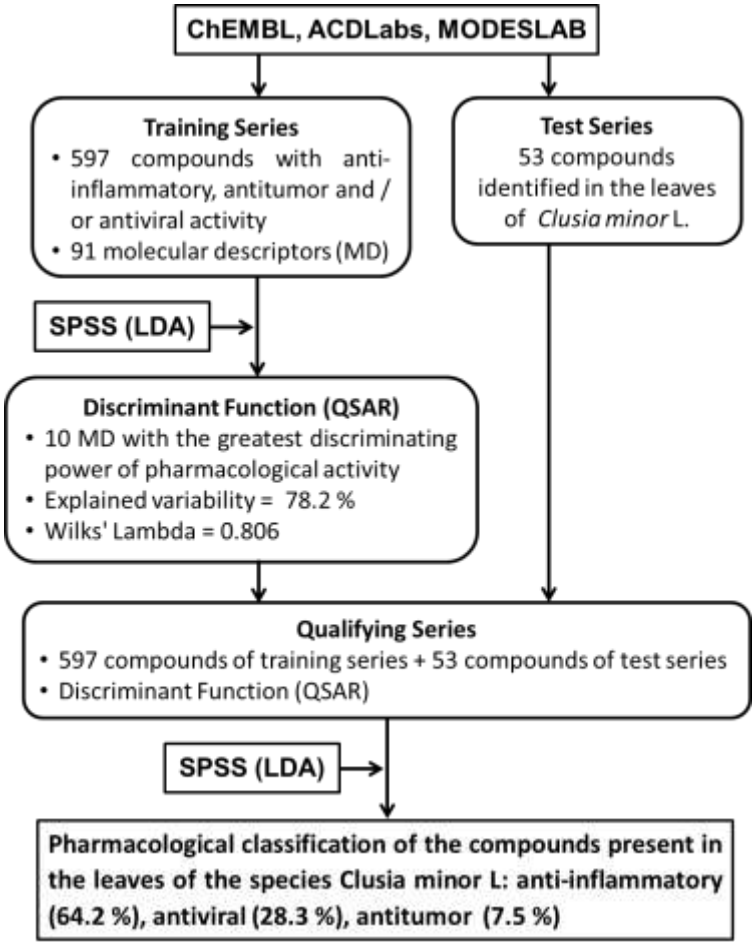
A QSAR MODEL FOR THE PREDICTION OF THE PHARMACOLOGICAL ACTIVITY OF THE COMPOUNDS PRESENT IN THE SPECIES *CLUSIA MINOR* L.

Juan C. Polo Vega[a,b], Raisa Mangas Marín[b], Luis A. Torres Gómez[b], Olivia A. Domínguez Alonso[b]

[a] Center for Genetic Engineering and Biotechnology, La Habana, Cuba.

[b] Department of Pharmacy, Institute of Pharmacy and Food Sciences, University of Havana, La Habana, Cuba.

Graphical Abstract



Abstract

The species *Clusia minor* L., family Clusiaceae, belongs to a group of higher plants of high research interest due to their uses in traditional medicine. In the literature, it has been reported that the metabolites present in the species have various properties such as anti-inflammatory, antirheumatic, antiviral, antitumor and antioxidant, among others. However, few studies on its chemical and pharmacological composition can justify its potential use. For this reason, the present work aims to classify theoretically the possible anti-inflammatory, antitumor and/or antiviral activity of the compounds present in the extract of its leaves. To achieve this objective, a training series made up of 597 compounds classified according to the aforementioned pharmacological activities and a test series with the 53 compounds identified in the plant were built, which formed the basis for obtaining, by means of Linear Discriminant Analysis, of a Quantitative Structure-Activity Relationship model that achieves a 52.9 % adequate classification of the compounds included in the training series. The model obtained was considered viable taking into account the limitations of the classification according to belonging to a single group of pharmacological activity. Finally, the pharmacological classification of the compounds present in the extract of the leaves of the species *Clusia minor* L. was carried out, which presented, for the most part, anti-inflammatory (64.2 %) and antiviral (28.3 %) activity.

Introduction

Natural resources provide a huge and highly diversified chemical bank from which we can explore for potential therapeutic agents by bioactivity-targeted screenings. Opportunity exploration of medicinal plants is still very wide open in line with the development of herbal industry, herbal medicine, and phytopharmacy [1-4].

