

Article

Evaluation of Antioxidant Properties and Molecular Design of Lubricant Antioxidants Based on QSPR Model

Jianfang Liu ^{1,*} , Yaoyun Zhang ¹, Chenglingzi Yi ¹, Rongrong Zhang ¹, Sicheng Yang ¹, Ting Liu ¹, Dan Jia ², Qing Yang ¹ and Shuai Peng ¹

¹ School of Life Science and Technology, Wuhan Polytechnic University, Wuhan 430023, China; zyaoyun0310@126.com (Y.Z.); x020000@163.com (C.Y.); zhangrr990@126.com (R.Z.); 15587749812@163.com (S.Y.); liuting571126@163.com (T.L.); yqx0645@163.com (Q.Y.); 18207196638@163.com (S.P.)

² State Key Laboratory of Special Surface Protection Materials and Application Technology, Wuhan Research Institute of Materials Protection, Wuhan 430030, China; jiadan0510@163.com

* Correspondence: jianfang66@126.com

Abstract: Two quantitative structure–property relationship (QSPR) models of hindered phenolic antioxidants in lubricating oils were established to help guide the molecular structure design of antioxidants. Firstly, stepwise regression (SWR) was used to filter out essential molecular descriptors without autocorrelation, including electronic, topological, spatial, and structural descriptors, and multiple linear regression (MLR) was used to construct QSPR models based on the screened variables. The two models are statistically sound, with R^2 values of 0.942 and 0.941, respectively. The models' reliability was verified by the frontier molecular orbital energy gaps of the antioxidants. A hindered phenolic additive was designed based on the models. Its antioxidant property is calculated to be 20.9% and 11.0% higher than that of typical commercial antioxidants methyl 3-(3,5-di-tert-butyl-4-hydroxyphenyl) propionate and 2,2'-methylenebis(6-tert-butyl-4-methylphenol), respectively. The structure–property relationship of hindered phenolic antioxidants in lubricating oil obtained by computer-assisted analysis can not only predict the antioxidant properties of existing hindered phenolic additives but also provide theoretical basis and data support for the design or modification of lubricating oil additives with higher antioxidant properties.

Keywords: QSPR; lubricating oil antioxidant; antioxidant property; MLR; HOMO–LUMO



Citation: Liu, J.; Zhang, Y.; Yi, C.; Zhang, R.; Yang, S.; Liu, T.; Jia, D.; Yang, Q.; Peng, S. Evaluation of Antioxidant Properties and Molecular Design of Lubricant Antioxidants Based on QSPR Model. *Lubricants* **2024**, *12*, 3. <https://doi.org/10.3390/lubricants12010003>

Received: 7 November 2023

Revised: 13 December 2023

Accepted: 20 December 2023

Published: 22 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Antioxidants are one of the indispensable additives in the field of material chemistry, which are important for extending the service life of materials [1–3]. Similarly, antioxidants play an important role in lubricants, which protect lubricants by effectively sacrificing themselves. The lack or depletion of antioxygen will accelerate the autoxidation of the lubricating oil, and the oxidation products that follow may interfere with the previously well-balanced friction system, resulting in an elevated friction coefficient and severe material wear [4]. In general, the main types of antioxidants in lubricating oils include hindered phenols, amines, phenolamines, phosphites, zinc-dialkyl dithiophosphates (ZDDPs), and sulfides. Hindered phenolic antioxidants have excellent antioxygenic function, are widely employed, satisfy the needs of green environmental protection and have a higher potential of application [5]. The high efficiency and safety of antioxidants have always been the common goal of scientists, and at this stage, the “trial-and-error” experimental method is the primary way to explore efficient antioxidants; but without scientific and effective laws to guide the experiment, it is undoubtedly time-consuming, labor-intensive, and costly. Computer-aided data analysis and molecular design have become a new trend in the development of lubricating oil additives to shorten the research cycle and lower trial costs.

The QSPR model is a mathematical model that describes the potential relationship between the molecular structure information of the compounds and their properties. Currently, QSPR models have been widely used in fields of medicinal chemistry, toxicology, biology, and materials, etc. [6–12]. The QSPR model, created using strong machine learning algorithms and integrated modeling software, can identify and optimize the experimental direction, reducing needless experimental measurement, shortening the experimental time, and lowering experiment expenses [13]. This model is quite popular among scientific researchers and entrepreneurs. At present, the commonly used statistical methods for constructing QSPR models include MLR, principal component regression (PCR), partial least squares (PLS), artificial neural network (ANN), random forest (RF), support vector machine (SVM), etc. MLR offers the advantages of being simple to use and computationally efficient, especially when a small number of descriptors are involved. Therefore, MLR is typically the chosen model to build [14]. Furthermore, MLR can intuitively gain information on the degree and direction of the influence of the independent factors on the dependent variables, making the results easier to interpret and predict.

In recent years, many scientists have devoted themselves to applying QSPR technology to investigate the relationship between the chemical structure and performance of additives in lubricating oil. For example, Yu et al. [15] used ANN based on the backward propagation of errors to construct a QSPR model of the maximum nonseizure loads (P_B) and molecular descriptors of extreme pressure additives in lubricants. Liu et al. [16] attempted to develop an MLR model to correlate the tribological properties of lubricant additives with their molecular structure. They found that additives with benzimidazole and xanthate groups had good anti-wear properties, and the model had high statistical quality. In addition, researchers have also carried out beneficial exploration on antioxidants. For example, Abdulfatai et al. [17] designed five novel antioxidant lubricant additives without sulfated ash, phosphorus, and sulfur (SAPS) by using a QSPR model; the results showed that the newly designed additives' performance and dynamic binding energy were improved, and they had better dynamic binding on the simulated steel coated surface than the commercially sold additives, ZDDP. Furthermore, Abdulfatai et al. [18] successfully linked the decomposition temperature of thiamine additives with their chemical structure by various physicochemical descriptors and a genetic function algorithm (GFA), opening up a new avenue for the discovery of new lubricant additives that can withstand high dynamic working temperatures while also resisting wear and friction.

Although investigations on QSPR of lubricating oil additives have been conducted, there are few reports on the structure–activity relationship of hindered phenolic antioxidants. Therefore, the research team decided to use MLR to establish a QSPR model that would analyze the relationship between antioxidant efficiency and molecular structure of hindered phenolic antioxidants, validate the model's reliability using the frontier molecular orbital (HOMO–LUMO) energy gaps of these compounds, and then design new hindered phenolic antioxidant molecules based on the model that have better antioxygenic properties than common lubricating oil antioxidants on the market. It is evident that the QSPR method and the structure–property relationship law obtained can provide effective data support and theoretical advice for the future design and modification of novel phenolic antioxidants in lubricating oils.

2. Materials and Methods

2.1. Data Sources of Antioxidants

The oxidation onset temperature (OOT) of lubricating oil can reflect the antioxidant performance of the antioxidant lubricant additives to some extent, so this study attempted to establish the QSPR model of the OOT and the structure of antioxidants. In this study, polyalphaolefin (PAO4) was employed as the lubricating base oil, and phenolic antioxidants were used as the research object, whose structures are shown in Table 1. Among them, the antioxidants 1–13 were purchased in Wuhan, Hubei Province, China, with a purity of 98%,

and their OOTs were obtained by experiment. The OOTs of antioxidants 14–30 in the same base oil were derived from the literature [19–21].

Table 1. Molecular structures and onset temperature of some phenolic antioxidant additives.

