

In silico development, validation and comparison of predictive QSAR models for lipid peroxidation inhibitory activity of cinnamic acid and caffeic acid derivatives using multiple chemometric and cheminformatics tools

Indrani Mitra · Achintya Saha · Kunal Roy

Received: 9 January 2012 / Accepted: 22 February 2012 / Published online: 21 March 2012
© Springer-Verlag 2012

Abstract The design and development of antioxidant molecules have lately gained a great deal of focus which is attributed to their immense biomedical importance in combating the free radical associated health hazards. In a situation to replenish the endogenous antioxidant loss, synthetic molecules with potent antioxidant activity is demanded. The present work thus aims at *in silico* modeling of antioxidant molecules that may facilitate in searching and designing of new chemical entities with enhanced activity profile. A series of cinnamic acid and caffeic acid derivatives having the ability to inhibit lipid peroxidation have been modeled in the present work. Three different types of models were developed using different chemometric and cheminformatics tools to identify the essential structural attributes: (a) descriptor based QSAR models, (b) 3D pharmacophore models and (c) HQSAR (hologram QSAR) models. For the conventional QSAR modeling, descriptors belonging to different categories [quantum chemical descriptors (Mulliken charges of the common atoms of the molecules), thermodynamic descriptors, electronic descriptors, structural descriptors and spatial descriptors] were calculated for the development of statistically significant as

well as well interpretable quantitative structure-activity relationship (QSAR) models. Two different chemometric tools [genetic function approximation (GFA) and genetic partial least squares (G/PLS)] were employed for the development of the QSAR models. The 3D pharmacophore model focused on the essential pharmacophoric features while the HQSAR model implicated the prime structural fragments that were necessitated for the optimal anti-lipid peroxidative activity of the molecules. All the models were validated based on internal, external and overall validation statistics. Randomization was performed in order to ensure the absence of chance correlation in the developed models. Among all models, the descriptor-based model developed using the GFA-spline technique yielded the most satisfactory results. The results obtained from all the models corroborate well with each other and chiefly signify the importance of the ketonic oxygen of the amide/ acid fragment and the ethereal oxygen substituted on the parent phenyl ring of the molecules under study. Thus the models can efficiently be utilized for extensive screening of large datasets and their subsequent activity prediction.

Keywords Antioxidant · GFA · HQSAR · Pharmacophore · QSAR

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

K. Roy
e-mail: kroy@pharma.jdvu.ac.in
URL: <http://sites.google.com/site/kunalroyindia/>

A. Saha
Department of Chemical Technology, University of Calcutta,
92 A P C Road,
Kolkata 700 009, India

Introduction

Free radicals are chemically reactive species having unpaired molecules, which play an important role in a number of vital life processes like intracellular killing of bacteria by phagocytic cells such as granulocytes and macrophages. They have also proved to be essential in certain cell signaling processes [1]. However, when produced in large excess they participate in a series of unwanted side reactions resulting in cell damage.

Free radical overload within the system leads to a condition of oxidative stress which can be implicated as the prime causative reason for most life threatening diseases. Free radicals are produced in cells by cellular metabolism and/or induction of exogenous agents. These free radicals damage biomolecules especially DNA involving hydrogen abstractions and addition reactions leading to carbon-centered sugar radicals and OH- or H-adduct radicals of heterocyclic bases [2]. Such oxidative damage to DNA may in turn lead to carcinogenesis, aging and can even result in detrimental biological consequences such as the initiation and promotion of cancer [3, 4]. Besides these, trace amounts of copper and iron present within the human system further catalyse free radical reactions aggravating their deleterious effects. Reactive oxygen free radical species (ROS) have also been implicated in the pathogenesis of degenerative joint disease [5] and exert the potential to initiate cellular damage to joint tissues [6]. The ROS oxidatively modify lipids and proteins with deleterious consequences for vascular function and lipid peroxidation. Such dysfunctional vasculature is characterized by lipid peroxidation and aberrant lipid deposition followed by inflammation, thrombus formation, disturbed hemodynamic flow etc. [7].

In vivo attenuation of the oxidative modification of lipoproteins by natural and synthetic antioxidants is considered as a possible way of combating oxidative stress related diseases. The free radicals are regulated by a balance between tissue oxidant and antioxidant activity [8]. The endogenous antioxidant system includes albumin, uric acid, total bilirubin and endogenous enzymes like superoxide dismutase, catalase and glutathione peroxidase. An imbalance in the oxidant-antioxidant reaction either due to excess free radical formation or insufficient removal by antioxidants leads to oxidative stress. Thus to balance the overall antioxidant status of the human body, dietary antioxidant supplementation becomes essential. They have a positive effect on the general health as well as on mortality associated with cardiovascular diseases and cancer [9–11]. To cope with the systemic demand for antioxidants, several synthetic antioxidants are also being marketed. Viewing the immense biomedical importance of these molecules, design and synthesis of such chemical entities with improved potency have gained remarkable attention over the last few decades. Such designing of active chemical compounds employing a time effective approach involves the incorporation of QSAR methodology [12]. The QSAR technique aims to determine a mathematical correlation between biological activity of the molecules and their structure through the utilization of descriptors. These descriptors are the numerical representation of molecular properties, atomic fragments, topology, spatial features etc. [13]. Development of QSAR models enables preliminary screening of databases for selection of optimally active molecules. Thus, the QSAR technique extensively reduces the time and cost

involved in synthesizing and analyzing all the available molecules in the databases.

Several researchers have utilized the QSAR technique as an efficient tool in the field of rational drug design. Among the varied types of molecules with different therapeutic activity data screened using this QSAR tool, research related to the design and development of antioxidant molecules with improved activity profile have been undertaken by different groups of authors. Worachartcheewan et al. [14] predicted the DPPH (2,2-diphenyl-1-picrylhydrazyl) free radical scavenging activity of curcumin derivatives by performing both classification and QSAR analyses and concluded that compounds having high chemical stability together with low dipole moments and higher number of hydroxyl groups could be ranked as potent free radical scavengers. Li et al. [15] reported a QSAR study of antioxidative peptides containing either histidine or tyrosine residue. Partial least squares (PLS) analysis performed by the authors revealed the importance of the central amino acid and the N-terminal amino acids over the C-terminal amino acid for antioxidative action of the tripeptides. A series of morpholine derivatives were studied by Nikitakis et al. [16] for their antioxidant and squalene synthase inhibitory activity based on the development of significant QSAR models suggesting that electron affinity along with molecular shape and electrostatic effects played a significant role, which provided some insight on the molecular mechanism of action of these derivatives. With easy accessibility of software, the QSAR methodology has been extrapolated to the task of determining the molecular pharmacophores and structural features responsible for their antioxidant activity profile through the development of 3D-pharmacophore models and comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) techniques [17, 18]. Further, importance of the different molecular fragments for the overall activity profile of the molecules can also be assessed using the HQSAR and group based QSAR (G-QSAR) methodologies [17]. Such multidimensional analysis has been recently reported by our group [17] which involves the development of multiple QSAR models (developed using different techniques like 3D pharmacophore mapping, CoMSIA, HQSAR, and G-QSAR) for consensus predictions and unified mechanistic interpretations of the free-radical scavenging activities of chromone derivatives. Besides these, essential pharmacophoric features for the free radical scavenging activity of the arylamino-substituted benzo[b]thiophenes have also been assessed by our group [18]. The vivid importance of antioxidant molecules in maintaining the normal physiological function has oriented our research toward detecting new chemicals with enhanced antioxidant activity which may be exhibited either by their free radical scavenging activity or by their anti-lipid peroxidative activity. The present work

deals with derivatives of cinnamic acid and caffeic acids, which were modeled for their ability to inhibit lipid peroxidation within the biological system. The essential molecular attributes contributing to the activity profile of the molecules were determined based on different QSAR models developed using various chemometric tools like GFA and G/PLS. Additionally, the pharmacophore and crucial molecular fragments of the compounds were also determined based on 3D pharmacophore mapping and HQSAR techniques.

Materials and methods

Dataset selection

The model dataset used for the present work was obtained by clubbing three different datasets reported by Jung *et al.* [19], Kang *et al.* [20] and Rajan *et al.* [21]. The dataset reported by Jung *et al.* [19] comprised of a series of 4-hydroxyphenylacetic acid amides and 4-hydroxycinnamides with the ability to inhibit lipid peroxidation. The second dataset as reported by Kang *et al.* [20] also comprised of a similar series of 4-hydroxycinnamide derivatives with dialkoxy substitutions having the potency to inhibit lipid peroxidation. Lastly, a series of amides of caffeic acids evaluated by Rajan *et al.* [21] as lipid peroxidation inhibitors was selected as the third dataset. The data reported in all the papers were estimated using the same experimental protocol, i.e., inhibition of lipid peroxidation. The dataset thus used for the present work comprised of derivatives of 4-hydroxyphenylacetic acid amide [19], 4-hydroxycinnamide [19, 20] and amides of caffeic acid [21] having the ability to inhibit lipid peroxidation. For the development of the QSAR models and the HQSAR analysis, 50% inhibitory concentration (IC_{50}) of the molecules was converted to the negative logarithmic scale (pIC_{50}). For the development of the pharmacophore models, the dependent variable IC_{50} (nM) of the molecules was used. All the molecules with their activity data are summarized in Table 1.

