Proceedings of the International Conference on Food, Nutrition, Health and Lifestyle, Vol. 1, Issue. 1, 2022, pp.10-18

ISSN 2682-7026 online
DOI: https://doi.org/10.17501/26827026.2022.1102

QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP (QSAR) STUDY OF EUGENOL DERIVATIVES AS ANTIOXIDANT COMPOUNDS

Triatmaja K[1*], Prabawati SY[2] and Widiakongko PD[2]

[1]*Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada*

[2]*Chemistry Study Program, Faculty of Science and Technology, UIN Sunan Kalijaga*

Abstract: Non-Communicable Diseases (NCDs) are one of the main problems in the health sector. This problem is shown by 70% of the world's deaths caused by NCDs. One way to overcome this is with Eugenol which is an antioxidant, indicated by an IC50 value of 4,38 µg/mL. The development of derivatives of eugenol compounds resulted in lower antioxidant activity values than eugenol activity. A study of the Quantitative Structure and Activity Relationship (QSAR) was carrying out on eugenol derivative compounds with antioxidant activity. This study aims to analyze the QSAR of eugenol-derived compounds and determine the computational model of the QSAR equation for eugenol-derived compounds. The data used are 21 data on eugenol-derived compounds that have been tested for their antioxidant activity using the 2,2-diphenyl-1-picrylhydrazyl (DPPH) radical experimentally (in vitro) from various literature. The data calculates computationally using the Hartree Fock (HF) calculation with a basis set of 6-311G. The computational results obtained are then analyzed using the multilinear regression method. The results of the QSAR analysis showed that the linear correlation coefficient (R) was 0,973, which means the predictive activity value was close to the experimental activity value. The best QSAR equation model obtained was compiled by an electronic descriptor with the energy parameter HOMO and the energy difference HOMO-LUMO (Gap), a hydrophobic descriptor with a log P parameter, and a steric descriptor with a Balaban index, with the equation model:

Log IC50 = 6,442 + (26,257) HOMO + (9,231) GAP + (0,056) LOGP – (0,433) BALABAN

Keywords: QSAR, Eugenol, antioxidants, Hartree Fock, 6-311G

Introduction

Cardiovascular disease, cancer, diabetes, stroke, and chronic respiratory disease are diseases that are included in NCDs. World Health Organization (WHO) data in 2020 showed that NCDs were the leading cause of death worldwide. Deaths caused by NCDs amount to 41 million people or more than 70% of the cause of death worldwide. The cause of the onset of NCDs is an unhealthy lifestyle. An unhealthy lifestyle comes from food, drink, and air that carry free radicals. Free radicals are compounds that have unpaired electrons and are highly reactive (Winarsi, 2007). The activity of free

*Corresponding Author's Email: karismatriatmaja@mail.ugm.ac.id

radicals can be suppressed by antioxidant compounds. Antioxidants are compounds that can provide electrons or reductants to radical compounds (Winarsi, 2007). Antioxidants can react faster with radicals so that they can protect cellular components from direct interaction with radical compounds.

One of the compounds that have the potential to be an antioxidant is eugenol (da Silva et al., 2018). The antioxidant potential in eugenol is due to the presence of a phenol functional group. Phenol groups can act as electron-givers or hydrogen atoms to stabilize free radicals (Aini et al., 2007). Antioxidant activity that is influenced by the phenol group in eugenol is shown by the value of its activity. The antioxidant activity of eugenol has been studied by da Silva et al. (2018) in vitro against DPPH radicals expressed as IC50, namely the concentration required by eugenol compounds to stabilize 50% of DPPH radicals. The antioxidant activity of eugenol compounds has an IC50 value is 4,38 μg/mL. The development of derivatives of eugenol compounds was also carried out by Silva et al. (2018) to increase its antioxidant activity. The synthesized derivatives compounds totaled 20 compounds, but the result of the activity of the modification of eugenol derivative compounds was only 7 compounds that were close to the activity value of eugenol (IC50: 4,38 μg/mL). This shows that research on the modification of eugenol-derived compounds in vitro has low effectiveness in the search for new compounds.

