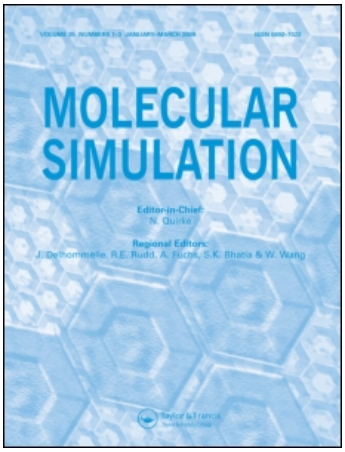


Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Exploring quantitative structure-activity relationship studies of antioxidant phenolic compounds obtained from traditional Chinese medicinal plants

Indrani Mitra[a]; Achintya Saha[b]; Kunal Roy[a]
[a] Division of Medicinal and Pharmaceutical Chemistry, Drug Theoretics and Cheminformatics Lab, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India [b] Department of Chemical Technology, University College of Science and Technology, University of Calcutta, Kolkata, India

Online publication date: 24 November 2010

To cite this Article Mitra, Indrani , Saha, Achintya and Roy, Kunal(2010) 'Exploring quantitative structure-activity relationship studies of antioxidant phenolic compounds obtained from traditional Chinese medicinal plants', Molecular Simulation, 36: 13, 1067 — 1079
To link to this Article: DOI: 10.1080/08927022.2010.503326
URL: http://dx.doi.org/10.1080/08927022.2010.503326

Molecular Simulation
Vol. 36, No. 13, November 2010, 1067–1079

Taylor & Francis
Taylor & Francis Group

Exploring quantitative structure–activity relationship studies of antioxidant phenolic compounds obtained from traditional Chinese medicinal plants

Indrani Mitra[a], Achintya Saha[b] and Kunal Roy[a]*

[a]*Division of Medicinal and Pharmaceutical Chemistry, Drug Theoretics and Cheminformatics Lab, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700032, India;* [b]*Department of Chemical Technology, University College of Science and Technology, University of Calcutta, 92, A.P.C. Road, Kolkata 700009, India*

(*Received 30 March 2010; final version received 19 June 2010*)

In the present work, quantitative structure–activity relationship (QSAR) models have been built for a wide variety of antioxidant phenolic compounds obtained from traditional Chinese medicinal plants, with their Trolox equivalent antioxidant capacity measured using 1,1-diphenyl-2-picrylhydrazyl (DPPH) radical and 2,2′-azinobis-(3-ethylbenzothiazoline-6-sulphonic acid) radical ($ABTS^{\cdot+}$) assay methods. Non-linear models obtained using genetic partial least-squares technique were acceptable both in terms of internal and external predictivity. Validation of developed models using r^2_m metrics and randomisation technique yielded results indicating the predictivity and robustness, respectively, of the developed models. The models signify that the presence of ketonic oxygen within the molecular structure favours their antioxidant activity. In addition, the number of hydroxyl groups, extent of branching, degree of methoxylation and the number of methyl and methylene substituents also dictate the antioxidant activity of these molecules. Thus, the QSAR models developed here can be utilised for the antioxidant activity prediction of a new series of molecules.

Keywords: antioxidants; phenolic compounds; QSAR; GFA; G/PLS

1. Introduction

Free radicals arise in the human body normally during metabolism [1]. Human cells purposely create them to neutralise viruses and bacteria. However, environmental factors such as pollution, radiation, cigarette smoke and herbicides [2] can also spawn free radicals. The free radicals thus produced in excess accumulate within the human system and participate in a series of biochemical events. Consequently, a cascade of chain reactions follow, resulting in an overload of reactive oxygen species. Polyunsaturated fatty acids present in abundance in cellular membranes and in low-density lipoproteins are prone to free radical-induced damage [3]. Such free radical damage within the cells has been linked with a range of oxidative stress-induced disorders including cancer, arthritis, atherosclerosis, Alzheimer's disease and diabetes. Free radical damage has also been implicated in phagocytosis, inflammation and apoptosis [4]. The large excess of reactive oxygen species generated within the mitochondria during respiration causes severe damage to the mitochondrial DNA, resulting in mutation. Extensive mitochondrial DNA damage accumulating over time shuts down mitochondria, causing cells to die and organisms to age [5].

Antioxidants are chemical entities that regulate free radical formation. These molecules function primarily by blocking free radical chain reaction, in addition to metal ion chelation. Since antioxidants are chiefly reducing agents, it can be inferred that hydrogen abstraction is their prime mechanism of action. Thus, at the molecular level, the mode of action of antioxidants can be explained based on the three different mechanisms: (a) hydrogen atom transfer; (b) single-electron transfer followed by proton transfer; and (c) sequential proton loss electron transfer [6,7]. The human system has its own antioxidant supply in the form of several endogenous enzymes [8], which are utilised to combat toxic free radicals. However, the surplus flow of these free radicals may saturate the available supply of systemic antioxidants and such a condition necessitates external antioxidant supplementation. Externally, antioxidants may be consumed either from food [9], as a regular diet, or in the form of medication. Although various fruits and vegetables are overloaded with several antioxidants, very few synthetic antioxidants have been reported to date. Hence, presently, a great deal of research has been directed to the design and development of antioxidant molecules with high efficacy and reduced toxicity.

Quantitative structure–activity relationship (QSAR) serves as an efficient tool for the design of new molecules with a definite response parameter. QSAR is a computational drug design approach which serves as the basis of rational drug design. QSAR is based on the notion that biological activity depends on the structure (C) and physico-chemical properties (P) of the molecules. It enables to determine the structural attributes required for

*Corresponding author. Email: kunalroy_in@yahoo.com

ISSN 0892-7022 print/ISSN 1029-0435 online

DOI: 10.1080/08927022.2010.503326
http://www.informaworld.com

maximal activity of the molecules under study [10]. This QSAR technique has been well utilised by several researchers for the designing of new antioxidant molecules with improved efficacy. QSAR models for antiradical and antioxidant activities of flavonoids have been developed by Ray and co-workers [11] using electrotopological state (E-state) atom parameters. Mitra and others [12] developed significant QSAR models for predicting antioxidant activities of hydroxybenzalacetone derivatives such as their ability to inhibit *t*-butyl hydroperoxide and γ-irradiation-induced lipid peroxidation and scavenge 1,1-diphenyl-2-picrylhydrazyl (DPPH) free radical. QSAR models were developed by Roy and co-workers [13] for antioxidant activity as well as squalene synthase inhibitory activities of aromatic tetrahydro-1,4-oxazine derivatives based on their molecular shape analysis. Pharmacophore mapping of arylamino-substituted benzo[b]thiophenes, as free radical scavengers, has also been carried out by Mitra and co-workers [14] in order to determine the key chemical features imparting antioxidant activity to this class of molecules. Recently, a review regarding QSAR of various classes of chemicals with antioxidant activity has also been reported by Roy and Mitra [15].

Since time immemorial, Chinese medicinal plants have been used for treating and preventing various chronic (cancer, atherosclerosis, ageing, etc.) and degenerative diseases (rheumatoid arthritis, etc.) [16]. Many of these plants have been used as food in a variety of countries since ancient days. Plants such as *Prunella vulgaris*, *Sanguisorba officinalis*, etc. are used in salad, and leaves of plants such as *Artemisia capillaris*, *Camellia sinensis*, *Sophora japonica*, etc. are used to produce tea. Again, *Morus alba* is used as a thickener in soup, bark of *Cinnamomum cassia* is used as a spice, while others are used for a variety of purposes in the preparation of food. It was suggested that phenolic antioxidants in these medicinal plants might play a significant role in combating many of these fatal diseases. In the present work, a variety of natural phenolic and flavonoid compounds (derived from traditional Chinese medicinal plants) with vivid structural features and possessing widely different antioxidant activities have been modelled for their ability to scavenge free radicals based on the QSAR technique. Different chemometric tools have been used for the development of QSAR models and the models thus developed have been validated based on various strategies. The models developed in this work emphasise on structural determinants regulating free radical scavenging activity of these molecules.