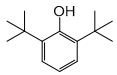
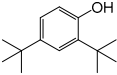
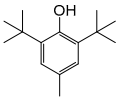
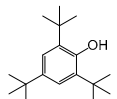
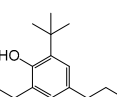
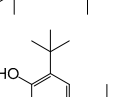
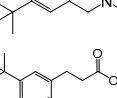
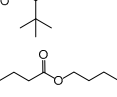
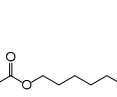
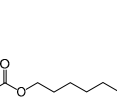
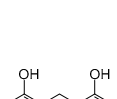
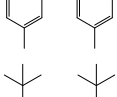
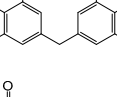
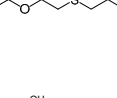
Compound	Molecular Structure	Onset Temperature T (°C)	lgAP
1		250.6	2.6749
2		256.6	2.7531
3		249.3	2.6845
4		246.0	2.7081
5		242.2	2.6389
6		247.5	2.7343
7		245.5	2.7466
8		243.1	2.8290
9		242.8	2.8950
10		242.7	2.9546
11		258.4	2.9916
12		248.6	2.9588
13		249.5	3.1525
14		246.8	3.1874

Table 1. Cont.

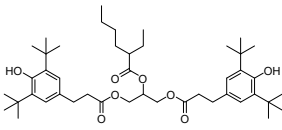
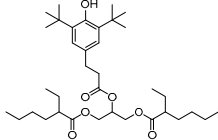
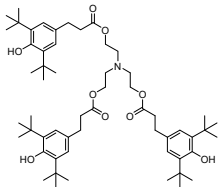
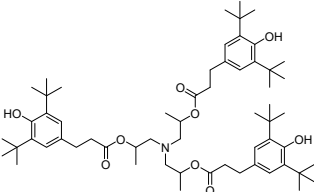
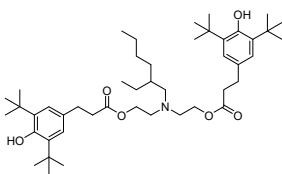
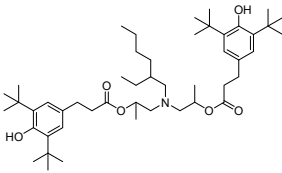
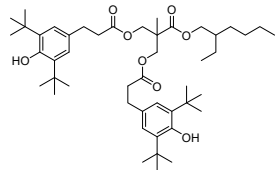
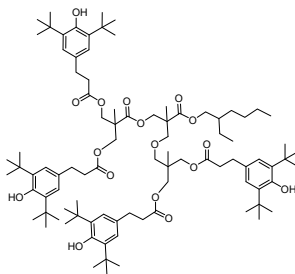
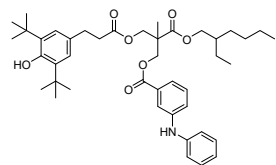
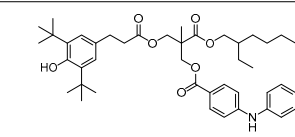
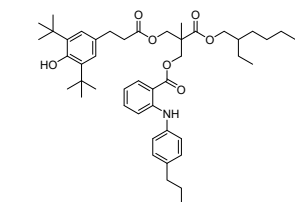
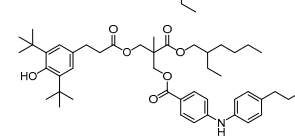
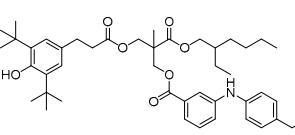
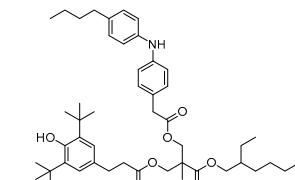
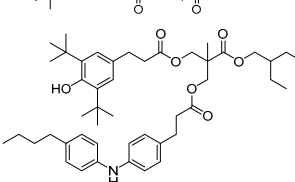
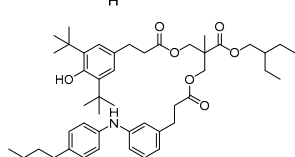
Compound	Molecular Structure	Onset Temperature <i>T</i> (°C)	lg <i>AP</i>
15		248.6	3.1461
16		245.3	2.9996
17		251.0	3.2858
18		245.0	3.1998
19		245.0	3.0801
20		249.0	3.1692
21		249.4	3.1769
22		249.6	3.4686
23		250.6	3.1575

Table 1. Cont.

Compound	Molecular Structure	Onset Temperature T (°C)	lgAP
24		251.0	3.1629
25		242.9	3.0485
26		250.3	3.1862
27		251.1	3.1980
28		253.4	3.2413
29		254.4	3.2630
30		254.3	3.2607

The OOT of the antioxidant was determined as follows: The lubricant samples were prepared by blending of 0.5% *w/w* of each antioxidant into 50 mL of PAO4. The samples were then analyzed using a Pressure Differential Scanning Calorimeter (a NETZSCH DSC 204 HP instrument (Bavarian, Germany)). An amount of 3–4 mg of sample was added to an aluminum crucible and pressurized to 30 bars with cylinder oxygen at a flow rate of 100 mL/min. The temperature was raised to 60 °C and allowed to stabilize before heating at 60 °C/min to 350 °C. The temperature at which the oxidation exotherm occurred was reported.

In the modeling calculation, the onset oxidation temperature was converted using Equation (1) in order to appropriately reduce the influence of molecular weight and synergistic effects of antioxidants for the base oil [22].

$$AP = \frac{T - T_0}{T_0} \times \frac{M_W}{N_A}, \quad (1)$$

where *AP* (antioxidant percent) represents the inhibition efficiency of antioxidant molecules in lubricants; *T* is the oxidation onset temperature of the lubricant; *T*₀, which is taken as

220.2 °C, is the oxidation onset temperature of base oil PAO4; M_W is the molecular weight of antioxidants; and N_A is the Avogadro constant. When calculating $\lg AP$, in order to convert the $\lg AP$ values to their positive form, N_A was taken as 0.06022. The calculation results are listed in Table 1.

Data including chemical structures, the oxidation onset temperature (°C), and oxidation onset temperature conversion values are recorded.

2.2. Construction of QSPR Model

The antioxidant dataset was randomly divided into a training set and a test set. The training set was used to construct the QSPR model, while the test set examined the predictive capability of the generated model. Usually, almost 80–90% of the samples are used for training purposes. In this paper, the compounds **7**, **13**, **19** and **26** were randomly chosen as test sets while the others as training sets. The model was built according to the training sets and the test sets which did not participate in the model building process. The flowchart for establishing the model is shown in Figure 1. An important step in obtaining a QSPR model is the numerical representation of molecular structural features, that is, molecular descriptors. A total of 117 groups of molecular descriptors were calculated using E-Dragon [23]. Each antioxidant was characterized by molecular descriptors as independent variables and $\lg AP$ values as the dependent variable, but not all independent variables had a statistically significant effect on dependent variables. Therefore, in order to establish the regression model with the best prediction performance, SWR [24] was used to screen out key independent variables with significant effects and eliminate those that do not influence dependent variables. The types and correlation coefficients of the 9 selected descriptors are shown in Table 2, including electronic, topological, spatial, and structural descriptors. It is shown that the antioxidant capacity of the antioxidant is related to its molecular descriptors Total dipole (X_1); Balaban index JX (X_2); Estate keys (sums): S_{SOH} (X_3); Estate keys (sums): S_{dO} (X_4); Estate keys (sums): S_{ssO} (X_5); Shadow ratio (X_6); Total energy (X_7); Inversion energy (X_8); and Electrostatic energy (X_9).

