Structure-Activity Relationship of Chalcones Against Influenza H1N1: Quantum Chemical Chemiometric Evaluation

Marcos Vinícius C. S. Rezende, Flávio O. S. Neto, Ademir J. Camargo & Lilian T. F. M. Camargo

É bem conhecido que o vírus da influenza causa epidemias anuais provocando milhares de mortes todos os anos. Embora já exista uma vacina para a influenza, o vírus sofre mutação e adquire resistência sazonal a medicamentos. Portanto, devido à necessidade urgente de novos compostos com potencial inibitório do vírus, muitos estudos têm apresentado várias substâncias para evitar a atividade inibitória à enzima responsável pela liberação do vírus no trato respiratório. Aqui, foi realizada uma demonstração de cálculo teórico e estudos quimiométricos de um conjunto de oito chalconas com potencial inibição da enzima neuraminidase. Foram realizados cálculos de estrutura eletrônica para determinar os parâmetros geométricos e eletrônicos usando a teoria da função densidade no cálculo de nível M06-2X/6-311++G(d,p). A relação entre a estrutura dos compostos e sua atividade contra o vírus da influenza H1N1 foi acessada por análise de componentes principais e análise hierárquica de cluster. Esperamos que nossos resultados possam ajudar a explicar as atividades de análogos de chalcona para modelar novos compostos com atividade contra o vírus da influenza.

Palavras-chave: *pandemias globais; fármacos; DFT; PCA.*

It is well known that the influenza virus causes annual epidemics provoking thousands of deaths annually. Although there is already a vaccine for influenza, the virus mutates and acquires seasonal resistance to medicines. Therefore, due to the urgent need for new compounds with the inhibitory potential of the virus, many studies have presented several substances to avoid inhibitory activity to the enzyme responsible for the release of the virus into the respiratory tract. Here, a demonstrative calculation of theoretical and chemometric was performed studies of a set of eight chalcones with potential inhibition of the neuraminidase enzyme. It was performed calculations of electronic structure to determine the geometric and electronic parameters using the theory of density functional at M06-2X/6-311++G(d,p) level calculation. The relationship between the structure of the compounds and their activity against the H1N1 influenza virus was accessed by principal component analysis and hierarchical cluster analysis. We hope that our results can help to explain the activities of chalcone analogs to model new compounds with influenza virus activity.

Keywords: *global pandemics; drugs; DFT; PCA.*

Introduction

The search for compounds with high inhibitory power in viruses with potential for epidemics and pandemics is unequivocally important.[1–3] Recently, a new virus, known as COVID-19[4] China, was caused by a novel betacoronavirus, the 2019 novel coronavirus (2019-nCoV originating in southern China, classified as a pandemic by the World Health Organization (WHO)[5], had a substantial impact in most countries[6], causing collapses in the economy and health system[2]. Therefore, understanding the mechanisms of action of certain viruses, specifically those with high transmissibility capacity, is crucial for the development of new compounds active to control these microorganisms.[1, 8–11]

Other global pandemics occurred in the last century (1918, 1957, and 1960) caused by a virus with genes of avian origin and due to its genetic recombination.[12] In this century, a similar emergency situation occurred in 2009, with the emergence of the H1N1 virus[11, 13], resulting in international outbreaks. The influenza virus expresses two major glycoproteins on its surface, a haemagglutinin (HA) and neuraminidase (NA).[8] Since HA mediates the binding of viruses to target cells by sialic acid residue in glycoconjugates.[14] NA is responsible for the release of virion progeny and the general mobility of the virus in the respiratory tract[14]. Since amino acid residues from the NA active site have been shown to be highly conserved, it has been considered a good target for anti-influenza research.[15]

NA inhibitors such as oseltamivir[16], zanamivir[17], have been reported as effective against the H1N1 influenza virus. However, these drugs presented some clinical limitations due to the rapid dissemination, mutation, increased pathogenicity, and resistance acquired by the virus.[18] Therefore, the development of new and potential anti-influenza agents is necessary to fight H1N1 influenza since these virus remains to reach millions of people each year.

