

QSAR of Lipid-Coated Virus Inhibition by Fatty Acids and Monoglycerides Using Electrotological Descriptors

Florentino C. Sumera *

Environmental-Health laboratory Services Cooperative, 21 Chestnut St. West Fairview, Quezon City 1118; 40 Mt. Fairweather, Filinvest 1, Batasan Hills, Quezon City, 1126

*Author to whom correspondence should be addressed; email: florentino.sumera@gmail.com

ABSTRACT

E-State QSAR models were developed for the inhibition of a lipid-coated virus with a set of 19 fatty acids (FA) and their monoglycerides (MG). The FA and MG were also divided into two subsets to improve correlations, the saturated ($n = 12$) and the unsaturated set ($n = 7$). QSAR of the sets yielded better correlation statistics; the saturated FA and MG gave $r^2 = 0.87$, $s = 0.26$ while the QSAR of the unsaturated FA and MG gave $r^2 = 0.9999$, $s = 0.0041$. The combined set model however yielded only $r^2 = 0.81$, $s = 0.32$. Statistically satisfactory four and five variable models for the saturated and the unsaturated respectively and three variable model for the combined set FA and MG were developed. Structure interpretation is given for each variable emphasizing their effects on the resulting viral inhibition (pIC values). A leave-one-out method (LOO) was applied to investigate the predictive quality of each of the QSAR equation. This predictive quality of the multiple regression equation will be useful in the prediction of new potential antiviral compounds against lipid-coated virus. This paper hopes that the study could also be applied to SARS-CoV-2 virus, which is also a lipid-coated virus.

Keywords: *electrotological state; quantitative structure activity relationships; lipid-coated virus; fatty acids and monoglycerides; virus inhibition; SARS-CoV-2 virus*

INTRODUCTION

Our country is faced with a great problem of eliminating the new coronavirus SARS-CoV-2. While vaccines have arrived, the arrival of enough supply to protect the whole population is very slow even to just the level that herd immunity could be felt. Drugs such as remdesivir, hydroquinone and steroid drugs like dexamethasone have been recommended for use to alleviate the suffering of those affected by the virus. But the plague seems to be a mythical hydra whose octopus-like arms spring back every time they are cut, proliferating whenever and wherever the population relax their guard. We have therefore to find our own solution in inhibiting the proliferation of the

virus one way or the other as it appears from place to place while we wait for the vaccine. Even the drugs recommended are not easily available to the majority of our citizens. There is a need to try any solution that is possible like searching for already known safe drugs or molecules that may possess anti-coronavirus properties.

The objective of this paper is to provide a means of finding new or known molecules that potentially have antiviral properties and can be directly applied where the virus concentrates. The search could be facilitated if we know molecules that have been studied against a similar virus, like for example, the case of fatty acids (FA) and monoglycerides (MG) that have been reported to attack lipid-coated virus resulting to viral inhibition. Fortunately, the coronavirus SARS-CoV-2 belongs to a class of lipid-coated virus that proliferates from person to person by penetrating and then assuming the same double layer lipid membrane of their host's cell as their own. Because of this behavior, the FA and MG, we are talking about, can block the virus at the first stage of entry, by intercalating with the lipid membrane of the virus, thus, inducing changes in the lipid coating and dampening the virus infectivity even without lysis [O'Donnell et al., 2020]. While there is not yet a report of a direct study on the SARS-CoV-2 viral inhibition by these FA and MG, we can take advantage of their anti-viral activity related to their structure and develop a kind of a SAR (structure activity relationship). We can use the same SAR in searching for other antiviral compounds that may be appropriate for particular applications against the SARS-CoV-2 virus, for example, as component of mouth wash [O'Donnell et al., 2020; Carrouel et al., 2021]; Meiller et al., 2005], nasal spray [Hilmarsson et al., 2013; Schroder and Svenson, 1999] or simple inhaler [Patne et al., 2020]. A study in the 1970's showed that FA and MG of 16 – 18 carbon chain length are highly effective *in vitro*, reducing survival of *Herpes simplex* virus to around 50% at micromolar concentrations [Sands, Auperin et al., 1979]. Thormar et al., [1987] have reported a more or less complete anti-viral activity of FA and MG against the lipid-coated virus VSV (*Vesicular stomatitis virus*), HSV-I (*Herpes simplex virus*) and VV (*visna virus*). This antiviral activity is specific only to lipid coated virus (other known lipid coated viruses are the well-known SARS, MERS and H₁N₁ and of course the SARS-CoV-2 virus. These FA and MG were found to destroy the lipid coating envelope of the VSV, HSV, and VV virus leading to leakage and then complete disintegration of the envelope and viral particles. They thus affect cell lysis followed by death of the virus [Thormar et al, 1987]. Although there are other proposed mechanisms, there is evidence, as other authors reported that the same process of inhibition by FA and MG happens with the different lipid coated virus [Pfaender et al., 2013].