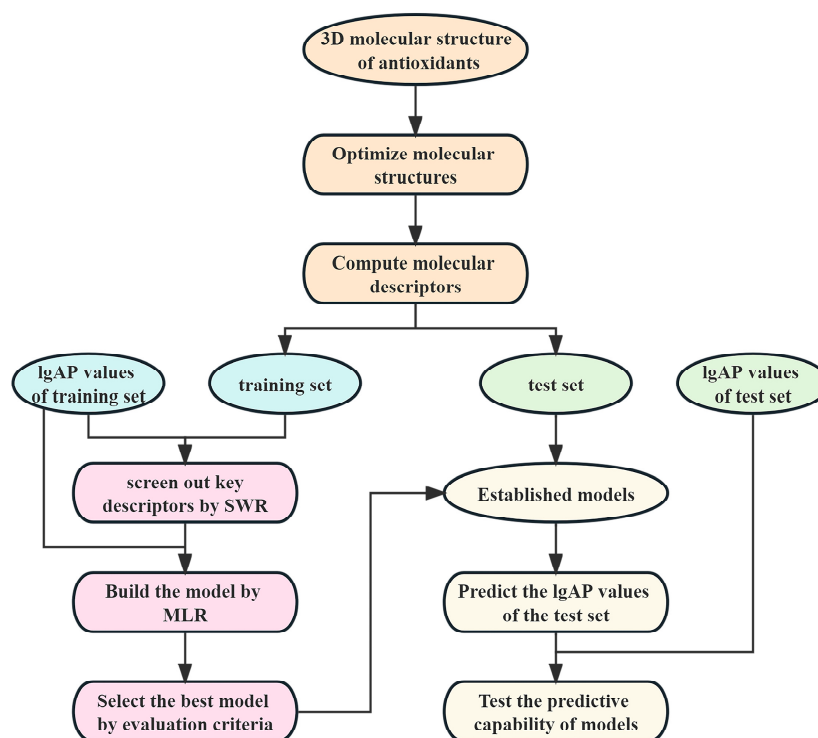


Figure 1. Flowchart for establishing the model.

Table 2. Types and correlation screening results of molecular descriptors.

Independent Variables	Molecular Descriptors	Species	Correlation Coefficient
X ₁	Total dipole	Electronic	0.7193
X ₂	Balaban index JX	Topological	−0.9182
X ₃	Estate keys (sums): S _s OH	Topological	0.5827
X ₄	Estate keys (sums): S _d O	Topological	0.8998
X ₅	Estate keys (sums): S _{ss} O	Topological	0.8847
X ₆	Shadow ratio	Spatial	0.4474
X ₇	Total energy	Structure	0.6026
X ₈	Inversion energy	Structure	0.5296
X ₉	Electrostatic energy	Structure	−0.3133

The screened molecular descriptors related to $\lg AP$ were used to construct a QSPR model using MLR. MLR is one of the classical QSPR modelling methods; its goal is to model the linear relationship between multiple independent variables and one dependent variable. The relationship between independent and dependent variables is shown in Equation (2).

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots \beta_n X_{in}, \quad (2)$$

where β_0 is the regression constant, X_i is the independent variable, β_i is the regression coefficient, and Y_i is the dependent variable.

The adjusted coefficient of determination (R^2), the existence of multicollinearity, the root mean square error (RMSE), and the value of Durbin–Watson (DW) were used as evaluation criteria for selecting the best prediction model [25]. It is generally considered that $R^2 > 0.6$ indicates that the model is acceptable and $R^2 > 0.8$ indicates that the model has a high degree of fit. The variance expansion factor (VIF) was used to judge whether there is multicollinearity among the variables in the model. In fact, the VIF strictly defined as less than 5 implies that there is no collinearity between the variable and one or more other variables in the model. RMSE was used to reflect the deviation between the predicted and true values in the model. The Durbin–Watson test is a test for autocorrelation of a dataset. Generally, the dataset with DW values between 1.5 and 2.5 meets the independence requirements.

2.3. Quality Evaluation of the Model

The normal distribution of data is closely related to the quality of the MLR model. The research results are valid only when the data distribution is relatively normal [26]. The normalized residual P-P plot is a graph drawn based on the relationship between the cumulative proportions of the variables and the cumulative proportions of the specified distribution. Whether the data satisfy the normal distribution can be checked by plotting the P-P plot of the model.

The validation analysis of the model includes internal and external validation. Internal validation mainly considers the robustness of a model, and external validation mainly tests the predictive ability of the established model. The leave-one-out cross-validation (LOO-CV) method was used for internal validation. The test set in the sample was used for external validation. The LOO-CV evaluates the model by treating each data of the training set in the sample as the test set data separately. The leave-one-out cross-validation coefficient (Q^2_{LOO}) (Equation (3)) and the cross-validation root mean square error (RMSE) (Equation (4)) were calculated to evaluate the robustness of the model. The external validation coefficients Q^2_{F1} (Equation (5)), Q^2_{F2} (Equation (6)), and Q^2_{F3} (Equation (7)) were used as indicators for evaluating the predictive ability of the model [27]. Their calculation formulas are shown in (3), (4), (5), (6), and (7), respectively.

$$Q^2_{LOO} = 1 - \frac{\sum (y_i - y_p)^2}{\sum (y_i - \bar{y}_{tr})^2}, \quad (3)$$

$$RMSE = \sqrt{\frac{\sum (y_i - y_p)^2}{n}}, \quad (4)$$

$$Q_{F1}^2 = 1 - \frac{\sum (y_i - y_p)^2}{\sum (y_i - \bar{y}_{tr})^2}, \quad (5)$$

$$Q_{F2}^2 = 1 - \frac{\sum (y_i - y_p)^2}{\sum (y_i - \bar{y}_{EXT})^2}, \quad (6)$$

$$Q_{F3}^2 = 1 - \frac{\sum (y_i - y_p)^2 / n_{EXT}}{\sum (y_i - \bar{y}_{EXT})^2 / n_{TR}}, \quad (7)$$

where y_i and y_p are the experimental and predicted $\lg AP$ of the samples, respectively; $\bar{y}_{tr}$ and $\bar{y}_{EXT}$ are the average values of the experimental $\lg AP$ of the training set and the test set respectively; n_{EXT} and n_{TR} are the sample numbers of the test set and the training set, respectively.

In addition to proving the accuracy of the model using the above validation parameters, the reliability of the research results was also confirmed from the perspective of quantum chemistry. According to the frontier molecular orbital theory, the higher is the energy of the highest occupied molecular orbital (HOMO), the more likely that the orbital is to lose electrons. Conversely, the lower is the energy of the lowest unoccupied molecular orbital (LUMO), the easier it is to obtain electrons [28,29]. The frontier molecular orbital energy gap ($\Delta E = E_{LUMO} - E_{HOMO}$) is an important theoretical parameter to characterize the molecular activity. The smaller the value of ΔE , the easier it is for electrons in the molecule to transition, and the stronger the reaction activity. Conversely, the reactivity is weaker [30]. Psi4 software was used to calculate the frontline molecular orbital energy and ΔE of antioxidants to verify their antioxidant effects.

2.4. Molecular Design of Lubricant Antioxidants

The molecular descriptors in the established MLR equations represent the structural characteristics of the hindered phenolic antioxidants. Based on these structural characteristics, template molecules were selected and structurally modified to design hindered phenolic antioxidants with better antioxidant properties. The molecular descriptors of the designed molecules were calculated by using the same method with the dataset molecules, and then their antioxidant properties were predicted by the models and compared with typical lubricant commercial antioxidants.