The development of new antioxidant compounds can be more effective using computational chemical modeling approaches or in silico. One of the applications of computational chemistry that can be applied is the study of quantitative relationships of structure and activity (QSAR). This study studied the quantitative correlation between molecular structure and the value of the biological activity that was measured experimentally. The correlation obtained is a QSAR regression equation. The equation can then be used to evaluate new compounds that are close to modeling a series of compounds (Siswandono, 2016).

Based on this background, this research uses computational modeling of eugenol compounds and their derivatives and their antioxidant activities. This study uses a QSAR study to obtain the best equation model. This equation can be used to predict the activity of the eugenol derivatives. This equation can be used as the basis for the development of derivative compounds that have higher antioxidant activity than eugenol.

Research Methods

The tools used in the study were in the form of hardware, namely a set of computers with AMD Ryzen 5 Processor specifications, AMD Radeon Graphics, 512 Gb Solid State Drives (SSD), and 8 GB Random Access Memory (RAM), and software namely the Windows 10 operating system, Avogadro, GaussView 5.0.8, MarvinSketch Trial, Microsoft Office Excel 2010, and IBM SPSS 26. The materials used in the research are shown in Table 1. in the form of eugenol derivative compounds along with IC50 data that have been experimentally tested by da Silva et al. (2018), Sohilait and Kainama (2019), and Dhiman et al., (2019).

The research materials used were visualized using Avogadro. Geometry optimization was then carried out using GaussView 5.0.8 on the Optimization menu using the HF calculation method and base set 6-311G. The electronic descriptor is calculated using GaussView 5.0.8 on the Frequency menu. Hydrophobic descriptors and steric descriptors were calculated using the MarvinSketch Trial with

partition-log P menus, geometry-molecular surface area (3D), and geometry-topology analysis. The analysis was performed by multilinear regression analysis by entering bound variable data and free variable data in IBM SPSS 26 applications on the Analyze-Regression-Linear menu. The method used is the Backward method looking for free variables that are strongly related to bound variables to obtain the QSAR equation.

The QSAR equation is then validated against the value of R, the value of the Fisher criterion (F) count, the value of F of the table, and cross-validation of Leave One Out (LOO) with the indicator of the square value of the cross-validation coefficient (q2). The value of q2 is obtained through the formula:

$$q2 = 1 - \frac{\Sigma(y-y')^2}{\Sigma(y-\bar{y})^2} \qquad (1)$$

y is the IC50 log value of the experiment, y' is the IC50 log value of the LOO cross-validation equation, and is the average value of the experiment IC50 log.

Table 1. Eugenol analog compounds and log IC50 value data

No	Compound	log IC50 value (µg/mL)
1	Eugenol	4,38
2	Isoeugenol	50,7
3	2-methoxy-4-((2-phenyl-1,3-dioxosane-4-il)methyl)phenol	30,37
4	2-methoxy-4-(oxyran-2-ilmethyl)phenol	19,3
5	3-(4-hydroxy-3-methoxyphenyl)propan-1,2-diol	20
6	4-((2,2-dimethyl-1,3-diococulant-4-il)methyl)-2-methoxyphenol	32
7	4-(2-hydroxypropyl)-2-methoxyphenol	51,12
8	Isometil eugenol	8,53
9	Methyl eugenol	7,80
10	1-(2-(4-alyl phenoxy)acetyl)-4-(4-hydroxy benzoyl) thiosemikarbazide	9,152
11	1-(2-(4-alyl phenoxy)acetyl)-4-(3,4,5-trihydroxy benzoyl) thiosemikarbazide	8,304
12	(E)-1-(2-(4-alyl phenoxy)acetyl)-4-cinnamoyl thiosemikarbazide	9,786
13	N′-(2-(4-alyl-2-methoxy phenoxy)acetyl)-2-hydroxy benzohydrazide	9,514
14	N′-(2-(4-alyl-2-methoxy phenoxy)acetyl)-2-mercapto benzohydrazide	9,293
15	N′-(2-(4-alyl-2-methoxy phenoxy)acetyl)pycolinohydrazide	11,21
16	4-alyl-2-methoxy phenyl 4-hydroxy-3-methoxy benzoate	20,66
17	4-alyl-2-methoxy phenyl 4-hydroxy benzoate	12,47
18	4-alyl-2-methoxy phenyl 2-hydroxy benzoate	16,04
19	4-alyl-2-methoxy phenyl 2-mercaptobenzoate	14,88
20	(Z)-4-alyl-2-methoxy phenyl 3-(4-hydroxy-3-methoxyphenyl) acrylate	27,7
21	(Z)-4-alyl-2-methoxyphenyl 3-phenyl acrylate	31,3