Many new compounds have been isolated from plants in the Clusiaceae family, and their potent biological properties have been the subject of several studies [5,6]. In the Cuban flora, four species belonging to the *Clusia* genus have been identified, of which only the *Clusia rosea* species has been studied exhaustively [7,8]. In the species *Clusia minor*, abundant in Cuba, numerous compounds with possible pharmacological activity have been identified, however, so far there are few studies that justify their use as phytopharmaceuticals [9-11].

Based on these premises, the present work aims to theoretically predict, by means of a QSAR model, the possible anti-inflammatory, antitumor and/or antiviral activity of the compounds identified in the extract of the leaves of the species *Clusia minor* L..

Materials and Methods

Construction of training and test series. To form the classification model of the pharmacological activity of the compounds of interest, a training series was built consisting of 597 organic compounds extracted from the ChEMBL database, classified into three sub-groups of 199 compounds each, according to the pharmacological properties to be considered in the study (anti-inflammatory, antitumor and antiviral). Their corresponding abbreviated representation codes (SMILES) were generated using the ACD-Labs version 2020.1.2 program. The corresponding molecular descriptors (MD) were calculated using the TOPS-MODE approximation of the MODESLAB computer program, which allowed obtaining a matrix with spectral moments from μ^0 to μ^{15} per graph. The molecular parameters that were considered for the calculation of the MD were: bond distance (Std), dipole moment (Dip), hydrophobicity (Hyd), polarizability (Pol), Van der Waals atomic radius (Van) and atomic weight (Ato).

Obtaining of the predictive model of pharmacological activity. The Linear Discriminant Analysis (LDA) technique, by means of the statistical program IBM SPSS Statistics version 26, was used to determine, within the set of calculated MD, those with the greatest ability to differentiate the belonging of each compound of the training series to the pharmacological classification sub-series. The statistical quality of the discriminant functions was evaluated based on the statistical parameters Wilks' Lambda (λ), canonical correlation, Chi-square (Lambda transformed value) and the percentage of correct classification, among others.

Assignment of the predominant pharmacological activity in the compounds of interest. For the assignment of the predominant pharmacological activity in the compounds of interest identified in the leaves of the species *Clusia minor* L., a classification series was constructed by combining the training and prediction series. In this new series, all the compounds of interest were assigned to an unknown group, different from those pre-established in the training series. In order to obtain the most probable assignment of each of the compounds of interest to one of the subgroups, according to the pharmacological properties considered in the study, a new linear discriminant analysis was developed by introducing all the molecular descriptors included in the QSAR model in a single step.

Results and Discussion

The TOPS-MODE approach of the MODESLAB software generated for each compound in the training series 91 MD, from order zero to order 15, related to the selected molecular properties: bond distance (Std); dipole moment (Dip); hydrophobicity (Hyd); polarizability (Pol); Van der Waals radius (Van) and atomic weight (Ato).

The use of the linear discriminant analysis (LDA) technique made it possible to select, among the 91 DM, those with the greatest discriminant power of the pharmacological activity of the compounds included in the training series. Table I shows the resulting discriminant functions.

Table I. Standardized coefficients of the MD included in the discriminant functions (DF).

DF	Molecular descriptors (MD)									
	μ0	$\mu(Std)^1$	$\mu(Std)^2$	$\mu(Std)^9$	$\mu(Dip)^1$	$\mu(Dip)^5$	$\mu(Hyd)^1$	$\mu(Hyd)^9$	μ(Pol)1	$\mu(Van)^{11}$
1	1.553	-7.772	8.639	0.907	-0.349	0.113	0.344	-0.475	-1.285	-0.880
2	4.108	-7.057	-1.511	3.935	1.157	1.000	1.057	-3.080	1.210	-0.479

As can be seen in table 1, from the initial set of independent variables, made up of 91 MD, only 10 were included in the discriminant functions, which constituted a significant reduction in the number of independent variables to be considered for the formulation of the predictive model of the pharmacological activity of the compounds present in the leaves of the plant. The main statisticians that determine the statistical quality of the discriminant functions obtained are summarized in table II.

Table II. Statisticians for the evaluation of the discriminant functions generated.