Descriptor calculation for the descriptor based QSAR model

The QSAR analysis was performed based on a series of descriptors of different categories such as: (a) quantum chemical descriptors (Mulliken charges of the common atoms of the molecules), (b) thermodynamic descriptors, (c) electronic descriptors, (c) structural descriptors and (d) spatial descriptors. The molecules were prepared using the Gauss View 3.0 software [22] and the common atoms (Fig. 1) were marked for the calculation of the Mulliken atomic charges. Subsequently, the energy minimization of the molecules was performed using the GAUSSIAN 03W software [23]. Energy calculation was performed at three

different levels of theory: (i) the semi-empirical AM1 method (Austin method), (ii) the Hartree–Fock method at HF/3-21G(d) level and (iii) Hartree–Fock method at HF/6-31G(d) level [23]. The output from each level was used as the input for the next level. The charges of the common atoms thus calculated at each level were then correlated with the lipid peroxidation inhibitory data of the compounds involved in this study. Although charges were calculated at three levels of theory, better correlation statistics were obtained when models were developed employing the charge descriptors calculated at the HF/6-31G(d) level compared to the other two levels. Hence, Mulliken atomic charges calculated at this level were further utilized for final QSAR model development. Besides these, Cerius 2 software version 4.1 [24] was employed for further calculation of the remaining descriptors. All the different types of descriptors thus calculated are detailed in Table 2.

Splitting of the dataset

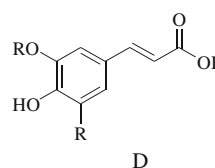
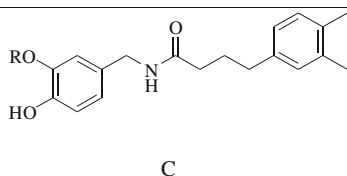
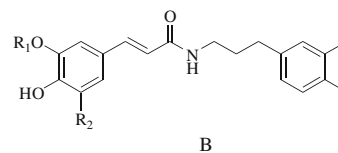
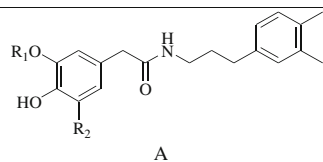
Selection of training and test sets plays a crucial role in the development of a QSAR model. The selection should be such that the test set molecules encompass the activity range of the total dataset and at the same time lie within the chemical domain of the training set data. A QSAR model exhibits poor predictivity for a test set molecule dissimilar from the training set ones, while molecule appreciably similar to the training molecules are ideally predicted. In the present work, 25% of the total dataset has been selected as the test set based on the cluster analysis technique [25, 26] using SPSS software [27]. This method ensures uniform selection of the test set bearing molecules covering the entire range of the chemical space as that of the total dataset. In order to further ascertain the uniformity in the distribution of the training and test set molecules, a principal component analysis (PCA) [27] of the descriptor matrix was performed using SPSS software [27]. The PCA score plot shows the distribution of the training and test set compounds in the 3D space with respect to the first three principal components obtained from the total descriptor matrix. The obtained plot (Fig. 2) shows that the test set molecules lie in close proximity to at least one training set compound, ensuring that the test set thus selected captures all of the essential features of the entire dataset of molecules.

Model development and validation

Three different types of models were built in the present work employing the training set molecules: (a) descriptor based QSAR models, (b) 3D pharmacophore models and (c) HQSAR model. Different techniques utilized for the development of the models are given in Fig. 3. The statistical fitness and the predictive quality of the developed models

Table 1 Structures of the 54 molecules under study along with their observed and calculated/predicted lipid peroxidation inhibitory activity data

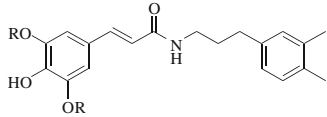
Sl. No.	Structure type	R1	R2	Observed activity ^y	Calculated/Predicted activity
1	A	CH ₃	H	1.587	1.333
2*	A	CH ₂ CH ₃	H	1.621	2.185
3	A	CH(CH ₃)CH ₃	H	1.713	1.671
4*	A	(CH ₂) ₃ CH ₃	H	2.032	2.136
5	A	C ₆ H ₅	H	1.495	1.679
6	A	CH ₃	OCH ₃	1.929	2.205
7	A	H	H	2.585	2.395
8*	A	H	OH	2.742	2.737
9	B	CH ₃	H	2.003	1.951
10	B	CH ₃	OCH ₃	2.467	2.418
11	B	H	H	2.955	2.802
12	C	CH ₃		1.243	1.437
13*	D	CH ₃		2.321	1.754
14	D	CH ₂ CH ₂ CH ₂ CH ₃		2.434	2.455
15	D	CH(CH ₃)(CH ₂ CH ₂ CH ₃)		2.315	2.459
16	D	CH(CH ₂ CH ₃) ₂		2.656	2.440
17	D	(CH ₂) ₅ CH ₃		2.650	2.349

^y [19–21]

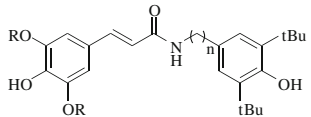
* Compounds selected as the test set compounds

^a Activity calculated/predicted based on the best GFA-spline equation [Eq. 3]

Table 1 (continued)

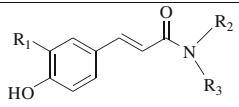


E



F




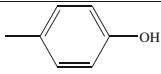
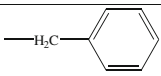
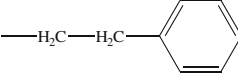
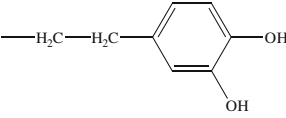
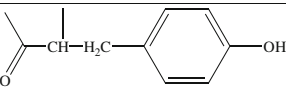


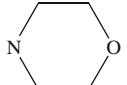
Sl. No.	Structure type	R	n	Observed activity ^y	Calculated/Predicted activity
18	E	CH(CH ₃)(CH ₂ CH ₂ CH ₃)	-	2.876	3.047
19	E	CH(CH ₂ CH ₃) ₂	-	2.762	2.947
20	E	(CH ₂) ₅ CH ₃	-	3.538	3.253
21	F	CH ₃	1	2.836	2.684
22*	F	CH ₃	2	3.161	3.150
23	F	(CH ₂) ₂ CH ₃	1	3.569	3.426
24	F	(CH ₂) ₂ CH ₃	2	3.585	3.566
25	F	(CH ₂) ₃ CH ₃	1	3.377	3.474
26	F	(CH ₂) ₃ CH ₃	2	3.387	3.465
27*	F	CH(CH ₃)(CH ₂ CH ₂ CH ₃)	1	3.244	3.439
28*	F	CH(CH ₃)(CH ₂ CH ₂ CH ₃)	2	3.444	3.452
29	F	CH(CH ₂ CH ₃) ₂	1	3.086	3.068
30	F	CH(CH ₂ CH ₃) ₂	2	3.357	3.262
31	F	(CH ₂) ₅ CH ₃	1	2.453	2.602
32	F	(CH ₂) ₅ CH ₃	2	2.562	2.441



G

Sl. No.	Structure type	R1	R2	R3	Observed activity ^y	Calculated/Predicted activity ^a
33*	G	OH	H	CH ₂ CH=C(CH ₃)CH ₃	2.469	2.735
34	G	OH	H	H	2.658	2.920
35	G	OH	H	OH	2.678	2.697
36	G	OH	H	CH ₃	2.222	2.255
37	G	OH	H	CH ₂ CH ₃	2.569	2.725
38	G	OH	H	CH(CH ₃)CH ₃	2.409	2.780
39*	G	OH	H	CH ₂ CH(CH ₃)CH ₃	2.658	2.736
40	G	OH	H	CH ₂ CH ₂ CH(CH ₃)CH ₃	2.854	2.755
41	G	OH	H	-CH ₂ CH=CH ₂	2.658	2.663

Table 1 (continued)

