J Mol Model (2010) 16:951–964
DOI 10.1007/s00894-009-0596-9

ORIGINAL PAPER

Docking and 3D-QSAR studies of acetohydroxy acid synthase inhibitor sulfonylurea derivatives

Kunal Roy · Somnath Paul

Received: 1 August 2009 / Accepted: 14 September 2009 / Published online: 20 October 2009

Abstract Docking and three dimensional quantitative-structure activity relationship (3D-QSAR) studies were performed on acetohydroxy acid synthase (AHAS) inhibitor sulfonylurea analogues with potential herbicidal activity. The 3D-QSAR studies were carried out using shape, spatial and electronic descriptors along with a few structural parameters. Genetic function approximation (GFA) was used as the chemometric tool for this analysis. The whole data set (n=45) was divided into a training set (75% of the data set) and a test set (remaining 25%) on the basis of the K-means clustering technique on a standardised topological, physicochemical and structural descriptor matrix. Models developed from the training set were used to predict the activity of the test set compounds. All models were validated internally, externally and using the Y-randomisation technique. Docking studies suggested that the molecules bind within a pocket of the enzyme formed by some important amino acid residues (Met351, Asp375, Arg377, Gly509, Met570 and Val571). In QSAR studies, molecular shape analysis showed that bulky substitution at the R_1 position may enhance AHAS inhibitory activity. Charged surface area descriptors suggested that negative charge distributed over a large surface area may enhance this activity. The hydrogen bond acceptor parameter supported the charged surface area descriptors and suggested that, for better activity, the number of electronegative atoms present in the molecule should be high. The spatial descriptors show that, for better activity, the molecules should possess a bulky substituent and a small substitution at the R_2 and R_3 positions, respectively.

K. Roy (✉) · S. Paul
Drug Theoretics and Cheminformatics Laboratory,
Department of Pharmaceutical Technology, Jadavpur University,
Kolkata 700032, India
e-mail: kunalroy_in@yahoo.com

Keywords Quantitative-structure activity relationship · Genetic function approximation · K-means cluster · Docking · Acetohydroxy acid synthase · Sulfonylurea

Introduction

Agriculture has played a crucial role in the development of human civilisation. It is widely believed that domestication of plants allowed humans to settle in a place and give up their previous migrant lifestyle. Until the industrial revolution, the majority of the human population laboured in agriculture. Development of agricultural techniques and the constituents of fertilisers have steadily increased agricultural productivity. However, this productivity may be greatly diminished under threats from certain agents such as weeds, fungus, pests, insects, etc., of which weeds have played a key role. Weeds are generally considered as unwanted plants in human-made settings like agricultural areas, gardens, etc. because they can (1) restrict the amount of light available to the desirable plants, (2) take nutrients from the soil leaving the desired plant unfed and making it less productive, (3) spread plant pathogens that infect and diminish the quality of the crop [1]. Thus, control measures are required to protect agricultural products from the above-mentioned harmful threats either chemically or biologically (genetically). Genetic control is very complicated and expensive, making chemical control by herbicides the method of choice.

Herbicides have been broadly classified into two groups according to their activity: (1) contact herbicides—destroy only the plant tissue in contact with the chemical; and (2) systemic herbicides—chemicals translocated through the plant circulation system, either from foliar application down to the roots, or from soil application up to the leaves [2].

Rational design strategies, especially in silico-based approaches, have emerged as a promising alternative or complementary tool towards the effective screening of potential chemicals. In silico approaches include, for example, quantitative structure-activity relationship (QSAR) modelling techniques, which are increasingly attracting the attention of scientists in the chemical as well as the pharmaceutical industries [3–8]. Computer-aided chemical design has been applied extensively in the area of modern drug discovery, ecotoxicological modelling and design of agrochemicals for its high efficiency in the design of new compounds and optimisation of lead compounds, thus saving both time and economic costs in large-scale experimental synthesis and biological tests [9]. QSAR helps us to understand structure–activity relationships (SAR) in a quantitative manner and is one of the most important applications of chemometrics, providing information useful for the design of new compounds acting on specific targets. QSAR attempts to find a consistent relationship between biological activity or toxicity and molecular properties. Thus, QSAR models can be used to predict the activity of new compounds.

QSAR models have been reported by different groups of researchers for agrochemicals such as herbicides, fungicides and insecticides. Peng et al. [10] performed molecular docking and three-dimensional (3D) QSAR studies on the herbicidal activity of 1-(substituted phenoxyacetoxy) alkylphosphonates that bind to the E1 component of pyruvate dehydrogenase. Zhang et al. [11] developed a density functional theory (DFT)-based QSAR study of protoporphyrinogen oxidase inhibitors for the class of phenyl triazolinones. Duggleby et al. [12] performed comparative molecular field analysis (CoMFA) and comparativemolecular similarity indices analysis (CoMSiA) analyses of a new family of sulfonylurea herbicides. Yang et al. [13] performed docking studies of α-hydroxy-substituted 3-benzylidenepyrrolidene-2,4-dione derivatives against plant 4-hydroxy phenyl pyruvate dioxygenase enzyme, and have also reported classical QSAR.

Xi et al. [14] recently synthesised 45 sulfonylurea derivatives and also performed a DFT-based QSAR study of sulfonylurea analogues for herbicidal activity using general quantum chemical descriptors. Sulfonylurea herbicides inhibit the plant enzyme acetohydroxy acid synthase (AHAS, E.C. 2.2.1.6). AHAS belongs to the family of enzymes using the co-factor thiamin diphosphate (ThDP). AHAS—the key enzyme in branched-chain amino acid metabolism in plants, fungi, bacteria, and archaea—catalyses the condensation of pyruvate with either another pyruvate molecule or, alternatively, 2-ketobutyrate, yielding acetolactate or acetohydroxybutyrate as precursors in the biosynthesis of valine, leucine, and isoleucine. A number of different studies have collectively [15] suggested that the herbicide sulfonylurea binds to the AHAS enzyme near thiamine pyrophosphate and flavin adenine dinucleotide (FAD), with its binding site overlapping that of the pyruvate (substrate for biosynthesis of amino acids) binding domain. Sulfonylurea herbicides must be accommodated on AHAS at a site distinct from that of the substrate or co-factor binding sites. In 2006, Duggleby et al. [16] reported the crystal structure of AHAS from *Arabidopsis thaliana* complexed with the sulfonylurea herbicide metsulfuron methyl.

In this paper, we report docking of 45 sulfonylurea derivatives into the AHAS enzyme from *A. thaliana* to explore important interactions between the ligands and the active site of the AHAS enzyme. Furthermore, we performed 3D QSAR analysis to obtain a clearer insight into the SAR of this class of compounds.

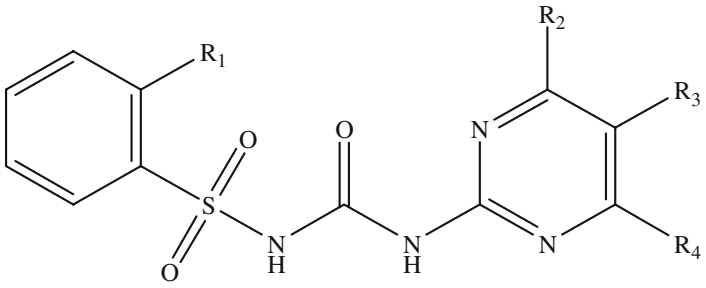
Materials and methods

In the present study, 45 sulfonylurea derivatives were employed, including 2 commercial herbicides (compounds **12**, **17**) and 43 derivatives [14]. Structures and experimental inhibitory activities of the compounds against wild-type *Escherichia coli* AHAS isoenzyme II [14] are listed in Table 1. There are four regions (R_1, R_2, R_3 and R_4) of structural variations in these compounds. The present QSAR study explores the impact on AHAS inhibitory activity of substitutional variations at these four positions. The range of AHAS inhibitory activity values is quite wide (4.93 log units).