Results and Discussion

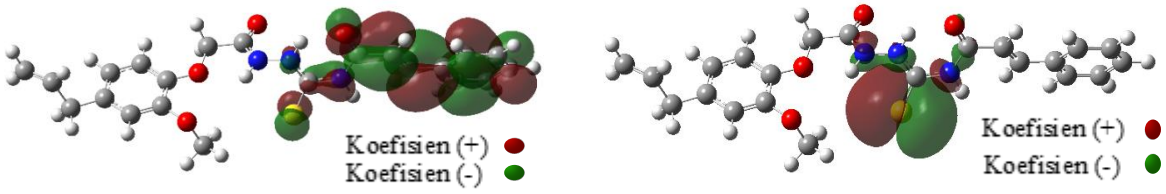
The compound data used in the calculation of QSAR in this study are eugenol analog compounds that have tested the structure of the compound together with testing its activity with IC50 value data from the experimental results. Structural modifications were carried out on the hydroxyl and allyl groups of

eugenol compounds. Activity testing was carried out on antioxidant activity testing to inhibit 50% of DPPH radicals (IC50) in Vitro.

Calculations of electronic descriptors, hydrophobic descriptors, and steric descriptors are carried out on all eugenol analogous compounds. Such descriptors are used to explain the structure and properties of eugenol analogous compounds. Hydrophobic descriptors are used to explain the solubility of drug compounds in biological membranes to be able to interact with receptors. The interaction of drug compounds with receptors is explained through electronic and steric descriptors, namely in the form of electron density, molecular size, and stereochemical effects (Siswandono, 2016).

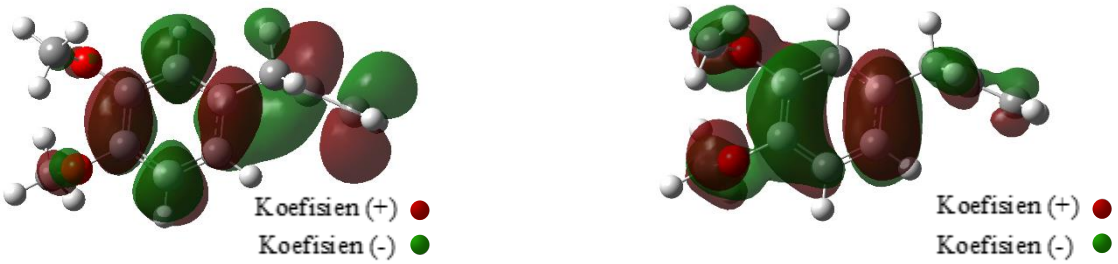
The parameters used in electronic descriptors are homo energy parameters (Highest Occupied Molecular Orbital), LUMO (Lowest Unoccupied Molecular Orbital) energy, difference between HOMO – LUMO (Gap) energy, and dipole moment. The parameter of the difference between the energy of HOMO – LUMO (Gap) is used to explain the process of antioxidant mechanisms in capturing free radicals by giving away their electrons. The smaller the difference in HOMO – LUMO energy (the higher the HOMO energy and the lower the LUMO energy) can indicate that the compound has a higher reactivity. This is because the smaller the energy required for electrons to move from HOMO to LUMO. This is also true the opposite the greater the difference in HOMO energy – LUMO (the lower the HOMO energy and the higher the LUMO energy) can indicate that the compound has a lower (stable) reactivity. This is because the greater the energy required for electrons to move from HOMO to LUMO (Asmuruf et al., 2017).