42	G	OH	H		3.420	2.716
43*	G	OH	H		3.538	3.199
44*	G	OH	H		3.432	2.904
45	G	OH	H		3.201	3.427
46*	G	OH	H		2.991	2.688
47	G	OH	H		3.071	2.703
48	G	OH	H		3.229	3.278
49*	G	OH	OCH ₃		2.495	2.286
50	G	OH	CH ₂ CH ₃	CH ₂ CH ₃	2.387	2.566
51	G	OH			2.620	2.613
52	G	OH			2.444	2.586
53	G	OH			2.215	2.416
54	G	H	H	CH ₂ CH=C(CH ₃)CH ₃	1.536	1.642

were assessed based on different internal and external validation statistics. The QSAR models enable to quantify the importance of the different molecular properties, arising due to the variation in substitutional patterns of the molecules, for their anti-lipid peroxidative ability. In the present work, QSAR models were built using genetic function approximation (GFA) [28] and genetic partial least squares (G/PLS) [29] techniques employing both linear and spline options with the aid of Cerius 2 software [24]. On the other hand, a

3D pharmacophore refers to a collection of chemical features in space essential for a desired biological activity. Thus a 3D pharmacophore model quantitatively determines the features that are required to obtain optimum activity of the molecules under study [30, 31] and provides a unique ligand-based approach for drug design. In the present work, the 3D pharmacophore model was built using HypoGen module, implemented in Discovery Studio 2.1 [31]. To assess the quality of the generated pharmacophore

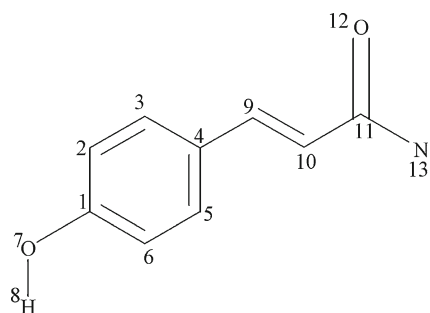


Fig. 1 Common atoms of the molecules under study that have been numbered for calculation of Mulliken charges at different levels of theory

hypotheses, cost functions [32] (represented in bit units) were calculated during hypothesis generation. Further, the test set compounds were mapped and predicted based on the best selected pharmacophore model. Besides these, hologram QSAR (HQSAR) model was also developed to determine the essential molecular fragments contributing to the optimum activity profile of the molecules. HQSAR [33, 34] is a modern 2D QSAR approach based on specialized molecular fragments providing an extended form of fingerprint encoding all possible molecular fragments which include linear, branched, cyclic and overlapping features of the molecules. Each molecule in a training set is broken down into several unique fragments which are arranged to form a molecular hologram representing the molecular fingerprint. The molecular hologram thus formed takes into account both the type and number of the various molecular fragments arranged in bins of fixed length array using a pre-defined set of rules derived from parameters that specify the size and type of fragment substructures that are to be

Table 2 Descriptors used for the present work

Category of descriptors	Descriptors used
Quantum chemical descriptors	C ₁ , C ₂ , C ₃ , C ₄ , C ₅ , C ₆ , O ₇ , H ₈ , C ₉ , C ₁₀ , C ₁₁ , O ₁₂ , N ₁₃ (Mulliken charges on the 13 common atoms of the phenolic derivatives)
Electronic	Dipole moment, HOMO (Highest occupied molecular orbital energy), LUMO (Lowest unoccupied molecular orbital energy), Superdelocalizability (Sr).
Topological indices	Wiener, Zagreb, Balaban, connectivity indices, kappa shape indices, E-state parameters
Structural	Hbond acceptor, Hbond donor, Rotlbonds, Chiral centers
Spatial	Jurs descriptors, Shadow indices, Radius of Gyration, Molecular surface area, Density, Principal moment of inertia, Molecular volume.
Thermodynamic	LogP, AlogP, AlogP98, Molar refractivity

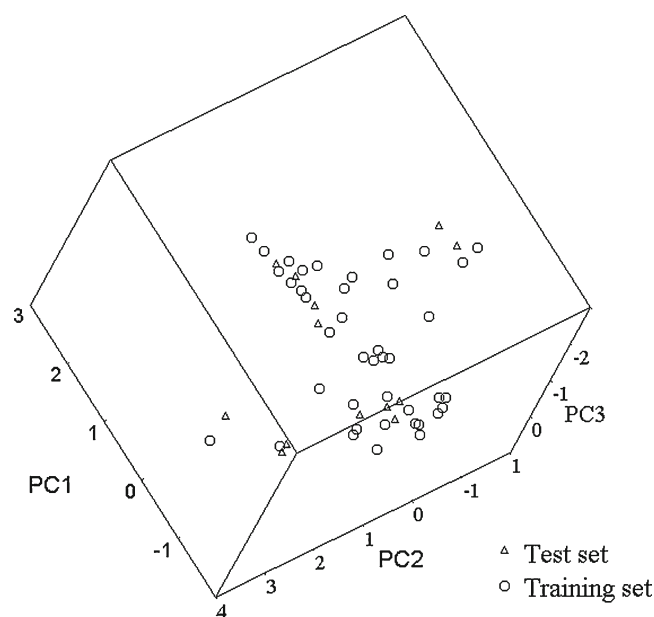


Fig. 2 PCA score plot of first three components for the total descriptor matrix

encoded. In the present work, HQSAR model was derived based on various combinations of fragment distinction and fragment generation parameters for each hologram length using the Sybyl software [35] and the best PLS model was selected based on the maximum value of Q^2 and a minimum value of cross-validated standard error (SE_{cv}). The “5% rule” (the rule allowed addition of latent variable only when it resulted in an increment in the value of Q^2 by 5% or more) [36] was applied to check the optimum component number. The final PLS model was thus developed with the optimum component number based on the specific fragment distinction parameters, fragment size and bin length.

All the models developed were validated using the internal validation, external validation and randomization techniques. The leave-one-out cross-validation technique was employed for the internal validation of the models. The statistical significance of the models was assessed based on the value of $LOO-Q^2$ (cross-validated squared correlation coefficient) which was calculated using the predicted activity data of each of the training set compounds that were deleted once in each of the cycles of LOO cross-validation. To further ascertain the proximity in the values of the predicted and observed activity data of the training set compounds, the r_m^2 metrics [$\overline{r_m^2(LOO)}$ and $\Delta r_m^2(LOO)$], developed by the present group of authors, (Eqs. 1 and 2) were calculated [37, 38].