In this paper we shall use topological methods to provide an excellent basis for development of models for viral inhibition more particularly by FA and MG with the intention of finding and predicting new candidate molecules possessing viral inhibition property. These topological models will directly give us specific structure information to guide the design of each molecule that are antiviral and suitable for a desired application. Topological structure descriptors will serve as the basis for structural representation in the model. These electrotopological indices have been used in model equations for predicting physico-chemical and biological property of molecules. The topological structure representation approach has been used successfully in QSAR [Contrera et al., 2005; Maw and Hall, 2000; Hall and Vaughn, 1997; Abou-Shaaban et al., 1996] as well as physico-chemical QSPR [Hall and Story, 1996; Hall and Story, 1997; Votano et al., 2004; Huuskonen, 2001] data collection. Topological approach clearly indicates their usefulness in drug design and property estimation. The method based on topological structure representation is considered less time-consuming and hence less costly to use for prediction than the 3-D modeling [Rose and Hall, 2003]. This topologically based method can lead us to an easy computation of pIC values (viral inhibition) for candidate molecules and thus can act as a preliminary antiviral bioassay. The QSAR modeling will be followed by a statistical validation such as, the leave-one-out (LOO) statistics that will be used as a means of indicating predictive ability.

METHODS

Data Input. Viral inhibition data have been measured by Thormar et al. [1987] and expressed here from the molar values to its negative logarithm, $pIC = -\log(1/C)$. See Table 1 for the molecular representations and the experimental pIC values. The available combined data ($n = 19$) of FA and their MG were then subdivided into two sub-sets, the saturated ($n = 12$) and unsaturated ($n = 7$) FA and MG respectively to improve the SAR relationships.

The E-State values of each molecule and E-State indices for atoms and groups of atoms were determined by

1. Assigning the intrinsic values I to atoms or group of atoms in the molecule
2. Solving for the electrotopological state S of each atom i in the molecule
3. Combining the atom-type or group-type E-State indices for each molecule to serve as the E-State descriptors or variables [e.g. $S^T(-CH_2-)$] in the multiple regression

Once all the E-State indices for atoms and group of atoms for each molecule are determined, atom-type or group-type E-State indices are combined and the process is ready for regression with the observed activity for the set of molecules. The actual statistical analysis (multiple regression analysis) was carried out using an EXCEL spreadsheet and its data analysis command for the mathematical regression. Atom type E-State and hydrogen E-State indices, along with molecular connectivity chi indices were calculated.

Table 1. Viral inhibition by incubation with fatty acids and monoglycerides at 37° C for 30 minutes [Thormar et al., 1987].

Fatty Acids (FA) & Monoglycerides (MG)	Inhibition conc. ^a (mM)	pIC^b [$\log(1/C)$]
Fatty Acids		
Butyric (4:0) ^c	113	0.942
Caproic (6:0)	88	1.07
Caprylic (8:0)	69	1.16
Capric (10:0)	22	1.66
Lauric (12:0)	10	2.00
Myristic (14:0)	16	1.80
Palmitic (16:0)	78	1.11
Stearic (18:0)	70	1.15
Palmitoleic (16:1)	15	1.82
Oleic (18:1) cis	7	2.15
Linoleic (18:2)	3.5	2.46
Linolenic (18:3)	3.6	2.44
Arachidonic (20:4)	1.6	2.80
Monoglycerides		
Monocaprylin (8:0)	9	2.05
Monocaprin (10:0)	2	2.70
Monolaurin (12:0)	0.9	2.54
Monomyristin (14:0)	13.0	1.89
Monoolein (18:1)	2.8	2.55
Monolinolein (18:2)	0.7	3.15

^a Concentration of fatty acid in virus mixture incubated at 37°C for 30 min.

^b Carbon atoms : double bonds.

^c In molar concentrations.

Before the multiple regression was performed the indices were first paired and correlations were performed to determine if pairs have correlation coefficient greater than 0.80. For each pairing process that produce such correlation coefficient, one of the pair of correlated variables was eliminated before preparing the QSAR equation. Spotting this pair or pairs was easily done using a correlation matrix for easy reference. The E-State indices or variables finally selected are primarily considered to be more easily interpreted in terms of molecular structure.

Statistical Analysis. All possible subset regressions were performed to examine every QSAR equation from one to five variables, listing the top three equations possessing coefficients, which are all significant ($p < 0.05$). All possible sets of variables are then considered and ranked on the r^2 values. The model QSAR equation is then selected from these requirements.

Model Development. The descriptors found to be relevant for modeling, making conservative use of structure information and giving satisfactory regression statistics for a combined FA and MG ($n = 19$) are the following;

- $^1\chi^v$: first order molecular connectivity valence chi index.
- $S^T(-OH)$: atom type E-State index for the hydroxyl oxygen atom.
- $S^T(-CH_2)$: atom type E-State index for the methylene carbon atom.

Since we have also subdivided the combined FA and MG into two data subsets, saturated FA/MG and unsaturated FA/MG, additional descriptors were also used as follows,

- $HS^T(\text{other})$: atom type hydrogen E-State for non-polar hydrogen atoms.
- $S^T(O=)$: atom type E-State index for the carbonyl oxygen atom.
- $S^T(-CH_3)$: atom type E-State index for the methyl carbon atom
- $S^T(-CH=)$: atom type E-State index for the vinyl carbon atom.