The calculation obtained by compound 12 is a compound that has the lowest HOMO energy difference value and LUMO energy, which is 0,35459 eV and compound 9 is a compound that has the largest HOMO energy difference value and LUMO energy, which is 0,44838 eV. This shows that compound 12 has the highest reactivity or tends to be unstable and compound 9 has the lowest reactivity or tends to be more stable. Visualization of the orbitals of HOMO and LUMO is shown in Figure 1. of compound 12 and Figure 2. of compound 9.



LUMO (0,03718 eV) HOMO (-0,31741 eV)

Figure 1. LUMO and HOMO molecular orbitals of compound 12



LUMO (0,13081 eV) HOMO (-0,31757 eV)

Figure 2. HOMO and LUMO molecular orbitals of compound 9

HOMO energy is related to the ionization potential of compounds, namely the ability of compounds to carry out nucleophilic attacks, while LUMO energy is related to electron affinity, namely the susceptibility of compounds to nucleophilic attacks (Santos et al., 2014). Figure 1. shows in compound 12 the orbitals of the HOMO molecule tend to be sulfur atoms and nitrogen atoms, so that sulfur atoms and nitrogen atoms can carry out nucleophilic attacks. The LUMO molecular orbitals in compound 12 tend to be oxygen atoms and benzene rings, so oxygen atoms and benzene rings have the susceptibility of compounds to nucleophilic attacks. Figure 2. shows in compound 9 that in the orbitals the HOMO molecule tends to the methyl group and the benzene ring so that the methyl group and the benzene ring can carry out nucleophilic attacks. The LUMO molecular orbitals in compound 9 tend to be on the allyl group and the benzene ring, so the allyl group and benzene ring have the susceptibility of the compound to nucleophilic attacks.

The dipole moment parameter can be used to estimate the polarisability properties of a compound. The greater dipole moment indicates that the compound has high polarizability (polar) and is increasingly electronegative (Male et al., 2018). The calculation obtained by compound 3 is a compound that has the highest dipole moment value of 4,6409 and compound 9 is a compound that has the lowest dipole moment value of 0,1326. This shows that compound 3 tends to be more polar and compound 9 tends to be more non-polar.

The parameters used in the hydrophobic descriptor are the P and Molecular Surface Area (MSA) log parameters. The log parameter P is used to explain the tendency to hydrophobic properties or hydrophilic properties of a compound. Hydrophobic properties or hydrophilic properties are related to the application of compounds as drugs to interactions in the body. Compounds that have hydrophobic properties can interact with receptors in the body and compounds that have hydrophilic properties can interact in the blood or cannot interact with receptors in the body. Hydrophobic properties are indicated by log values P getting more positive and hydrophilic properties are indicated by log values P getting negative (Mardianingrum et al., 2018).

All calculated compounds have a positive P log value which means that all compounds tend to have hydrophobic properties. Compound 21 is a compound that has the largest log P value of 5,114 and compound 5 is a compound that has the smallest log P value of 0,403. This suggests that compound 21 tends to be more hydrophobic (can interact with receptors) and compound 5 tends to be more hydrophilic (can interact in the blood). Research that has been carried out by Lipinski et al. (2012)

indicates that the compound used as a drug must have a log value of P more than 5. A log value of more than 5 is intended for the drug compound to have good absorption in the body (to interact with receptors). This shows that compound 21 is the compound that has the best absorption in the body.