$$\overline{r_m^2} = (r_m^2 + r'_m{}^2)/2 \quad (1)$$

$$\Delta r_m^2 = |r_m^2 - r'_m{}^2| \quad (2)$$

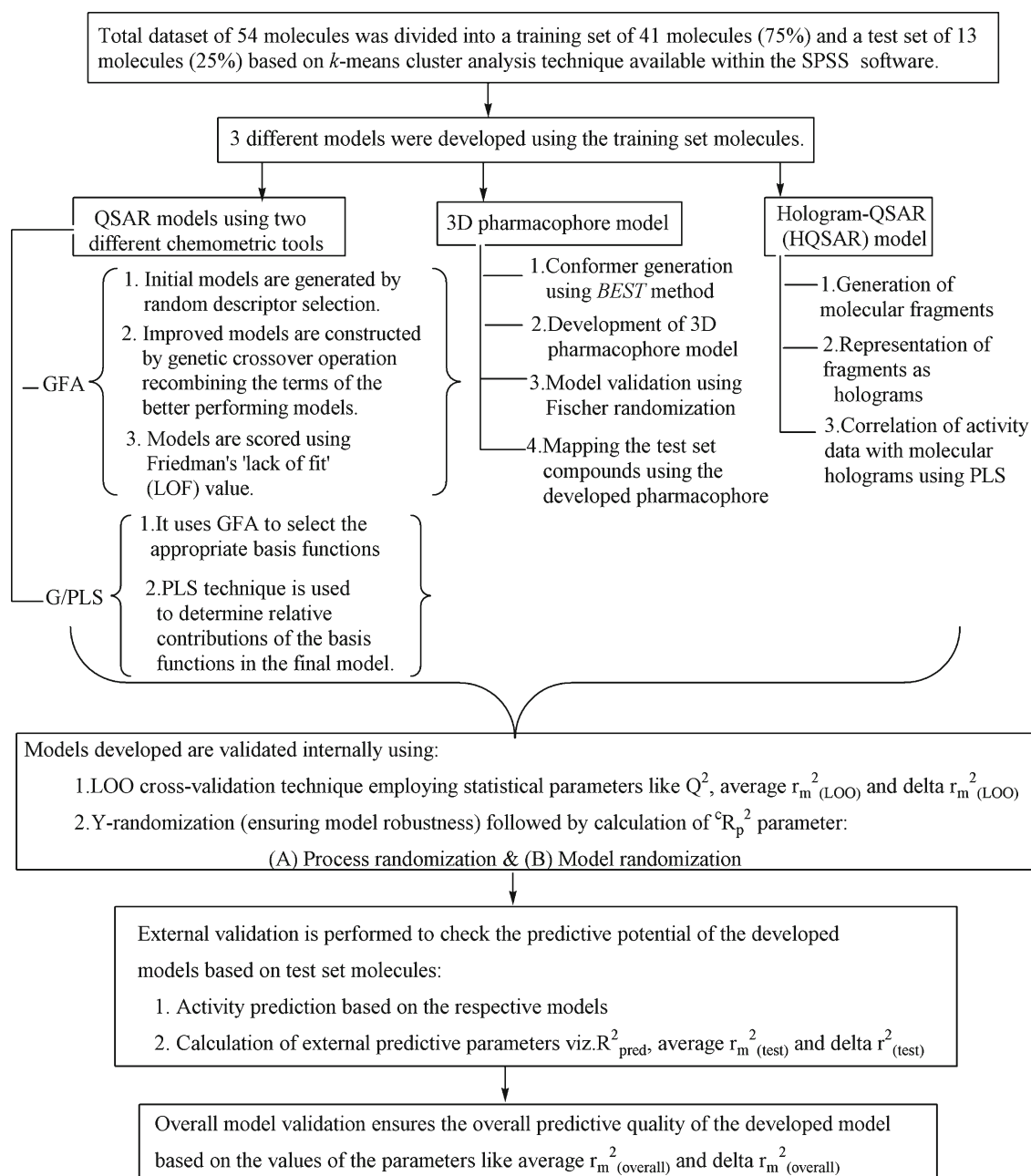


Fig. 3 Schematic representation of different QSAR techniques performed in the present study

Here, $r_m^2 = r^2 * (1 - \sqrt{(r^2 - r_0^2)})$ and $r_m'^2 = r^2 * (1 - \sqrt{(r^2 - r_0'^2)})$. 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 for determination of r_m^2 . Change of the axes gives the value of $r_0'^2$ and the $r_m'^2$ metric was calculated based on the value of $r_0'^2$.

Further, the external predictive ability of the models was assessed based on the predictions of the test set compounds. The external predictive parameter (R_{pred}^2) [39, 40] thus calculated reflect the ability of the models to predict the activity

of new series of molecules belonging to the same class. The calculation of the r_m^2 metrics for the test set data

$[r_m^2$ (test) and Δr_m^2 (test)] additionally estimated the fitness between the values of the predicted and the corresponding observed activity data. The overall performance of the QSAR models was also checked using the overall validation parameters like r_m^2 (overall) and Δr_m^2 (overall). The r_m^2 metrics have also been used by other groups of authors [41, 42] and also implemented in the freeware CORAL [43]. The robustness of the models was checked based on the randomization

technique where the activity data of the training set molecules was scrambled keeping the descriptor matrix unchanged and new models were built based on the permuted activity data. For a robust model, the squared correlation coefficient (R^2) of the non-randomized model exceeds the average correlation coefficient of the randomized model (R_r^2) by far. In case of the QSAR models, process randomization (the response parameter was permuted keeping the total descriptor matrix unchanged followed by variable selection) was performed at 90% while the model randomization (the Y variable was scrambled based on the unaltered model descriptors only) was performed at 99% confidence level followed by calculation of the ${}^{\circ}R_p^2$ parameter [40] that penalizes model R^2 for small differences in the values of R^2 and R_r^2 . Additionally, Fischer randomization technique at 95% confidence level was employed to judge whether the pharmacophore model was a significant one or a mere outcome of chance only.

Results and discussion

The developed QSAR models were analyzed to estimate the structural prerequisites of the molecules in order to exhibit optimum anti-lipid peroxidative activity. The statistical results of the different types of QSAR models developed in the present work are summarized in Table 3. Although all the different models yielded statistically significant results, the GFA model developed using the spline option displayed the most satisfactory results for all the internal and external validation metrics. Besides these, the 3D pharmacophore model and the HQSAR model are also analyzed in Table 3. Table 1 also summarizes the calculated/predicted activity of the molecules based on the most significant models developed here.

GFA model developed using the spline option

The GFA-spline model, being the most satisfactory one among all the developed models, has been analyzed further.

$$\begin{aligned}
 pC &= 33.1 + 1.88 \times C_4(\pm 3.081) + 2.82 \times HOMO(\pm 0.412) \\
 &+ 0.271 \times AtypeH_50(\pm 0.053) - 0.112 \times \langle AtypeH_46 - 28 \rangle (\pm 0.021) \\
 &+ 0.446 \times \langle 1 - AtypeC_5 \rangle (\pm 0.097) \\
 n_{training} &= 41, F(df) = 50.78(5, 35), R^2 = 0.879, Q^2 = 0.841, \overline{r_m^2(LOO)} = 0.745, \\
 \Delta r_m^2(LOO) &= 0.115, n_{test} = 13, R^2_{pred} = 0.710, \overline{r_m^2(test)} = 0.585, \Delta r_m^2(test) = 0.139, \\
 r_m^2(overall) &= 0.748, \Delta r_m^2(overall) = 0.141
 \end{aligned} \tag{3}$$

Here, $n_{training}$ and n_{test} refer to the number of compounds in the training and test sets respectively. Besides these, R^2 refers to the squared correlation coefficient and F indicates the variance ratio at a specified degree of freedom (df). The threshold value for all the parameters except Δr_m^2 is 0.5 and for the Δr_m^2 parameter, the value should be less than the stipulated value of 0.2. All the internal and external validation parameters for the developed model bear statistically significant values that lie within the acceptable limit. Predictive potential of the developed model in terms of internal and external validation tests is reflected from the acceptable values of the Q^2 (0.841) and R_{pred}^2 (0.710) metrics respectively. Further, satisfactory values for all the r_m^2 metrics account for least possible deviations of the predicted activity data from the corresponding observed ones.

The descriptors thus appearing in Eq. 3 obey the following order of significance based on their standardized coefficients: (a) HOMO, (b) C_4 , (c) $\langle AtypeH_46-28 \rangle$, (d) $AtypeH_50$ and (e) $\langle 1-AtypeC_5 \rangle$. The HOMO descriptor refers to the highest occupied molecular orbital energy which is crucially important in governing molecular reactivity and properties. The HOMO energy refers to the ability of the molecules to donate electrons during bond formation

and thus measures the nucleophilicity of the molecules. Since the value of the HOMO descriptor of a compound bears negative value, a positive coefficient for the descriptor in Eq. 3 denotes an increase in the anti-lipid peroxidative ability of the molecules for less negative values of the HOMO descriptor. This is well evident in case of compound nos. 20, 24, 26 and 30 that bear low negative values for the HOMO descriptor and thereby exhibit higher activity range. On the contrary, least activity profile of compound no. 12 may be attributed to the maximum negative value for the HOMO descriptor. Again, a reduction in the negative value of the HOMO descriptor refers to a decrease in nucleophilicity of the molecule with a subsequent increase in the electrophilic nature. Thus a decrease in the nucleophilic character of the antioxidant molecule favors their interaction with the electron rich reactive free radicals. The second important descriptor belongs to the class of quantum chemical descriptors and refers to the charge on C_4 atom calculated using the Hartee-Fock method at the HF 6-31 level of theory [23]. A positive coefficient of the charge descriptor denotes that the anti-lipid peroxidative activity of the molecules is favored with an increase in positive charge on the C_4 atom as in the case of compound nos. 20, 23 and 25. However, compound nos. 1, 3 and 5 having

Table 3 Comparison of the statistical quality of all the developed models

Descriptors/ features/ fragments	QSAR analysis		3D pharmacophore model		HQ SAR analysis	
	GFA		G/PLS			
	linear	spline	linear	spline		
Jurs FNSA_2, C ₄ , HOMO, Atype C_5		C ₄ , HOMO, Atype H_50, <Atype H_46-28>, <1- Atype C_5>	Atype C_5, Atype H_51, Jurs FPSA_1, C1, HOMO, Hbond donor	<3- Atype H_50>, Atype H_51, <Jurs WPSA_1 -903.515>, <-11.2006-HOMO>, <1- Atype C_5>, Jurs WNSA_3	HBA, HBA, hydrophobic	Atoms, bonds, donor & acceptor
R ²	0.780	0.879	0.774	0.874	0.571	0.819
Q ²	0.710	0.841	0.705	0.822	–	0.586
Avg r _m ² (LOO)	0.624	0.745	0.590	0.747	–	–
Δr _m ² (LOO)	0.179	0.115	0.189	0.101	–	–
R _{pred} ²	0.733	0.710	0.817	0.739	0.568	0.736
Avg r _m ² (test)	0.626	0.585	0.638	0.595	0.516	0.605
Δr _m ² (test)	0.214	0.139	0.178	0.234	0.166	0.168
Avg r _m ² (overall)	0.647	0.748	0.613	0.717	–	–
Δr _m ² (overall)	0.185	0.141	0.226	0.140	–	–
Process randomization (90%)						
R _r ²	0.233	0.479	0.346	0.483	–	–
c _r R _p ²	0.653	0.593	0.576	0.585	–	–
Model randomization (99%)						
R _r ²	0.085	0.125	0.007	0.002	0.199	–
c _r R _p ²	0.736	0.814	0.770	0.873	0.461	–

negative charge on C_4 atom exhibit lowest activity profile.