2. Method and materials

2.1 Quantitative structure–activity relationship

In the QSAR study, variations in biological activity (BA) within the compounds of a congeneric series are correlated with changes in measured or computed molecular features of the compounds [10]. These features measure properties such as hydrophobic, steric and electronic characteristics of the molecules which may influence their BA. In the data table of QSAR analysis, each row represents a candidate compound and each column represents an experimental or computational feature. Using the QSAR technique, initial models are built based on the available experimental data of a specific class of chemicals with a definite response (BA). The model thus built is validated using stringent statistical parameters and the final model thereby obtained is used for predicting the activity of a new series of chemicals belonging to the same class. Following the path laid by Hansch in 1962 [17] in the form of linear free energy-related models, QSAR has travelled a long way into the present date traversing multidimensional features of vivid classes of chemicals. Figure 1 shows a schematic representation of the QSAR model development methodology.

2.2 Data-set

The data-set comprises of a series of phenolic acids and flavonoids together with other phenolics such as quinines, coumarins, lignans, stilbenes, curcuminoids and tannins derived from Chinese medicinal plants bearing potent antioxidant activity as reported by Cai and co-workers [18]. The free radical scavenging activities (antioxidant activity) of these compounds were reported using both 1,1-diphenyl-2-picrylhydrazyl (DPPH) radical and 2,2′-azino-bis-(3-ethylbenzothiazoline-6-sulphonic acid) radical ($ABTS^{\cdot+}$) assay methods. Due to the lack of activity data for some of the compounds in the series, only a selected number of compounds were modelled for the two different response parameters. Out of the total series of 90 compounds, 87 compounds were modelled for their DPPH free radical scavenging activity (pC1), while 89 compounds were modelled for their $ABTS^{\cdot+}$ radical scavenging activity (pC2). Trolox equivalent antioxidant capacities (TEAC) of the compounds measured using the two methods were converted to negative logarithmic scale (molar) for the QSAR model development. All the molecules together with their observed and corresponding leave-one-out (LOO)-predicted/predicted activity data are listed in Table 1.

2.3 Descriptor calculation

The molecular structures of all the compounds were obtained from Merk Index [19] and were also downloaded from the site http://pubchem.ncbi.nlm.nih.gov/. Some of the structures were not available from the site and hence were drawn using Gauss View 4.10 software. The structures were then exported to the Cerius2 software version 4.10 [20] and subsequently descriptors were calculated. Prior to the calculation of 3D descriptors,

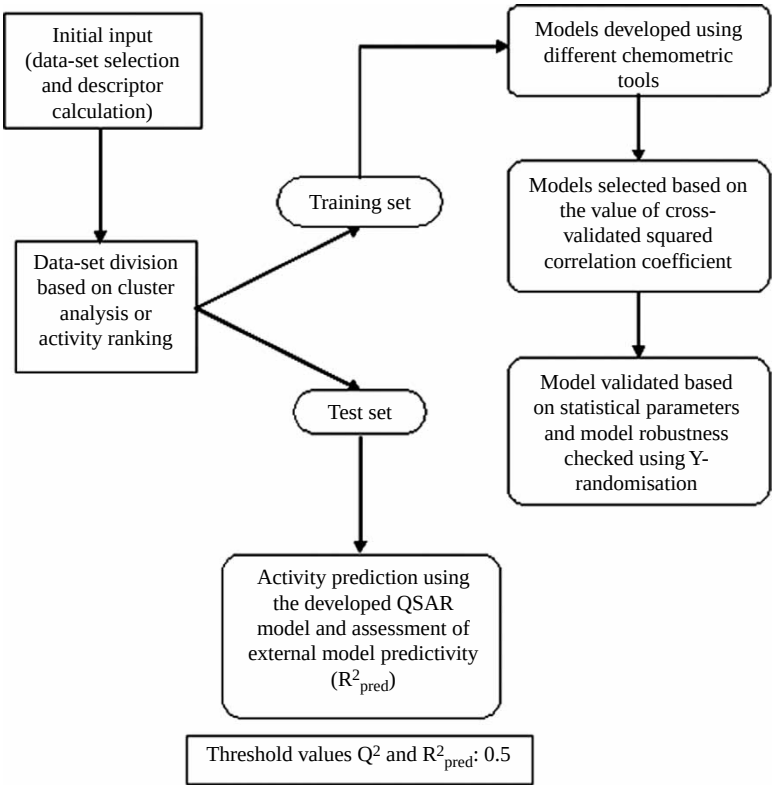


Figure 1. Schematic representation of the QSAR model development methodology.

conformational analysis of the molecules was performed using 'optimal search method' for conformer search with Cerius2 software. The lowest energy conformers were selected using the smart minimiser under open force field and the subsequent charge calculation of the lowest energy conformer was performed using the Gasteiger method [20]. A large number of descriptors belonging to a variety of categories were calculated using Descriptor+ module of the Cerius2 software version 4.10 [20]. The calculated topological indices include descriptors such as Wiener, Zagreb, Balaban J, connectivity indices ($^0\chi$, $^1\chi$, $^2\chi$, $^3\chi_P$, $^3\chi_C$, $^0\chi^v$, $^1\chi^v$, $^2\chi^v$, $^3\chi^v_P$, $^3\chi^v_C$), kappa shape indices ($^1\kappa$, $^2\kappa$, $^3\kappa$, $^1\kappa_{am}$, $^2\kappa_{am}$, $^3\kappa_{am}$) and E-state parameters. In addition, spatial (*Jurs* charged partial surface area descriptors and shadow indices), structural and physico-chemical descriptors were also calculated. After excluding those descriptors having variance lower than 0.0001, a total of 86 descriptors were chosen. The descriptors used in the present work are listed in Table S1 of the Supplementary Material, available online.

2.4 Selection of training and test sets

The data-set was divided into training and test sets, which were used for the development and validation, respectively, of the QSAR model. However, the training-set compounds are to be selected in such a way that they span the activity range of the entire data-set. The training set thus selected should capture all the features and chemical characteristics of the test-set compounds [21]. A test-set molecule that is very much alike to the training-set ones is predicted well, while a test-set molecule that is much different from the training-set molecules is poorly predicted. Since a single QSAR model can never predict the activity of an entire universe of chemicals, there should be sufficient structural similarity between the training and test-set molecules so as to obtain a high efficacy of prediction. In the present work, the splitting of the data-set was performed based on activity ranking of the whole data-set. Initially, all the compounds were sorted in ascending order of their activity data and thereafter every alternate compound was selected as a member of the training set. In order to ensure that the training set captures all the features of the test-set compounds, the whole data-set was split in such a way that 50% of the compounds were selected as the training set and the remaining 50% as the test set. For the development of QSAR models, different statistical tools were employed [22,23]. Non-linear models were built using spline option with two prime chemometric tools, namely, genetic function approximation (GFA) and genetic partial least-squares (G/PLS) techniques (discussed in detail in the Supplementary Material, available

Table 1. Observed and predicted activity data of all the compounds used for modelling.