Validation Studies. Leave-one-out method. To demonstrate that the model is useful for prediction we carried out a leave-one-out method (LOO) where one equation is left out from a set and a multiple regression was done on the reduced set (Hall et al., 1991). The resulting coefficients from the regression were then substituted to the left -out equation to give the predicted pIC values. This was done X times as the number of the equations in a set.

Randomization Study. To determine the possibility that the model is no different from a simple random model, the pIC values were scrambled 10 times. For each randomized model, the r^2 and F values were listed and then averaged. The results were then compared with the model.

RESULTS

The QSAR models developed for the combined set ($n = 19$) FA and MG, the saturated set ($n=12$) and the unsaturated set ($n = 7$) are given in the following equations. Quantities in parentheses are the standard deviations of the coefficients. The variables in the models (eqns. 1-3) are statistically independent (their individual correlation coefficients with each other are above $r^2 = 0.80$).

1. Combined Set ($n = 19$)

$$pIC = 0.466(\pm 0.074) S^T(^1X^v) - 0.153(\pm 0.032) S^T(-CH_2-) - 0.077(\pm 0.026) S^T(-OH) - 1.03(\pm 0.28) \quad (1)$$

$r^2 = 0.81 \quad s = 0.32 \quad F = 21 \quad n = 19$

2. Saturated FA/MG Set (n = 12)

$$\begin{aligned} \text{pIC} = & 7.2(\pm 2.2)\text{S}^{\text{T}}(\text{O}=\text{O}) - 7.06(\pm 2.2)\text{HS}^{\text{T}}(\text{other}) + 2.29(\pm 0.67)\text{S}^{\text{T}}(-\text{CH}_3) \\ & + 2.00(\pm 0.63)\text{S}^{\text{T}}(-\text{CH}_2-) - 64(\pm 19) \end{aligned} \quad (2)$$

$$r^2 = 0.87 \quad s = 0.26 \quad F = 12 \quad n = 12$$

3. Unsaturated FA/MG Set (n = 7)

$$\begin{aligned} \text{pIC} = & 2.039(\pm 0.084)\text{S}^{\text{T}}(-\text{CH}_3) - 0.4548(\pm 0.0049)\text{HS}^{\text{T}}(\text{other}) + \\ & 0.4544(\pm 0.0040)\text{S}^{\text{T}}(-\text{CH}=\text{C}) + 0.3635(\pm 0.0034)\text{S}^{\text{T}}(-\text{OH}) + \\ & 0.3412(\pm 0.0040)\text{S}^{\text{T}}(-\text{CH}_2-) - 8.76(\pm 0.17) \end{aligned} \quad (3)$$

$$r^2 = 0.9999 \quad s = 0.0041 \quad F = 13214 \quad n = 7$$

Plots of the calculated pIC values versus the observed pIC values are shown in Figures 1-3. The observed pIC values, calculated, and residual are listed in Tables 2 - 4. An examination of the plot of residual versus observed pIC (not shown) revealed no trends and appear randomly distributed.

The summary data for the LOO method are listed on Table 5 - 7. For the combined data set, no residual exceeds two standard deviations (0.31). For the saturated subset, no residual exceeds two standard deviation (0.21) and for the unsaturated subset no residual also exceeds two standard deviation (0.004) except only that of the coefficient of $\text{S}^{\text{T}}(-\text{CH}_3)$.

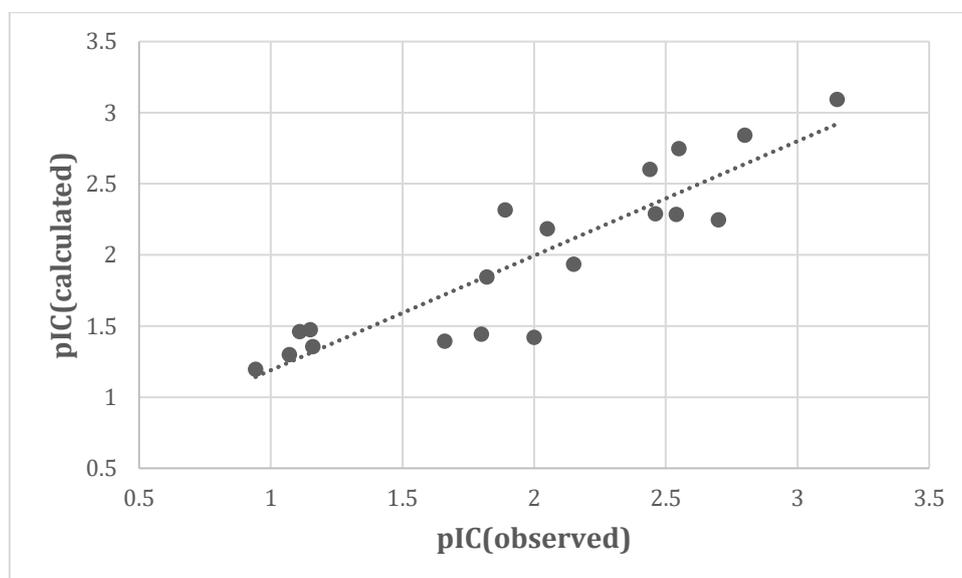


Figure 1. Plot of calculated pIC values versus observed pIC values for the combined saturated and unsaturated data set using equation (1)