3. Results and Discussion

3.1. Established QSPR Model

In order to avoid overfitting, only a portion of the molecular descriptors is selected to build the MLR model. In general, the number of compounds in the system (sample size) should exceed 2^n (where n represents the number of descriptors in the model). For this reason, while considering the balance between MLR model quality and descriptor quantity, the model containing only four independent variables is chosen to be established. Finally, by combining the structural information of 26 compounds, two QSPR models are successfully constructed using the training set. The expressions for models 1 and 2 are shown as Equations (8) and (9), respectively:

$$\lg AP = 3.5379 - 0.2267 \times \text{Balaban index JX} + 0.0034 \times \text{Estate keys (sums)} : S_{sOH} + 0.0031 \times \text{Estate keys (sums)} : S_{dO} - 0.0461 \times \text{Shadow ratio} \quad (8)$$

$$\lg AP = 3.5778 - 0.2362 \times \text{Balaban index JX} + 0.0031 \times \text{Estate keys (sums)} : S_{sOH} + 0.0065 \times \text{Estate keys (sums)} : S_{ssO} - 0.0482 \times \text{Shadow ratio} \quad (9)$$

The statistical parameters of these models are shown in Table 3. Model 1 is built based on four independent variables: X_2 , X_3 , X_4 , and X_6 ; while model 2 is constructed with X_2 , X_3 , X_5 , and X_6 . With R^2 values of 0.942 and 0.941, respectively, both of these two models explain the data at a very high level (above 90%). Through T -tests, it is found that the significance index (p -value) of each variable in the models is less than 0.05, which shows that all variables are statistically significant at the 95% confidence interval. The DW values of models 1 and 2 are close to 2, and the VIFs of all variables are less than 5, indicating that there is no collinearity between the variables in the model equations. Taken together, these statistical parameters indicate that both models 1 and 2 are statistically significant. Molecular descriptors calculated by quantum chemical methods and the predicted $\lg AP$ values along with relative deviations for antioxidants calculated based on models 1 and 2 are shown in Table 4. The relative deviations between the predicted $\lg AP$ values and the experimental values for the test sets including compounds 7, 9, 19, and 26, as calculated by models 1 and 2, fall within 0.18% to 1.16% and 0.29% to 1.08% respectively. The consistency between the models' predicted values and the experimental values confirms the reliability and accuracy of the two models.

Table 3. Statistical parameters of QSPR-MLR model.

Model	R^2	Durbin–Watson	X	NS	S	p	VIF
1	0.9416	1.6962	constant	3.5379			
			X_2	−0.2267	−0.7211	0.0000	4.9330
			X_3	0.0034	0.1460	0.0222	1.4987
			X_4	0.0031	0.3003	0.0035	3.5777
			X_6	−0.0461	−0.1484	0.0493	2.1675
2	0.9413	1.6603	constant	3.5778			
			X_2	−0.2362	−0.7512	0.0000	4.3860
			X_3	0.0031	0.1320	0.0421	1.5832
			X_5	0.0065	0.2863	0.0038	3.2930
			X_6	−0.0482	−0.1550	0.0403	2.1395

R^2 is the adjusted coefficient of determination of the models. X is the dependent variable of the models. NS and S are the non-standardized and standardized regression coefficients of the models, respectively. p is the significance index.

Table 4. Molecular descriptors calculated by quantum chemical methods and $\lg AP$ of antioxidants predicted by models 1 and 2 along with relative deviations.

Compound	X_2	X_3	X_4	X_5	X_6	$\lg AP$	$\lg AP_1$	RD_1 (%)	$\lg AP_2$	RD_2 (%)
1	3.3405	10.2600	0.0000	0.0000	1.7560	2.6749	2.7346	2.20	2.7360	2.30
2	3.2750	9.8363	0.0000	0.0000	1.6750	2.7531	2.7517	−0.08	2.7540	0.05
3	3.4805	10.3845	0.0000	0.0000	1.7496	2.6845	2.7035	0.67	2.7036	0.73
4	3.6529	10.6502	0.0000	0.0000	1.6965	2.7081	2.6678	−1.52	2.6662	−1.52
5	3.5486	10.6390	0.0000	0.0000	1.7935	2.6389	2.6869	1.78	2.6861	1.81
6	3.4284	10.6050	0.0000	0.0000	1.8455	2.7343	2.7117	−0.86	2.7119	−0.80
7	3.2467	10.6164	11.3441	4.7024	2.0444	2.7466	2.7789	1.16	2.7759	1.08
8	2.6007	10.7909	12.1106	5.4007	3.2199	2.8290	2.8741	1.59	2.8769	1.70
9	2.3075	10.8488	12.2588	5.4762	2.9273	2.8950	2.9547	2.06	2.9609	2.28
10	2.0994	10.8854	12.3357	5.5133	4.7248	2.9546	2.9194	−1.19	2.9238	−1.04
11	2.5597	21.6358	0.0000	0.0000	1.9074	2.9916	2.9433	−1.64	2.9483	−1.45
12	2.7146	22.1110	0.0000	0.0000	1.4815	2.9588	2.9294	−1.02	2.9337	−0.84
13	1.7033	21.8336	24.8405	10.8766	3.1520	3.1525	3.1577	0.18	3.1619	0.29
14	1.7619	33.5276	40.0022	17.2228	2.3389	3.1874	3.2687	2.58	3.2648	2.40
15	2.1141	22.1536	39.2605	17.0618	2.9880	3.1461	3.1179	−0.86	3.1140	−1.04
16	2.9421	11.0231	38.5657	16.9106	1.6429	2.9996	2.9522	−1.55	2.9478	−1.73
17	1.6315	33.4346	39.2911	17.1374	3.0121	3.2858	3.2647	−0.61	3.2623	−0.75
18	1.7386	33.6506	40.3083	17.9739	3.1610	3.1998	3.2374	1.21	3.2359	1.10

Table 4. Cont.

Compound	X ₂	X ₃	X ₄	X ₅	X ₆	lgAP	lgAP ₁	RD ₁ (%)	lgAP ₂	RD ₂ (%)
19	1.9735	22.1344	25.8748	11.4882	3.1717	3.0801	3.0998	0.66	3.1021	0.70
20	2.0784	22.2411	26.4243	11.9754	2.6829	3.1692	3.1006	−2.15	3.1043	−2.06
21	2.1794	22.2293	40.1104	17.2650	2.6044	3.1769	3.1237	−1.64	3.1186	−1.85
22	1.6867	45.7926	85.8960	42.8523	2.2933	3.4686	3.4718	0.18	3.4893	0.54
23	1.6854	11.0853	40.0083	17.1862	2.5119	3.1575	3.2017	1.44	3.2047	1.48
24	1.6597	11.0787	39.8197	17.1236	2.7595	3.1629	3.1955	1.07	3.1984	1.11
25	1.7172	11.1257	40.6188	17.4085	3.1046	3.0485	3.1692	4.00	3.1702	3.98
26	1.6497	11.1065	40.0747	17.2220	2.5195	3.1862	3.2097	0.78	3.2131	0.83
27	1.6824	11.1154	40.3005	17.2993	3.1033	3.1980	3.1762	−0.64	3.1777	−0.64
28	1.6130	11.1086	39.9162	17.2076	2.8364	3.2413	3.2030	−1.14	3.2064	−1.09
29	1.5786	11.1108	39.8302	17.2032	2.7446	3.2630	3.2148	−1.44	3.2189	−1.36
30	1.6066	11.1176	39.9594	17.2474	3.2148	3.2607	3.1872	−2.22	3.1899	−2.18

lgAP is the true value. lgAP₁ and lgAP₂ are the predicted values by models 1 and 2, respectively. RD% is the relative deviation between the predicted and true values.