MSA parameters relate to the area of a compound surface (Perwira et al., 2015). The surface area of a compound is related to the interaction that can occur on a certain surface when it encounters another compound. The calculation obtained by compound 12 is a compound that has the largest surface area of 573,475 and compound 1 is a compound that has the smallest surface area of 257,956.

Parameters – the parameters used in the steric descriptor are topology index parameters. The topology indices used are Platt Index, Randic Index, Balaban Index, Harary Index, Hyper Wiener Index, Szeged Index, and Wiener Index. The topology index relates to the size of the turmoil of a compound. The magnitude of the turmoil of a compound is influenced by a large number of constituent atoms of a compound. The greater the turmoil of a compound, the greater the price of the topology index (Mardianingrum et al., 2018). The structure between similar compounds produces the same topology index value. This is because changes in the location of the bonds between atoms are not calculated. Senyawa 1 and compound 2 tend to have the same topology index value this is due to the location of the bonds between different atoms in the double bond in the alyl group but the same structure.

The parameter calculation data on the descriptor obtained is then analyzed to determine the relationship of the structure of a compound with its activity. These parameters are then used as free variables. The activity value of a compound has also been obtained through the experimental results of each compound, namely the IC50 value. The value of that activity is then used as a bound variable. The data of the entire compound were analyzed using the multilinear regression method. The multilinear regression analysis method is used to determine the best equation of the relationship between bound variables and their free variables. The best equation is obtained by looking for a free variable that affects the value of the bound variable. The method used to obtain the most influential free variable is the backward method. The backward method contained in the IBM SPSS application can perform calculations repeatedly to obtain some calculation results.

Table 2. Equation validation results

No	Equation	R	R2	Fcalc/Ftab	q2 Value
1	Log IC50 = 6,442 + (26,257) HOMO + (9,231) GAP + (0,056) LOGP – (0,433) BALABAN	0,973	0,947	10,879	0,851
2	Log IC50 = 7,323 + (20,292) HOMO + (5,428) GAP – (0,002) MSA – (0,657) BALABAN	0,960	0,922	7,328	0,368

Equations are obtained in Table 2. then the calculation of statistical requirements is carried out, namely the value of R, the value of R2, and the comparison of the calculated value of F with the value of F in the table. The R-value that must be met is more than 0,8 or with an R2 value more than 0,6. The comparison of calculated F values with the table F values to be met is more than 1 (La Kilo and La Kilo, 2019). The comparison of the calculated F value with the table F value in both equations has a value greater than 1 (it has met the requirements). Requirements of the value of R, the value of R2

and the comparison of the calculated value of F with F of the table show that in both equations the free variable has a significant influence on the bound variable and has a confidence level of 95%.

Both equations are then validated by the QSAR equation. Equation validation is used to ensure the QSAR equation can accurately predict the value of the biological activity of compounds or have the smallest possible error. Validation is carried out by the Leave One Out (LOO) cross-validation method. Validation with the LOO cross-validation method is carried out using one compound data as a test set data and the rest as training set data. The data test set is carried out on all compound data and is only used once (Soleh and Aunuddin, 2013). The test set data was omitted to obtain the multilinear regression analysis equation from the training set (Arba et al., 2016).

The data obtained are then calculated indicators in cross-validation of LOO. The indicator used in such validation is the value of q2. The value of q2 that must be met is more than 0,5 (Tjahjono and Fadhilah, 2012). Validation performed by the LOO cross-validation method shows that the equation that has met the value of q2 is the model of equation no. 1 in Table 2. where the value of q2 is 0,851. The results of the validation show that the best equation model selected and has complied with statistical requirements is equation model no. 1 in Table 2. that is:

Log IC50 = 6,442 + (26,257) HOMO + (9,231) GAP + (0,056) LOGP – (0,433) BALABAN (2)

The free parameters or variables included in the equation model (2) are HOMO energy parameters, HOMO – LUMO energy differences (Gap), P logs, and Balaban indices. The coefficients of the energy parameters HOMO and Gap have the greatest value. The HOMO energy parameter coefficient is 26,257 and the Gap parameter coefficient is 9,231. This shows that the HOMO energy parameter and the Gap parameter play a significant role in the activity value of the compound. The coefficient of the HOMO energy parameter and the Gap value indicates that the parameter is directly proportional to the value of its activity. The greater the value of HOMO energy, the greater the activity value as well as the Gap value parameter where the greater the Gap value, the greater the activity.