Sl. no.	Compound	DPPH assay: TEAC (mM)	*p*C1[a] [logTEAC (M)]	*p*C1[b]	$ABTS^{\cdot+}$ assay: TEAC (mM)	*p*C2[a] [logTEAC (M)]	*p*C2[c]
01	Caffeic acid	1.24	2.907	2.600	1.31	2.883	3.178
02[d]	Chlorogenic acid	1.75	2.757	4.645	1.56	2.807	3.543
03	*o*-Coumaric acid	0.84	3.076	3.261	0.93	3.032	3.399
04[d]	*m*-Coumaric acid	0.75	3.125	2.854	0.82	3.086	3.258
05	*p*-Coumaric acid	1.44	2.842	2.850	1.96	2.708	3.255
06[d]	Ferulic acid	1.49	2.827	2.909	1.92	2.717	2.865
07	Isoferulic acid	1.24	2.907	2.917	1.53	2.815	2.811
08[d]	*trans*-Cinnamic acid	0.002	5.699	4.605	0.007	5.155	4.500
09	Gallic acid	3.92	2.407	2.597	3.52	2.453	3.394
10[d]	Protocatechuic acid	1.29	2.889	2.798	1.15	2.939	2.960
11	2,4-Dihydroxybenzoic acid	1.27	2.896	2.786	1.22	2.914	2.957
12[d]	*o*-Hydroxybenzoic acid	0.052	4.284	3.621	0.037	4.432	4.160
13	*m*-Hydroxybenzoic acid	0.069	4.161	3.249	0.025	4.602	3.766
14[d]	*p*-Hydroxybenzoic acid	0.059	4.229	3.325	0.028	4.553	3.769
15	Syringic acid	1.33	2.876	3.128	1.39	2.857	2.955
16[d]	Vanillic acid	0.056	4.252	3.072	0.092	4.036	3.157
17	Benzoic acid	0.006	5.222	5.093	0.005	5.301	4.704
18[d]	(−)-Epigallocatechin gallate (EGCG)	6.09	2.215	2.298	5.95	2.225	2.859
19	(−)-Epicatechin gallate	5.26	2.279	2.605	5.29	2.277	2.612
20[d]	(−)-Epigallocatechin	3.56	2.449	2.638	3.71	2.431	3.038
21	(−)-Epicatechin	3.18	2.498	2.940	3.08	2.511	2.631
22[d]	(+)-Catechin	2.95	2.530	2.908	3.04	2.517	2.564
23	Myricetin	1.38	2.860	2.514	1.31	2.883	2.989
24[d]	Quercetin	4.60	2.337	2.809	4.42	2.355	2.735
25	Morin	2.75	2.561	2.821	2.68	2.572	2.750
26[d]	Kaempferol	1.32	2.879	3.077	1.59	2.799	3.078
27	Galangin	0.705	3.152	3.361	1.12	2.951	2.805
28[d]	Quercetin-3-glucoside	2.16	2.666	3.580	2.39	2.622	3.022
29	Quercetin-3-rutinoside (rutin)	2.33	2.633	3.205	2.02	2.695	2.580
30[d]	Quercetin-3-rhamnoside (quercitrin)	2.57	2.590	2.714	2.18	2.662	3.110
31	Kaempferol-3-glucoside (astragalin)	0.151	3.821	3.766	0.138	3.860	3.284
32[d]	Quercetin-3-glucoside-7-rhamnoside	1.63	2.788	3.572	1.56	2.807	2.549
33	Flavonol	0.484	3.315	3.962	0.707	3.151	4.057
34[d]	Phloretin	1.24	2.907	2.797	1.79	2.747	3.235
35	Sappanchalcone	1.82	2.740	3.040	1.93	2.714	3.015
36[d]	Carthamin	1.36	2.866	3.223	1.43	2.845	3.185
37	*trans*-Chalcone	–	–	–	0.001	6.000	5.471
38[d]	Luteolin	2.24	2.650	3.072	2.18	2.662	3.342
39	Baicalein	2.74	2.562	3.402	2.56	2.592	3.109
40[d]	Apigenin	0.041	4.387	3.335	0.086	4.066	3.081
41	Chrysin	0.053	4.276	3.577	0.081	4.092	3.403
42[d]	Luteolin-7-glucoside	1.39	2.857	3.628	1.47	2.833	3.258
43	Apigenin-8-glucoside (vitexin)	0.205	3.688	3.847	0.216	3.666	4.196
44[d]	Apigenin-7-glucoside (apigetrin)	0.050	4.301	3.850	0.083	4.081	3.582
45	Baicalein-7-glucuronide (baicalin)	1.79	2.747	2.917	1.55	2.810	3.303
46[d]	Flavone	0.005	5.301	5.762	0.003	5.523	5.420
47	Naringenin	0.136	3.866	3.415	0.217	3.664	3.299
48[d]	Hesperetin	0.268	3.572	3.465	0.403	3.395	3.306
49	Naringenin-7-rutinoside (naringin)	0.077	4.114	3.527	0.098	4.009	4.235
50[d]	Hesperetin-7-rutinoside (hesperidin)	0.075	4.125	3.833	0.104	3.983	3.455
51	Flavanone	0.005	5.301	5.946	–	–	–
52[d]	Genistein	0.095	4.022	3.364	0.123	3.910	3.030
53	Daidzein	0.033	4.481	3.590	0.101	3.996	3.272
54[d]	Glycitein	0.020	4.699	3.677	0.097	4.013	3.561
55	Genistein-7-glucoside (genistin)	0.026	4.585	3.774	0.077	4.114	3.453
56[d]	Daidzein-7-glucoside (daidzin)	0.035	4.456	4.044	0.072	4.143	3.704
57	Isoflavone	0.001	6.000	5.693	0.005	5.301	5.269
58[d]	Procyanidin B-1 (dimer)	5.94	2.226	2.343	6.14	2.212	2.082
59	Procyanidin C-1 (trimer)	7.93	2.101	1.454	8.29	2.081	1.462
60[d]	Corilagin (monomeric ellagitannin)	6.98	2.156	3.420	7.76	2.110	2.219
61	Piceatannol	2.35	2.629	2.668	2.53	2.597	3.035

Table 1 – *continued*

Sl. no.	Compound	DPPH assay: TEAC (mM)	*p*C1[a] [logTEAC (M)]	*p*C1[b]	$ABTS^{\cdot+}$ assay: TEAC (mM)	*p*C2[a] [logTEAC (M)]	*p*C2[c]
62[d]	Resveratrol	1.71	2.767	2.895	2.14	2.670	2.923
63	*trans*-Stilbene	–	–	–	0.002	5.699	6.359
64[d]	Curcumin	2.02	2.695	3.196	2.24	2.650	3.619
65	Demethoxycurcumin	1.48	2.830	3.099	1.63	2.788	3.744
66[d]	Bisdemethoxycurcumin	1.02	2.991	2.942	1.18	2.928	3.698
67	Esculetin	2.08	2.682	3.360	2.38	2.623	3.029
68[d]	Scopoletin	0.210	3.678	4.674	0.383	3.417	3.735
69	Esculetin-6-glucoside	0.172	3.764	3.765	0.164	3.785	3.844
70[d]	5-Methoxyfuranocoumarin	0.002	5.699	5.342	0.003	5.523	3.706
71	Coumarin	–	–	–	0.001	6.000	5.234
72[d]	Matairesinol	0.207	3.684	3.912	0.253	3.597	3.408
73	Arctigenin	0.085	4.071	4.118	0.104	3.983	3.331
74[d]	Magnolol (neolignans)	0.197	3.706	5.032	0.209	3.680	3.608
75	Purpurin	2.00	2.699	3.550	1.93	2.714	3.645
76[d]	Pseudopurpurin	1.70	2.770	3.244	1.62	2.790	3.539
77	Alizarin	1.05	2.979	3.829	1.07	2.971	3.737
78[d]	Quinizarin	0.035	4.456	3.746	0.548	3.261	3.629
79	Emodin	0.029	4.538	3.486	0.095	4.022	3.162
80[d]	Chrysazin	0.057	4.244	3.746	0.068	4.167	3.675
81	Rhein	0.056	4.252	3.452	0.076	4.119	3.324
82[d]	Chrysophanol	0.042	4.377	3.870	0.069	4.161	3.586
83	Physcion	0.073	4.137	3.865	0.068	4.167	3.601
84[d]	Aloe-emodin	0.044	4.357	3.813	0.077	4.114	3.321
85	1,5-Dihydroxyanthraquinone	0.073	4.137	3.692	0.076	4.119	3.609
86[d]	2,6-Dihydroxyanthraquinone	0.076	4.119	3.748	0.072	4.143	3.687
87	Ruberythric acid	0.070	4.155	4.727	0.073	4.137	3.653
88[d]	Anthraquinone	0.002	5.699	5.573	0.009	5.046	5.747
89	Juglone	0.088	4.056	4.020	0.105	3.979	4.484
90[d]	Shikonin	0.083	4.081	3.463	0.124	3.907	3.721

[a] Observed activity of the compounds obtained from the data-set of Cai et al. [18]. [b] LOO-predicted/predicted activity obtained using Equation (4). [c] LOO-predicted/predicted activity obtained using Equation (5). [d] Compounds selected as test-set members.

online). The efficiency of the developed QSAR model is determined on the basis of several statistical parameters [24] (discussed in the Supplementary Material, available online).