3.2. Reliability of Model

If the scatter points in the P-P diagram are closer to a diagonal linear distribution, it indicates that the data is more satisfied with the normal distribution [25]. The points in the P-P plots (Figure 2) of models 1 and 2 are approximately distributed in a diagonal straight line, and the dispersion of both models is diagonally distributed, which proves that the two models are reliable. In order to further verify the reliability of the QSPR models, Q^2_{LOO} and RMSE values are calculated by using the training set for internal validation. The external validation of the models is performed with the test set, and Q^2_{F1} , Q^2_{F2} , and Q^2_{F3} are calculated. The results of the calculations are listed in Table 5. The Q^2_{LOO} values for models 1 and 2 are 0.8674 and 0.8723, respectively, which are both greater than 0.8, indicating that the established models have strong stability. The three external validation coefficients for both models are all greater than 0.8 and numerically similar, and the results all prove that the predictive abilities of the models are very good and can be used to predict the antioxidant capacities of new antioxidants.

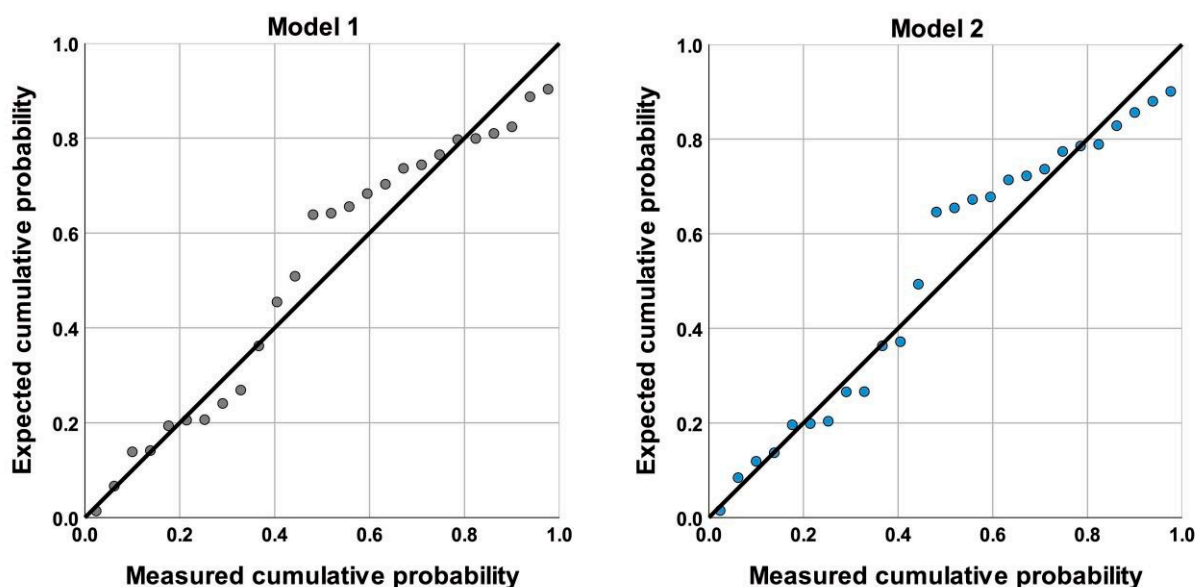


Figure 2. P-P Plot of normalized residuals for models 1 and 2. The dependent variable is lgAP.

Table 5. Validation coefficients of models.

Model	Q^2_{LOO}	RMSE	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}
1	0.8674	0.0555	0.983	0.983	0.889
2	0.8723	0.0557	0.983	0.982	0.886

Q^2_{LOO} is the leave-one-out cross-validation coefficient. RMSE is the cross-validation root mean square error. Q^2_{F1} , Q^2_{F2} , and Q^2_{F3} are the external validation coefficients.

It has been shown that the ΔE of polyphenols is related to their reactivity. The lower the ΔE of a compound, the more unstable its molecules are and the higher its reactivity, which indicates a stronger antioxidant capacity [31,32]. The HOMO and LUMO energies of antioxidants 1–30, as well as their ΔE values, can be found in Table 6. The reliability of the predictive ability of the QSPR models is further demonstrated by analyzing the correlation between the ΔE of 1–30 and the $\lg AP$ predicted by the models. The relationship between ΔE and the predicted $\lg AP$ of phenolic antioxidants is shown in Figure 3. The predicted values for models 1 and 2 are $\lg AP_1$ and $\lg AP_2$, respectively (Table 4). The Pearson correlation coefficient (R_1) between ΔE and $\lg AP_1$ of 1–30 is -0.793 , and $0.6 < |R_1| < 0.8$, indicating that there is a strong correlation, but not a high one. Figure 3a shows that there are eight outliers (represented by black dots), which correspond to compounds 23–30. This result may be related to the fact that 23–30 contain aniline groups. The aniline groups of 23–30 are also located in the HOMO and LUMO orbitals, which leads to their low ΔE values. This indicates that they are not suitable for the analysis of ΔE together with phenolic antioxidants. Model 2 is the same as model 1. Compounds 1–22 are selected for correlation analysis of ΔE and $\lg AP_1$, and the results are shown in Figure 3b. $R_2 = -0.878$, indicating that ΔE is highly negatively correlated with $\lg AP_1$. The correlation of ΔE with $\lg AP_2$ for 1–22 is shown in Figure 3d. $R_4 = -0.875$, indicating that ΔE is also highly negatively correlated with $\lg AP_2$. The reliability of models 1 and 2 is further verified.

Table 6. Frontier molecular orbital energy and energy level difference value of antioxidants.

Compound	E_{HOMO} (eV)	E_{LUMO} (eV)	ΔE (eV)
1	−4.6690	−0.1545	4.5145
2	−4.7185	−0.2672	4.4513
3	−4.4956	−0.1131	4.3826
4	−4.5163	−0.2292	4.2871
5	−4.5194	−0.1205	4.3989
6	−4.2683	−0.1487	4.1196
7	−4.6543	−0.7434	3.9109
8	−4.5852	−0.6348	3.9504
9	−4.6105	−0.6260	3.9845
10	−4.5795	−0.6422	3.9372
11	−4.2939	−0.5011	3.7927
12	−4.4267	−0.3731	4.0535
13	−4.7269	−1.0246	3.7023
14	−4.6105	−1.0300	3.5806
15	−4.5967	−1.0378	3.5588
16	−4.6897	−0.9724	3.7173
17	−4.5853	−0.8846	3.7007
18	−4.3167	−0.8064	3.5103
19	−4.2479	−0.7582	3.4896
20	−4.0158	−0.6519	3.3639
21	−4.5637	−1.0154	3.5483
22	−4.4459	−1.0850	3.3609
23	−4.5799	−1.8055	2.7744
24	−4.5878	−1.6386	2.9492
25	−4.4673	−1.8783	2.5890
26	−4.6075	−1.4051	3.2023

Table 6. Cont.