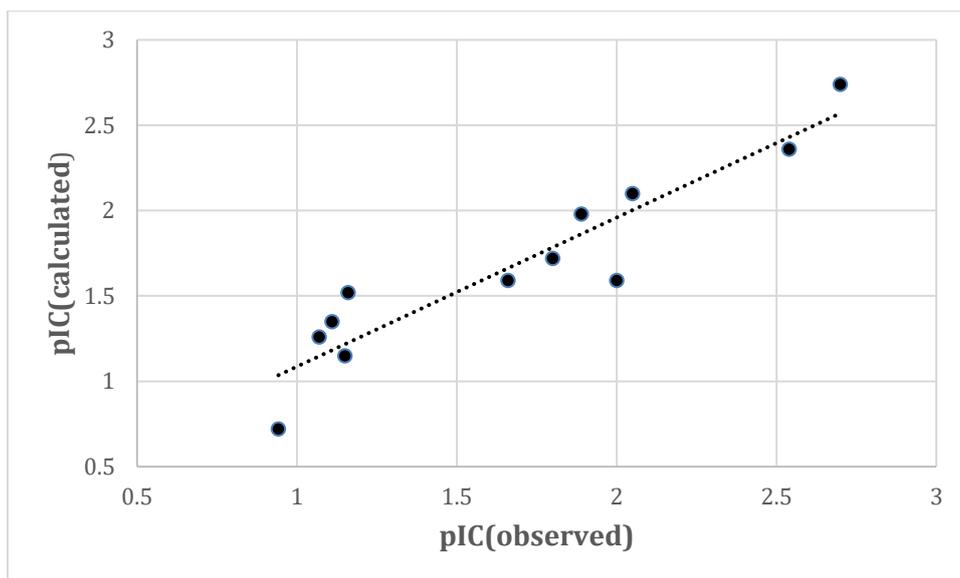


Figure 2. Plot of calculated pIC values versus observed pIC for the saturated FA/MG data set using equation (2)

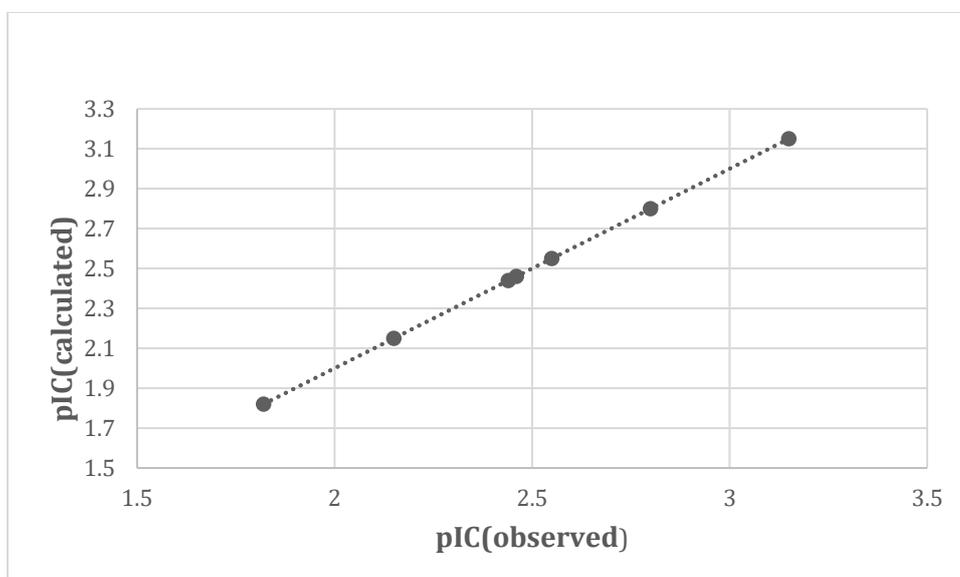


Figure 3. Plot of calculated pIC values versus observed pIC values for the unsaturated FA/MG data set using equation (3)

Table 2. Observed, Calculated, and Residual Viral Inhibition Data for Combined Saturated and Unsaturated FA/MG with the Electrotopological Values.