DF	Canonical Correlation	Wilks' Lambda	Chi-square (p)
1	0.391	0.806	127.113 (0.000)
2	0.219	0.952	29.076 (0.001)

For both functions, the canonical correlation is moderate and the Wilks' Lambda value is very high, although its transformed value (Chi-square) was significant ($p < 0.05$). This means that, of the 597 compounds present in the training series, only 55.8 % were correctly classified by the model. These indicators could lead to the conclusion that any predictive model of pharmacological activity formulated from these functions would be ineffective and would fail when attempting to classify the identified compounds based on the information provided by the molecular descriptors selected through discriminant analysis.

However, in the opinion of the authors, these results do not in themselves constitute an inherent limitation of the model, rather they are due to the fact that although a compound in the literature is classified in a specific group according to the pharmacological activity that it predominantly exhibits, this does not deny that in practice it manifests more than one type of pharmacological activity. The same drug can interact with different receptors or exert its action at different sites in the body, allowing different pharmacological responses to be triggered, some being more pronounced than others. This situation is reflected in the combined graph of the values of the discriminant functions obtained for the compounds of the training series (See Figure 1) where the overlapping of the theoretically assigned pharmacological functions can be observed.

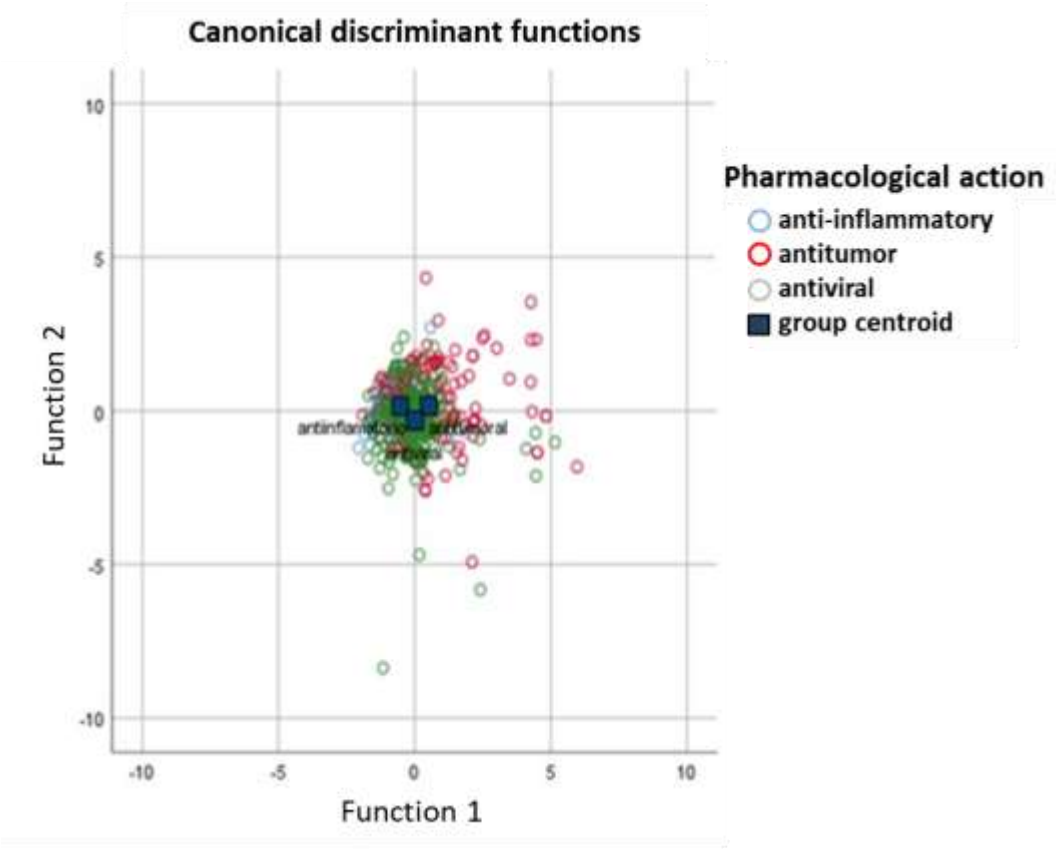


Figure 1. Combination of the values of the discriminant functions obtained for the compounds of the training series.