The various atom type descriptors signify the contribution of the different fragments to the hydrophobicity profile of the molecules. *AtypeH_46* refers to an atom centered fragment descriptor measuring the number of fragments bearing hydrogen atom attached to a sp^3 hybridized carbon atom ($C_{sp^3}^0$) having no heteroatom (denoted as X representing O, N, S, P, Se, and halogens) attached to the next carbon. Thus a negative coefficient for the *<AtypeH_46-28>* descriptor indicates that the activity data of the molecules increase with a decrease in the value of the spline term as evident from a decrease in the number of *AtypeH_46* fragments. The spline function, *<AtypeH_46-28>*, exerts zero contribution for values of the *AtypeH_46* parameter lower than that of the knot of the spline, i.e., 28 since a negative value within a spline term denotes zero contribution of the corresponding descriptor [28]. Compound nos. **23**, **24**, **25** and **26** having zero values for the *<AtypeH_46-28>* descriptor show maximum activity range. However, moderate activity profile of compound nos. **34**, **35**, **45** and **48** is attributed to the unfavorable values of the higher ranking descriptors, even though these compounds bear zero values for the *<AtypeH_46-28>* descriptor. On the contrary, although having non-zero values for the *<AtypeH_46-28>* descriptor, compound no. **30** exhibits high activity range due to the acceptable values of the more significant HOMO descriptor. *AtypeH_50* refers to an atom centered fragment descriptor which measures the number of fragments bearing hydrogen atom attached to a heteroatom. A positive coefficient of the *AtypeH_50* descriptor indicates that an increase in the number of fragments comprising of a hydrogen atom attached to a heteroatom favors the activity profile of the molecules. Compound nos. **34**, **35**, **45** and **48** although bearing maximum values for the *AtypeH_50* descriptor exert moderate activity profile that may be attributed to large negative values and reduced positive values for the HOMO and C_4 descriptors respectively having higher weightage. Additionally, compound nos. **24**, **26** and **30** attaining acceptable values for all the significant descriptors bear a maximum activity profile. Compound no. **24** although bearing moderate values for the C_4 and *AtypeH_50* descriptor, shows maximum activity due to the ideal values of the remaining more significant descriptors. *AtypeC_5* is another atom centered descriptor providing a measure of the number of fragments bearing sp^3 hybridized carbon atom substituted with a heteroatom (CH_3X). As suggested by Eq. 3, the activity of the molecules improves with a decrease in the number of such fragments. Since the knot of the spline is 1, it may be inferred that total absence of such fragments favor the activity profile of the molecules. However, compound nos. **1**, **6**, **9**, **10**, **12**, **21** and **36** although lacking the CH_3X fragment lies in the lower

activity domain due to the inappropriate values of the other more significant descriptors.

Thus the model describes that the prime important factor governing the activity profile of the molecules is their HOMO energy, indicating an increase in activity with a decrease in the nucleophilic character of the molecules. Additionally, reduced positive charge or negative charge on the C_4 atom adversely affects the activity profile of the molecules. Besides these, specific molecular fragments also influence the optimal activity of the molecules. Decrease in the number of fragments bearing hydrogen atom attached to a sp^3 hybridized carbon atom ($C_{sp^3}^0$) having no heteroatom attached to the next carbon and those bearing sp^3 hybridized carbon atom substituted with a heteroatom (CH_3X) are conducive for the anti-lipid peroxidative activity of the molecules. Conversely, an increase in the number of atom centered fragments bearing hydrogen atom attached to a heteroatom (-OH) further enhances the activity profile of the molecules.

3D pharmacophore model

A set of 10 pharmacophore hypotheses (Table 4) were developed using 41 training set compounds based on the conformers obtained using the *BEST* method of conformer generation. All the hypotheses yielded acceptable results in terms of cost functions and correlation coefficients. The values of fixed cost and null cost, expressed in bits, differed significantly from each other by 211.801 bits and such a difference implies existence of more than 90% chance of true correlation for the developed 3D pharmacophore. As required for a satisfactory hypothesis, the value of total cost of each hypothesis was close to the fixed cost values (Table 4).

Table 4 Results obtained from the Pharmacophore Hypotheses using *BEST* method

Hypothesis no	Total cost	Correlation (R)	Features
1	205.001	0.778	HBA HBA HBD
2	212.261	0.756	HBA HBA HYDROPHOBIC
3	212.822	0.786	HBA HBA HYDROPHOBIC
4	213.59	0.788	HBA HBA HYDROPHOBIC
5	214.44	0.751	HBA HBA HBA
6	215.545	0.747	HBA HBA HYDROPHOBIC
7	216.622	0.752	HBA HBA HYDROPHOBIC
8	222.153	0.726	HBA HBA HYDROPHOBIC
9	222.709	0.728	HBA HBA HBA
10	228.867	0.710	HBA HBA HBD

Configuration cost: 15.690

Fixed cost: 113.855

Null cost: 325.656

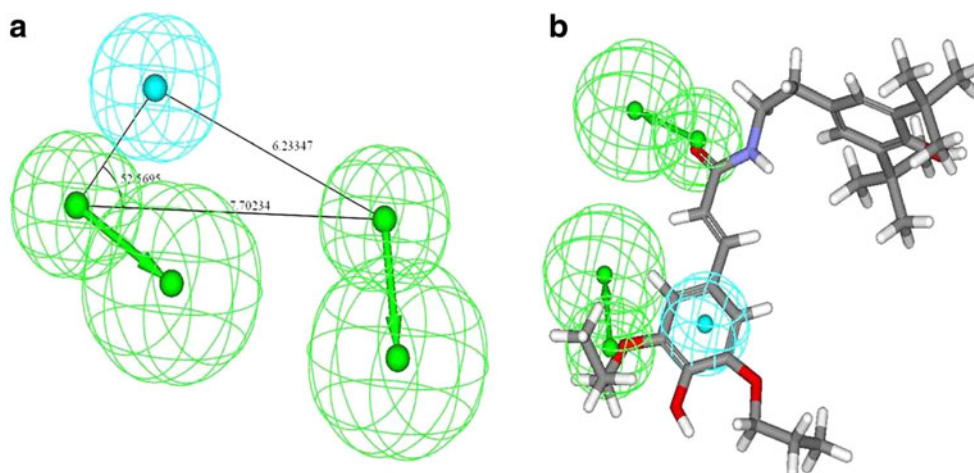
Again good overall correlation between the observed and calculated activity data was implicated from the correlation coefficients of the 10 hypotheses which ranged between 0.788–0.710. Moreover, a value of configuration cost much lower than the threshold of 17 accounts for the lack of any chance correlation in the developed hypotheses. Out of the 10 generated hypotheses, six revealed the importance of a combination of three pharmacophore features (HBA, HBA, HY). Two other hypotheses consisted of a different combination of three pharmacophore features (HBA, HBA, HBD). The remaining two comprised of a repetition of three hydrogen bond acceptor features (HBA, HBA, HBA). Further, the fitness of the developed pharmacophore models was checked using the Fischer validation technique at the 95% confidence level. For hypothesis 2, the value of the average correlation coefficient of the randomized models (R_r) was much lower than the corresponding correlation coefficient (R) of the unrandomized matrix ($R_r=0.461$, $R=0.756$), implicating the robustness of the developed model and denoting the existence of a true correlation. Further, the randomization results obtained for the cost functions showed that the total cost for hypothesis 2 was much closer to the fixed cost than that of the null cost as compared to values obtained based on the scrambled activity data. Thus, the randomization results additionally proved lack of any chance correlation for hypothesis 2. Subsequently, hypothesis 2 (Fig. 4) was selected as the best-ranking pharmacophore and was analyzed further.