Docking study

Molecular docking is an application wherein molecular modelling techniques are used to predict how a protein (enzyme) interacts with small molecules (ligands) [17]. The ability of a protein (enzyme) to interact with small molecules plays a major role in the dynamics of that protein, which may enhance/inhibit its biological function. In the current study, we performed docking of 45 sulfonylurea derivatives into the active site of the enzyme AHAS. The crystal structure of AHAS (E.C. 2.2.1.6, 1YHY.pdb) was obtained from the RCSB protein data bank (http://www.pdb.org). We have worked on the AHAS enzyme complexed with metsulfuron methyl isolated from *A. thaliana* by Duggleby et al. [16]. The AHAS enzyme has been co-crystallised with flavin adenine dinucleotide (FAD). The 667-amino-acid AHAS enzyme is complexed with a sulfonylurea herbicide, metsulfuron methyl. We performed the docking studies using LigandFit in the Receptor-Ligand Interactions protocol section of Discovery Studio 2.0 [18]. An initial pretreatment for both ligands and the AHAS enzyme was performed. For ligand preparation, all duplicate structures were removed and options for ionisation change, tautomer generation, isomer generation, Lipinski filter and 3D generator were set true. For enzyme

Table 1 Structural features and acetohydroxy acid synthase (AHAS) inhibitory activity (observed and calculated) of 45 sulfonylurea derivatives



Compound	Substitutions				Activity		
					Observed [14]	Calculated	
	R_1	R_2	R_3	R_4		M1[b]	M2[b]
1	Cl	CH_3	H	H	5.22	4.85	5.13
2	Cl	OCH_3	H	OCH_3	6.93	6.52	6.66
3	Cl	OCH_3	Br	OCH_3	5.52	5.47	6.06
4	Cl	OCH_3	Br	Cl	4.78	4.99	4.16
5[a]	Cl	CH_3	Br	CH_3	4.67	3.97	4.36
6[a]	Cl	$CH_2OC_2H_5$	Br	H	4.27	4.78	4.17
7	Cl	CH_2OOCCH_3	Br	H	4.7	5.08	5.56
8	COOEt	CH_3	H	H	6.69	6.90	6.29
9	COOEt	$CH_2OC_2H_5$	Br	H	6.51	6.92	6.69
10	COOEt	OCH_3	Br	OCH_3	7.03	7.26	6.64
11	COOEt	OCH_3	H	OCH_3	9.2	9.12	8.82
12	COOEt	OCH_3	H	Cl	8.34	8.05	7.66
13	COOEt	CH_3	H	Cl	7.09	7.51	7.58
14	COOEt	OCH_3	Br	Cl	6.8	6.78	5.84
15	COOEt	$CH_2OOCCH=CH_2$	Br	H	7.11	6.74	6.67
16	COOMe	OCH_3	Br	OCH_3	6.31	6.76	6.31
17	COOMe	CH_3	H	CH_3	7.69	6.83	7.38
18	COOMe	OCH_3	H	OCH_3	8.63	8.67	8.48
19	COOMe	OCH_3	H	Cl	7.77	7.88	7.38
20[a]	COOMe	$CH_2OCH_2CH=CH_2$	Br	H	6.04	5.62	6.53
21	COOMe	CH_2OOCCH_3	Br	H	5.2	5.30	5.54
22	COOMe	OCH_3	H	H	6.66	7.41	7.60
23[a]	NO_2	$CH_2OC_2H_5$	Br	H	4.68	5.58	4.99
24[a]	NO_2	CH_3	Br	CH_3	4.8	4.74	5.14
25	NO_2	OCH_3	Br	OCH_3	5.3	5.61	5.74
26[a]	NO_2	OCH_3	H	OCH_3	8	7.46	7.97
27[a]	NO_2	OCH_3	H	H	5.82	6.49	7.25
28	NO_2	$CH_2OOCCH=CH_2$	Br	H	5.17	5.39	5.30
29[a]	OCH_2CH_2Cl	OCH_3	H	OCH_3	7.46	7.14	8.05
30	OCH_2CH_2Cl	CH_3	H	Cl	6.58	6.06	7.48
31	OCH_2CH_2Cl	CH_2SCN	Br	H	5.58	5.01	5.45
32	OCH_2CH_2Cl	$CH_2OC_2H_5$	Br	H	5.14	5.67	5.73
33	OCH_2CH_2Cl	$CH_2OCH_2CH=CH_2$	Br	H	5.31	5.16	5.75
34	OCH_2CH_2Cl	OCH_3	Br	OCH_3	5.45	5.37	5.65
35	OCH_2CH_2Cl	OCH_3	Br	Cl	4.77	4.66	4.18
36	Cl	CH_3	H	CH_3	4.71	5.13	4.65
37[a]	Cl	CH_2OCH_3	Br	H	5.66	4.61	6.03

Table 1 (continued)

Compound	Substitutions				Activity		
					Observed [14]	Calculated	
38	Cl	CH_2SCN	Br	H	4.63	4.75	4.79
39	COOEt	CH_2OCH_3	Br	H	6.7	6.28	6.96
40	COOEt	CH_3	Br	CH_3	6.78	6.39	6.35
41	COOMe	CH_3	Br	CH_3	5.94	5.73	6.17
42[a]	NO_2	CH_2OCH_3	Br	H	5.54	5.04	4.79
43	NO_2	OCH_3	H	Cl	7.69	6.92	6.92
44[a]	NO_2	CH_2OOCCH_3	Br	H	4.53	5.12	4.76
45	OCH_2CH_2Cl	CH_3	H	H	5.33	6.08	5.71

[a] Test set members

[b] See Eqs. M1 and M2 below

preparation, the whole enzyme (after preparing the ligand) was selected and hydrogen atoms were added to it. The pH of the protein was set within the range 6.5–8.5. We then defined the AHAS enzyme as the receptor and the active site was selected based on the ligand binding domain of metsulfuron methyl. The preexisting ligand (metsulfuron methyl) was then removed and freshly prepared ligand (sulfonylurea derivative) prepared by us was used for docking. LigandFit was then chosen from the Receptor–Ligand Interaction section of the program. We used preprocessed receptor and ligand as inputs. PLP1 was selected as the energy grid. The conformational search of ligand poses was performed by the Monte Carlo trial method. The torsional step size for polar hydrogen was set at five. Docking was performed taking electrostatic energy into account. The maximum internal energy was set at 10,000 cal. Pose saving and interaction filters were set as default. Fifty poses were docked for each compound. During the docking procedure, no attempt was made to minimise the ligand–enzyme complex (rigid docking). After completion of docking, the docked enzyme (protein–ligand complex) was analysed to investigate the type of interactions. The 50 docking poses saved for each compound were ranked according to their dock score function. The pose (conformation) having the highest dock score was selected for further analysis.

Descriptors for QSAR study

We also performed QSAR studies on the data set with three-dimensional (shape, spatial and electronic) descriptors along with a few structural descriptors. The categorical list [19] of descriptors used in the development of QSAR models is shown in Table 2.

Cluster analysis

The ultimate target of any QSAR modelling is that the developed model should be strong enough to be capable of making accurate and reliable predictions of biological activities of new compounds. The models were cross-validated using the leave-one-out (LOO) method. However, internal validation does not guarantee that the model will

Table 2 Categorical list of descriptors used in the development of quantitative structure-activity relationship (QSAR) models

Category of descriptors	Name of the Descriptors
Shape	DIFFV, COSV, F_o, NCOSV, ShapeRMS
Electronic	Dipole-mag, Sr
Spatial	RadOfGyration, Jurs_SASA, Jurs_PPSA_1, Jurs_PNSA_1, Jurs_DPSA_1, Jurs_PPSA_2, Jurs_PNSA_2, Jurs_DPSA_2, Jurs_PPSA_3, Jurs_PNSA_3, Jurs_DPSA_3, Jurs_FPSA_1, Jurs_FNSA_1, Jurs_FPSA_2, Jurs_FNSA_2, Jurs_FPSA_3, Jurs_FNSA_3, Jurs_WPSA_1, Jurs_WNSA_1, Jurs_WPSA_2, Jurs_WNSA_2, Jurs_WPSA_3, Jurs_WNSA_3, Jurs_RPCG, Jurs_RNCG, Jurs_RPCS, Jurs_RNCS, Jurs_TPSA, Jurs_TASA, Jurs_RPSA, Jurs_RASA, Shadow_XY, Shadow_XZ, Shadow_YZ, Shadow_XYfrac, Shadow_XZfrac, Shadow_YZfrac, Shadow_nu, Shadow_Xlength, Shadow_Ylength, Shadow_Zlength, Area, Vm, Density, PMI_mag
Structural	Rotlbonds, Hbond acceptor, Hbond donor

perform well on a new set of data. In most cases, an appropriate external data set is not available for prediction purposes. Hence, the whole data set is divided into a training set and a test, or external evaluation, set. In the present study, the models developed from the training set (subset of the original set) were validated externally using a test set. The predictive capacity of a model for new chemical entities is influenced by the chemical nature of the training set molecules used for development of the model [20–22]. In reality, the test set molecules will be predicted well when these molecules are structurally very similar to the training set molecules, the reason being that the model has considered all features common to the training set molecules. There are different techniques available for dividing the data set into training and test sets, e.g. statistical molecular design, self-organising map, clustering, Kennard–Stone selection, sphere exclusion, etc. [23]. In the present case, we used the clustering technique as the method for training set selection. Cluster analysis [24] is a technique whereby objects can be arranged into groups.