2.5 *Validation*

Internal validation was performed using the leave-one-out (LOO) cross-validation technique. The model predictivity is then judged based on the value of cross-validated squared correlation coefficient (Q^2) and predicted residual sum of squares (PRESS) [25]. To determine the reproducibility of the model in predicting the activity of a new series of molecules of this class, the external predictivity of the model needs to be examined. A comparison of the observed activity values and the model predictions (activity calculation of molecules not included in the development of the model) is done through calculation of a parameter referred to as root mean square error in prediction (rmsep) [26]. Further external validation of the developed model is performed on the test set molecules (molecules not involved in the QSAR model development) based on the determination of two different external predictive parameters, namely $Q^2_{\text{ext (F1)}}$ and $Q^2_{\text{ext (F2)}}$ [27]. Still, another parameter ($Q^2_{\text{ext (F3)}}$) for the validation of a QSAR model has been proposed by Consonni and others [26]. The details of validation parameters have been discussed in the Supplementary Material, available online.

2.6 *Validation using the* r^2_m *metrics*

Traditionally used metric for expressing internal validation results is Q^2 and that for external validation results is predictive R^2 ($Q^2_{\text{ext (F1)}}$) [27]. The value of R^2_{pred} calculated using the external validation technique has been extensively used as the parameter for selecting the statistically significant QSAR model. However, a significantly large value of this parameter, much above the threshold value of 0.5, may not necessarily indicate that the predicted activity values lie in close proximity to the observed data. This may be explained by the fact that

the equation for calculating the value of R^2_{pred} bears a denominator with the term $Y_{obs\,(test)} - \bar{Y}_{training}$. This indicates that as the difference between the observed activity of the test-set compounds and the mean observed activity of the training-set compounds increases, the value of R^2_{pred} increases. If this difference is reasonably large, then the value of R^2_{pred} increases irrespective of the predicted activity values of the test-set compounds. Thus, there may be a considerable difference between the observed and predicted activity values of the test-set compounds, although they might maintain an overall good correlation among themselves. In order to nullify this error and to better indicate the predicted activity values of the test-set compounds, values of modified r^2 (r^2_m) [28] with the threshold value of 0.5 were calculated (given by Equation (1))

$$r^2_m = r^2\left(1 - \sqrt{\left(r^2 - r^2_0\right)}\right). \quad (1)$$

In Equation (1), r^2 and r^2_0 are the squared correlation coefficient values between the observed and predicted activity data [LOO predicted activity of the training-set compounds in the case of $r^2_{m\,(LOO)}$ and the predicted activity of the test-set compounds in the case of $r^2_{m\,(test)}$] with and without the intercept, respectively. As the above equation suggests that the value of r^2_m solely depends on the observed and predicted activity data of the molecules, hence, any large deviation between these values is well reflected through the r^2_m parameter. Similarly, based on the predicted activity values of both the training and test sets, values of $r^2_{m\,(overall)}$ were calculated.

2.7 *Randomisation*

The developed QSAR models were further validated using the randomisation technique in order to examine their robustness. The Y-randomisation technique proceeds with scrambling of the Y-column data, keeping the descriptor matrix (X-matrix) unchanged. Each time, the models are built using the scrambled data and the values of correlation coefficients are calculated. If the squared correlation coefficient of the original QSAR model (R^2) surpasses the average of the squared correlation coefficient of the randomised models (R^2_r), then the developed model might be considered to be robust enough. In the present work, two types of randomisation techniques were used, namely process randomisation and model randomisation performed at 90 and 99% confidence levels, respectively. In the process randomisation, Y-scrambling is performed with the whole descriptor matrix, while in the model randomisation, Y-scrambling is performed with the descriptors present in the developed QSAR model only. Thus, process randomisation examines the robustness of the model building process while the model randomisation examines the clarity of the developed QSAR model. However, no guideline is given as to how much the difference in the values of R^2 and R^2_r should be, for a model to be statistically reliable. Thus, to determine the extent of the difference in the values of R^2 and R^2_r that signifies the reliability of the developed QSAR model, we used another parameter named R^2_p [12,28]. The R^2_p parameter penalises the model R^2 for small differences in the values of R^2 and R^2_r. The threshold value of R^2_p is 0.5 and a QSAR model exceeding this stipulated value might be rightly considered to be robust and not the outcome of mere chance only. To express the value of R^2_p, so far we have used the following formula:

$$R^2_p = R^2\sqrt{\left(R^2 - R^2_r\right)}. \quad (2)$$

However, in an ideal case, the average value of R^2 for the randomised models should be zero, i.e. R^2_r should be zero. Consequently, in such a case, the value of R^2_p should be equal to the value of R^2 for the developed QSAR model. Thus, the corrected formula of R^2_p ($^cR^2_p$), as proposed by Todeschini [29], is given by:

$$^cR^2_p = R\sqrt{\left(R^2 - R^2_r\right)}. \quad (3)$$

3. Results and discussion

Different chemometric tools were employed for the development of the QSAR models. A detailed report of the statistical quality of various models is elaborated in Table 2. Among the various models developed, for both kinds of activities, the best results were obtained from those developed using the G/PLS technique. G/PLS was performed with 1000 iterations, scaled variables and with the option of no fixed length of equation. The maximum number of components or latent variables (LVs) fixed for variable selection was 4. These components are the functions of the original descriptors, and they encode data as represented by the descriptors. The models thus developed are non-linear and the spline terms are expressed as truncated power splines and denoted with angular brackets. For example, $\langle f(x) - a\rangle$ is equal to zero if the value of $f(x) - a$ is negative, else it is equal to $f(x) - a$. The constant a is called the knot of the spline. Although the G/PLS equations yielded the most significant results, the GFA equations obtained were also statistically significant in terms of external predictivity. The GFA models were developed using 5000 iterations considering both linear and spline options. However, for brevity, we have discussed here only the G/PLS models.

Table 2. Comparison of the statistical quality of the developed QSAR models.

Sl. no.	Statistical tool	Descriptors	LVs	$n_{training}$	s	R^2	R^2_a	PRESS	Q^2	$r^2_{m\,(LOO)}$	n_{test}	rmsep	$Q^2_{ext\,(F1)}$	$Q^2_{ext\,(F2)}$	$Q^2_{ext\,(F3)}$	$r^2_{m\,(test)}$	$r^2_{m\,(overall)}$
Results for DPPH free radical scavenging activity																	
1	GFA	*S_dO*, $^3\kappa_{am}$, *S_aasC*, $^0\chi^v$, ⟨*Jurs FNSA2* + 0.568525⟩	–	42	0.505	0.735	0.698	11.891	0.656	0.489	45	0.643	0.579	0.576	0.749	0.502	0.491
2	G/PLS	⟨*Jurs FNSA2* + 0.557925⟩, $^3\kappa$, $^2\chi^v$, *S_sOH* ⟨ − 0.235811 − *S_ssCH*$_2$⟩	4	42	0.482	0.752	0.725	10.965	0.683	0.642	45	0.674	0.537	0.533	0.447	0.501	0.564
3	FA-MLR	*AlogP98*, *S_dsCH*, *Sr*, HOMO, Dipole-mag, *Jurs RPSA*	–	42	0.634	0.593	0.523	19.424	0.438	0.347	45	0.619	0.609	0.605	0.533	0.531	0.431
4	FA-PLS	*Jurs FPSA3*, *Jurs RPSA*, *Jurs RASA*, Dipole-mag, *Sr*, *S_dsCH*, *AlogP98*	2	42	0.611	0.579	0.557	18.027	0.479	0.453	45	0.633	0.592	0.589	0.513	0.534	0.523
Results for ABTS^{+} free radical scavenging activity																	
1	GFA	*S_aasC*, *Sr*, *S_dsCH*, *S_dO*, ⟨*Jurs FNSA2* + 0.568525⟩	–	44	0.479	0.802	0.776	12.596	0.714	0.558	45	0.541	0.639	0.576	0.708	0.630	0.573
2	G/PLS	⟨58.3592 − *Jurs DPSA3*⟩, ⟨*Jurs TPSA* − 327.273⟩, *Sr*, *S_dO*, ⟨0.031901 − *S_dssC*⟩, ⟨2.0218 − *AlogP*98⟩	2	44	0.514	0.754	0.742	13.612	0.691	0.675	45	0.583	0.581	0.533	0.661	0.566	0.635
3	FA-MLR	*Jurs RPCG*, *Jurs TPSA*, *Jurs RASA*, LUMO, *Sr*, *S_dsCH*, *AlogP98*	–	44	0.644	0.642	0.595	21.233	0.518	0.407	45	0.566	0.605	0.605	0.681	0.589	0.456
4	FA-PLS	*Jurs RPCG*, *Jurs FPSA*3, *Jurs RPSA*, *Jurs RASA*, *Sr*, *S_sOH*, *AlogP98*, Hbonddonor	5	44	0.636	0.651	0.605	20.587	0.533	0.500	45	0.568	0.603	0.589	0.678	0.598	0.544