Three different chemical features are displayed in hypothesis 2: HBA, HBA and HY. The hydrophobic feature maintains a distance of 6.233 Å from one of the hydrogen bond acceptor groups while the HBA features are separated from each other by 7.702 Å (Fig. 4a). The vectors for the HBA features indicate the direction of formation of the hydrogen bond between the electronegative atom of the antioxidant molecules and the electropositive hydrogen atom of the free radicals. Additionally, the hydrophobic feature denotes the regions favorable for substitution by

hydrophobic groups. The existence of hydrogen-bond acceptor groups in the developed pharmacophore indicates that the molecules function by the mechanism of single electron transfer followed by deprotonation [44]. Mapping of the most active compound (compound no. 24) with the developed pharmacophore (Fig 4b) revealed that the presence of ketonic and the ethereal oxygen atoms is well corroborated with the two hydrogen bond acceptor features that appeared in the 3D pharmacophore model developed in this work. The ethereal oxygen for the substituent at C₆ position of the parent nucleus and the =O fragment for the amide/carboxylic acid group constitute the primary hydrogen bond acceptor sites for interaction with the nocive free radicals. However, the presence of a hydroxy –OH fragment at the C₆ position could also map the HBA feature and thereby favored the activity profile of the molecules (compound nos. 42, 45, 47 and 48). Thus these features constitute the essential structural fragments for molecules exhibiting optimum anti-lipid peroxidative activity. Besides these, mapping of the parent phenyl moiety with the hydrophobic feature implies its significance in imparting the necessary biological activity to the molecules. Molecules bearing hydrophobic substituents develop an area of transient electron deficiency [45], which may in turn interact with nearby free radicals that have transient electron-rich areas. Compound nos. 20, 23, 24 and 25 bearing all the essential features were efficiently mapped by the pharmacophore model developed in hypothesis 2. On the contrary, compound nos. 1, 3, 5 and 12 lacking the necessary spatial orientation of the molecule could not fit accurately into the developed pharmacophore and thus exhibited poor activity profile. Again, the absence of the essential ethereal linkage in the parent moiety accounted for the poor mapping and subsequently reduced activity profile of compound no. 54.

The pharmacophore was further validated based on its external predictive ability. Here the remaining 13 test set

Fig. 4 Pharmacophore obtained from hypothesis 2 (a) monitoring the distances among the different features and (b) mapping the most active compound (compound no. 24) to the developed pharmacophore. [Shown are hydrophobic group (cyan) and hydrogen bond acceptor (green) features with vectors in the direction of putative hydrogen bonds]



molecules were mapped based on the model obtained in hypothesis 2. Based on the estimated activity data of the test set molecules, the value of predictive R^2 (R_{pred}^2) was determined. Existence of a good overall correlation between the observed and predicted activity data was reflected in the value of the R_{pred}^2 (0.568) parameter which was much higher than the threshold value of 0.5. Further, the proximity between the observed and the predicted activity data was checked based on the value of the $\overline{r_m^2(\text{test})}$ and Δr_m^2 metrics. Close fitness between the observed and predicted activity data is ensured more precisely by values of these parameters higher than the threshold of 0.5 and lower than the optimum value of 0.2 respectively compared to the traditional parameter of R_{pred}^2 . Acceptable values for these parameters ($\overline{r_m^2(\text{test})} = 0.516$ and $\Delta r_m^2 = 0.166$) for the selected hypothesis indicates that the pharmacophore obtained reflects statistical significance and bears improved external predictive potential.

Development of HQSAR model

The results of the HQSAR analysis are reported in Tables 5, 6 and 7. The analyses were performed based on the training set molecules and optimization of the fragment features and the hologram length was fixed based on the maximum Q^2 and minimum cross-validated standard error (SE_{cv}) for different PLS runs. Initially the best fragment combination [A (atom type), B (bond type) and D&A (donor and acceptor)] was selected using the default fragment length followed by the selection of the most suitable fragment size. The fragment size and the fragment combination thus optimized were utilized for the selection of the significant hologram length and the component number was optimized using the 5% rule. The “5% rule” was employed to reduce the noise and obtain a more robust model. The final model was obtained by repeating the analysis using the specific fragment contribution, fragment size (5-9), hologram length (307) and optimum component number (4). The model thus obtained was validated externally using the test set molecules and the activity predicted for the test set molecules

Table 5 HQSAR analysis for various fragment distinction using default fragment size (4-7)

Fragment distinction	Q^2	SE_{cv}	R^2	SE	LVs	Length
A/B	0.518	0.448	0.838	0.260	6	97
A/B/C	0.439	0.470	0.700	0.344	4	83
A/B/H	0.365	0.515	0.795	0.292	6	151
A/B/C/H	0.369	0.513	0.792	0.294	6	59
A/B/D&A	0.563	0.421	0.835	0.259	5	257
A/B/C/ D&A	0.510	0.439	0.782	0.293	4	257

Table 6 HQSAR analysis for the influence of various fragment sizes using the best fragment distinction (A/B/D&A)

Fragment size	Q^2	SE_{cv}	R^2	SE	LVs	Length
2-6	0.481	0.446	0.714	0.331	3	151
3-7	0.556	0.424	0.830	0.262	5	257
4-8	0.581	0.412	0.839	0.256	5	257
5-9	0.629	0.393	0.871	0.232	6	307
6-10	0.623	0.396	0.864	0.238	6	307
7-11	0.563	0.427	0.869	0.234	6	307
8-12	0.522	0.447	0.868	0.234	6	151

was correlated with the observed activity data thereby yielding a significantly high value for the R_{pred}^2 parameter (0.736). The acceptable values for the $\overline{r_m^2}$ metrics ($\overline{r_m^2(\text{test})} = 0.605$ and $\Delta r_m^2 = 0.168$) further confirmed a close proximity between the observed and predicted activity data. Existence of a high degree of external predictive potential and absence of any chance correlation was thus reflected from the acceptable values for all the internal and external predictive parameters.

The results of the HQSAR analysis are represented in the form of contribution map (Fig. 5) where the color of the atom or fragment determines its overall contribution to the activity profile of the molecules under study. The contributions of the different colors are listed as follows: (i) white color indicates an average contribution ranging from -0.097 to 0.102, (ii) red color denotes unfavorable contribution and ranges below -0.034, (iii) red-orange color also implicates similarly bad impact ranging between -0.034 to -0.020, (iv) yellow indicates a good contribution of 0.102 to 0.153 and (v) green signifies maximum contribution of 0.254 and above. For the purpose of discussion, contribution of the different fragments with respect to the most active compound (compound no. **24**) has been shown here. The green colored fragments indicating maximum contributing fragments of the parent moiety include: (i) the ethylene (-CH = CH-) linkage of the cinnamic acids/amides and (ii) the ketonic (=O) fragment of the acid/amide derivatives. Besides these, some of the methyl hydrogen atoms of the t-butyl fragment also contribute significantly to the enhanced activity profile of the molecules. The fragments contributing moderately (yellow colored) to the activity profile constitute C_5 atom of the parent moiety and the secondary carbon atom

Table 7 Selection of best model with less number of LVs using 5% rule

LVs	Q^2	SE_{cv}	R^2	SE	Length	% increase in Q^2
6	0.629	0.393	0.871	0.232	307	3.624
5	0.607	0.399	0.845	0.251	307	3.584
4	0.586	0.404	0.819	0.267	307	–

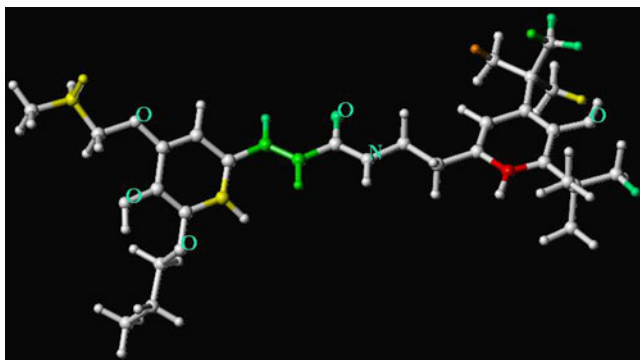
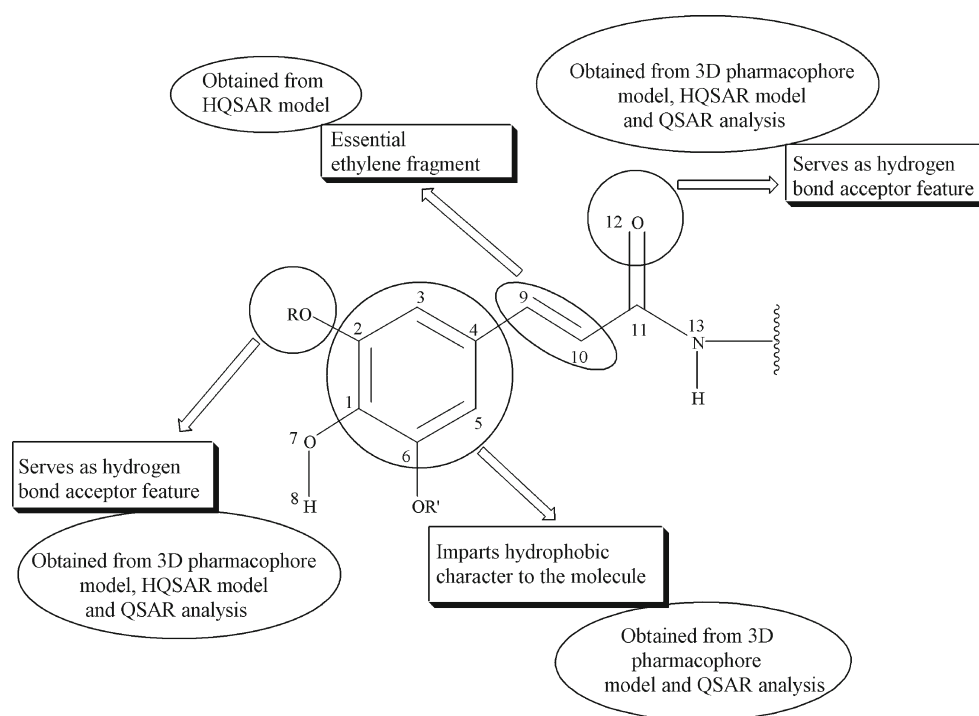