In the present work, the total data set (n=45) was divided into training set (n=34) and test (external evaluation) set (n=11; 75% and 25%, respectively, of the total number of compounds) based on clusters obtained from K-means clustering [25] applied to a standardised topological, physicochemical and structural descriptor matrix. The whole data set was clustered into three subgroups, from each of which were selected approximately 25% of compounds as members of the test set. The serial numbers of compounds in different clusters are shown in Table 3.

Molecular shape analysis

Molecular shape analysis (MSA) was used as a 3D QSAR technique. In our study, the following steps were used to perform MSA [26]:

1) Conformational analysis. The first operation in MSA is the conformational analysis of the analogues. The conformers were generated with the "optimal search method" option followed by energy minimization.
2) Hypothesizing an active conformer. The aim of this step is to select a structure that is present in the rate-limiting step for the activity in a biological reaction. The minimum-energy conformer (global minimum) of the most active compound **11** was taken as the active conformer.
3) Selection of a candidate shape reference compound. The shape reference compound is the molecule that is used when shape descriptors are calculated. MSA compares all other molecules to the shape reference compound (global minimum of compound **11**) and provides information about each comparison.
4) Performing pair-wise molecular superposition. Each study molecule was aligned to the shape reference compound using the maximum common sub graph (MCSG) method to calculate the shape descriptors.
5) Measurement of molecular shape commonality. After alignment, various shape descriptors, based on relative shape similarity with the shape reference compound, were calculated for each study molecule.
6) Other molecular descriptors. Determination of other molecular features by calculating spatial, structural and electronic parameters was done in addition to the shape descriptors.
7) Construction of QSAR. QSAR equations were generated using genetic function approximation (GFA) with linear option as the statistical tool.

For comparison, we have performed another MSA using the docked conformations of the ligands instead of using global energy minimum conformers.

Table 3 Serial numbers of compounds under different clusters

Cluster number	Serial number of compounds
1	1,2,5,30,36,37,45
2	8,11,12,13,17,18,19,22,26,27,29,43
3	9,10,15,20,21,28,33,39,44
4	3,4,7,14,16,23,24,25,34,35,40,41,42
5	6,31,32,38

Genetic function approximation—multiple linear regression

Genetic algorithms are derived from an analogy with the evolution of DNA [27]. The GFA algorithm was initially anticipated by (1) Holland's genetic algorithm, and (2) Friedman's multivariate adaptive regression splines (MARS) algorithm. In this algorithm, a model is represented as a one-dimensional string of bits. A distinctive feature of GFA is that it produces a population of models (e.g. 100), instead of generating a single model as do most other statistical methods. Genetic algorithm methods lead to models superior to those developed using stepwise regression technique because they select the basis functions genetically. Descriptors selected by this algorithm were subjected to multiple linear regression for generation of models. A "fitness function" or lack of fit (LOF) is used to estimate the quality of an individual or model, so that the best individual or model receives the best fitness score. The error measurement term LOF is determined by the following equation:

$$LOF = \frac{LSE}{\left(1 - \frac{c + d^{*}p}{M}\right)^{2}} \quad (1)$$

In Eq. 1, c is the number of basis functions (other than constant term); d is a smoothing parameter (adjustable by the user); M is the number of samples in the training set; LSE is the least squares error and p is the total number of features contained in all basis functions.

Once models in the population have been rated using the LOF score, the genetic cross-over operation is repeatedly performed. Initially, two good models are probabilistically selected as parents and each parent is cut randomly into two pieces and a new model (child) is generated using a piece from each parent. After many mating steps, i.e. genetic crossover type operations, the average fitness of models in the population increases as good combinations of genes are discovered and spread through the population. This procedure can build not only linear models but also higher-order polynomials, splines and Gaussians. In the present work, linear terms have been used. For the development of GFA models, we used Cerius2 version 4.10 [19]. The mutation probabilities were kept at 50% with 5,000 iterations. Smoothness (d) was kept at 1.00. Initial equation length value was selected as four and the length of the final equation was not fixed.

Validation methods

The robustness of the models should be verified by using different types of validation criteria. For validation of QSAR models, four strategies [28] are usually adopted: (1) internal validation or cross-validation, (2) validation by dividing the data set into training and test compounds, (3) data randomisation or Y-scrambling, and (4) true external validation by application of the developed model to new external data. In the case of the present data set, due to the lack of a true external evaluation set, the total data set was divided into an internal evaluation (training) set and an external evaluation (test) set. Thus, we performed only the first three validation techniques. Most QSAR modelling methods implement the LOO or leave-many-out (LMO) cross-validation procedures, which are internal validation techniques. The outcome from the cross-validation procedure is cross-validated R^2 (LOO-Q^2 or LMO-Q^2), which is used as a criterion of both robustness and of the predictive ability of the model. In this paper, we used the LOO validation method as the internal validation tool. The cross-validated squared correlation coefficient R^2 (LOO-Q^2) is calculated according to this equation:

$$Q^2 = 1 - \frac{\sum \left(Y_{obs(training)} - Y_{pred(training)}\right)^2}{\sum \left(Y_{obs(training)} - \overline{Y}_{training}\right)^2} \quad (2)$$

In Eq. 2, $\overline{Y}_{\text{training}}$ represents the average activity value of the training set while $Y_{\text{obs(training)}}$ and $Y_{\text{pred(training)}}$ represents the observed and predicted activity values of training set compounds, respectively. Often, a high Q^2 value (Q^2>0.5) is considered as a proof of the high predictive ability of the model [29].

Models are generated based on training set compounds and predictive capacity of the models is judged based on the predictive R^2 ($R_{\text{pred}}{}^2$) values calculated according to the following equation [30]:

$$R^2_{pred} = 1 - \frac{\sum \left(Y_{obs(test)} - Y_{pred(test)}\right)^2}{\sum \left(Y_{obs(test)} - \overline{Y}_{training}\right)^2} \quad (3)$$

In Eq. 3, $Y_{\text{pred(test)}}$ and $Y_{\text{obs(test)}}$ indicate predicted and observed activity values, respectively, of the test set compounds and $\overline{Y}_{\text{training}}$ indicates the mean activity value of the training set compounds. The value of $R_{\text{pred}}{}^2$ for an acceptable model should be more than 0.5.

Further statistical significance of the relationship between AHAS inhibitory activity and the descriptors was obtained from randomisation (Y-randomisation) of the developed models. This test was carried out by repeatedly scrambling the activity values to generate QSAR models, and comparing the resulting scores with the score of the original QSAR model generated from non-randomised activity values. If the score of the non-random QSAR model is significantly better than that of the random models then that model should be considered as a statistically robust model [31].

Software

MINITAB [32] was used for linear regression and partial least squares methods. Cerius2 version 4.10 [19] was used for MSA and GFA analyses. SPSS [33] was used for cluster analysis and deriving the intercorrelation matrix of the descriptors. LigandFit of the receptor-ligand interactions protocol available under Discovery Studio 2.0 [18] was used to dock the inhibitor molecules into the active site of the enzyme AHAS.

Results and discussion

The results obtained from docking and QSAR studies are described below.