A series of studies have shown that compounds containing hydrazone fragments[8,19], derived from flavonoids[15] and methyl ester of 1- (1'-hydroxyalkyl) rupestonic acid[10] show good NA-inhibiting activity. Dao and collaborators[20] also analyzed the ability of chalcone derivatives to inhibit the NA enzyme, and 5-prenylbutein was the most potent compound with an IC_{50} around 5.80. Although it has been reported that chalcone analogues have inhibitory effects on the NA enzyme, an information limited was acquires from structure-activity relationship (SAR) of these compounds[15].

Therefore, considering a reduced number of molecules, quantum chemistry calculations combined with chemometric methods were used demonstratively may be advantageous to explain the activity of various compounds.[21–25] Here, Density Functional Theory (DFT)[26] were carried out to optimize and get the parameters the chalcones analogs. Principal Component Analise (PCA)[27] and Hierarchical Cluster Analysis(HCA)[27] were the methods used to elucidate the relationship between the chemical structure and the activity of the compounds.

Computational Procedure

MOLECULAR DESCRIPTORS

Chalcone derivatives are compounds that present many degrees of freedom, which allows for the existence of many stable conformation structures. In the absence of experimental structural data, it is important to carry out a conformational search on the compounds to find the conformation with the lowest energy. This search was carried out using the semi-empirical method pm3[28] implemented in HyperChem 8.5.[29] The conformation of lowest energy of each compound was optimized using the Density Functional Theory (DFT)[26] at the M062X/6-311++G(d,p)[30] level of theory using Gaussian G09 programs.[31]

From the optimized structures, the following molecular descriptors were calculated: the partial atomic charges (*C*) (derived from electrostatic potential (CHELPG) according to the method proposed by Breneman and Wiberg,[32] bond order indexes from NBO program, Kohn-Sham frontier molecular orbital energies (E_{HOMO} and E_{LUMO}), energy Gap ($E_{LUMO} - E_{HOMO}$), and hardness (η). [22, 33, 34] Also, bond angles (A), dihedral angles (D), dipole moment, molecular Volume (V), polarizability (α), and partition coefficient were calculated.[35]

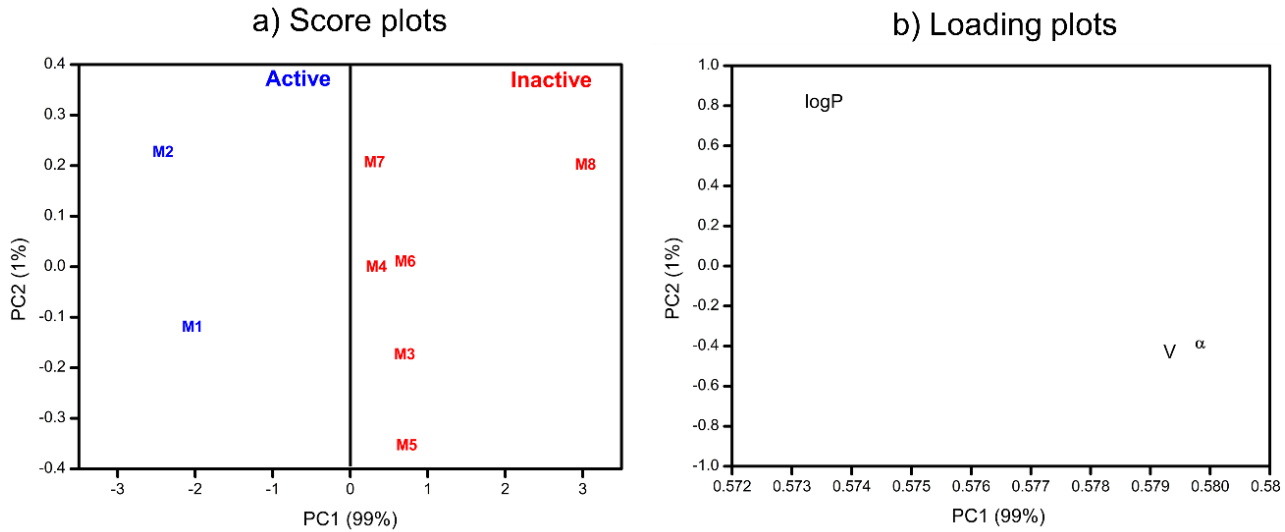


Figure 1. Graphical representations of the: a) scores for the chalcones derivates and b) loadings for the molecular descriptors selects by the PCA model that can discriminate the chalcones derivates in active and inactive classes. In addition, the variance explained by the two main components is represented in parentheses