3.1 G/PLS model developed for DPPH free radical scavenging activity

$$
\begin{aligned}
p\mathrm{C1} &= 3.388 + 11.075 \times \langle \mathit{Jurs\ FNSA2} + 0.557925 \rangle \\
&\quad - 0.203 \times {}^{3}\kappa \ + 0.338 \times {}^{2}\chi^{\mathrm{v}} + 2.386 \\
&\quad \times \langle -0.235811 - \mathit{S\_ssCH_2} \rangle - 0.032 \times \mathit{S\_sOH} \\
n_{\mathrm{training}} &= 42,\ \mathrm{LVs} = 4,\ s = 0.482,\ R^2 = 0.752, \\
R^2_{\mathrm{a}} &= 0.725,\ F = 28.02\,(\mathrm{d}f\,4, 37), \\
\mathrm{PRESS} &= 10.965,\ Q^2 = 0.683,\ r^2_{\mathrm{m\,(LOO)}} = 0.642, \\
n_{\mathrm{test}} &= 45,\ Q^2_{\mathrm{ext\,(F1)}} = 0.537, \\
Q^2_{\mathrm{ext\,(F2)}} &= 0.533,\ Q^2_{\mathrm{ext\,(F3)}} = 0.447, \\
r^2_{\mathrm{m\,(test)}} &= 0.501,\ r^2_{\mathrm{m\,(overall)}} = 0.564
\end{aligned}
\tag{4}
$$

In the above equation, n is the number of compounds used for developing the model. The statistical significance of the developed model is reflected from the acceptable value of Q^2 (0.683) and $Q^2_{\mathrm{ext\,(F1)}}$ (0.537). In addition to these, the values of all the r^2_{m} metrics exceed the stipulated value of 0.5, inferring that the predicted activity values for these molecules, calculated using the above equation, are in close proximity to the experimental data. Quite identical values for the $Q^2_{\mathrm{ext\,(F1)}}$ and $Q^2_{\mathrm{ext\,(F2)}}$ parameters indicate that the test set selected for the QSAR model development captures features of the whole data-set. Thus, the model may be considered statistically significant and satisfactory for predicting the activity of a new set of such molecules and screening of potential antioxidant compounds. Moreover, the activity values of all the compounds calculated/predicted using Equation (4) were plotted against the observed activity data and the resulting graph (Figure 2(a)) showed that the points were minutely scattered about the line of fit. This again implicated the predictive efficacy of the developed QSAR model.

The ${}^{3}\kappa$ index refers to a molecular shape index proposed by Hall and Kier [30]. It encodes information about a specific kind of branching. The ${}^{3}\kappa$ values are larger when molecular branching is non-existent or when it is located at the extremities of a graph. According to the above equation, reduced values of this index favour the activity profile since the ${}^{3}\kappa$ index bears a negative coefficient. Compounds such as benzoic acid (**17**), flavanone (**51**) and isoflavone (**57**) with low-range values of the ${}^{3}\kappa$ index show a high order of activity. On the contrary, compounds such as (−)-epicatechin gallate (ECG) (**19**), quercetin-3-rutinoside (**29**), baicalein-7-glucuronide (**45**), procyanidin C-1 (**59**) and demethoxycurcumin (**65**) with molecular branching at the extremities bear large values of this index and consequently exhibit reduced DPPH free radical scavenging activity.

Jurs FNSA2 refers to fractional charged partial negative surface area and it is obtained by dividing the atomic charge weighted negative surface area (which is given by partial negative solvent-accessible surface area (SASA) multiplied by the total negative charge) by the total molecular SASA. According to Equation (4), the spline function with the *Jurs FNSA2* descriptor bears a positive coefficient and hence the activities of these molecules increase for any negative numerical value of the *Jurs FNSA2* descriptor lower than 0.557925 as in the case of compounds such as benzoic acid (**17**), flavanone (**51**) and isoflavone (**57**). This can be explained by the fact that a spline term exerts a zero contribution for any negative value of the corresponding spline function. On the contrary, the low-activity profile of compounds such as (−)-ECG (**19**), morin (**25**), baicalein-7-glucuronide (**45**) and procyanidin C-1 (**59**) can be attributed to the large negative values of the *Jurs FNSA2* descriptor. Thus, the equation implies that molecules with moderate negative charge on the respective atoms show desirable antioxidant activity.

The *S_ssCH₂* descriptor refers to the summation of E-state values for the —CH_2— fragments. A positive coefficient of the spline term bearing the *S_ssCH₂* descriptor indicates that an increment in activity is observed for any negative numerical value of the *S_ssCH₂* descriptor more than the knot of the spline, −0.235811. On the contrary, any positive value or zero value of the *S_ssCH₂* descriptor imparts zero contribution of the $\langle -0.235811 - \mathit{S\_ssCH_2} \rangle$ term and hence leads to a decline in the activity profile of these molecules. Thus, as the value of this descriptor gradually decreases from positive to more and more negative, the activity of the molecules gradually improves. Compounds

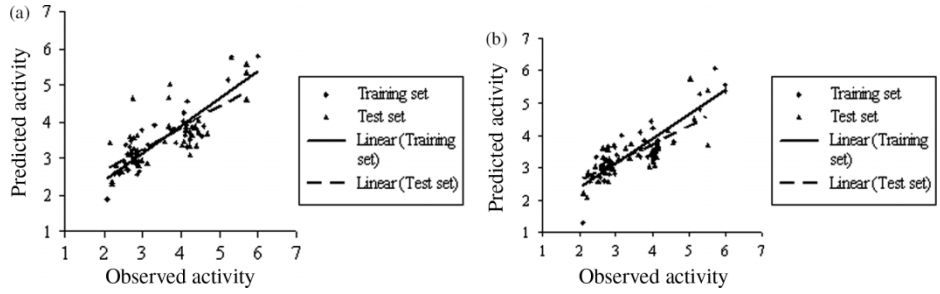


Figure 2. Observed vs. calculated/predicted activity plot for (a) G/PLS model developed using DPPH free radical scavenging activity and (b) G/PLS model developed using $ABTS^{\bullet+}$ free radical scavenging activity.

such as naringenin-7-rutinoside (**49**), genistein-7-glucoside (**55**) and ruberythric acid (**87**) with positive values of the $\langle -0.235811 - S\_ssCH_2 \rangle$ term show enhanced activity. This indicates that a reduction in the number of —CH_2— fragments favours activity of these molecules.

The *S_sOH* descriptor refers to the summation of E-state values for the —OH fragments. Since the *S_sOH* descriptor bears a negative coefficient, it can be inferred that a reduction in the value of this descriptor favours activity. The increased activity profile of compounds such as benzoic acid (**17**), flavanone (**51**) and isoflavone (**57**) may be explained by their reduced value of the *S_sOH* descriptor, while compounds such as quercetin-3-rutinoside (**29**), kaempferol-3-glucoside (**31**), apigenin-8-glucoside (**43**) and procyanidin C-1 (**59**) with significantly large values of this descriptor show reduced activity data.

