fig. 2 Bottom left image). Also, Pi-Pi interactions were observed with Val121, Tyr7, His96, Leu198, Thr200, His94 and Ala65. Asn244 interaction with methoxy group at R4 position favored bioactivity which was predicted by QSAR model descriptor S_513, which symbolizes required less steric groups at the position.

Compound 28 displayed H-bond interactions with residues Thr199, His119 and Gln92. Interaction of quinazolinone ring oxygen atom with Gln92 residue at a bonding distance of 1.79 Å is notable due to variations in interaction distance in other compounds. For instance in compound 31 the Gln92-quinazolinone oxygen bonding distance was found to be 1.70 Å, also compound 21 showed 1.80 Å and
compound 1 displayed 1.96 Å. It was noteworthy that substitution at R2 position will surely have an influence on carbonyl oxygen interaction with Gln92.

Fig.-2: 2D-interactions of the Compound 1 (Top Left), 21 (Top Right), 31 (Bottom Left) and Cocrystallized Ligand (Bottom Right) in the Active Site of hCA XII enzyme, 6EBE.

Meaning; favourable electropositive groups at R2 position will favour Gln92 interaction at a bioactive distance and thereby increase the potency of the compound. The above statement was again in relation to QSAR model prediction with E_459 descriptor. Gln92 bonding distance may be one reason for fluctuations in zn301 bonding distance with sulfonamide nitrogen. Observations in compounds 31, 21 and 28 clearly state lesser the Gln92 bonding distance with carbonyl oxygen of quinazolinone ring, better the zn301 bonding distance with sulfonamide nitrogen (Fig.-3).

Compounds 23 displayed interactions in a typical conformation by bonding with Pro202 and Phe131. Pro202 interacted with nitrogen of quinazolinone ring and Phe131 interacted with sulphur atom of mercapto group. It is noticeable that due to the presence of chlorine atom at R2 position there may be a conformation change which totally avoided Gln92 interaction. Similar observations were seen with compound 30 in which Glu69 interacted with methoxy group at R2 position and Trp5 interacted with mercapto sulphur atom. Supplementary Fig.-1 displays bonding interactions of compound 23 (left image) and 30 (right image).
**CONCLUSION**

The 3D-QSAR study on a set of mercapto-quinazolinone based benzene sulfonamides highlights the importance of the structural features responsible for the human carbonic anhydrase XII inhibitory activity. Simulated annealing and the genetic algorithm have been applied as variable selection methods and the models were developed by the kNN–MFA and MLR methods. The reliability of the models was confirmed by validated statistical analyses. The findings of 3D QSAR studies provided an overall substitution pattern (electrostatic and steric fields) required around the mercapto-quinazolinone ring for hCA XII inhibitory activity. The 6th and 8th position substitutions on the mercapto-quinazolinone ring were the key aspects of SA-kNN molecular field analysis. The visualization of the 3DQSAR model provides explicit indications for the design of better analogs. Simulated annealing method coupled with the machine learning algorithm method, kNN-MFA provided insights into the design of novel hCA XII inhibitors. Molecular docking studies of the dataset active compounds exposed structural information in conjunction with QSAR model predicted descriptors E_459, S_513 and E_797. Thus the current study revealed useful insights for improving selectivity and bioactivity of human carbonic anhydrase XII inhibition. These understandings will certainly help researchers in developing selective inhibitors for hypoxia dependent hCA XII enzyme.
ACKNOWLEDGEMENT

We thank the Research Council of SRMIST for their support.