Compound	E_{HOMO} (eV)	E_{LUMO} (eV)	ΔE (eV)
27	−4.4912	−1.8784	2.6128
28	−4.4272	−0.9747	3.4524
29	−4.3505	−1.1136	3.2369
30	−4.2340	−1.1280	3.1059

E_{HOMO} is the highest occupied molecular orbital energy. E_{LUMO} is the lowest unoccupied molecular orbital energy. ΔE is the frontier molecular orbital energy gap.

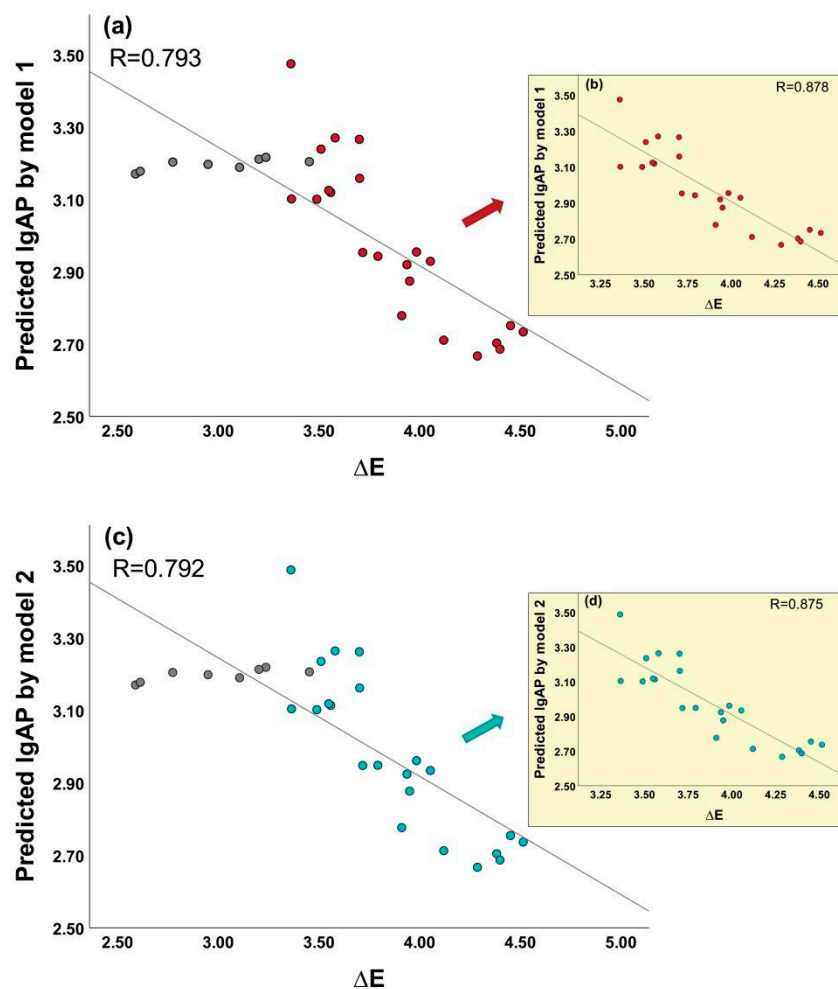


Figure 3. Correlation between (a) ΔE and $\lg AP_1$ for 1–30, (b) ΔE and $\lg AP_1$ for 1–22, (c) ΔE and $\lg AP_2$ for 1–30, and (d) ΔE and $\lg AP_2$ for 1–22. ΔE is the frontier molecular orbital energy gap. $\lg AP_1$ and $\lg AP_2$ are the predicted values by models 1 and 2, respectively. R is the Pearson correlation coefficient. The red and green dots refer to compounds 1–22 in models 1 and 2, respectively.

3.3. Interpretation of Model

Models 1 and 2 consist of molecular descriptors: Balaban index JX, electrical topological state keys (Estate keys), and Shadow ratio. The Balaban indices JX are highly discriminating descriptors whose values do not increase substantially with molecule size or the number of rings present. The Balaban index measures the ramification, and it tends to increase with molecular ramification [33]. The longer the main chain and the shorter and fewer the branched chain of the molecule, the smaller its Balaban index JX. The standardized regression coefficient (Beta value) refers to the standardized unit of change in the dependent variable when the independent variable changes by one unit. By comparing Beta values, it is possible to determine the influence of the independent variables on the

model predictions. The Beta values of X_2 for models 1 and 2 are -0.721 and -0.751 , respectively. The absolute value of the Beta value of X_2 is larger than that of other independent variables, indicating that X_2 contributes the most to the QSPR models. The negative correlation between X_2 and $\lg AP$ indicates that longer main chains, shorter substituent chains, and fewer branches are conducive to improving the antioxidant properties of phenolic antioxidants. Estate keys are numerical values, computed for each atom in a molecule, which encode information about both the topological environment of that atom and the electronic interactions due to all other atoms in the molecule [34]. X_3 , X_4 , and X_5 are the sum of Estate keys of hydroxyl ($-OH$), double-bonded oxygen ($=O$), and single-bonded oxygen ($-O-$) in the molecular structure, respectively. Extensive studies have shown that the antioxygenic properties of phenolic antioxidants are inextricably linked to the number of phenolic hydroxyl groups contained in their structures. Effectively increasing the number of phenolic hydroxyl groups will significantly enhance the antioxidant properties of these compounds [35]. In models 1 and 2, X_3 is positively correlated with $\lg AP$, which is consistent with the above views. X_4 in model 1 and X_5 in model 2 contribute positively to the results, indicating that double-bonded oxygen ($=O$) and single-bonded oxygen ($-O-$) groups can enhance the antioxygenic properties of hindered phenolic antioxidants. The shadow ratio is a space descriptor that is the ratio of the maximum dimension of a molecule to its minimum dimension. The shadow ratio presents a negative contribution to the prediction result. The smaller the shadow ratio of the molecule, the larger the $\lg AP$, indicating that the more concentrated the spatial structure of the antioxidant molecule, the stronger the antioxidant capacity.

3.4. Molecular Structure Design of Antioxidants

QSPR research aims to design the molecular structure based on the relationships between the molecular structure information and the properties implied in the constructed models. From the analysis in Section 3.3, it is clear that double-bonded oxygen ($=O$) and single-bonded oxygen ($-O-$) groups are conducive to improving the antioxidant properties of hindered phenolic additives. So, the molecular structure of Figure 4a serves as a template for the structural design of hindered phenolic antioxidants. The models' conclusions show that increasing the number of hindered phenolic hydroxyl groups effectively improves the antioxygenic properties of hindered phenolic antioxidants. Therefore, a triary structure containing three hindered phenolic hydroxyl groups is selected to design hindered phenolic antioxidants. The new antioxidant structure **31** is designed by appropriately increasing the length of the main chain of the molecule based on controlling the small shadow ratio of the molecule, as shown in Figure 4b. Five molecular descriptors (X_2 – X_6) for **31** are calculated, and the $\lg AP$ of this antioxidant is predicted using models 1 and 2. The results are listed in Table 7. Compared to a typical commercial monophenol antioxidant methyl 3-(3,5-di-tert-butyl-4-hydroxyphenyl)propionate (**7**) and a bisphenol antioxidant 2,2'-methylenebis(6-tert-butyl-4-methylphenol) (**11**), the antioxidant properties of **31** are improved by 20.9% and 11.0%, respectively. Therefore, when designing new antioxidant molecules, it can be considered to introduce double-bonded oxygen ($=O$) and single-bonded oxygen ($-O-$) groups, effectively increase the number of hindered phenolic hydroxyl groups, and appropriately increase the length of the main chain of molecule on the basis of controlling the shadow ratio of the molecule.