Entry	FA/MG	S(¹ χ ^v)	S(-CH ₂ -)	S(-OH)	pIC	calc	res
1	Butyric (4:0)	1.989	1.023	7.913	0.942	1.20	-0.26
2	Caproic (6:0)	2.989	3.277	8.140	1.07	1.30	-0.23
3	Caprylic (8:0)	3.989	5.884	8.270	1.16	1.36	-0.20
4	Capric (10:0)	4.989	8.642	8.354	1.66	1.39	0.27
5	Lauric (12:0)	5.989	11.480	8.413	2.00	1.42	0.58
6	Myristic (14:0)	6.989	14.359	8.457	1.80	1.44	0.36
7	Palmitic (16:0)	7.989	17.269	8.490	1.11	1.46	-0.35
8	Stearic (18:0)	8.989	20.199	8.517	1.15	1.48	-0.33
9	Monocaprylin (8:0)	5.649	- 1.013	21.337	2.05	2.18	-0.13
10	Monocaprin (10:0)	6.649	1.565	21.436	2.70	2.25	0.45
11	Monolaurin (12:0)	7.649	4.325	21.513	2.54	2.29	0.25
12	Monomyristin (14:0)	8.649	7.143	21.569	1.89	2.32	-0.43
13	Palmitoleic (16:1)	7.638	13.689	8.481	1.82	1.85	-0.03
14	Oleic (18:1)	8.638	16.579	7.619	2.15	1.94	0.21
15	Linoleic (18:2)	8.288	13.203	7.614	2.46	2.29	0.17
16	Linolenic (18:3)	7.938	10.091	7.610	2.44	2.60	-0.16
17	Arachidonic (20:4)	8.587	10.507	7.604	2.80	2.84	-0.04
18	Monoolein (18:1)	10.29	9.278	21.642	2.55	2.75	-0.20
19	Monolinolein (18:2)	9.948	5.981	21.635	3.15	3.09	0.06

Table 3. Observed, Calculated, Residual Inhibition Data for Saturated FA/MG with the Electrotopological Values

Entry	FA/MG (Satd)	S ^T (=O)	HS ^T (other)	S ^T (CH ₂)	S ^T (CH ₃)	pIC	calc	res
1	Butyric	9.599	1.566	1.023	1.841	0.942	0.72	0.22
2	Caproic	9.874	2.478	3.277	2.057	1.07	1.26	0.19
3	Caprylic	10.032	3.370	5.884	2.150	1.16	1.52	0.36
4	Capric	10.134	4.259	8.642	2.199	1.66	1.59	0.07
5	Lauric	10.206	5.144	11.478	2.228	2.00	1.59	0.41
6	Myristic	10.260	6.002	14.359	2.247	1.80	1.72	0.08
7	Palmitic	10.301	6.924	17.269	2.260	1.11	1.35	0.24
8	Stearic	10.333	7.816	20.199	2.270	1.15	1.15	0.00
9	Monocaprylin	15.872	7.116	-1.012	1.588	2.05	2.10	0.05
10	Monocaprin	15.974	8.056	1.565	2.191	2.70	2.74	0.04
11	Monolaurin	16.046	8.973	4.325	2.221	2.54	2.36	0.18
12	Monomyristin	16.099	9.885	7.143	2.241	1.89	1.98	0.09

Table 4. Observed, Calculated and Residual Inhibition for Unsaturated FA/MG with the Electrotopological Values

Entry	FA/MG (Unsaturated)	HS ^T (other)	S ^T (CH=)	S ^T (-OH)	S ^T (CH ₂)	S ^T (CH ₃)	pIC	calc	res
13	Palmitoleic	8.4723	4.6282	8.4814	13.6886	2.2423	1.82	1.82	0.000
14	Oleic	9.3347	4.6603	7.6192	16.5794	2.2588	2.15	2.15	0.000
15	Linoleic	10.411	9.0783	7.6137	13.2029	2.2347	2.46	2.46	0.000
16	Linolenic	11.8854	13.2326	7.61	10.0911	2.1479	2.44	2.44	0.000
17	Arachidonic	15.9581	17.4184	7.6042	10.5067	2.2314	2.80	2.80	0.000
18	Monoolein	14.1403	4.6453	21.6422	9.2777	2.2537	2.55	2.55	0.000
19	Monolinolein	14.6403	9.04927	21.6352	5.981	2.2296	3.15	3.15	0.000

Table 5. Summary of results in the Leave-One-Out method on virus Inhibition data of combined set of saturated and unsaturated FA/MG.

Quantity	Coefficients from the first model	Coefficients from LOO method
¹ X _v	+ 0.466 (± 0.074)	+ 0.466 (± 0.015)
S ^T (-OH)	- 0.076 (± 0.026)	- 0.077 (± 0.006)
S ^T (-CH ₂ -)	- 0.153 (± 0.032)	- 0.153 (± 0.006)
intercept	+ 1.03 (± 0.28)	+ 1.035 (± 0.069)
Average residual	0.246	0.312

Table 6. Summary of the Leave-One-Out method on virus Inhibition data set of saturated FA/MG.

Quantity	Coefficients from second model	Coefficients from LOO method
S ^T (O=)	+ 7.21 (± 2.21)	+ 7.51 (± 1.16)
HS ^T (other)	- 7.06 (± 2.21)	- 7.36 (± 1.15)
S ^T (-CH ₂ -)	+ 2.00 (± 0.63)	+ 2.05 (± 0.35)
S ^T (-CH ₃)	+ 2.29 (± 0.67)	+ 2.21 (± 0.48)
Intercept	- 64 (± 19)	-66.1 (± 9.6)
Average residual	0.159	0.480

Table 7. Summary of the Leave-One-Out method on virus inhibition data set of unsaturated FA/MG.