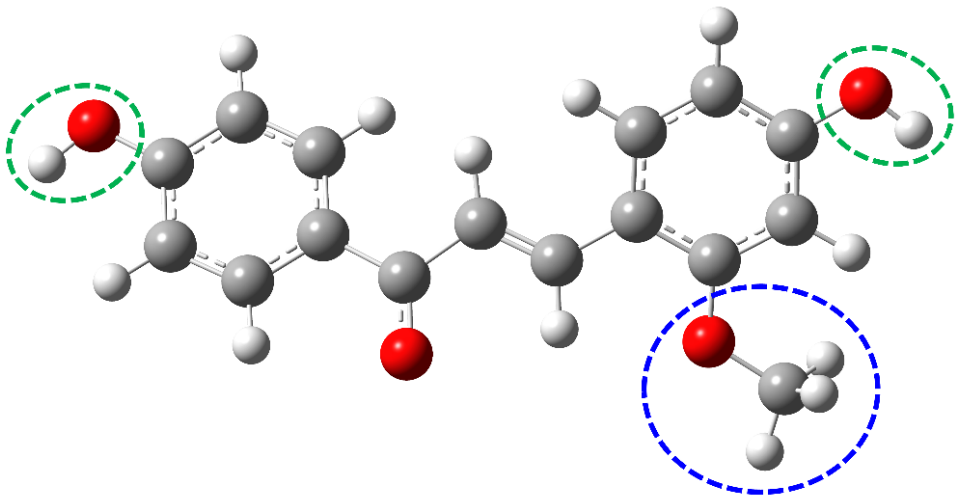


Figure 2. Representation of the structure with the highest activity (M1). Discussions of structure-activity relationships for the design of a new compound are in the text

DESCRIPTORS' SELECTION

The PCA is a method that creates new variables as linear combinations from the originals. The components (new variables) are orthogonal to each other, and, therefore, each component accounts for a dimension of the original data. The PCA analysis reduces the dimensionality of the data set that, in general, presents many interrelated variables. The components are ranked in such a way that the PC1 accounts for the largest variance; PC2 accounts for the second largest variance and so on.

Principal component analysis (PCA) was used to establish a relationship between quantum chemical descriptors and the activities of compounds against H1N1 disease. Before starting the PCA analysis, Fisher's weight analysis was performed on the descriptors to measure their power discrimination. Fisher's weight is defined as the ratios of the square of the interclass means to the sum of the intraclass variances:[27]

$$W_i\,(A,I) = \frac{[\bar{x}_i(A) - \bar{x}_i(B)]^2}{s_i^2(A) + s_i^2(B)} \qquad (1)$$

Where A and B are the classes considered in the investigation; represents the descriptors; is the average of the descriptor in class; is the average of the descriptor in class ; is the variance of the descriptor in class and is the variance of the descriptor in class. Descriptors with higher Fisher's weights are generally important in the PCA analysis to discriminate compounds in active and inactive classes.

The hierarchical cluster analysis (HCA) is also an exploratory technique that examines the similarity between the sample in a data set. Here, the HCA technique was carried out using the selected descriptors by the Fisher's weights and PCA, for the purposeto validate the PCA analysis.

Results and Discussion

Table 1 shows the molecular structure and the nomenclature used in the calculations. The IC_{50} values were obtained from th work of Dao and collaborators.[20] The discrimination of molecules also is presented: active (M1 and M2) and inactive (M3-M8) classes.

Before applying the PCA analysis, each variable was auto-scaled, *i.e.*, each variable was scaled to zero mean and unity variance. It was necessary to compare the descriptors on the same scale. From the descriptors selected by Fisher's weight, many PCA analyses were performed with the aim of always obtaining a classification with the least possible descriptors. The best separation was obtained considering three descriptors (Table 3): V, *logP*, and α .

Figure 1 shows the majority contributions for the classification into active and inactive classes. PC1 explains 99% of the total variance of the data, and the first two components together explain 100% of the total variance.