REFERENCES

1. A. Nocentini and C. T. Supuran, Expert Opinion on Therapeutic Patents, 28, 729(2018), https://doi.org/10.1080/13543776.2018.1508453
2. C. T. Supuran, Expert Opinion on Drug Metabolism & Toxicology, 3, 1(2016), https://doi.org/10.1517/17425255.2016.1154534
3. R. Gieling and K. Williams, Bioorganic & Medicinal Chemistry, 21, 1470(2013), https://doi.org/10.1016/j.bmc.2012.09.062
4. C. T. Supuran, Biochemical Journal, 473, 2023(2016), https://doi.org/10.1042/BCJ20160115
5. F. Pacchiano, F. Carta, P. C. McDonald, Y. Lou, D. Vullo, A. Scozzafava, S. Dedhar, C. T. Supuran, Journal of Medicinal Chemistry, 54, 1896(2011), https://doi.org/10.1021/jm101541x
6. S. Angapelly, P. V. Sri Ramya, A. Angeli, C. T. Supuran, M. Arifuddin, ChemMedChem, 12, 1578(2017), https://doi.org/10.1002/chem.201700446
7. M. Bozdag, A. M. Alafeefy, F. Carta, M. Ceruso, A. S. Al-Tamimi, A. A. Al-Kahtani, F. A. S. Alasmary, C. T. Supuran, Bioorganic & Medicinal Chemistry, 24, 4100–4107(2016), https://doi.org/10.1016/j.bmc.2016.06.052
8. A. S. El-Azab, A. A. Abdel-Aziz, H. A. Ghabbour, M. A. Al-Gendy, Journal of Enzyme Inhibition and Medicinal Chemistry, 32, 1229 (2017), https://doi.org/10.1080/14756366.2017.1368504
9. M. Bozdag, A. M. Alafeefy, D. Vullo, F. Carta, N. Dedeoglu, A. M. Al-Tamimi, N. A. Al-Jaber, A. Scozzafava, C. T. Supuran, Bioorganic & Medicinal Chemistry, 23, 7751 (2015), https://doi.org/10.1016/j.bmc.2015.11.023
10. M. Bozdag, A. M. Alafeefy, A. M. Altamimi, F. Carta, C. T. Supuran, D. Vullo, Bioorganic & Medicinal Chemistry, 25, 2782(2017), https://doi.org/10.1016/2017.03.054
11. P. Gopinath, M. K. Kathiravan, Journal of Chemometrics, 33, 1(2019), https://doi.org/10.1002/cem.3189
12. VLife MDS V4.6. Molecular design suite, VLife Sciences Technologies Pvt. Ltd., Pune, India, 2020, Software available at http://www.vlifesciences.com/
13. A. S. El-Azab, A. A. Abdel-Aziz, S. Bua, A. Nocentini, N. A. AlSaif, M. M. Alanazi, M. A. ElGendy, H. E. A. Ahmed, C. T. Supuran, Journal of Enzyme Inhibition and Medicinal Chemistry, 35, 733(2020), https://doi.org/10.1080/14756366.2020.1742117
14. S. El-Azab, A. A. Abdel-Aziz, H. E. A. Ahmed, S. Bua, A. Nocentini, N. A. AlSaif, A. J. Obaidullah, M. M. Hefnawy, C. T. Supuran, Journal of Enzyme Inhibition and Medicinal Chemistry, 35, 598(2020), https://doi.org/10.1080/14756366.2020.1722121
15. M. Bozdag, A. M. Alafeefy, F. Carta, M. Ceruso, A. S. Al-Tamimi, A. A. Al-Kahtani, F. A. S. Alasmary, C. T. Supuran, Bioorganic & Medicinal Chemistry, 24, 4100(2016), https://doi.org/10.1016/j.bmc.2016.06.052
16. K. S. Bhadoriya, N. K. Kumawat, S. V. Bhavthankar, M. H. Avchar, D. M. Dhumal, S. D. Patil, S. V. Jain, Journal of Saudi Chemical Society, 20, S256(2016), https://doi.org/10.1016/j.jscs.2012.11.001
17. K. S. Bhadoriya, C. Mukesh Sharma, S. Sharma, S. V. Jain, M. H. Avchar, Arabian Journal of Chemistry, 7, 924(2014), https://doi.org/10.1016/j.arabjchem.2013.02.002
18. M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, G. R. Hutchinson, Journal of Cheminformatics, 4, 17(2012), https://doi.org/10.1186/1758-2946-4-17
19. G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, Journal of Computational Chemistry, 30, 2785(2009), https://doi.org/10.1002/jcc.21256
20. B. Webb and A. Sali, Current Protocols in Bioinformatics, 54, 56.1(2016), https://doi.org/10.1002/cpbi.3
21. Discovery Studio Visualizer Version 20.1.0.19295, San Diego Dassault Systèmes: BIOVIA, 2020.
22. J. L. Velázquez-Libera, F. Durán-Verdugo, A. Valdés-Jiménez, G. Núñez-Vivanco, J. Caballero, Bioinformatics, 36, 2912(2020), https://doi.org/10.1093/bioinformatics/btaa018

[RJC-6767/2021]