The equation model (2) is then used to predict the activity value of the IC50 log. The expected result of the predicted activity value is close to or equal to the activity value of the experiment. The indicator used to determine that the value of the predicted activity is close to or equal to the activity value of the experiment is from the value of R. R-value is obtained by connecting the assessment activity value data with the prediction activity value data through a graph.

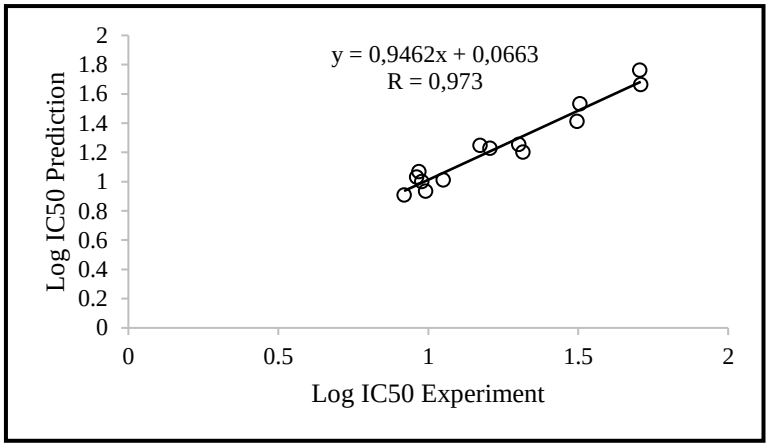


Figure 3. Graph of relationship of IC50 log values Experiment with IC50 log values Prediction

The relationship of the value of the experimental activity to the value of the predicted activity is shown in Figure 3. The result shown by the R-value obtained is more than 0,9 or close to 1, which is 0,973. This shows that the predictive activity value is close to the experimental activity value. The predictive activity value which is close to the experimental activity value is shown in equation (2) so that the equation can be used for further research by modifying the eugenol compound based on the equation studies that have been carried out to get better activity.

Conclusion

Based on the results and discussion, it can be concluded that the quantitative relationship between structure and activity can be carried out to analyze eugenol derivative compounds with antioxidant activity indicated by an R-value of 0,973, namely the value of the predicted activity has approached the value of the experimental activity. The best equation model of the Quantitative Relationship The structure and activity of eugenol-derived compounds are compiled by electronic descriptors with homo energy parameters and HOMO-LUMO energy differences (Gap), hydrophobic descriptors with log parameters P, and steric descriptors with a Balaban index, with the equation model:

Log IC50 = 6,442 + (26,257) HOMO + (9,231) GAP + (0,056) LOGP – (0,433) BALABAN

References

Aini, N., Purwono, B., and Tahir, I. (2007). Structure – Antioxidant Activities Relationship Analysis of Isoeugenol, Eugenol, Vanilin and Their Derivatives. *English Journal of Chemistry*, 7(1), 61–66. https://doi.org/10.22146/ijc.21714

Arba, M., Andriansyah, R., and Leonita, M. (2016). Quantitative Relationship Study of the Activity Structure of Meisoindigo Derivative Compounds as Cdk4 Inhibitors. *Journal of Research Chemistry*, 1(2), 129. https://doi.org/10.20473/jkr.v1i2.3090

Asmuruf, F., Futwembun, A., and Kopeuw, A. J. (2017). Electronic Structure Characteristics in Goniotalamin, Infecttokaryon And Kurzicalkolakton Compounds Using The Density Functional Theory (DFT) Method. *AVOGADRO Journal of Chemistry*, 1(1), 1–5.