The active compounds have negative scores, and the inactive compounds have a positive score on the PC1 (Figure 1). Loadings can be interpreted to show the weights of physicochemical properties that influence activity and that are important factors in understanding the behavior of compounds in the real world.[36, 37] The PC1 loadings for the Volume (V), logP, and polarizability (α) descriptors are 0.5792, 0.5732, and 0.5797, respectively, which is responsible for the separation.

Equation 2 shows the linear combination of the selected descriptors that are used to calculate the scores plotted in Figure 1. According to Figure 1, for a compound to be classified as active against H1N1, it must have a negative score. That means that a molecule will be classified as active when it presents a small volume, partition coefficient, and polarizability.

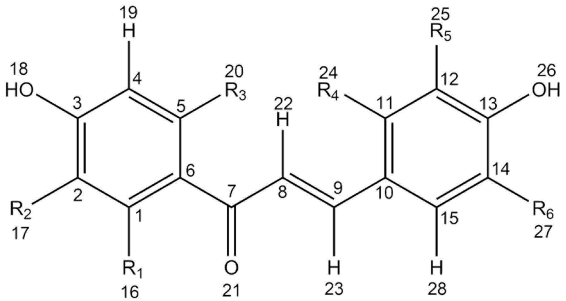
$$PC1 = 0.5792V + 0.5732logP + 0.5797\alpha \qquad (2)$$

Volume and polarizability can vary expressively with substitution of substituents, interfering with the linkage of the drug-receptor system. The is defined as the ratio of the concentration of a given substance between two immiscible solvents, i.e.,

$$logP_{octanol/water} = Log\left(\frac{[solute]_{octanol}}{[solute]_{water}}\right), \qquad (3)$$

Where $[solute]_{octanol}$ stands for the solute concentration in *n-octanol* and $[solute]_{water}$ stands for the concentration of the solute in water. *logP* From Eq. 2, the classification as active, requires small *logP*, hydrophilic compound as preferential structures.[36]

Table 1. Molecular structure and nomenclature adopted in calculations of chalcones derivatives and their respective activities



Molecules	Substituents	$IC_{50}\ \mu g/mL$	Activity
M1	R_1=R_2=R_3=R5=R_6=H R_4=O-CH_3	5.80±0.30	Active
M2	R_2=R_3=R_4=R_5=R_6=H R_1=OH	8.41±0.39	Active
M3	R_1=R_2=R_5=H R_3=OH; R_4=O-CH_3 R_6= 3,3-dimethyl-1-butyne	37.68±2.17	Inactive
M4	R_1=R_2=R_4=H R_3=R_5=OH R_6=2-methyl-2-pentene	19.09±1.10	Inactive
M5	R_1=R_3=R_6=H R_2=2-methyl-2-pentene R_4=O-CH_3, R_5=OH	25.87±2.03	Inactive
M6	R_1=R_2=R_3=R5=H R_4=O-CH_3 R_6= 3,3-dimethyl-1-butyne	28.62±1.67	Inactive
M7	R_1=R_2=R_4=R_5=H R_3=OH R_6=2-methyl-2-pentene	51.59±2.77	Inactive
M8	R_1=OH R_3=R_4=R5=H R_2=R_4=2-methyl-2-pentene	75.38±2.47	Inactive

The descriptors with Fisher's weight greater than 1.0 are listed in Table 2. This method was used for a preliminary assessment of the variables with the main contribution of the compound classifications. The selected variables from Fisher's weight analysis were: polarizability (α), refractibility (R), molar volume (V), surface area(A), atomic partial charge at 13 and 26 atoms (C13 and C26, respectively), partition coefficient (*logP*), and dihedral angles ($D_{16\_1\_6\_7}$, $D_{\_1\_2\_3\_18}$, and $D_{\_9\_10\_15\_14}$).