Fig. 5 Contribution map obtained using the HQSAR technique based on compound no. **24** (see text for details)

of the propoxy linkage attached at the C_2 position of the parent phenyl moiety. Red coloration of the carbon atom, *ortho* to one of the *t*-butyl groups, in phenyl ring signifies its worst impact toward activity profile of the compounds. Compound nos. **23**, **24**, **25**, **26** and **30** bearing all the necessary substitutions exhibit maximum antioxidant activity. Again, the high activity profile of compound nos. **42**, **45** and **48** is represented by their corresponding hologram map showing extensive contribution of the carbon atoms of the parent phenyl moiety of the cinnamic acid/ amide derivatives toward their anti-lipid peroxidative activity status. Compound nos. **1**, **3**, **5**, **6**, **9** (lacking the $-CH = CH-$ fragment in the parent nucleus), **12** (lacking the ene fragment of the parent moiety and bearing several unfavorable substitutions) and **54** (deficient in the essential substitutions) exhibit lowest range of activity profile.

Fig. 6 Schematic diagram showing different features at various positions favoring the anti-lipid peroxidative activity profile of the molecules obtained using different QSAR techniques



Overview and conclusions

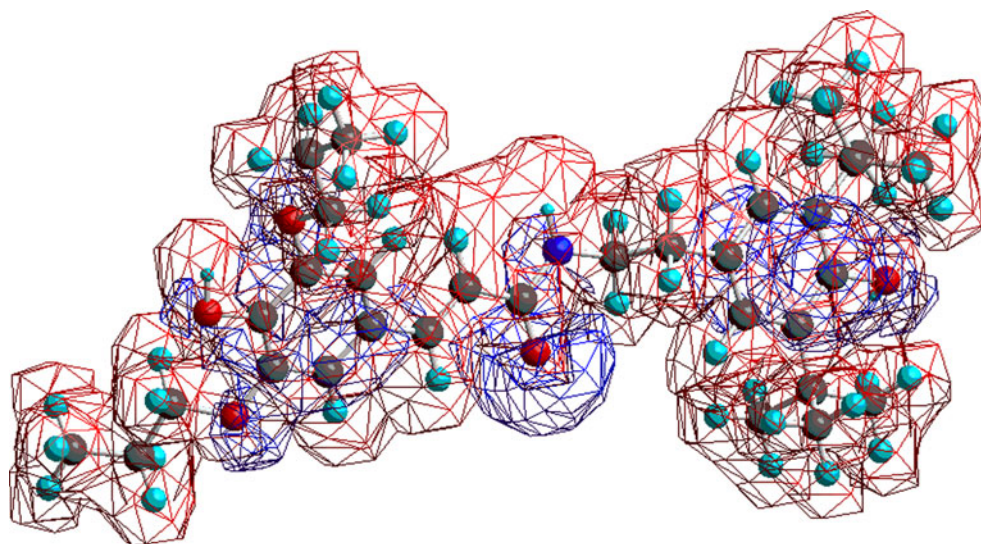
Based on the molecules belonging to the class of hydroxycinnamic acid and caffeic acid derivatives, QSAR models were developed to quantitatively define the relationship between molecular structure and their activity profile. Out of the total set of 54 compounds, 41 were utilized as the training set for model development and subsequent internal validation of the models. The remaining 13 molecules were classified as the test set for external validation of the developed models. The essential molecular fragments and their degree of contribution to the anti-lipid peroxidative activity of the molecules were determined from the HQSAR model. Additionally, the crucial features constituting the pharmacophore of the molecules were analyzed based on the 3D pharmacophore model developed. Although these models could qualitatively analyze the prime structural attributes, elucidation of the quantitative relationship between the molecular structure and their activity data was done using the descriptor based QSAR models developed using different chemometric tools (GFA, G/PLS). Among the different QSAR models developed in the present work, the GFA-spline model yielded the most satisfactory results for all the validation parameters and the inferences drawn from this model well corroborated with the HQSAR and 3D pharmacophore models. Figure 6 depicts the overall inferences drawn from the different models developed in this work. Both the HQSAR and the 3D pharmacophore models implicate the importance of the ketonic oxygen fragment of the acid/amide functional group attached to the ethylene

carbon which is in turn linked to the C₄ position of the phenyl ring. Thus the C₄ atom and the ethylene chain also constitute the essential structural features for the molecules so that the hydrogen bond acceptor fragment, =O, lies at an optimum distance from the remaining pharmacophoric features. The observation ideally matches with the interpretation of the QSAR model which signifies the importance of the charge on the C₄ atom while the HQSAR model emphasizes the utility of the ethylene fragment. The ketonic fragment constituting the hydrogen bond acceptor feature being in conjugation with the 9, 10-double bond enhances the electron transfer ability and the free radical scavenging action of the molecules through electron delocalization [46]. Molecular electrostatic potential surface (wire mesh) of the energy minimized geometry of the most active compound (**24**) is shown in Fig. 7 which indicates increased charge density over the =O (keto) fragment and the ethereal oxygen. Thus, it again reflects the ability of these fragments to serve as hydrogen bond acceptor groups. These fragments withdraw electrons from the rest of the molecule by delocalization mechanism decreasing the electron density on the rest of the molecular entity which in turn facilitates their interaction with the electron rich free radicals. Again, based on the interpretation of the HOMO descriptor appearing in the QSAR model, it may be further inferred that decrease in nucleophilicity of the molecules facilitates their interaction with the neighboring free radicals. The remaining hydrogen bond acceptor feature of the 3D pharmacophore maps with the ethereal oxygen linked at the C₂ position of the parent phenyl ring. Similar observation is also noted in the case of the HQSAR model that indicates the significant contribution of the substituent attached to the phenyl ring through ether linkage at the

C₂ position. Besides these, the parent phenyl ring mapping the hydrophobic feature serves as an electron deficient center that effectively interacts with the electron rich sites of the reactive free radicals and thereby facilitates their neutralization. The electron deficient nature of the phenyl ring is also reflected from the QSAR model that infers that a positive charge on the C₄ atom favors the activity of the molecules. In addition to these, the QSAR model also infers that an increase in the number of hydrogen fragments attached to a heteroatom and absence of sp³ hybridized carbon fragments substituted with a heteroatom are conducive for the anti-lipid peroxidative activity of the molecules. Further, a decrease in the number of hydrogen fragments attached to a tetrahedral carbon lacking any heteroatom substituent on the α-carbon results in an increment in the activity data of the molecules.

The QSAR models thus developed determine the crucial structural attributes of the molecules, belonging to the class of hydroxycinnamic acid and caffeic acid derivatives, which constitute their prime prerequisites in order to exhibit optimum lipid peroxidation inhibitory activity. All the models developed have been subjected to rigorous validation procedures. Acceptable internal validation parameters account for the high internal predictivity of the models, while improved external validation statistics for all the developed models implicate increased external predictive potential of the models. Thus the QSAR models may be efficiently utilized for prediction of lipid peroxidation inhibitory activity of untested molecules belonging to the class of cinnamic acid derivatives. Moreover, the 3D pharmacophore model and the HQSAR model developed in the present work may serve as efficient 3D query tools that may be utilized for screening of large databases and selecting the molecules bearing the necessary structural features for exhibiting optimal activity.