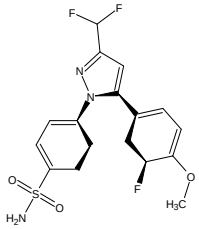
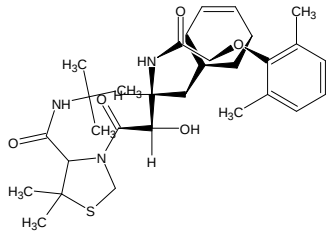
From this point of view, a QSAR model for the theoretical prediction of pharmacological action based on the previous discriminant functions could be viable, by revealing not only the predominant pharmacological activity according to which a drug appears reported in the literature, it would also show other possible therapeutic effects possible in correspondence with its chemical structure. The analysis from the structural point of view of the compounds whose classification according to the QSAR model does not coincide with that reported in the literature constitutes evidence in favor of the previous arguments. Table III shows the structure of two of these compounds and their classification according to pharmacological (literature) and statistical (LDA) criteria.

Compound 42, reported in the literature as AI, was classified as AT according to the LDA. The antitumor are generally fused polynuclear three-ring systems. In this case, the sulfonamide present, being large enough, is capable of forming hydrogen bonds and creating a condensed polynuclear system. In the spatial arrangement of the rings, there could apparently be free rotation between them. But given their structural characteristics, there is the possibility of the mesomeric or conjugation effect occurring, which enables the tendency to form double bonds and provides structural rigidity. Due to this, a greater resemblance to a condensed polynuclear system is generated, which is the pharmacophore group of antitumor drugs.

Compound 417 is pharmacologically classified as AV and as AI according to the LDA. Anti-inflammatory drugs must contain in their structure an acid group with a pKa low enough to generate a significant amount of anionic conjugate base at the pH of the inflamed tissue. Although when analyzing the structure of this compound no groups of this type are observed, the presence of the hydroxyl group could provide certain acidity to the molecule. In addition, several amide groups are distinguished that, although they are not acidic, could be hydrolyzed by the action of metabolism and provide the necessary acid functional groups. These amide groups

also help secure attachment to the allosteric site of cyclooxygenase. On the other hand, the presence of an aromatic ring favors the affinity for the enzyme and can bind to the aromatic and aliphatic residues of the active site of the enzyme, providing a therapeutically useful anti-inflammatory activity. Also the presence of lipophilic groups improves its distribution through the membranes to reach the cyclooxygenase receptors and exert its anti-inflammatory action.

Table III. Classification of case compounds according to their pharmacological action and according to LDA (fragment).

Case in Training Series	Chemical Structure	Pharmacological Classification	LDA Classification
42		AI	AT
417		AV	AI
AI: anti-inflammatory; AT: antitumor; AV: antiviral			

The elements indicated justify the use of the QSAR model obtained for the theoretical prediction of the pharmacological activity of the compounds extracted from the *Clusia minor* L. species, despite the fact that the statistical quality of the discriminant functions that comprise it is not optimal.

The LDA developed on the classification series, formed by combining the training and prediction series, taking the molecular descriptors included in the QSAR model as independent variables, yielded as main results the most probable assignments, according to their molecular structure, of each one of the compounds of interest to the groups considered in the study (anti-inflammatory, antitumor and antiviral), Table IV summarizes the classification made.

Table IV. Summary of the classification, by means of the QSAR model, of the compounds identified in the leaves of the species *Clusia minor* L.

Group		Frequency	Percentage
1	Anti-inflammatory (AI)	34	64,2
2	Antitumor (AT)	4	7,5
3	Antiviral (AV)	15	28,3
Total		53	100,0

As can be seen, most of the compounds identified in the leaves of the *Clusia minor* L. species were theoretically classified as anti-inflammatory (64.2 %), and to a lesser extent as antiviral (28.3 %). Only in four cases was the pharmacological activity identified as predominant by the QSAR model antitumor.

The parameters that justify, from a statistical point of view, the assignment of each molecule to the groups considered are:

- **P (D>d | G=g).** Conditional probability, which indicates the "rarity degree" of the case within the predicted group. A value close to 1 indicates that this case is close to the centroid of the group, thus decreasing the probability that it constitutes an outlier within it.
- **P (G=g | D=d).** A posteriori probability, probability of belonging to each predicted group.
- **Mahalanobis distance squared for centroid (DM^2).** Square of the distance from the case to each of the predicted group centroids.