Table 2. Fisher's weights obtained at M06-2X/6-311++G(d,p) level calculation

Variáveis	Peso Fisher
V	11.04
logP	8.22
α	12.29
R	11.27
A	9.04
C_{13}	3.40
C_{26}	8.76
$D_{\_1\_6\_7\_16}$	2.13
$D_{\_1\_2\_3\_18}$	2.00
$D_{\_9\_10\_14\_15}$	1.01

Table 3. Descriptors capable of classifying compounds into active and inactive classes and their respective values calculated at the M06-2X/6-311++G(d,p) level of theory

Molecules	V (Å3)	*logP*	α (Å3)
M1	800.81	5.00	29.24
M2	743.74	5.07	27.4
M3	1020.87	6.66	38.86
M4	986.23	6.59	37.02
M5	1045.76	6.52	38.86
M6	1016.26	6.83	38.22
M7	973.82	6.76	36.39
M8	1201.42	8.46	45.37

Previous studies have shown that the presence of OH groups play an important role in biological activity, since these groups can result in strong hydrogen bonds that presumably interfering with the linkage of the drug-receptor system.[38–40] On the other hand, the presence of methoxy groups may be associated with less activity.[39] Based on this information and rationalizing the design of a drug, the replacement of the methoxy group by an OH group may be promising as a potential inhibitor for H1N1 virus Neuraminidase. This proposal is consistent with Eq 2 since the resulting compound has a lower volume, polarizability, and possibly more hydrophilic.

Figure 3 shows the hierarchical cluster analysis (HCA) carried out on the descriptors selected by PCA analysis. The dendrogram classifies the chalcones derivatives into two clusters: active (compounds 1 and 2) on the left and inactive cluster (compounds from 3 to 8). HCA and PCA classified 8 chalcones compounds under study exactly in the same way.

It is not easy to infer about the drug-receptor interaction when the target receptor site is not known. The results indicate that geometrics parameters and solubility are essential to explain the activities of the chalcone derivatives against the H1N1 virus. The descriptors volume (V), logP, and polarizability (α) can guide us in modeling new compounds with activities against the H1N1 virus.

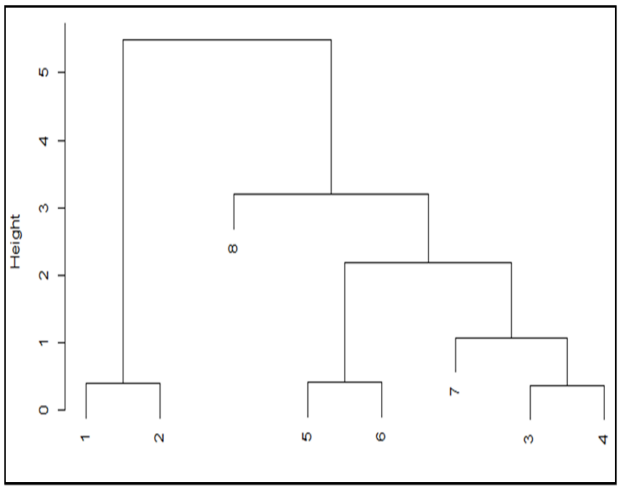


Figure 3. Dendrogram obtained from HCA analysis. The active compounds are clustered on the left

Conclusion

For demonstrative purposes Principal component analysis (PCA) and hierarchical cluster analysis (HCA) were applied on the molecular descriptor set obtained at M062X/6-311++(d,p) level of theory to classify eight chalcones derivatives into active and inactive classes against

the H1N1 virus. The molecular volume, polarizability, and LogP descriptors were able to discriminate the compounds into two classes: active (compounds 1 and 2) and inactive (compounds from 3 to 8). The PC1 alone was able to discriminate the compounds, and it accounts for 99% of the total data variance. For a compound to be classified as active, it must have lower volume, polarizability, and *logP.* These results indicate that smaller and more hydrophilic compounds as preferential more efficient in action against the virus and. Our results guide the design of novel pharmaceuticals against the H1N1 influenza virus.

Acknowledgments

The authors are grateful for the support given by the Brazilian agency CAPES. This research is also supported by the High-Performance Computing Center at the Universidade Estadual de Goiás, Brazil. Lilian T. F. M. Camargo thanks Foundation for Research Support of the State of Goiás (FAPEG).