Fig. 7 Molecular electrostatic potential surface (wire mesh) of the energy minimized geometry of compound no. **24** (blue points in the surface indicate negatively charged areas and red points indicate positively charged areas)



Acknowledgments This research work is supported in the form of a major research project to KR by Indian Council of Medical Research (ICMR), New Delhi and a senior research fellowship to IM by Council of Scientific & Industrial Research (CSIR), New Delhi.

References

- Pacher P, Beckman JS, Liaudet L (2007) Nitric oxide and peroxynitrite in health and disease. *Physiol Rev* 87:315–424
- Dizdaroglu M, Jaruga P, Birincioglu M, Rodriguez H (2002) Free radical-induced damage to DNA: mechanisms and measurement. *Free Radic Biol Med* 32:1102–1115
- Dreher D, Junod AF (1996) Role of oxygen free radicals in cancer development. *Eur J Cancer* 32:30–38
- Leonard SS, Wang S, Shi X, Jordan BS, Castranova V, Dubick MA (2000) Wood smoke particles generate free radicals and cause lipid peroxidation, DNA damage, NFkB activation and TNF- α release in macrophages. *Toxicology* 150:147–157
- Halliwell B, Gutteridge JMC (1990) Role of free radicals and catalytic metal ions in human disease: an overview. *Method Enzymol* 186:1–85
- Winrow VR, Winyard PG, Morris CJ, Blake DR (1993) Free radicals in inflammation: second messengers and mediators of tissue destruction. *Br Med Bull* 49:506–522
- Leopold JA, Loscalzo J (2009) Oxidative risk for atherothrombotic cardiovascular disease. *Free Radic Biol Med* 47:1673–1706
- Frei B, Stocker R, Ames BN (1988) Antioxidant defenses and lipid peroxidation in human blood plasma. *Proc Natl Inst Sci India* 85:9748–9752
- Gey KF, Brubacher GB, Stahelin HB (1987) Plasma levels of antioxidant vitamins in relation to ischemic heart disease and cancer. *Am J Clin Nutr* 45:1368–1377
- Block G, Patterson B, Subar A (1992) Fruit, vegetables and cancer prevention: a review of the epidemiological evidence. *Nutr Cancer* 18:1–29
- Kritchevsky SB (1999) Beta-carotene, carotenoids and the prevention of coronary heart disease. *J Nutr* 129:5–8
- Buchwald P, Bodor N (2002) Computer-aided drug design: the role of quantitative structure–property, structure–activity and structure–metabolism relationships (QSPR, QSAR, QSMR). *Drugs Future* 27:577–588
- Todeschini R, Consonni V (2000) Handbook of molecular descriptors. Wiley-VCH, Weinheim
- Worachartcheewan A, Nantasenamat C, Isarankura-Na-Ayudhya C, Prachayasittikul S, Prachayasittikul V (2011) Predicting the free radical scavenging activity of curcumin derivatives. *Chemom Intell Lab Syst* 109:207–216
- Li YW, Li B, He J, Qian P (2011) Quantitative structure–activity relationship study of antioxidative peptide by using different sets of amino acids descriptors. *J Mol Struct* 998:53–61
- Nikitakis A, Kourounakis AP (2011) QSAR of substituted morpholines with antioxidant and squalene synthase inhibitory activity. *Med Chem Res* 20:566–575
- Mitra I, Saha A, Roy K (2011) Development of multiple QSAR models for consensus predictions and unified mechanistic interpretations of the free-radical scavenging activities of chromone derivatives. *J Mol Model* <http://dx.doi.org/10.1007/s00894-011-1198-x>
- Mitra I, Saha A, Roy K (2010) Pharmacophore mapping of arylamino-substituted benzo[b]thiophenes as free radical scavengers. *J Mol Model* 16:1585–1596
- Jung YS, Kang TS, Yoon JH, Joe BY, Lim HJ, Seong CM, Park WK, Kong JY, Chob J, Park NS (2002) Synthesis and evaluation of 4-hydroxyphenylacetic acid amides and 4-hydroxycinnamides as antioxidants. *Bioorg Med Chem Lett* 12:2599–2602
- Kang TS, Jo HO, Park WK, Kim JP, Konishi Y, Kong JY, Park NS, Jung YS (2008) Synthesis and antioxidant activities of 3,5-dialkoxy-4-hydroxycinnamides. *Bioorg Med Chem Lett* 18:1663–1667
- Rajan P, Vedemikova I, Cos P, Berghe DV, Augustyns K, Haemers A (2001) Synthesis and evaluation of caffeic acid amides as antioxidants. *Bioorg Med Chem Lett* 11:215–217
- Semichem Inc (2003) GaussView3.0, Gaussian Inc, Pittsburgh, PA
- Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA, Vreven JT, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2003) Gaussian 03, Revision B.05, Gaussian Inc, Pittsburgh, PA
- Accelrys Inc (2005) Cerius2 Version 4.10. Accelrys Inc. San Diego, CA
- Dougherty ER, Barrera J, Brun M, Kim S, Cesar RM, Chen Y, Bittner M, Trent JM (2002) Inference from clustering with application to gene-expression microarrays. *J Comput Biol* 9:105–126
- Everitt B, Landau S, Leese M (2001) Cluster analysis. Arnold, London
- SPSS Inc. (2011) SPSS. SPSS Inc, Chicago. <http://www.spss.com>
- Rogers D, Hopfinger AJ (1994) Application of genetic function approximation to quantitative structure–activity relationships and quantitative structure–property relationships. *J Chem Inf Comput Sci* 34:854–866
- Wold S, Sjostrom M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab Syst* 58:109–130
- Smellie A, Teig SL, Towbin P (1995) Poling: promoting conformational variation. *J Comput Chem* 16:171–187
- Accelrys Inc (2010) Discovery Studio 2.1. Accelrys Inc, San Diego, CA
- Sutter J, Guner OF, Hoffman R, Li H, Waldman M (2000) HypoGen: an automated system for generating 3D predictive pharmacophore models. In: Guner OF (ed) Pharmacophore perception, development, and use in drug design. International University Line, La Jolla, pp 501–511
- Tong W, Lowis DR, Perkins R, Chen Y, Welsh WJ, Goddette DW, Heritage TW, Sheehan DM (1998) Evaluation of quantitative structure–activity relationship methods for large-scale prediction of chemicals binding to the estrogen receptor. *J Chem Inf Comput Sci* 38:669–677
- Waller CL (2004) A comparative QSAR study using CoMFA, HQSAR, and FRED/SKEYS paradigms for estrogen receptor binding affinities of structurally diverse compounds. *J Chem Inf Comput Sci* 44:758–765
- SYBYL 7.3 (2006) Tripos Inc, St. Louis. www.tripos.com
- Wold S, Johansson E, Cocchi M (1993) PLS: partial least squares projections to latent structures. In: Kubinyi H (ed) 3D QSAR in drug design: theory methods and applications. ESCOM, Leiden, pp 523–550
- Ojha PK, Mitra I, Das RN, Roy K (2011) Further exploring r_m^2 metrics for validation of QSPR models. *Chemom Intell Lab Syst* 107:194–205
- Roy K, Mitra I, Kar S, Ojha P, Das RN, Kabir H (2012) Comparative studies on some metrics for external validation of QSPR models. *J Chem Inf Model* 52:396–408
- Golbraikh A, Tropsha A (2002) Beware of q^2 ! *J Mol Graph Model* 20:269–276
- Mitra I, Saha A, Roy K (2010) Exploring quantitative structure–activity relationship (QSAR) studies of antioxidant phenolic compounds

- obtained from traditional Chinese medicinal plants. *Mol Simul* 36:1067–1079
41. Toropova AP, Toropov AA, Martyanov SE, Benfenati E, Gini G, Leszczynska D, Leszczynski J (2012) CORAL: QSAR modeling of toxicity of organic chemicals towards *Daphnia magna*. *Chemom Intell Lab Syst* 110:177–181
 42. Dashtbozorgi Z, Golmohammadi H (2010) Prediction of air to liver partition coefficient for volatile organic compounds using QSAR approaches. *Eur J Med Chem* 45:2182–2190
 43. CORAL (2011) CORAL freeware available at <http://www.insilico.eu/coral>
 44. Wright JS, Johnson ER, Dilabio GA (2001) Predicting the activity of phenolic antioxidants: theoretical method, analysis of substituent effects, and application to major families of antioxidants. *J Am Chem Soc* 123:1173–1183
 45. Patrick GL (2009) *An introduction to medicinal chemistry*. Oxford University Press, New York
 46. Pietta PG (2000) Flavonoids as antioxidants. *J Nat Prod* 63:1035–1042