Da Silva, F. F.M., Monte, F. J. Q., de Lemos, T. L. G., do Nascimento, P. G. G., de Medeiros Costa, A. K., and de Paiva, L.M.M. (2018). Eugenol derivatives: synthesis, characterization, and evaluation

of antibacterial and antioxidant activities. *Chemistry Central Journal*, 12(34), 1–9. https://doi.org/10.1186/s13065-018-0407-4

Dhiman, P., Malik, N., and Khatkar, A. (2019). Lead Optimization for Promising Monoamine Oxidase Inhibitor from Eugenol for the Treatment of Neurological Disorder: Synthesis and In Silico Based Study. *BMC Chemistry*, 13(38), 1–20. https://doi.org/10.1186/s13065-019-0552-4

La Kilo, J., and La Kilo, A. (2019). QSAR Antimalarial Study of Quinolon-4(1H)-imines Derivative Compounds Using mlr-ann method. *Jambura Journal of Chemistry*, 1(1), 21–26. https://doi.org/10.34312/jambchem.v1i1.2104

Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2012). Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Advanced Drug Delivery Reviews*, 64(SUPPL.), 4–17. https://doi.org/10.1016/j.addr.2012.09.019

Male, Y. T., Sutapa, I. W., and Pusung, Y. A. D. (2018). Prediction of anticancer Potential of Xanthon Derivative Compounds using Quantitative Relationship of Structure and Activity (QSAR). Chem. Prog. , 11(1). https://doi.org/https://doi.org/10.35799/cp.11.1.2018.27608

Mardianingrum, R., Nuraisah, A., and Ruswanto. (2018). Quantitative Relationship of Antimalarial Structure-Activity of Quinoline-Pyrazolopyridine Derivative Compounds. *Journal of Pharmacopolium*, 1(3), 136–142.

Perwira, G., Kasmui, and Hadisaputro, S. (2015). Quantitative Relationship Analysis of the Structure of Antioxidant Activity of Apigenin-Derived Compounds. *Indonesian Journal of Chemical Science*, 4(3).

Santos, C.B. R., Lobato, C.C., Braga, F. S., Morais, S. S. S., Santos, C. F., Fernandes, C. P., Brazil, D. S.B., Hage-Melim, L. I. S., Macêdo, W. J.C., and Carvalho, J.C. T. (2014). Application of Hartree-Fock Method for Modeling of Bioactive Molecules Using SAR and QSPR. *Computational Molecular Bioscience*, 04(01), 1–24. https://doi.org/10.4236/cmb.2014.41001

Siswandono. (2016). Medical Chemistry (Kedu Edition). Airlangga University Press.

Sohilait, J., and Kainama, H. (2019). Free Radical Scavenging Activity of Essential Oil of Eugenia caryophylata from Amboina Island and Derivatives of Eugenol. *Open Chemistry*, 17(1), 422–428. https://doi.org/10.1515/chem-2019-0047

Soleh, A.M., and Aunuddin. (2013). Lasso: Alternative Solutions to The Selection of Changers and Shrinkage Coefficients Of Linear Regression Models. *Indonesian Journal of Statistics*, 18(1), 21–27.

Tjahjono, H., and Fadhilah, Q. (2012). Quantitative Relationship of Structure and Activity of 3-Haloasilaminobenzoilurea Derivative Compounds as Inhibitors of Microtubule Formation. *Acta Pharmaceutica English*, 37(3), 76–82.

Winarsi, H. (2007). Natural Antioxidants and Free Radicals. Yogyakarta: Kanisius.

World Health Organiation. (2020). Non-Communicable Diseases Progress Monitor 2020. In World Health Organization (Issue Oct). https://doi.org/10.5005/jp/books/11410_18