Docking

In the present study, to understand the interactions between AHAS and its inhibitors, and to explore their binding mode, a docking study was performed using the LigandFit function available in Discovery Studio 2.0 [18]. These docking studies yielded crucial information concerning the orientation of the inhibitors in the binding pocket of AHAS. Ligand–enzyme interaction analysis shows that Met351,

Asp375, Arg377, Gly509, Met570 and Val571 are the most important residues present at the active site and are the main contributors to the receptor–ligand interaction. It has been observed that, for better AHAS inhibitory activity, four amino acid residues (Met351, Arg377, Met570 and Val571) should optimally interact with the substituted sulfonylurea derivatives. In the case of compound **36** (Fig. 1), the four amino acid residues mentioned above are in close vicinity to the molecule. The methyl ($–CH_3$) substitution at the R_2 position and the adjacent nitrogen atom of the pyrimidine ring system form ten intramolecular bumps with the chlorine atom at the R_1 position and the hydrogen atoms present in the phenyl ring. These bumps may disturb the optimal position of the molecule in the pocket and thus hinder the interaction with the amino acid residues; thus the herbicidal activity of compound **36** is poor (4.71). In compound **11** (Fig. 2), two intramolecular bumps are present between ethyloxy carbonyl (–COOEt) at the R_1 position and the oxygen atom of the sulfonyl (O=S=O) group. However, three hydrogen bonds are formed between the amino acid Arg377 and the oxygen atom of the sulfonyl group, the oxygen atom of the methoxy group at the R_2 position and the nitrogen atom present in the close vicinity of that methoxy group, respectively. These hydrogen bonds help the molecule to fit into the cavity of the enzyme. The AHAS inhibitory activity (9.20) is very high due to these three hydrogen bonds. In compound **16** (Fig. 3), four intermolecular hydrogen bonds form between the inhibitor molecule and the amino acid residue Arg377. But the bromine atom present at the R_3 position forms one intramolecular bump with the hydrogen atom of the methoxy group at R_2 and three intermolecular bumps with the amino acid Met570. Although the hydrogen bonds help to enhance the binding potential, the bumps disfavour the binding capability and the herbicidal activity is only of intermediate level (6.31). In compound **24** (Fig. 4), Arg377 forms two hydrogen bonds as well as two bumps with the inhibitor molecule. The bromine atom present at the R_3 position has formed three bumps with the amino acid Met570. There are also four more intramolecular bumps which preclude the molecule obtaining optimum binding and thus compound **24** cannot fit well into the pocket of the AHAS enzyme. As a consequence, the herbicidal activity (4.80) of the molecule is very much reduced. In compound **18** (Fig. 5), the hydrogen atom of the methoxy group present at the R_4 position has made a bump with the important amino acid Val571. But, Arg377 has formed four hydrogen bonds with the electronegative atoms present in the inhibitor molecule and the other two important amino acid residues (Met351, Met570) are also present in the close vicinity of the molecule. The negative effect of one bump has been overcome by the favourable effect of four hydrogen bonds, which favour the fitting of the molecule in the cavity of the AHAS enzyme. Thus, the AHAS inhibitory activity (8.63) of compound **18** is high. In compound **6** (Fig. 6), the

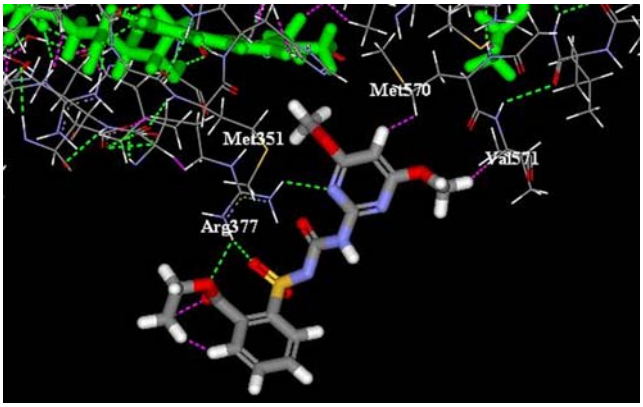


Fig. 2 Docked conformation of compound **11** showing important amino acid residues

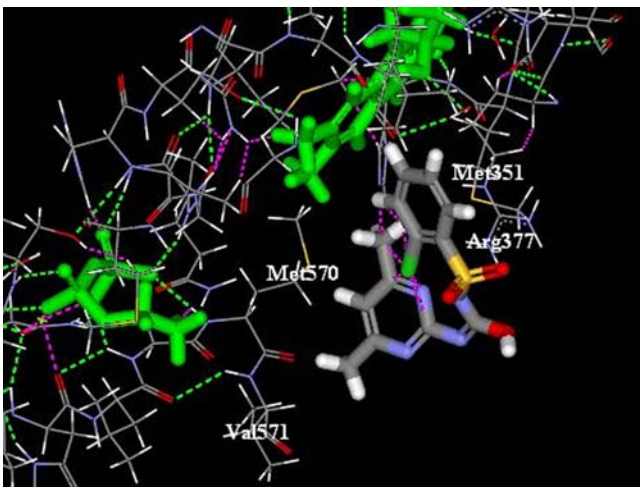


Fig. 1 Docked conformation of compound **36** showing important amino acid residues

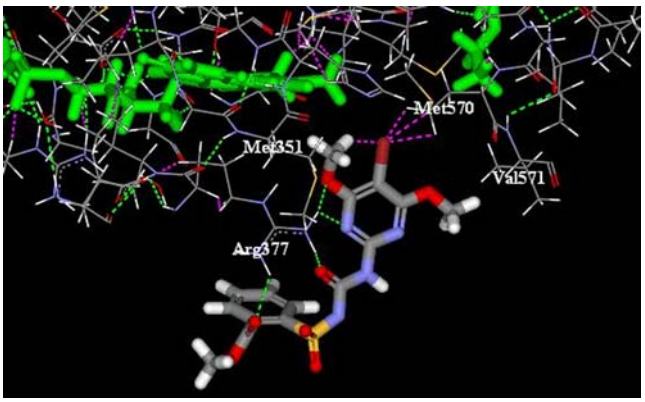


Fig. 3 Docked conformation of compound **16** showing important amino acid residues

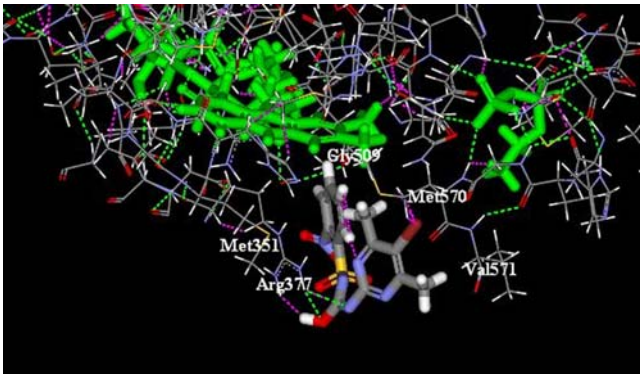


Fig. 4 Docked conformation of compound **24** showing important amino acid residues

substituent present at the R_2 position has formed a bump with the FAD co-enzyme present in the AHAS enzyme. Thus the inhibitor molecule cannot fit itself well into the pocket of the enzyme. Again, the activity of this inhibitor is low. This results also suggests that FAD plays a crucial role in the inhibitor–enzyme interaction.