Quantity	Coefficients from the third model	Coefficients from LOO method
HS ^T (other)	- 0.4548 (± 0.0049)	- 0.4584 (± 0.0092)
S ^T (-CH=)	+ 0.4544 (± 0.0040)	+ 0.4570 (± 0.0071)
S ^T (-OH)	+ 0.3635 (± 0.0034)	+ 0.3651 (± 0.0060)
S ^T (-CH ₂ -)	+ 0.3412 (± 0.0040)	+ 0.3425 (± 0.0113)
S ^T (-CH ₃)	+ 2.039 (± 0.084)	+ 2.27 (± 0.81)
Intercept	- 8.76 (± 0.17)	- 9.3 (± 1.7)
Average Res	0.0013	0.1032

Randomization. The independent variable, the pIC values (of the combined set), was randomized using the random function of EXCEL. The randomized data replaced the inhibition data in the data set. When the pIC values were randomly shuffled out ten times, the resulting r² values ranged from 0.013 to 0.40 with an average r² = 0.16 while the resulting F values ranged from 0.005 to 3.3 with an average F = 1.2

DISCUSSION

QSAR by E-State Modeling. Used in many QSAR studies is the E-State model whose approach is topological in structure representation [Hall et al., 2007]. This approach is very appropriate in obtaining useful information about structural features that influence property such as a particular biological activity. Models are expressed in E-State indices in both atom-type and atom-level forms. For structure representation to develop information, hydride groups are considered as atoms for simplicity. To compute for the E-State index of an atom in a molecule the equation below is used. S_i is known

$$S_i = I_i + \Delta I_i$$

as the E-State index for an atom or a group of atoms (hydride groups), I_j is the atom's intrinsic state, and ΔI_j is known as the perturbation term. The atom intrinsic state I is expressed as a function of the valence of the atom or group of atoms, over the number of nearest neighboring atoms or groups. Where $N = n^{\text{th}}$ row in the periodic table, $\delta^v =$ valence of atom i and $\delta =$ number of neighboring atoms.

$$I = [(2/N)^2 \delta^v + 1] / \delta \quad (4)$$

When the atom involved is an element in the second row of the periodic table, $N = 2$, giving,

$$I = (\delta^v + 1) / \delta \quad (5)$$

while,

$$\Delta I_i = \sum_{j=1}^n (I_j - I_i) / r_{ij}^2 \quad (6)$$

where r_{ij} = the number of atoms in the shortest path between atoms i and j , and I_j = intrinsic state of atom j , and ΔI_i = the perturbation term. In this way the atom's E-State index encodes the electronic state and topological structure from all other atoms within the structure. The nearer the neighboring atom to atom i , the stronger its influence on its E-State value, the farther it is by a path of several bonds the smaller its influence. This influence diminishes as the square of the number of atoms in the path.

Similar expression is used for the hydrogen E-State [Molecular Descriptor's Guide, 2008] as in,

$$HS_i = KHE_i + [KHE_i - KHE(H_i)] + \sum_{j \neq i} \frac{KHE_j - KHE(H_i)}{(d_{gj} + 1)^2} \quad (7)$$

where

$$KHE_i = \frac{\delta_i^v + \delta_i}{N_i^2} \quad (8)$$

is KHE = Kier-Hall electronegativity.

When the E-State indices of the same atom or group type are combined in a common skeletal core, they can be used in equation to serve as the E-State descriptors or variables in the multiple regression. The atom- type E-State descriptor then

$$S^T(X) = \sum_i S_i(X) \quad (9)$$

is the sum of the individual atom-level E-State values for a particular atom-type. This descriptor combines three important aspects of structural information, 1) electron accessibility at the atom, 2) presence/absence of the atom and 3) count of the atom. The hydrogen atom-type E-state

descriptor expresses very similar information as the atom-type E-State descriptor except that accessibility refers to proton accessibility. Similar expression is used for the hydrogen atom type E-State descriptor:

$$HS^T(X) = \sum_i HSi(X) \quad (10)$$

In this viral inhibition study, with nineteen straight chain FA and MG, the atom-level E-State and hydrogen E-state indices were used to develop QSAR models along with atom-type descriptors in addition to molecular connectivity χ indices.

Interpretation of the models. *Combined saturated and unsaturated FA/MG model.* The three-variable model eqn. 1 adequately represent the pIC data, based on direct statistics as well as by the validation method. Each of the variables is a descriptor of an aspect of the molecular structure and will be discussed for the three models to indicate the specific structure information encoded.

The ${}^1\chi^v$ Descriptor. The variable that makes the greatest contribution to calculated pIC is ${}^1\chi^v$, a representation of molecular variation [Molecular Descriptor's Guide, 2008]. The descriptor ${}^1\chi^v$ is the first order molecular connectivity valence chi index. The ${}^1\chi^v$ descriptor (Eqn. 1) contributes 58% on average to the calculated pIC value. The significance of this descriptor is further indicated by the fact that largest pIC values also correspond to largest ${}^1\chi^v$ values and the smallest pIC values correspond to smallest ${}^1\chi^v$ values. This structural descriptor provides molecular architecture information. The ${}^1\chi^v$ descriptor decreases with increase skeletal branching but increases with the number of skeletal atoms. Because of its positive coefficient, less branching leads to increase calculated pIC values.