Table 7. Molecular descriptors calculated by quantum chemical methods and antioxidant property parameters of antioxidant **31** predicted by models 1 and 2.

Compound	X_2	X_3	X_4	X_5	X_6	$\lg AP_1$	$\lg AP_2$
31	1.5444	33.3964	38.9782	17.2138	2.2545	3.3194	3.3188

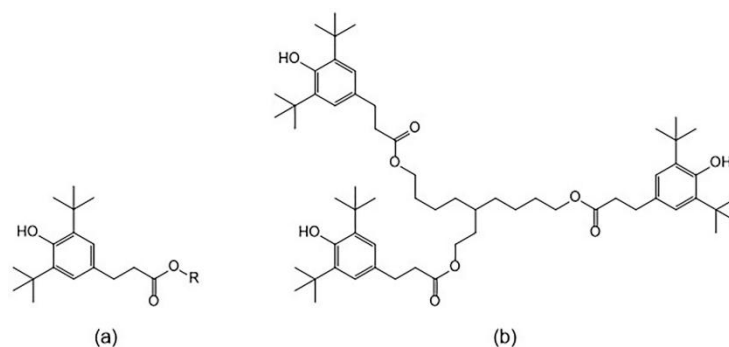


Figure 4. Structures of (a) template and (b) antioxidant **31**.

4. Conclusions

With the increasing demand for lubricating oil, the development of high-quality lubricants through the combination of computer-assisted and experimental approaches is becoming a future trend. In this work, 117 molecular descriptors of each lubricant antioxidant in the dataset are calculated using E-Dragon. Two QSPR models for predicting $\lg AP$ values are established to describe the relationship between the chemical structure of hindered phenolic antioxidants and their antioxygenic capacity. The models are also validated internally and externally, and the analysis of relevant statistical parameters confirms the models' stability and good predictive ability. Compared with existing literature, the novelty of this research lies in the fact that in addition to the usual model validation methods, the density functional theory is also used to calculate the frontier molecular orbital energy gaps of antioxidants to verify the accuracy of the models.

The results of the models demonstrate that the antioxidant properties of hindered phenolic additives can be enhanced by (1) suitably increasing the length of the main chain and decreasing and shortening the branch chain, and (2) introducing hydroxyl (-OH), double-bonded oxygen (=O) and single-bonded oxygen (-O-) groups. Based on the QSPR models, a lubricant antioxidant with improved structure and performance is designed. The calculated results show that the antioxygenic property of the designed antioxidant is 20.9% and 11.0% higher than that of typical commercial antioxidants methyl 3-(3,5-di-tert-butyl-4-hydroxyphenyl) propionate and 2,2'-methylenebis(6-tert-butyl-4-methylphenol), respectively.

In conclusion, the QSPR model developed in this study can predict the antioxidant capacity of phenolic additives and can guide the molecular design of antioxidants using the information on the structure–property relationship obtained by the model, which can reduce the cost of trial-and-error during the development of lubricating oil antioxidants. However, this paper still has deficiencies due to the limitations of the time cycle, simulation technology, and small sample size. Subsequently, a series of works, such as synthesis and performance testing of the designed antioxidant, can be carried out, and the predictive ability and reliability of the QSPR models can be confirmed further. Other categories of additives can also be designed to develop diversified additives that are more environmentally friendly, greener, and with better performance. This work also provides the theoretical basis and data support for the future design or modification of lubricating oil additives with higher antioxidant properties.

Author Contributions: Conceptualization, J.L.; Data curation, Y.Z., S.Y., T.L., Q.Y. and S.P.; Formal analysis, S.Y., T.L., Q.Y. and S.P.; Investigation, C.Y., R.Z. and D.J.; Methodology, J.L.; Resources, C.Y., R.Z. and D.J.; Validation, C.Y., R.Z. and D.J.; Visualization, Y.Z.; Writing—original draft, Y.Z.; Writing—review and editing, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 52075405).

Data Availability Statement: The authors confirm that the data supporting the findings of this study are available within the article, and some quoted data are openly available in reference number [19–21].

Conflicts of Interest: The authors have no relevant financial or non-financial interests to disclose.