Validation of the docking process

The ligand binding process was validated by docking the most active inhibitor molecule (compound **11**) with the enzyme AHAS using other two docking tools, i.e. LibDock and CDOCKER, available in the Receptor–Ligand Interactions section available under Discovery Studio 2.0. The docked geometries of the most active compound obtained using the LibDock and CDOCKER tools are shown in Figs. 7 and 8, respectively. The docked geometries obtained from the latter two docking tools were very similar to that obtained from the LigandFit tool of Discovery Studio 2.0. The important amino acids (Met351, Arg377, Met570 and Val571) in the docked geometries obtained with LibDock and CDOCKER interact with the ligand in a pattern very similar to that determined with LigandFit. This supports the

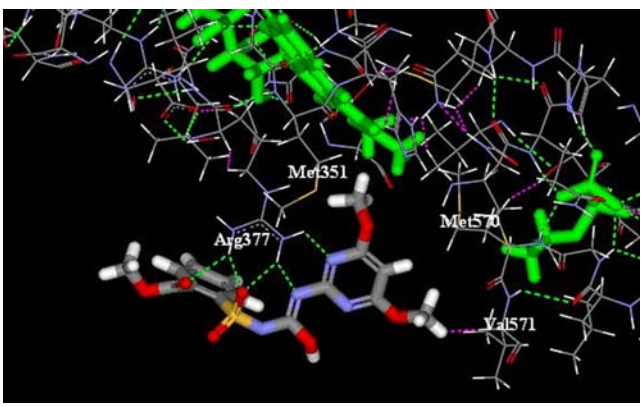


Fig. 5 Docked conformation of compound **18** showing important amino acid residues

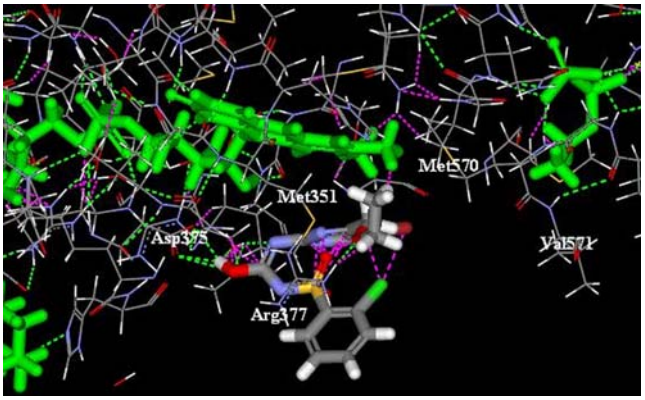


Fig. 6 Docked conformation of compound **6** showing important amino acid residues

view that our docking process is robust and reproducible. We performed another type of validation study to show the robustness of our docking process: we docked metsulfuron methyl (the original co-crystallised ligand) with the AHAS enzyme and then compared the docked geometry with that of the original crystal structure (E.C. 2.2.1.6, 1YHY.pdb) obtained from PDB. This analysis showed that the amino acid residues present close to the inhibitor molecule are the same as those in the enzyme-inhibitor complex in PDB, thus suggesting that our docking process is reproducible.

Molecular shape analysis

We performed 3D-QSAR to obtain information about the effect of shape, the spatial arrangement of atoms in 3D space, and the charge distribution of substituents on the biological activity. This study was conducted using MSA descriptors along with additional descriptors such as spatial and electronic parameters and a few structural descriptors. Figure 9 shows the aligned geometry of the training set compounds used in MSA.

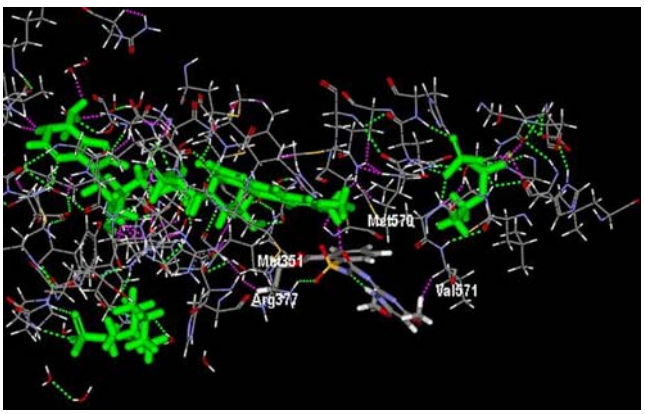


Fig. 7 Docked conformation of compound **11** (most active) showing the important amino acid residues (obtained using the LibDock algorithm)

Initially, we performed MSA by generating conformers using Cerius 2 version 4.10 software as detailed in Materials and methods. Models were generated with shape, spatial and electronic descriptors using GFA with spline option as the statistical tool. The mutation probability was kept at 50% with 5,000 iterations. In the case of GFA linear technique, the following best equation was obtained with acceptable LOO internal variance (Q^2) and external predicted variance ($R_{\text{pred}}{}^2$).

$$\begin{aligned} pK_i &= 5.494(\pm 1.099) - 0.003(\pm 0.0003)PMI\_mag + 0.031(\pm 0.004)DIFFV \\ &\quad -0.680(\pm 0.120) < 9 - HBondacceptor > +0.432(\pm 0.089)Shadow\_Xlength \\ n_{Training} &= 34, R^2 = 0.890, R_a^2 = 0.875, F = 58.7(df\,4, 29), Q^2 = 0.847, \\ PRESS &= 7.593, n_{Test} = 11, R_{pred}^2 = 0.785, r^2 = 0.718, r_0^2 = 0.713, r_m^2 = 0.667. \end{aligned} \quad \text{(M1)}$$

The above model could explain 87.5% of the variance (adjusted coefficient of variation). The LOO-predicted variance was found to be 84.7%. The predictive potential of this model was determined by predicted R^2 for the test set compounds, and was found to be 0.785. The squared correlation coefficient between the observed and predicted activity of the test set compounds was 0.718. The squared correlation coefficient between the observed and predicted activity of the test set compounds, setting intercept to zero, was 0.713.

Using the standardised variable matrix for regression, the significance level of the descriptors was found to be of the order: PMI_mag, DIFFV, <9-HBondacceptor> and Shadow_Xlength. PMI_mag is the moment of inertia, the resultant of the moment of inertia of three axes that are calculated for a series of straight lines through the centre of mass. These are associated with the principal axes of the ellipsoid. PMI_mag makes an unfavourable contribution towards herbicidal activity. This can be explained by compounds **9**–**11**. All these compounds have the same substituent (–COOEt) at the R_1 position. Compound **9** has ethoxy methyl ($–CH_2OC_2H_5$) and bromo (–Br) at the R_2 and R_3 positions, respectively. Thus the value of PMI_mag is high (due to the larger substituents) in compound **9**. This may reduce the herbicidal activity. In compounds **10** and **11** the R_1, R_2, and R_4 substituents are the same but compound **10** has a bromo (–Br) substitution at R_3 position whereas that position in compound **11** is unsubstituted. Thus the value of PMI_mag is lower in the case of compound **11** and it possesses greater herbicidal activity (9.20). This suggests that small substituents at the R_3 position enhance AHAS inhibitory activity.

DIFFV is the difference between the volume of the individual molecule and the volume of the shape reference compound. It makes a favourable contribution towards herbicidal activity, which is directly proportional to the volume of the compounds. This is illustrated by compounds **1** and **8**, which differ in only one substitutional feature at the R_1 position. Compound **1** has a chloro (–Cl) substitution at the R_1 position and compound **8** has an ethyloxy carbonyl (–COOEt) substitution in that location. Thus, the volume of compound **1** is lower than that of compound **8**. Thus compound **8** possesses higher AHAS inhibitory

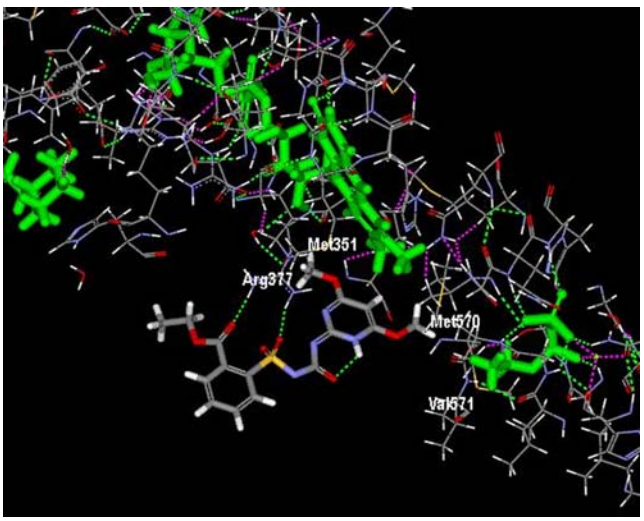


Fig. 8 Docked conformation of compound **11** (most active) showing the important amino acid residues (obtained using the CDOCKER algorithm)