The $S^T(-CH_2-)$ Descriptor. The second variable in the model is the atom type E-state descriptor for the skeletal carbon atoms of the methylene. The $S^T(-CH_2-)$, the second significant descriptor encodes the electron accessibility for the carbon in the methylene group. The $S^T(-CH_2-)$ descriptor contributes 24% on average to the calculated pIC values. Because of the negative coefficient on $S^T(-CH_2-)$ descriptor (Eqn. 1), larger values are related to decrease in pIC values and smaller values are related to increase in pIC values.

In the case of monoglycerides where the $-CH_2-$ groups are directly attached to electronegative atoms like oxygen of the hydroxyl groups, the $S^T(-CH_2-)$ assume large negative values and thus decrease the total $S^T(-CH_2-)$ values gained by the attached fatty acid, thereby increasing pIC. And, because larger values of $S^T(-CH_2-)$ descriptor mean increase also in size of the molecule, meaning also increase in the value of the $S^T({}^1\chi^v)$ descriptor, the overall effect is increase in the pIC value due to the larger positive coefficient of the $S^T({}^1\chi^v)$ descriptor. The value of the $S^T(-CH_2-)$ descriptor also decreases as the number of double bonds increases, which contribute to the increase of pIC values.

The $S^T(-OH)$ Descriptor. The third variable in the model is the atom type E-state descriptor $S^T(-OH)$ (Eqn. 1). It encodes the electron accessibility for oxygen atoms in each molecule. The $S^T(-OH)$ descriptor contributes 19% on average to the calculated pIC values. The value of the $S^T(-OH)$ descriptor can be divided into two sources, the $-OH$ of the FA and the two $-OH$ groups from the glyceride side of MG. Because of the negative coefficient on the $S^T(-OH)$ descriptor, a larger value of this descriptor relates to a diminished value of the pIC.

However, in MG the addition of the glyceride group increases also the size of the molecule thereby contributing to the value of ${}^1\chi^v$, which has a larger positive coefficient combined with a decrease in $S^T(-CH_2-)$ value, which has a negative coefficient, both of these effects are reflected as an overall increase in the value of the pIC.

The saturated FA/MG model. The $S^T(O=)$ Descriptor. The first variable of the equation, the $S^T(O=)$, an atom-type E-State index, encodes the electron accessibility for the carbonyl oxygen in the

molecule. Larger values of the positive coefficient are related to larger pIC values in the molecule. The $S^T(O=)$ descriptor is a significant variable in the model, contributing to 59% on the average to the pIC, making it the most significant contributor to pIC (Eqn. 2). Because of the positive coefficient on $S^T(O=)$ larger values of the positive coefficient are related to larger pIC values.

Increasing length of the alkyl chain also increases slightly the $S^T(O=)$ value. Moreover, the substitution of a glyceride group to the molecule provides an additional ester oxygen that renders its π electrons accessible to the carbonyl oxygen thereby increasing largely the positive value of $S^T(O=)$ descriptor and thus the pIC value.

The $HS^T(\text{other})$ Descriptor. The $HS^T(\text{other})$ descriptor is the sum of the hydrogen-atom level E-State indices for all non-polar hydrogen atoms in the molecule (from the hydrogens of the methyl to the methylene hydrogens of the fatty alkyl chain and to those in the glyceride group). This descriptor encodes the structure attributes of hydride groups that are non-polar. Such a descriptor represents parts of the structures that may participate in hydrophobic-like interaction. $HS^T(\text{other})$ is the second most significant variable in the model (Eqn. 2), contributing to 27% on the average to the calculated pIC. Because of the negative coefficient on $HS^T(\text{other})$, smaller values are related to larger pIC values. The values of the $HS^T(\text{other})$ descriptor increase with increasing size of the molecule and with the addition of the glyceride group. These increases diminish the pIC values.

The $S^T(-CH_2-)$ Descriptor. $S^T(-CH_2-)$ descriptor is the atom-type E-State index for the methylene carbon in the molecule. It encodes the electron accessibility for the methylene carbon atoms in the molecule. The $S^T(-CH_2-)$ descriptor contributes 10.8% on average to the saturated FA/MG model (Eqn. 2). Because of the positive coefficient on $S^T(-CH_2-)$, larger values are related to larger pIC values. Increasing size of the fatty acid increases the $S^T(-CH_2-)$ value but decreases with addition of the monoglyceride group affecting the pIC values as discussed above.

The $S^T(-CH_3)$ Descriptor. The $S^T(-CH_3)$ descriptor is the fourth variable in the model and is the atom-type E-State descriptor for the carbon atom of the methyl group. $S^T(-CH_3)$ encodes the electron accessibility for carbon atoms in each molecule. The $S^T(-CH_3)$ descriptor contributes 12.4% on average to the calculated pIC values. Because of the positive coefficient on the $S^T(-CH_3)$, larger values are related to larger pIC values.