The examination of these parameters for each compound in particular allows us to assess how predominant the pharmacological action predicted by the theoretical model is. Table V shows a selection of the results.

Table V. Classification of the compounds identified in the leaves of *Clusia minor* L. according to the QSAR model (fragment).

Case number in the classification series	Top group					Second top group			Discriminant punctuation	
	Forecast Group	P(D>d | G=g)		P(G=g | D=d)	Mahalanobis distance squared for centroid	Group	P(G=g | D=d)	Mahalanobis distance squared for centroid	Function 1	Function 2
		p	fd							
17	2	,883	2	,354	,248	1	,330	,389	,059	,369
21	3	,005	2	,752	10,656	2	,130	14,165	,126	-3,579
22	3	,361	2	,679	6,664	2	,196	9,152	,453	-2,861
23	3	,000	2	,861	22,799	2	,076	27,666	,219	-5,088
24	3	,000	2	,837	19,109	2	,087	23,634	,174	-4,686
25	1	,938	2	,384	,127	3	,369	,210	-,428	-,190
26	1	,857	2	,419	,309	3	,379	,511	-,700	-,374
31	1	,974	2	,420	,054	3	,310	,663	-,432	,365
Forecast Group: 1-AI; 2-AT; 3-AV										

The analysis of the possible pharmacological actions of the compounds identified in the leaves of the species *Clusia minor* L. based on their chemical structure, generally allows justifying the assignments to the different groups, thus supporting the validity of the QSAR model. As an example, case 17 is analyzed, whose chemical structure is represented in Figure 2.

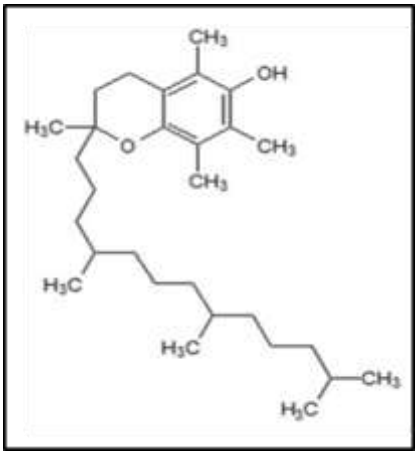


Figure 2. Chemical structure of case 17.

The compound corresponding to case 17 was classified by the LDA as antitumor in the first instance and as anti-inflammatory in second place, with similar posterior probabilities of belonging to each of them (See Table IV). Observing the molecular structure of this compound, the presence of the chroman heterocycle, to which redox properties are attributed, can be appreciated. The phenolic grouping present acts as a suppressor on the induction of reactive oxygen species, linked to the appearance of many cancers as a consequence of the cell damage they produce. In addition, the fusion between the aromatic ring and the dihydropyran heterocycle establishes a flat structure with an aromatic nature and a certain structural rigidity that would favor acting as an intercalating agent contributing to antitumor activity. Another element in favor of this classification may be the structural similarity that compound 17 presents with some coumarins with proven anticancer activity [12]. The anti-inflammatory activity attributed to this compound, according to the second predicted group, can be justified due to the presence of the phenolic hydroxyl group, which gives it a certain degree of acidity, necessary to achieve this type of activity. This type of grouping is present in a large number of molecules with recognized anti-inflammatory action [13-15].

Case 17 corresponds to vitamin E, with a known antioxidant action, in correspondence with the actions attributed by the QSAR model for this molecule. This fat-soluble vitamin produces a significant inhibition in the behavior of tumors and decreases the production of prostaglandins. In addition, it is known that oxidative stress, which this substance counteracts, is closely correlated with various inflammatory conditions and with the appearance of tumors, so both activities would be supported [16]

From what was discussed above, it could be concluded that the low statistical quality of the QSAR model obtained through ADL is not necessarily inherent to its nature. Rather, it responds to the limitations of a classification that is based only on the predominant pharmacological action of a drug, when in fact the same molecule can induce more than one therapeutic effect on the organism. Consequently, the methodology described constitutes a starting point to justify, from a scientific basis, the use of this plant species as a phytopharmaceutical, although like any theoretical prediction, the pharmacological activity assigned from the QSAR model for the metabolites identified in *Clusia minor* L needs to be verified in practice.