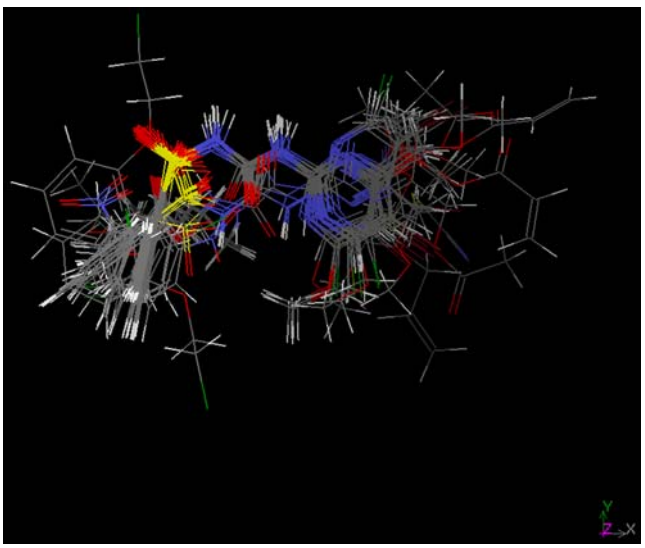


Fig. 9 Aligned geometry of the training set members used in molecular shape analysis (MSA)

activity. This suggests that, for better herbicidal activity, bulky substituents are required at the R_1 position.

HBondacceptor is the number of hydrogen bond acceptor atoms or groups present in a molecule. However, in the above model, it is present as a spline term and the negative coefficient of this term suggests that if the number of the hydrogen bond acceptors is less than 9, this has an unfavourable effect on AHAS inhibitory activity. An increase in the number of electronegative atoms may enhance the hydrogen bond acceptor count. This is illustrated by compounds **1** and **11**. In the case of compound **1**, there are fewer electronegative atoms than in compound **11**. Thus compound **1** possesses less herbicidal activity than compound **11**. The above parameter suggests that the presence of a large number of electronegative atoms in a molecule may enhance AHAS inhibitory activity.

Shadow_Xlength is the length of a molecule in the direction of the *x*-axis. This descriptor makes a favourable contribution toward activity as evidenced by the positive regression coefficient, and represents the length of the substituent present at the R_2 position. This can be illustrated by compounds **9** and **15**. All three substituents at the R_1, R_3 and R_4 positions are same in both the compounds. But the change in length of the substitution at the R_2 position may affect AHAS inhibitory activity. Compound **9** contains an ethoxymethyl ($–CH_2OEt$) group at the R_2 position and compound **15** contains a vinylcarbonyloxymethyl group ($–CH_2OOCCH{=}CH_2$) at this position. Thus the value of Shadow_Xlength is higher in the case of compound **15**. As a consequence, the herbicidal activity of compound **15** is greater than that of compound **9**.

We further performed MSA using docked conformers obtained from Discovery Studio 2.0. Figure 10 shows the aligned geometry of docked conformers of the training set compounds used in MSA. The model was generated with shape, spatial and electronic descriptors using GFA with spline option as the statistical tool. The mutation probability was kept at 50% with 5,000 iterations. In the case of GFA linear technique, the following equation was obtained with acceptable LOO internal variance (Q^2) and external predicted variance ($R_{\text{pred}}{}^2$).

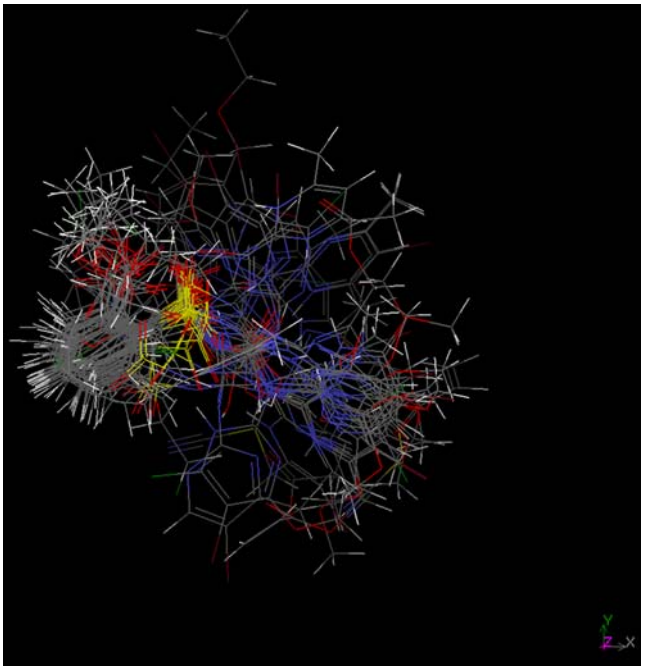


Fig. 10 Aligned geometry of docked conformers of the training set members used in MSA

$$
\begin{aligned}
pK_i = {} & 7.361(\pm 4.648) + 2.153(\pm 0.592)RadofGyration - 0.002(\pm 0.001)PMI\_mag \\
& -5.758(\pm 1.984)Density - 0.006(\pm 0.001) < \text{Jurs\_WNSA\_2} + 728.635 > \\
n_{Training} = {} & 34, R^2 = 0.837, R_a^2 = 0.815, F = 37.3(df\,4, 29), Q^2 = 0.776, \\
PRESS = {} & 11.119, n_{Test} = 11, R_{pred}^2 = 0.813, r^2 = 0.853, r_0^2 = 0.812, r_m^2 = 0.682.
\end{aligned}
\tag{M2}
$$

Using the standardised variable matrix for regression, the significance level of the descriptors was found to be of the order: RadofGyration, PMI_mag, Density and <Jurs_WNSA_2+728.635>. RadofGyration is a measure of the size of an object, a surface, or an ensemble of points. It is calculated as the root mean square distance of the objects' parts from either its centre of gravity or an axis. This can be calculated by the following equation:

$$
RadofGyration = \sqrt{\left(\sum \frac{(x_i^2 + y_i^2 + z_i^2)}{N}\right)} \tag{4}
$$

Here, N is the number of atoms and x, y, and z are the atomic coordinates relative to the centre of mass. It makes a favourable contribution towards herbicidal activity as evidenced by the positive regression coefficient. This reflects the fact that the shape of the molecules plays an important role in herbicidal activity. For better herbicidal activity, the molecules should have an asymmetrical shape, i.e. an increase in substituent length of any one direction may enhance AHAS inhibitory activity. This can be illustrated by compounds **12** and **13**. There are fewer atoms present in compound **12** than in compound **13**, thus the numerical value of radius of gyration is higher in

compound **12** than in compound **13**. Due to the higher value of radius of gyration, compound **12** possesses greater herbicidal activity.

Density is a 3D spatial descriptor that is defined as the ratio of molecular weight to molecular volume. The density reflects the types of atoms and how tightly they are packed in a molecule. It makes an unfavourable contribution to AHAS inhibitory activity. For better herbicidal activity, lower molecular weight and higher molecular volume would be required. This is illustrated by compounds **11** and **12**. The molecular weight of compound **11** is less than that of compound **12**, and the molecular volume of compound **11** is greater than that of compound **12**. Thus, the value of Density is higher in compound **12** than in compound **11**. As a consequence, compound **11** has higher activity. There is another example to support this position. The molecular weight of compound **24** is much higher than that of compound **26**, while the molecular volume of compound **24** is only a bit lower than that of compound **26**. As the weight:volume ratio is much higher for compound **24**, the density of compound **26** is lower than that of compound **24**. Thus the AHAS activity of compound **26** is higher than that of compound **24**. The molecular weight of compound **24** is higher due to the presence of the bromo (–Br) substitution at the R_3 position.

Jurs_WNSA_2 is the surface weighted charged partial negative surface area. It can be calculated from the total charge weighted negative surface area (PNSA_2) multiplied by the total molecular solvent-accessible surface area (SASA) divided by 1,000.

$$Jurs\_WNSA\_2 = \frac{PNSA\_2*SASA}{1000} \quad (5)$$

But, in the above model, Jurs_WNSA_2 is present as a spline term and its negative coefficient reflects that if the value of Jurs_WNSA_2 is greater than −728.635 then it makes an unfavourable contribution towards AHAS inhibitory activity. This implies that an increase in the total charge weighted negative surface area and total SASA will enhance activity. For better herbicidal activity, both the negatively charged surface area as well as the total surface area should be higher. This can be illustrated by compounds **11** and **36**. The total charge weighted negative surface area of compound **11** is lower than that of compound **36**, and the total SASA is higher in the case of compound **11**. Thus, the value of Jurs_WNSA_2 in compound **11** (−763.873) is lower than that of compound **36** (−397.631). As a consequence, the AHAS inhibitory activity of compound **11** is much greater than that of compound **36**. This occurs due to the presence of fewer electronegative atoms in compound **36** compared to compound **11**. Negative charge distributed over a large surface area may enhance the herbicidal activity.