The ${}^2\chi^v$ descriptor is a topological descriptor referring to the molecular valence connectivity index (χ) of the second order. It is summed over those bonds that are double-path fragments of the molecule. A positive coefficient of this descriptor suggests that the activity of the molecules increases with an increase in the value of this descriptor. Compounds such as flavanone (**51**), genistein-7-glucoside (**55**) and isoflavone (**57**) bearing moderately high values of the ${}^2\chi^v$ descriptor show higher antioxidant activity. Since the value of ${}^2\chi^v$ is inversely proportional to the number of hydrogen atoms bonded to the respective atoms, therefore, substitution at different positions raises the value of ${}^2\chi^v$ and hence favours activity.

However, the importance of the different descriptors depends on their respective order of significance. The five descriptors occurring in Equation (4) obey the following order of significance: (i) *S_sOH*, (ii) ${}^2\chi^v$, (iii) $\langle Jurs\ FNSA2 + 0.557925 \rangle$, (iv) $\langle -0.235811 - S\_ssCH_2 \rangle$, and (v) ${}^3\kappa$. However, compounds such as flavanone (**51**) and isoflavone (**57**) exhibit a very high activity profile despite bearing zero values of the $\langle -0.235811 - S\_ssCH_2 \rangle$ term since this descriptor bears a much lower order of significance compared to the remaining parameters. The moderate activity of compounds such as chrysin (**41**), daidzein (**53**), physcion (**83**) and 1,5-dihydroxyanthraquinone (**85**) can be explained by their acceptable values for the ${}^2\chi^v$ and ${}^3\kappa$ descriptors, despite the fact that these compounds have non-acceptable values of the remaining three descriptors. Compound such as procyanidin C-1 (**59**), although bearing the highest value for the ${}^2\chi^v$ descriptor (15.652) and a reasonably acceptable value for the $\langle -0.235811 - S\_ssCH_2 \rangle$ descriptor, exerts the lowest activity due to the very large unacceptable value of the most significant descriptor *S_sOH* (166.278). In spite of bearing moderate values for the ${}^2\chi^v$ descriptor, compounds such as (−)-ECG (**19**), morin (**25**), baicalein (**39**), baicalein-7-glucuronide (**45**) and purpurin (**75**) exhibit a low range of activity due to the lack of the structural requisites required for attaining acceptable values for most of the significant descriptors.

The model developed for DPPH free radical scavenging activity of these molecules thus infers the optimal structural requirements to exhibit potent antioxidant activity. Equation (4) signifies the importance of the topological descriptors, indicating that the degree of branching about the parent moiety and the length of the molecules chiefly determine their free radical scavenging activity. Again, the appearance of the *Jurs FNSA2* term implies that the activity of these molecules also depends on the number of negatively charged atoms present within their molecular structure.

3.2 G/PLS model developed for ABTS radical scavenging activity

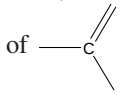
$$\begin{aligned} p\mathrm{C2} &= 2.470 + 0.097 \times \langle 58.3592 - Jurs\,DPSA3 \rangle \\ &\quad - 0.004 \times \langle Jurs\,TPSA - 327.273 \rangle \\ &\quad + 0.286 \times Sr + 0.037 \times S\_dO \\ &\quad - 0.231 \times \langle 0.031901 - S\_dssC \rangle + 0.175 \\ &\quad \times \langle 2.0218 - A\log P98 \rangle \\ n_{\mathrm{training}} &= 44,\ \mathrm{LVs} = 2,\ s = 0.514,\ R^2 = 0.754, \\ R^2_{\mathrm{a}} &= 0.742,\ F = 62.94\,(\mathrm{d}f\,2, 41), \\ \mathrm{PRESS} &= 13.612,\ Q^2 = 0.691,\ r^2_{\mathrm{m\,(LOO)}} = 0.675, \\ n_{\mathrm{test}} &= 45,\ Q^2_{\mathrm{ext\,(F1)}} = 0.581,\ Q^2_{\mathrm{ext\,(F2)}} = 0.577, \\ Q^2_{\mathrm{ext\,(F3)}} &= 0.661,\ r^2_{\mathrm{m\,(test)}} = 0.566,\ r^2_{\mathrm{m\,(overall)}} = 0.635 \end{aligned} \tag{5}$$

In the above equation, acceptable values for internal $[Q^2 = 0.691]$ and external $[Q^2_{\mathrm{ext\,(F1)}} = 0.581]$ predictive parameters account for excellent ability of the developed QSAR model to predict new molecules with potent free radical scavenging activity. Further calculation of the r^2_{m} parameters signifies that the activity predicted based on Equation (5) does not deviate much from the corresponding activity as observed experimentally. This in turn ensures the accuracy of prediction of the above developed QSAR model. The fact that the test set selected spans the characteristics of the total data-set is reflected through the close values of the $Q^2_{\mathrm{ext\,(F1)}}$ and $Q^2_{\mathrm{ext\,(F2)}}$ parameters. Similar to the previous model [Equation (4)], here also, a plot of estimated/predicted activity values (for both the training and test sets) against the observed activity data yielded a straight line with points well predicted about the line (Figure 2(b)). This further suggests the predictive potential of the developed QSAR model.

Jurs DPSA3 descriptor refers to the difference between atomic charge weighted surface area and is calculated as the difference between atomic charge weighted positive SASA (*Jurs PPSA3* is the sum of the products of atomic

SASAs and partial charges over all positively charged atoms) and atomic charge weighted negative SASA (*Jurs PNSA3* is the sum of the products of atomic SASAs and partial charges over all negatively charged atoms). The presence of a positive coefficient for the spline term with the *Jurs DPSA3* descriptor indicates that a reduction in the value of this descriptor below the stipulated value of 58.3592 favours the radical scavenging activity of these molecules. This infers that molecules with higher negative SASA, i.e. having a higher number of negatively charged atoms, exhibit increased activity profile. Compounds such as benzoic acid (**17**), *trans*-chalcone (**37**), isoflavone (**57**), *trans*-stilbene (**63**) and coumarin (**71**) bear a low range of values for the *Jurs DPSA3* descriptor and hence are highly active, while the lowest activity of procyanidin C-1 (**59**) may be attributed to the maximum value of this descriptor.

Jurs TPSA descriptor refers to the sum of SASAs of atoms with absolute value of partial charges greater than or equal to 0.2. In Equation (5), the $\langle Jurs\,TPSA - 327.273\rangle$ descriptor bears a negative coefficient inferring that a value of the *Jurs TPSA* descriptor below 327.273 favours the antioxidant activity of these molecules due to zero contribution of the spline function. Thus, it may be suggested that a lower range of partial charge over the individual atoms is conducive to the activity profile of the molecules. Compounds such as kaempferol-3-glucoside (**31**), apigenin-8-glucoside (**43**) and naringenin-7-rutinoside (**49**) with very low range values of the spline term show moderate activity profile, while compounds such as quercetin-3-rutinoside (**29**) and procyanidin C-1 (**59**) exhibit minimum activity due to significantly larger values of this spline term. In addition, zero values of this spline term account for the maximum activity range of compounds such as benzoic acid (**17**), *trans*-chalcone (**37**), isoflavone (**57**), *trans*-stilbene (**63**) and coumarin (**71**).

Similarly, a negative coefficient of the $\langle 0.031901 - S\_dssC\rangle$ descriptor infers that as the negative numerical value of the *S_dssC* descriptor decreases, the activity profile of the molecules improves and for any positive value of the *S_dssC* descriptor greater than the optimum value of 0.031901, the antioxidant activity of the molecules increases as in the case of compounds such as *trans*-chalcone (**37**) and isoflavone (**57**). As the value of this descriptor increases from negative to positive, an increase in antioxidant activity of the molecules is observed. Now, the *S_dssC* descriptor refers to the summation of E-state values for the  fragments. Thus, the equation suggests that the increase in the number of  fragments favours activity until it reaches an optimum numbers and beyond that, any value of the *S_dssC* descriptor exerts zero contribution to the activity profile of the molecules.