References

- Li, Y.; Zhou, W.D.; Xue, W.A.; Huang, Y.W.; Zhang, Q.; Han, J.B. The Enhancement of Overall Performance of Lubricating Grease by Adding Layered Double Hydroxides. *Lubricants* **2023**, *11*, 260. [\[CrossRef\]](#)
- Lacroix-Andrivet, O.; Hubert-Roux, M.; Bourhis, C.L.; Moualdi, S.; Siqueira, A.L.M.; Afonso, C. Characterization of Base Oil and Additive Oxidation Products from Formulated Lubricant by Ultra-High Resolution Mass Spectrometry. *Lubricants* **2023**, *11*, 345. [\[CrossRef\]](#)
- Gijsman, P. A review on the mechanism of action and applicability of Hindered Amine Stabilizers. *Polym. Degrad. Stab.* **2017**, *145*, 2–10. [\[CrossRef\]](#)
- Zheng, Y.Z.; Li, F.S.; Zhang, H.C.; Wang, S.; Zhao, Z.J.; Wang, W.C.; Chen, R.Y. Fishhook characteristics of biodiesel lubricity during autoxidation. *Fuel* **2023**, *331*, 125897. [\[CrossRef\]](#)
- Orives, J.R.; Galvan, D.; Pereira, J.L.; Coppo, R.L.; Borsato, D. Experimental Design Applied for Cost and Efficiency of Antioxidants in Biodiesel. *J. Am. Oil Chem. Soc.* **2014**, *91*, 1805–1811. [\[CrossRef\]](#)
- Parveen, S.; Awan, N.U.; Farooq, F.B.; Fanja, R.; Anjum, Q.U. Topological Indices of Novel Drugs Used in Autoimmune Disease Vitiligo Treatment and Its QSPR Modeling. *Biomed Res. Int.* **2022**, *2022*, 6045066. [\[CrossRef\]](#)
- Osman, M.S.; Awad, T.A.; Shantier, S.W.; Garelnabi, E.A.; Osman, W.; Mothana, R.A.; Nasr, F.A.; Elhag, R.I. Identification of some chalcone analogues as potential antileishmanial agents: An integrated and evaluation. *Arab. J. Chem.* **2022**, *15*, 103717. [\[CrossRef\]](#)
- Khan, K.; Kumar, V.; Colombo, E.; Lombardo, A.; Benfenati, E.; Roy, K. Intelligent consensus predictions of bioconcentration factor of pharmaceuticals using 2D and fragment-based descriptors. *Environ. Int.* **2022**, *170*, 107625. [\[CrossRef\]](#)
- Ni, Z.H.; Li, F.S.; Wang, H.; Xiao, H. Prediction of physical parameters of Jatropha biodiesel-ethanol dual fuel based on topological indices. *Appl. Energy* **2022**, *328*, 120202. [\[CrossRef\]](#)
- Koné, A.; Es-Sabar, A.; Do, M.T. Application of Machine Learning Models to the Analysis of Skid Resistance Data. *Lubricants* **2023**, *11*, 328. [\[CrossRef\]](#)
- Navabi, A.; Isfahani, T.M.; Ramazani, M.; Alimoradi, M. QSPR Models for Predicting Retention Indices of Polygonum minus Huds. Essential Oil Composition Using GA-BWMLR and GA-BPANN Methods. *J. Essent. Oil Bear. Plants* **2021**, *24*, 879–896. [\[CrossRef\]](#)
- Wang, H.W.; Zhang, C.H.; Yu, X.W.; Li, Y.Y. Evaluating Wear Volume of Oligoether Esters with an Interpretable Machine Learning Approach. *Tribol. Lett.* **2023**, *71*, 43. [\[CrossRef\]](#)
- Tao, C. Quantitative Structure-Activity Relationship on Adsorption Coefficient of Organic Pollutants between Microplastics and Water Environment. Master's Thesis, Yangzhou University, Yangzhou, China, 2023.
- Liu, P.X.; Long, W. Current Mathematical Methods Used in QSAR/QSPR Studies. *Int. J. Mol. Sci.* **2009**, *10*, 1978–1998. [\[CrossRef\]](#)
- Yu, X.L.; Zhan, R.M.; Deng, J.Y.; Huang, X.W. Prediction of the maximum nonseizure load of lubricant additives. *J. Theor. Comput. Chem.* **2017**, *16*, 1750014. [\[CrossRef\]](#)
- Liu, Y.L.; Deng, J.Y.; Yu, X.L. Quantitative Structure-Property Relationships of Antiwear Performance of Lubricant Additives. *Rev. Roum. Chim.* **2017**, *62*, 849–855.
- Abdulfatai, U.; Uzairu, A.; Uba, S.; Shallangwa, G.A. Molecular design of antioxidant lubricating oil additives via QSPR and analysis dynamic simulation method. *Heliyon* **2019**, *5*, e02880. [\[CrossRef\]](#)
- Abdulfatai, U.; Uzairu, A.; Uba, S.; Shallangwa, G.A. Quantitative Structure-Properties Relationship of Lubricating Oil Additives and Molecular Dynamic Simulations Studies of Diamond-Like-Carbon (DLC). *Iran. J. Chem. Chem. Eng.-Int. Engl. Ed.* **2020**, *39*, 281–295. [\[CrossRef\]](#)
- Higgins, C.L.; Filip, S.V.; Afsar, A.; Hayes, W. Synthesis, characterisation, and performance evaluation of tri-armed phenolic antioxidants. *Tetrahedron Lett.* **2020**, *61*, 152127. [\[CrossRef\]](#)
- Higgins, C.L.; Filip, S.V.; Afsar, A.; Hayes, W. Evaluation of thermal and oxidative stability of three generations of phenolic based novel dendritic fuel and lubricant additives. *React. Funct. Polym.* **2019**, *142*, 119–127. [\[CrossRef\]](#)
- Higgins, C.L.; Filip, S.V.; Afsar, A.; Hayes, W. Increasing the antioxidant capability via the synergistic effect of coupling diphenylamine with sterically hindered phenol. *Tetrahedron* **2019**, *75*, 130759. [\[CrossRef\]](#)
- Zhan, S.P.; Duan, H.T.; Hua, M.; Xu, H.P.; Shang, H.F.; Jin, Y.L.; Jia, D.; Tu, J.S.; Li, J. Studies of antioxidant performance of amine additives in lubricating oil using 3D-QSAR. *Sci. China-Technol. Sci.* **2017**, *60*, 299–305. [\[CrossRef\]](#)
- Tetko, I.V.; Gasteiger, J.; Todeschini, R.; Mauri, A.; Livingstone, D.; Ertl, P.; Palyulin, V.; Radchenko, E.; Zefirov, N.S.; Makarenko, A.S.; et al. Virtual computational chemistry laboratory—Design and description. *J. Comput.-Aided Mol. Des.* **2005**, *19*, 453–463. [\[CrossRef\]](#)
- Wang, Y.Q.; Zhou, P.P.; Lin, Y.; Shu, M.; Hu, Y.; Xia, Q.Y.; Lin, Z.H. Quantitative Prediction of Class I MHC/Epitope Binding Affinity Using QSAR Modeling Derived from Amino Acid Structural Information. *Comb. Chem. High Throughput Screen.* **2015**, *18*, 75–82. [\[CrossRef\]](#)

25. Liu, J.F.; Zhang, R.R.; Yang, S.C.; Liu, T.; Yi, C.L.Z.; Zhang, Y.Y.; Jia, D. Evaluation of Anti-wear Properties of Different Vegetable Oils Based on QSPR Model. *Tribol. Lett.* **2023**, *71*, 35. [\[CrossRef\]](#)
26. Lai, I.K.W.; Hitchcock, M. A consideration of normality in importance-performance analysis. *Curr. Issues Tour.* **2015**, *18*, 979–1000. [\[CrossRef\]](#)
27. Tropsha, A. Best Practices for QSAR Model Development, Validation, and Exploitation. *Mol. Inf.* **2010**, *29*, 476–488. [\[CrossRef\]](#)
28. Wang, G.R.; Xue, Y.S.; An, L.; Zheng, Y.G.; Dou, Y.Y.; Zhang, L.; Liu, Y. Theoretical study on the structural and antioxidant properties of some recently synthesised 2,4,5-trimethoxy chalcones. *Food Chem.* **2015**, *171*, 89–97. [\[CrossRef\]](#)
29. Xue, Y.S.; Liu, Y.P.; Luo, Q.Q.; Wang, H.; Chen, R.; Liu, Y.; Li, Y. Antiradical Activity and Mechanism of Coumarin-Chalcone Hybrids: Theoretical Insights. *J. Phys. Chem. A* **2018**, *122*, 8520–8528. [\[CrossRef\]](#)
30. Alasalvar, C.; Soylu, M.S.; Guder, A.; Albayrak, C.; Apaydin, G.; Dilek, N. Molecular structure, quantum mechanical calculation and radical scavenging activities of (E)-4,6-dibromo-2- (3,5-dimethylphenylimino)methyl -3-methoxyphenol and (E)-4,6-dibromo-2- (2,6-dimethylphenylimino)methyl -3-methoxyphenol compounds. *Spectrochim. Acta Part A-Mol. Biomol. Spectrosc.* **2014**, *130*, 357–366. [\[CrossRef\]](#)
31. Boulmouk, Y.; Belguidoum, K.; Meddour, F.; Amira-Guebailia, H. Investigation of antioxidant activity of epigallocatechin gallate and epicatechin as compared to resveratrol and ascorbic acid: Experimental and theoretical insights. *Struct. Chem.* **2021**, *32*, 1907–1923. [\[CrossRef\]](#)
32. Szewczuk, N.A.; Duchowicz, P.R.; Pomilio, A.B.; Lobayan, R.M. Resonance structure contributions, flexibility, and frontier molecular orbitals (HOMO-LUMO) of pelargonidin, cyanidin, and delphinidin throughout the conformational space: Application to antioxidant and antimutagenic activities. *J. Mol. Model.* **2023**, *29*, 2. [\[CrossRef\]](#)
33. Eliasi, M.; Taeri, B. Balaban index of zigzag polyhex nanotorus. *J. Comput. Theor. Nanosci.* **2007**, *4*, 1174–1178. [\[CrossRef\]](#)
34. Cash, G.G.; Anderson, B.; Mayo, K.; Bogaczyk, S.; Tunkel, J. Predicting genotoxicity of aromatic and heteroaromatic amines using electrotopological state indices. *Mutat. Res.-Genet. Toxicol. Environ. Mutagen.* **2005**, *585*, 170–183. [\[CrossRef\]](#)
35. Wang, J.; Lv, C.S.; Li, J.; Li, H.Y.; Li, C.Q.; Chen, S. Study on kinetics and antioxidant activity of hindered phenolic antioxidant by oxygen uptake measurement. *Chem. Eng.* **2015**, *29*, 4–7. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.