Additional external validation test

In the early 1980s, Unger and Hansch [34] stated that "...without a quality perspective, one can generate statistical unicorns, beasts that exist on paper but not in reality". Validation of QSAR models is a very important task. In the present paper, the models were subjected to a test for criteria of external validation as suggested by Golbraikh and Tropsha [35]. To find the predictive potential of the models, squared correlation coefficient values between the observed and predicted values of the test set compounds with intercept (r^2) and without intercept (r_0^2) were calculated. Interchange of the axes gives the value of r'^2_0. According to Golbraikh and Tropsha [35], models are considered acceptable if they satisfy all of the following conditions:

(1) $Q^2>0.5$
(2) $r^2>0.6$
(3) $(r^2 - r_0^2)/r^2 < 0.1$ or $(r^2 - r'^2_0)/r^2 < 0.1$
(4) $0.85 \le k \le 1.15$ or $0.85 \le k' \le 1.15$

When the observed values of the test set compounds (y-axis) are plotted against the predicted values of the compounds (x-axis) setting the intercept to zero, the slope of the fitted line gives the value of k. Interchange of the axes gives the value of k'. A list of values of different validation parameters defined above for different models is given in Table 4.

It has been shown previously [36] that $R_{pred}{}^2$ may not be sufficient to indicate the external predictivity of a model. The value of R^2_{pred} is controlled mainly by $\sum \left(Y_{obs(test)} - \overline{Y}_{training}\right)^2$, i.e. the sum of squared differences between observed values of test set compounds and the mean observed activity values of the training data set. Thus, R^2_{pred} may not truly reflect the predictive capability on a new dataset. Furthermore, the squared regression coefficient (r^2) between observed and predicted values of the test set compounds does not necessarily mean that the predicted values are very near to observed activity (despite maintaining an overall good intercorrelation, there may be considerable numerical difference between the values). So, for better external predictive potential of the model, a modified r^2 ($r^2_{m(test)}$) was introduced via the following equation [36]:

$$r^2_{m(test)} = r^{2*}\left(1 - \sqrt{r^2 - r_0^2}\right) \quad (6)$$

Table 4 External validation criteria of all the models according to Golbraikh and Tropsha [35]. *GFA* Genetic function approximation

Equation no.	Model type	r^2	Q^2	$(r^2-r_0^2)/r^2$	k
M1	GFA- spline (undocked)	0.718	0.847	0.007	1.012
M2	GFA- spline (docked)	0.853	0.776	0.047	0.951

In Eq. 6, r_0^2 is the squared correlation coefficient between the observed and predicted values of the test set compounds with the intercept set to zero. The value of $r^2_{m(test)}$ should be greater than 0.5 for an acceptable model. The values of $r^2_{m(test)}$ for the different models is reported in Table 5.

Initially, the concept r^2_m was applied only to the test set prediction [36], but it can equally well be applied to the training set if one considers the correlation between observed and LOO predicted values of the training set compounds [37]. More interestingly, this can be used for the whole set if LOO-predicted values for the training set and predicted values of the test set compounds are considered. The advantages of such consideration are: (1) unlike external validation parameters (R^2_{pred}, etc.), the $r^2_{m(overall)}$ statistic is based not only on a limited number of test set compounds, but includes prediction for both test set and training set (using LOO predictions) compounds. Thus, this statistic is based on prediction of a comparably large number of compounds. In many cases, the size of the test set is quite small, and regression-based external validation parameters may be less reliable and highly dependent on individual test set observations. In such cases, the $r^2_{m(overall)}$ statistic may be advantageous. (2) In many cases, comparable models are obtained, with some models showing comparatively better internal validation parameters and others showing relatively superior external validation parameters. This may create problems in selecting the final model. The $r^2_{m(overall)}$ statistic may be used for selection of the best predictive model from among comparable models. For the present QSAR study, we determined r^2_m values for both training (based on LOO-predicted values) and test sets and also for the whole set of reported models (results shown in Table 5).

When different models having different numbers of predictor variables are compared, it may be very difficult to determine which is the best model as r_m^2 does not consider the number of predictor variables used. To solve this problem, we developed another parameter [$r^2_{m(overall)}$ (adjusted)] in a manner similar to the adjusted R^2 (R^2_a). This newly introduced parameter can be calculated as follows:

$$r^2_{m(overall)}(adjusted) = \frac{(n-1)*r^2_{m(overall)} - p}{n-p-1} \quad (7)$$

In Eq. 7, n is the total number of compounds being predicted and p is the number of predictor variables. The values of the new parameter [$r_m^2{}_{(overall)}$(adjusted)] for Eqs. M1 and M2 are shown in Table 5.

Process randomisation

Robustness of the models relating AHAS inhibitory activity with selected descriptors was judged by randomisation (Y-randomisation) of the model development process. The test

Table 5 Comparison of statistical quality parameters and validation parameters of the models. *LOO* Leave one out

Equation no.	Type of descriptors	Model type	Model quality				Internal validation parameters			External validation parameters		Overall validation parameters		Model randomisation	
			R^2	R_c^2	F	s	Q^2	PRESS	$r^2_{m(LOO)}$	R^2_{pred}	$r^2_{m(test)}$	$r^2_{m(overall)}$	$r^2_{m(overall)}$ (adjusted)	R^{2a}_r	R^2_p
M1	Shape+ Spatial+ Electronic+ Structural	GFA- spline[b]	0.890	0.875	58.72	0.434	0.847	7.593	0.694	0.785	0.667	0.694	0.663	0.107	0.788
M2	Shape+ Spatial+ Electronic+ Structural	GFA- spline[a]	0.837	0.815	37.29	0.528	0.776	11.119	0.607	0.813	0.682	0.638	0.602	0.111	0.713

[a] Squared mean R for random models

[b] Using conformers generated by conformational analysis

[c] Using docked conformers

Table 6 Results of randomisation test applied to the model development process

Equation no.	Type of descriptors	Model type	R^2	R_r^2	R_p^2
M1	Shape+Spatial+Electronic+Structural	GFA-spline[a]	0.890	0.222	0.727
M2	Shape+Spatial+Electronic+Structural	GFA-spline[b]	0.837	0.419	0.541

[a] Using conformers generated by conformational analysis

[b] Using docked conformers

was done by repeatedly scrambling the activity values to generate QSAR models from the whole pool of descriptors and then comparing the resulting scores with the score of the original QSAR model generated from non-randomised activity values. In each case, the average correlation coefficient (R_r) of randomised models were significantly less than the correlation coefficient (R) of the non-randomised model. The results of process randomisation are shown in Table 6.

Model randomisation

Further statistical significance of the relationship between AHAS inhibitory activity and descriptors was checked by randomisation test (Y-randomisation) of the models. This technique ensures the robustness of the model. The values of dependent variables were scrambled randomly and new QSAR models were developed keeping the independent variable matrix unchanged. The randomisation test for the models was performed at 99% confidence level. The test was performed by shuffling the AHAS inhibitory activity values and the average value of the correlation coefficient (R_r) of random models was calculated. For an acceptable QSAR model, the average correlation coefficient (R_r) of randomised models should be less than the correlation coefficient (R) of the non-randomised model.