On the contrary, positive coefficients of the *Sr* and *S_dO* descriptors suggest that the antioxidant activity of these molecules increases with an enhancement in the values of these descriptors. *Sr* is an index of reactivity in aromatic hydrocarbons and is referred to as superdelocalisability of the molecules. *Sr* gives a metric for electrophilicity of the molecules and may be used to predict their relative reactivity. Consequently, proper substitution of the parent moiety determines the electronic environment of the molecules and thus controls their reactive ability with the neighbouring free radicals through proper overlapping of the molecular orbital. The high range of *Sr* values accounts for the high activity profile of compounds such as *m*-hydroxybenzoic acid (**13**), *trans*-chalcone (**37**), isoflavone (**57**) and *trans*-stilbene (**63**). The *S_dO* descriptor refers to the summation of E-state values for the =O fragments. Hence, according to Equation (5), an increase in the number of ketonic oxygen (=O) fragments within the molecular structure leads to an increase in their antioxidant activity profile. Compounds such as (−)-epicatechin (EC) (**21**), procyanidin C-1 (**59**) and piceatannol (**61**) exhibit reduced activity range due to the absence of =O in their molecular structure, indicating zero contribution of the *S_dO* descriptor, while compounds such as benzoic acid (**17**), *trans*-chalcone (**37**), isoflavone (**57**) and coumarin (**71**) with higher values for the *S_dO* descriptor show enhanced antioxidant activity profile.

The *AlogP98* descriptor refers to the log of the partition coefficient calculated using latest parameters (Ghose, Viswanadhan & Wendoloski, 1998). The *logP* descriptor measures a balance between the lipophilic and hydrophilic character of a molecule which determines the ability of the compound to penetrate the biological membrane barriers and reach its target site. In Equation (5), the *AlogP98* descriptor is present within a spline term bearing a positive coefficient. Since a negative value of a spline term corresponds to zero contribution of that term, a value of this descriptor below the knot of the spline (2.0218) or any negative value is conducive for potential antioxidant activity of these molecules as in the case of compounds such as *m*-hydroxybenzoic acid (**13**), benzoic acid (**17**) and coumarin (**71**).

A relative order of significance of the descriptors occurring in Equation (5) reveals that the descriptors can be arranged according to the following order: (i) $\langle 58.3592 - Jurs\,DPSA3\rangle$, (ii) *S_dO*, (iii) *Sr*, (iv) $\langle Jurs\,TPSA - 327.273\rangle$, (v) $\langle 0.031901 - S\_dssC\rangle$, and (vi) $\langle 2.0218 - AlogP98\rangle$. Compounds such as *m*-hydroxybenzoic acid (**13**), benzoic acid (**17**), *trans*-chalcone (**37**), isoflavone (**57**), *trans*-stilbene (**63**) and coumarin (**71**) exhibit the highest activity range since they fulfil the requisites of most of the significant descriptors. Com-

Table 3. Results of randomisation tests for the developed QSAR models.

Activity	Statistical tool	R	R^2	R_r	R_r^2	${}^cR_p^2$
Process randomisation						
*p*C1	G/PLS	0.867	0.752	0.6	0.360	0.543
*p*C2	G/PLS	0.868	0.754	0.585	0.342	0.557
Model randomisation						
*p*C1	GFA	0.857	0.735	0.332	0.110	0.678
	G/PLS	0.867	0.752	0.089	0.008	0.748
	FA-MLR	0.770	0.593	0.373	0.139	0.519
	FA-PLS	0.761	0.579	0.065	0.004	0.577
*p*C2	GFA	0.896	0.802	0.315	0.099	0.751
	G/PLS	0.868	0.754	0.041	0.002	0.753
	FA-MLR	0.801	0.642	0.314	0.099	0.591
	FA-PLS	0.807	0.651	0.268	0.072	0.614

pounds such as emodin (**79**), rhein (**81**), physcion (**83**), 1,5-dihydroxyanthraquinone (**85**) and ruberythric acid (**87**) despite bearing a zero value for the most significant descriptor, ⟨58.3592 − *Jurs DPSA3*⟩, show moderate activity profile due to acceptable values of most of the remaining descriptors. Compound such as piceatannol (**61**), although bearing a high value of the *Sr* descriptor, shows a low-activity range since it fails to satisfy the requirements of the most significant descriptors. In spite of the presence of the ketonic oxygen fragment, (−)-ECG (**19**) shows low activity due to non-acceptable values of the remaining descriptors. The lowest activity of procyanidin C-1 (**59**) may be attributed to the lack of structural necessities needed to attain an acceptable value for the significant descriptors.

Thus, the model developed for the ABTS free radical scavenging activity infers that in addition to the number of negatively charged atoms and ketonic oxygen fragments present within the molecular structure, the electrophilicity of the molecules plays the prime role for the proper interaction of these molecules with the neighbouring free radicals. The presence of an optimum number of methylene substituents adds to the activity profile of these molecules. Additionally, the partition coefficient of the molecules also controls their free radical scavenging activity by enabling the molecules to penetrate the biological membrane and reach the site of action.

3.3 Validation by randomisation technique

Further validation of the models was carried out using the randomisation technique. In order to examine the reliability of the process employed for developing the QSAR model, process randomisation was performed at 90% confidence level. In addition to this, the robustness of the developed model was ensured using the model randomisation technique at 99% confidence level. The R_r^2 values obtained for the two methods (process and model randomisations) were lower than the corresponding R^2 values of the developed QSAR models for both types of activity (*p*C1 and *p*C2). However, in order to ensure whether a significant difference exists between the two values so as to consider the model as a robust one, values of corrected R_p^2 (${}^cR_p^2$) were also calculated. The value of ${}^cR_p^2$ penalises the model R^2 for small differences in the values of R^2 and R_r^2. The results obtained for both process and model randomisation techniques are tabulated in Table 3. The process randomisation was performed for the G/PLS models and the results show that both the G/PLS models satisfactorily exceed the stipulated value for ${}^cR_p^2$. In the case of the model randomisation for both DPPH and ABTS free radical scavenging activities, maximum values of ${}^cR_p^2$ were obtained for Equation (4) (0.748) and Equation (5) (0.753), respectively. Thus, it can be inferred that the two QSAR models (Equations (4) and (5)) developed in the present work are robust enough and not the outcome of chance only.

4. Overview and conclusion

A wide variety of representative phenolic compounds identified from traditional Chinese medicinal plants, as reported by Cai et al. [18] were used for modelling their free radical scavenging activity. Two types of activity have been modelled in the present work: DPPH (*p*C1) and ABTS (*p*C2) radical scavenging activities. Due to the lack of available data for some of the compounds, out of the 90 data-set compounds, 87 and 89 compounds were modelled for the respective type of response. The data-set was split in such a way that each of the training and test sets constituted 50% of the total data-set compounds. Among the various QSAR models built in the present work, the models developed using the G/PLS technique were the most acceptable ones in terms of overall predictive ability. The G/PLS model developed with the DPPH radical scavenging activity yielded statistically significant values for the internal and external predictive parameters ($Q^2 = 0.683$, $Q^2_{\text{ext(F1)}} = 0.537$), together with acceptable values for the r_m^2 parameters. Similarly, the QSAR model developed using the G/PLS technique for ABTS radical scavenging activity also revealed acceptable values of both the internal

($Q^2 = 0.691$) and external predictive [$Q^2_{\text{ext (F1)}} = 0.581$] parameters as well as the r^2_{m} metrics. Moreover, quite identical values for the two external validation parameters ($Q^2_{\text{ext (F1)}}$ and $Q^2_{\text{ext (F2)}}$) in both the cases imply that the training and test sets selected for the present work span the features of the whole data-set molecules.