References

1. Girard MP, Tam JS, Assossou OM, Kieny MP. The **2009** A (H1N1) influenza virus pandemic: A review. *Vaccine* **2010**;28:4895–4902.
2. Hancock K, Veguilla V, Lu X, Zhong W, Butler EN, *et al.* Cross-reactive antibody responses to the **2009** pandemic H1N1 influenza virus. *N Engl J Med* **2009**;361:**1945**–**1952**.
3. Le TT, Andreadakis Z, Kumar A, Roman RG, Tollefsen S, *et al.* The COVID-19 vaccine development landscape. *Nat Rev Drug Discov* **2020**;19:305–306.
4. Huang C, Wang Y, Li X, Ren L, Zhao J, ***et al.*** Clinical features of patients infected with **2019** novel coronavirus in Wuhan, China. *Lancet* **2020**;395:497–506.
5. Organization WH, others. Coronavirus disease **2019** (COVID-19) pandemic.
6. Organization WH, others. Coronavirus disease **2019** (COVID-19): situation report, 72.
7. Qiu Y. Impacts of Social and Economic Factors on the Transmission of Coronavirus Disease **2019** (COVID-19) in China ∗. *medRxiv* **2020**;**2019**:1–27.
8. Cui MY, Nie JX, Yan ZZ, Xiao MW, Lin D, *et al*. Design, synthesis, bioactivity, and DFT calculation of 2-thiazolyl-hydrazone derivatives as influenza neuraminidase inhibitors. *Med Chem Res* **2019**;28:938–947.
9. Li H, Li M, Xu R, Wang S, Zhang Y, *et al. Synthesis, structure activity relationship and in vitro anti-influenza virus activity of novel polyphenol-pentacyclic triterpene conjugates*. Elsevier Masson SAS. Epub ahead of print **2019**. DOI: 10.1016/j.ejmech.**2018**.12.006.
10. Li G, Zhao JY, Niu C, Nie LF, Dong CZ, et al. Structure–activity relationship studies of 1-(1′-hydroxyalkyl)rupestonic acid methyl esters against influenza viruses. *Bioorganic Med Chem Lett* **2017**;27:1484–1487.
11. Neumann G, Noda T, Kawaoka Y. Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature* **2009**;459:931–939.
12. Organization WH, others. Ten things you need to know about pandemic influenza (update of 14 October **2005**). *Wkly Epidemiol Rec Relev épidémiologique Hebd* **2005**;80:428–431.
13. Garten RJ, Davis CT, Russell CA, Shu B, Lindstrom S, *et al.* Antigenic and genetic characteristics of swine-origin **2009** A(H1N1) influenza viruses circulating in humans. *Science (80-)* **2009**;325:197–201.
14. Mercader AG, Pomilio AB. QSAR study of flavonoids and biflavonoids as influenza H1N1 virus neuraminidase inhibitors. *Eur J Med Chem* **2010**;45:1724–1730.
15. Gao L, Zu M, Wu S, Liu AL, Du GH. 3D QSAR and docking study of flavone derivatives as potent inhibitors of influenza H1N1 virus neuraminidase. *Bioorganic Med Chem Lett* **2011**;21:5964–5970.
16. Taylor NR, von Itzstein M. Molecular modeling studies on ligand binding to sialidase from influenza virus and the mechanism of catalysis. *J Med Chem* **1994**;37:616–624.
17. Kim CU, Lew W, Williams MA, Liu H, Zhang L, *et al.* Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. *J Am Chem Soc* 1**997**;119:681–690.
18. Ison MG. Antivirals and resistance: influenza virus. *Curr Opin Virol* **2011**;1:563–573.
19. Mu JX, Shi YX, Wu HK, Sun ZH, Yang MY, ***et al.*** Microwave assisted synthesis, antifungal activity, DFT and SAR study of 1,2,4-triazolo[4,3-a]pyridine derivatives containing hydrazone moieties. *Chem Cent J* **2016**;10:1–9.
20. Dao TT, Nguyen PH, Lee HS, Kim E, Park J, ***et al.*** Chalcones as novel influenza A (H1N1) neuraminidase inhibitors from Glycyrrhiza inflata. *Bioorganic Med Chem Lett* **2011**;21:294–298.
21. Alves CN, Pinheiro JC, Camargo AJ, Ferreira MMC, Romero RAF, ***et al.*** A multiple linear regression and partial least squares study of flavonoid compounds with anti-HIV activity. *J Mol Struct THEOCHEM* **2001**;541:81–88.

22. Camargo LTFM, Sena MM, Camargo AJ. A quantum chemical and chemometrical study of indolo[2,1-b]quinazoline and their analogues with cytotoxic activity against breast cancer cells. *SAR QSAR Environ Res* **2009**;20:537–549.