No clear-cut recommendation was found in the literature for the desired difference between the average correlation coefficient (R_r) of randomised models and the correlation coefficient (R) of a non-randomised model. We used the parameter R_p^2 [38], which penalises the model R^2 for the small difference between squared mean correlation coefficient (R_r^2) of randomised models and the squared correlation coefficient (R^2) of a non-randomised model. The above-mentioned novel parameter can be calculated by the following equation:

$$R_p^2 = R^2 * \sqrt{R^2 - R_r^2} \tag{8}$$

This novel parameter R_p^2 ensures that the models developed are not obtained by chance. For an acceptable QSAR model, the value of R_p^2 should be greater than 0.5. The values of R^2, R_r^2 and R_p^2 based on model randomisation for different models are reported in Table 5. The values of R^2, R_r^2 and R_p^2 based on process randomisation for different models are reported in Table 6.

Comparison with previous work

Xi et al. [14] worked on the same dataset and performed a DFT-based QSAR study of sulfonylurea analogues with herbicidal activity using quantum chemical charge descriptors along with the molecular volume of the electron cloud. They developed models from training set compounds and used these to predict the activity of test set compounds. The predictive potential (R_{pred}^2, not reported by Xi et al. [14]) of their best model (0.573) is somewhat lower than that of our models (Table 5). As the main objective of QSAR is to predict the activity of new sets of compounds, our models are of more potential practical use than the previous model. In addition, we also performed a docking study of all 45 AHAS inhibitor molecules.

Overview and conclusions

In the present work, we performed docking of 45 AHAS inhibitors into the active site of AHAS enzyme. We also performed QSAR studies for the compounds using three dimensional (shape, spatial and electronic) descriptors along with a few structural descriptors. The whole dataset (n=45) was divided into a training set (75% of the dataset) and a test set (remaining 25%) on the basis of the K-means clustering technique. Models developed from training set compounds were used to predict the activity of test set compounds. A comparison of statistical quality of different models was given in Table 5.

The docking study showed that Met351, Asp375, Arg377, Gly509, Met570 and Val571 are the important residues present at the active site, and that these are the main contributors to the receptor–ligand interaction. These amino acid residues form a pocket to which the inhibitors bind. It has been observed that, for better AHAS inhibitory activity, four amino acid residues (Met351, Arg377, Met570 and Val571) should optimally interact with the substituted sulfonylurea derivatives. The co-enzyme FAD plays a major role in the receptor binding of the inhibitors. The inhibitors form hydrogen bonds with some of the

amino acid residues to bind properly with the enzyme. However, steric bumps have a detrimental effect on AHAS inhibition activity. As compounds **11**, **12** form intramolecular or intermolecular hydrogen bonds, the AHAS inhibitory activity of these compounds is very high. In contrast, compounds **35**, **36** and **38** form steric bumps (either intermolecular or intramolecular), and their AHAS inhibitory activity is consequently lower.

On the other hand, models generated from MSA reflect the importance of structural (HBondacceptor), shape (DIFFV) and spatial (RadofGyration, PMI_mag, Shadow_X-length, Density and Jurs_WNSA_2) descriptors. According to the internal variance (Q^2=0.847), Eq. M1 is the best model. However, if we consider external predictive variance, Eq. M2 (R^2_{pred}=0.813) is better To avoid this type of contradiction, we developed a novel parameter ($r^2_{m(overall)}$) [36]. Again, according to both $r^2_{m(overall)}$, and the newly introduced parameter $r^2_{m(overall)}$(adjusted), Eq. M1 is the best model, giving values of 0.694 and 0.663, respectively. Considering the randomisation test as a validation criterion, Eq. M1 (0.788) is the best one according to the newly introduced parameter (R^2_p). Shape parameter shows that a bulky substitution at R_1 position may enhance AHAS inhibitory activity. The charged surface area descriptors suggest that negative charge distributed over a large surface area may enhance activity. Structural parameters support the charged surface area descriptors in that, for better activity, the number of electronegative atoms present in a molecule should be high. The spatial descriptors show that, for better activity, the molecules should possess bulky substituents and small substitution at *the* R_2 and R_3 positions, respectively.

The results of the present study may be useful in the design and development of novel compounds having better AHAS inhibitory activity with the potential to be used against the unwanted herbs and weeds that reduce agricultural productivity.

Acknowledgement Financial assistance from the Ministry of Human Resource Development, Government of India, New Delhi, in the form of a scholarship to S.P. is gratefully acknowledged.

References

1. http://en.wikipedia.org/wiki/Weed
2. http://en.wikipedia.org/wiki/Herbicide
3. Estrada E, Uriarte E, Montero A, Teijeira M, Santana L, De Clercq E (2000) J Med Chem 43:1975–1985
4. González MP, Dias LC, Helguera AM, Rodríguez YM, De Oliveira LG, Gómez LT, Díaz HG (2004) Bioorg Med Chem 12:4467–4475
5. González MP, González Díaz H, Molina RR, Cabrera MA, Ramos de Armas R (2003) J Chem Inf Comput Sci 43:1192–1199
6. González MP, Terán C, Teijeira M, Helguera AM (2006) Curr Med Chem 13:2253–2266
7. Xiao Z, Xiao YD, Feng J, Golbraikh A, Tropsha A, Lee KH (2002) J Med Chem 45:2294–2309
8. González-Díaz H, Torres-Gómez LA, Guevara Y, Almeida MS, Molina R, Castañedo N, Santana L, Uriarte E (2005) J Mol Model 11:116–123
9. Yang GF, Huang X (2006) Curr Pharm Des 12:4601–4611
10. Peng H, Wang T, Xie P, Chen T, He H, Wan J (2007) J Agric Food Chem 55:1871–1880
11. Zhang L, Wan J, Yang G (2004) Bioorg Med Chem 12:6183–6191
12. Wang JG, Li ZM, Ma N, Wang BL, Jiang L, Pang SS, Lee YT, Guddat LW, Duggleby RG (2005) J Comput Aided Mol Des 19:801–820
13. Zhu YQ, Liu P, Si XK, Zou XM, Liu B, Song HB, Yang HZ (2006) J Agric Food Chem 54:7200–7205
14. Xi Z, Yu Z, Niu C, Ban S, Yang G (2006) J Comput Chem 27:1571–1576
15. Schloss JV (1990) Pest Manag Sci 29:283–292
16. McCourt JA, Pang SS, King-Scott J, Guddat LW, Duggleby RG (2006) Proc Natl Acad Sci USA 103:569–573
17. Kirkpatrick P (2004) Nat Rev Drug Discov 3:299
18. Discovery Studio 2.0 is a product of Accelrys Inc, San Diego, CA
19. Cerius2 Version 4.10 is a product of Accelrys Inc, San Diego, CA
20. Eriksson L, Jaworska J, Worth AP, Cronin MTD, McDowell RM (2003) Environ Health Perspect 111:1361–1375
21. Guha R, Jurs PC (2005) J Chem Inf Model 45:65–73
22. Leonard JT, Roy K (2006) QSAR Comb Sci 25:235–251
23. Roy K (2007) Expert Opin Drug Discov 2:1567–1577
24. Everitt B, Landau S, Leese M (2001) Cluster analysis. Arnold, London
25. Dougherty ER, Barrera J, Brun M, Kim S, Cesar RM, Chen Y, Bittner M, Trent JM (2002) J Comput Biol 9:105–126
26. Hopfinger AJ, Tokarsi JS (1997) In: Charifson PS (ed) Practical applications of computer-aided drug design. Dekker, New York
27. Rogers D, Hopfinger AJ (1994) J Chem Inf Comput Sci 34:854–866
28. Roy PP, Leonard JT, Roy K (2008) Chemom Intell Lab Sys 90:31–42
29. Kubinyi H, Hamprecht FA, Mietzner T (1998) J Med Chem 41:2553–2564
30. Marshall GR (1994) In: Kubinyi H (ed) 3D QSAR in drug design—theory, methods and applications. ESCOM, Leiden
31. Deswal S, Roy N (2006) Eur J Med Chem 41:1339–1346
32. MINITAB is a statistical software of Minitab Inc, State College, PA
33. SPSS is a statistical software of SPSS Inc., Chicago, IL
34. Unger SH, Hansch C (1973) J Med Chem 16:745–749
35. Golbraikh A, Tropsha A (2002) J Mol Graph Model 20:269–276
36. Roy PP, Roy K (2008) QSAR Comb Sci 27:302–313
37. Roy PP, Roy K (2008) Chem Biol Drug Des 72:370–382
38. Roy K, Paul S (2009) QSAR Comb Sci 28:406–425