The developed QSAR models suggest the prime structural requirements for potent antioxidant activity of these molecules. As the equations explain, a reduction in the number of hydroxyl groups to an optimum value and the addition of ketonic groups favour the antioxidant activity profile of these molecules. The equations also suggest that excessive branching at the extremities of the molecule and a number of methyl (—CH_2—) fragments beyond the optimum value reduce the radical scavenging activity of these molecules. On the contrary, an increase in the number of methylene fragments favours the antioxidant activity. Replacement of the hydrogen atoms with specific substituents at different positions of the molecular structure leads to the design of systemically active antioxidant molecules. Moreover, the addition of hydrophobic substituents maintains the partition coefficient value of the molecules at an optimum range and hence facilitates permeation of the molecules through the systemic membrane barriers. In addition to these, the charged spatial descriptors indicate that an optimum number of negatively charged atoms throughout the molecular structure add to the antioxidant activity values of these molecules. These observations are in good agreement with the work carried out by Cai et al. [18], which suggest that the differences in the radical scavenging activity of these molecules are attributed to the variations in hydroxylation, glycosylation and methoxylation. For the designing of new molecules with potent activity, the parent moiety may be substituted with ketonic groups at favourable positions. Incorporation of extensive branching at the extremities of these molecules and the addition of methylene fragments to the parent moiety may result in newly designed molecules with improved activity. However, the substitution of the molecules should be such that the compounds have low partition coefficient values. Thus, the QSAR models built in the present work may be satisfactorily utilised for screening and designing of a new series of antioxidant molecules with increased efficacy and potency. These models focus more systemically and quantitatively on the scientific understanding regarding the improvisation of newer molecules with antioxidant activity.

Acknowledgements

This work was supported in the form of a major research project to K.R. and a senior research fellowship to I.M. by the Indian Council of Medical Research (ICMR), New Delhi.

References

[1] M. Genestra, *Oxyl radicals, redox-sensitive signaling cascades and antioxidants*, Cell Signal. 19 (2007), pp. 1807–1819.
[2] W. Dröge, *Free radicals in the physiological control of cell function*, Physiol. Rev. 82 (2002), pp. 47–95.
[3] R.C. Murphy, *Free-radical-induced oxidation of arachidonoyl plasmalogen phospholipids: Antioxidant mechanism and precursor pathway for bioactive eicosanoids*, Chem. Res. Toxicol. 14 (2001), pp. 463–472.
[4] G.M. Rosen, S. Pou, C.L. Ramos, M.S. Cohen, and B.E. Britigan, *Free radicals and phagocytic cells*, FASEB J. 9 (1995), pp. 200–209.
[5] T. Ozawa, *Oxidative damage and fragility of mitochondrial DNA*, in *Understanding the Process of Aging*, E. Cadenas and L. Packer, eds, Marcel Dekker, New York, 1999, pp. 265–292.
[6] J.S. Wright, E.R. Johnson, and G.A. DiLabio, *Predicting the activity of phenolic antioxidants: Theoretical method, analysis of substituent effects, and application to major families of antioxidants*, J. Am. Chem. Soc. 123 (2001), pp. 1173–1183.
[7] M. Musialik and G. Litwinienko, *Scavenging of DPPH˙ radicals by vitamin E is accelerated by its partial ionization: The role of sequential proton loss electron transfer*, Org. Lett. 7 (2005), pp. 4951–4954.
[8] L.A. Kirshenbaum, M. Hill, and P.K. Singal, *Endogenous antioxidants in isolated hypertrophied cardiac myocytes and hypoxia-reoxygenation injury*, J. Mol. Cell Cardiol. 27 (1995), pp. 263–272.
[9] J. Pokorný, *Natural antioxidants for food use*, Trends Food Sci. Technol. 2 (1991), pp. 223–227.
[10] A.M. Helguera, R.D. Combes, M.P. Gonzalez, and M.N. Cordeiro, *Applications of 2D descriptors in drug design: A DRAGON tale*, Curr. Top. Med. Chem. 8 (2008), pp. 1628–1655.
[11] S. Ray, C. Sengupta, and K. Roy, *QSAR modeling of antiradical and antioxidant activities of flavonoids using electrotopological state (E-State) atom parameters*, Cent. Eur. J. Chem. 5 (2007), pp. 1094–1113.
[12] I. Mitra, A. Saha, and K. Roy, *Quantitative structure–activity relationship modeling of antioxidant activities of hydroxybenzalacetones using quantum chemical, physicochemical and spatial descriptors*, Chem. Biol. Drug Des. 73 (2009), pp. 526–536.
[13] K. Roy, I. Mitra, and A. Saha, *Molecular shape analysis of antioxidant and squalene synthase inhibitory activities of aromatic tetrahydro-1,4-oxazine derivatives*, Chem. Biol. Drug Des. 74 (2009), pp. 507–516.
[14] I. Mitra, A. Saha, and K. Roy, *Pharmacophore mapping of arylamino-substituted benzo[b]thiophenes as free radical scavengers*, J. Mol. Model. (2010), in press, DOI: 10.1007/s00894-0010-0661-4.
[15] K. Roy and I. Mitra, *Advances in quantitative structure–activity relationship models of antioxidants*, Expert Opin. Drug Discov. 4 (2009), pp. 1157–1175.
[16] Q.M. Bo, Z.Y. Wu, Q.S. Shun, X.S. Bao, Z.S. Mao, S.Q. Ha, S.Y. Lu, and J.M. Huang, *A Selection of the Illustrated Chinese Anti-Cancer Herbal Medicines*, Shanghai Science and Technology Literature Press, Shanghai, 2002.
[17] C. Hansch, P.P. Maloney, T. Fujita, and R.M. Muir, *The correlation of the biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients*, Nature 194 (1962), pp. 178–180.
[18] Y.Z. Cai, M. Sun, J. Xing, Q. Luo, and H. Corke, *Structure–radical scavenging activity relationships of phenolic compounds from traditional Chinese medicinal plants*, Life Sci. 78 (2006), pp. 2872–2888.
[19] S. Budarari, M.J. O'Neil, and P.E. Heckelman, *The Merck Index*, Merck and Co., Rahway, NJ, 1989.
[20] *CERIUS2 Version 4.10*, Accelrys Inc. San Diego, CA, USA; available at http://www.accelrys.com.
[21] P. Gramatica, *Principles of QSAR models validation: Internal and external*, QSAR Comb. Sci. 26 (2007), pp. 694–701.
[22] D. Rogers and A.J. Hopfinger, *Application of genetic function approximation to quantitative structure–activity relationship and*

quantitative structure–property relationship, J. Chem. Inf. Comput. Sci. 34 (1994), pp. 854–866.

[23] S. Wold, M. Sjostrom, and L. Eriksson, *PLS-regression: A basic tool of chemometrics*, Chemom. Intell. Lab. Syst. 58 (2001), pp. 109–130.

[24] G.W. Snedecor and W.G. Cochran, *Statistical Methods*, Oxford & IBH, New Delhi, 1967.

[25] S. Wold and L. Eriksson, *Validation tools*, in *Chemometric Methods in Molecular Design*, H. van de Waterbeemd, ed., VCH, Weinheim, 1995, pp. 312–317.

[26] V. Consonni, D. Ballabio, and R. Todeschini, *Evaluation of model predictive ability by external validation techniques*, J. Chemom. 24 (2010), pp. 104–201.

[27] G. Schüürmann, R.-U. Ebert, J. Chen, B. Wang, and R. Kühne, *External validation and prediction employing the predictive squared correlation coefficient – Test-set activity mean vs training set activity mean*, J. Chem. Inf. Model. 48 (2008), pp. 2140–2145.

[28] P.P. Roy, S. Paul, I. Mitra, and K. Roy, *On two novel parameters for validation of predictive QSAR models*, Molecules 14 (2009), pp. 1660–1701.

[29] R. Todeschini, *Milano Chemometrics*, University of Milano-Bicocca, Milano, Italy (personal communication), 2010.

[30] L.H. Hall and L.B. Kier, *The molecular connectivity chi indexes and kappa shape indexes in structure–property modeling*, Rev. Comput. Chem. 2 (1991), pp. 367–422.