23. Worachartcheewan A, Nantasenamat C, Isarankura-Na-Ayudhya C, Prachayasittikul V. QSAR Study of H1N1 Neuraminidase Inhibitors from Influenza a Virus. *Lett Drug Des Discov* **2014**;11:420–427.

24. Reed G. Feature Selection Methods in QSAR Studies. *J AOAC Int* **2012**;95:5–24.

25. Vieira I, Camargo LTFM, Ribeiro L, Rodrigues ACC, Camargo AJ. Structure–activity relationship of tacrine and its analogues in relation to inhibitory activity against Alzheimer's disease. *J Mol Model*;25. Epub ahead of print **2019**. DOI: 10.1007/s00894-019-3993-8.

26. Kohn W, Sham LJ. Self-consistent equations including exchange and correlation effects. *Phys Rev* **1965**;140:1133–1138.

27. Sharaf MA, Illman DL, Kowalski BR. *Chemometrics*. Wiley; **1986**.

28. Stewart JJP. Optimization of parameters for semiempirical methods II. Applications. *J Comput Chem* **1989**;10:221–264.

29. Hypercube Inc. HyperChem, Release 7.0.

30. Zhao Y, Truhlar DG. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other function. *Theor Chem Acc* **2008**;120:215–241.

31. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, *et al.* Gaussian 09, Revision A.02. *Gaussian Inc Wallingford CT* **2009**;34:Wallingford CT.

32. Breneman CM, Wiberg KB. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J Comput Chem* **1990**;11:361–373.

33. Camargo AJ, Honório KM, Mercadante R, Molfetta FA, Alves CN, *et al.* A Study of Neolignan Compounds with Biological Activity Against Paracoccidioides brasiliensis by Using Quantum Chemical and Chemometric Methods. *J Braz Chem Soc* **2003**;14:809–814.

34. Martins GR, Napolitano HB, Camargo LTFM, Camargo AJ. Structure-activity relationship study of rutaecarpine analogous active against central nervous system cancer. *J Braz Chem Soc* **2012**;23:2183–2190.

35. Karelson M, Lobanov VS, Katritzky AR. Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem Rev* **1996**;96:1027–1044.

36. Bhal SK. Understanding When to Use Log P & Log D. *ACD/Labs - Adv Chem Dev Inc Toronto, Canada* **2019**;3–6.

37. Österberg T, Norinder U. Prediction of drug transport processes using simple parameters and PLS statistics - The use of ACD/logP and ACD/ChemSketch descriptors. *Eur J Pharm Sci* **2001**;12:327–337.

38. Rice-Evans C, Miller N, Paganga G. Antioxidant properties of phenolic compounds. *Trends Plant Sci* **1997**;2:152–159.

39. Rosa GP, Seca AML, Barreto M do C, Silva AMS, Pinto DCGA. Chalcones and flavanones bearing hydroxyl and/or methoxyl groups: Synthesis and biological assessments. *Appl Sci*;9. Epub ahead of print **2019**. DOI: 10.3390/app9142846.

40. Yaeghoobi M, Frimayanti N, Chee CF, Ikram KK, Najjar BO, *et al.* QSAR, in silico docking and in vitro evaluation of chalcone derivatives as potential inhibitors for H1N1 virus neuraminidase. *Med Chem Res* **2016**;25:2133–2142.

Marcos Vinícius C. S. Rezende[1,2], Flávio O. S. Neto[3], Ademir J. Camargo[2] & Lilian T. F. M. Camargo[1,2]*

[1] Instituto Federal de Educação, Ciência e Tecnologia de Goiás, Câmpus Anápolis, Av. Pedro Ludovico, S/N - Residencial Reny Cury, 75131-457, Anápolis - GO, Brazil

[2] Grupo de Química Teórica e Estrutural de Anápolis (QTEA), Câmpus de Ciências Exatas e Tecnológicas. Universidade Estadual de Goiás, CP 459, 75001-970 Anápolis, GO Brazil.

[3] Instituto de Química, Universidade de Brasília, Caixa Postal 4478,70904-970, Brasília, Brazil

*E-mail: lilianthaty@yahoo.com.br