Referens

1. Süntar I. (2020). Importance of ethnopharmacological studies in drug discovery: role of medicinal plants. Phytochem Rev 19, 1199–1209 https://doi.org/10.1007/s11101-019-09629-9
2. Elya B, Basah K, Mun'im A, Yuliastuti W, Bangun A, Septiana EK. (2012). Screening of α-glucosidase inhibitory activity from some plants of Apocynaceae, Clusiaceae, Euphorbiaceae, and Rubiaceae. J Biomed Biotechnol.
3. Rasool H. (2012). Medicinal Plants (Importance and Uses). Pharmaceut Anal Acta 3: e139. doi:10.4172/2153-2435.1000e139
4. Escalona LJ, Tase A, Estrada A, Almaguer ML. (2015). Uso tradicional de plantas medicinales por el adulto mayor en la comunidad serrana de Corralillo Arriba. Guisa, Granma. Revista Cubana de Plantas Medicinales [Citado 07 octubre 2021]; 20 (4). www.scielo.com.
5. Acuna M, Jancovski N, Kennelly E. (2009). Polyisoprenylated Benzophenones from Clusiaceae: Potential Drugs and Lead Compounds. Current Topics in Medicinal Chemistry, Volume 9, Number 16, pp. 1560-1580(21). Bentham Science Publishers. https://doi.org/10.2174/156802609789909830
6. Medina E, Aguiar G, Gómez M, Medina JD. (2004). Patterns of leaf epicuticular waxes in species of *Clusia*: taxonomical implications. Interciencia. 29(10): 579-582.
7. Cuesta-Rubio O, Vélez H, Frontana B, Cárdenas J. (2001). Nemorosone, the major constituent of floral resins of *Clusia rosea*. Phytohemistry. 57(2): 279-283.

8. Rojas N, Cuesta O, Avilas A, Lugo D, Avellaneda S. (2001). Actividad antimicrobiana de nemorosona aislada de *Clusia rosea*. Rev. Cubana Farm. 35 (Suplemento especial): 197-199.
9. Mangas R, Leyva JC, Becerra RM, Bello A. (2018). Phytochemical screening and isolation of triterpenes and sterols from *Clusia minor* L. leaves. Rev Cubana Plant Med, 23(3). http://www.revplantasmedicinales.sld.cu/index.php/pla/rt/printerFriendly/797/317
10. Mangas R, Reynaldo G, Dalla MT, Aver K, Piovesan LG, Bello A, Rodeiro I, Malheiros A, de Souza MM, Menéndez R. (2018). Gas Chromatography/Mass Spectrometry characterization and antinociceptive effects of the ethanolic extract of the leaves from Clusia minor L. J. Pharm Pharmacogn Res, 7(1): 21-2
11. Mangas R, Montes de Oca R, Herrera ME, Bello A, Hernández I, Menéndez R, Paz MT, Rodeiro I. (2018). GC/MS analysis and bioactive properties of extracts obtained from *Clusia minor* L. leaves J. Mex. Chem. Soc. 62(4). DOI: http://dx.doi.org/10.29356/jmcs.v62i4.544Yun B, Lee I, RyooI J, Yoo I. (2001). Coumarins with monoamine oxidase inhibitory activity and antioxidative coumarino-lignans from hibiscus s yriacus. Journal of natural products, 64(9), 1238-1240.
12. Kumbar, S, Shettar A, Joshi S, Patil, S. (2021). Design, synthesis, molecular docking and biological activity studies of novel coumarino-azetidinones. Journal of Molecular Structure, 1231, 130016.
13. Lee J, Jang Y, Kang H, Moon H, Sim S, Kim C. (2006). Anti-inflammatory action of phenolic compounds from Gastrodia elata root. Archives of Pharmacal Research, 29(10), 849-858.
14. Ruiz-Ruiz J, Matus-Basto A, Acereto-Escoffié P, Segura-Campos M. (2017). Antioxidant and anti-inflammatory activities of phenolic compounds isolated from Melipona beecheii honey. Food and Agricultural Immunology, 28(6), 1424-1437.
15. Zhang T, Qiu F, Chen L, Liu R, Chang M, Wang X. (2021). Identification and in vitro anti-inflammatory activity of different forms of phenolic compounds in *Camellia oleifera* oil. Food Chemistry, 344, 128660.
16. Niki E, Noguchi N. (2004). Dynamics of antioxidant action of vitamin E. Accounts of chemical research, 37(1), 45-51.