The unsaturated FA/MG model. The $S^T(-CH_3)$ Descriptor. The $S^T(-CH_3)$ descriptor is the atom-type E-State descriptor for the carbon atom of the methyl group in the molecule. $S^T(-CH_3)$ encodes the electron accessibility for carbon atom of the methyl group in each molecule. The $S^T(-CH_3)$ descriptor is a significant variable in the model since it contributes 21% on average to the calculated pIC values (Eqn. 3). Because of the positive coefficient on the $S^T(-CH_3)$, larger values are related to larger pIC values.

The $HS^T(\text{other})$ Descriptor. The $HS^T(\text{other})$ of the unsaturated FA and MG is the sum of the hydrogens described above as in the saturated model except the addition of the vinyl hydrogens from the additional double bond(s). This descriptor encodes the structure attributes of hydride groups that are non-polar. It contributes 25% on the average to the calculated pIC values (Eqn. 3). It is the most significant variable in the model equation. Because of the negative coefficient on $HS^T(\text{other})$, smaller values are related to larger pIC values.

The descriptor $HS^T(\text{other})$ can be partitioned among three atom-type hydrogen E-State indices: $HS^T(\text{other}) = HS^T(C_{\text{sats}}) + HS^T(C_{\text{satu}}) + HS^T(\text{vinyl})$. These descriptors represent hydrogen atoms in three different environments; $HS^T(C_{\text{sats}})$, hydrogens on saturated carbon atoms bonded to saturated carbon atoms; $HS^T(C_{\text{satu}})$, hydrogens on saturated carbon atoms bonded to unsaturated carbon atoms; $HS^T(\text{vinyl})$, hydrogens on vinylic carbon atoms. When each of these three variables replaces $HS^T(\text{other})$ in the model the correlation, such as, if $HS^T(\text{other})$ is replaced by $HS^T(C_{\text{sats}})$, $r^2 = 0.98$ and $F = 8.3$; if by $HS^T(C_{\text{satu}})$, $r^2 = 1.00$ and $F = 9500$; if by $HS^T(\text{vinyl})$, $r^2 = 0.87$ and $F = 1.8$.

The analysis indicates that $HS^T(C_{\text{sat}})$ of these three regions of carbon skeleton, dominates the $HS^T(\text{other})$ descriptor.

The $S^T(-CH=)$ Descriptor. The $S^T(-CH=)$ descriptor is the atom-type E-State index for the vinyl carbon in the molecule. It encodes the electron accessibility of the vinyl carbon in the molecule. The $S^T(-CH=)$ descriptor contributes 12.4% on average to the pIC values. Because of the positive coefficient on the $S^T(-CH=)$ descriptor, larger values are related to larger pIC values. Increasing number of double bonds, increases the value of the $S^T(-CH=)$ and therefore the pIC also.

The $S^T(-OH)$ descriptor. The $S^T(-OH)$ descriptor encodes the electron accessibility for oxygen atoms in each molecule. The $S^T(-OH)$ descriptor contributes 9.94% on average to the calculated pIC. The value of the $S^T(-OH)$ descriptor can be divided into two sources, the -OH of the FA and the two -OH groups in the glyceride of MG. Because of the positive coefficient on the $S^T(-OH)$ descriptor, larger values lead to larger pIC values.

The $S^T(-CH_2-)$ Descriptor. $S^T(-CH_2-)$ descriptor is the atom-type E-state index for the methylene carbon of the unsaturated molecule. It encodes the electron accessibility for the methylene carbon atom. The $S^T(-CH_2-)$ descriptor contributes 9.47% on average to the pIC of the unsaturated FA/MG model. Because of the positive coefficient on $S^T(-CH_2-)$, larger values are related to larger pIC values (Eqn. 3). These percent contribution values serve to guide the reader to the calculated significance of each descriptor. For each descriptor, the absolute value of each contribution was taken to compute the percentage contribution.

To summarize, the combined model equation 1 demonstrates that inhibition of lipid coated virus is increased for FA and MG with less skeletal branching, larger size, possession of a glyceride chain, and greater number of double bonds. This model allows the prediction of pIC values for organic compounds and blends the structure information of the two subsets, the saturated and unsaturated FA and MG. It can be noticed however, that some indicators counteract the property of other indicators and the positive or negative sign of the values are not in the direction of the expected pIC of the molecule but when the coefficients are applied the resulting calculation results to the nearest or correct value expected.

Validation Studies. The predictive quality of each of the equations is shown in Tables 5-7. The leave-one-out method applied, clearly seen from Table 5, shows that the coefficients are very stable and the predicted values are quite acceptable for the combined set study. The difference of the average residuals from LOO is essentially the same as in the full regression and is quite acceptable (0.25→0.31). This quality of the combined equation is not seen in the two remaining models where the coefficients are only moderately similar, although acceptable overall. While their correlation coefficients are relatively higher (Table 6-7) and impressive as in the unsaturated set, (Table 7 and Fig. 3), the differences of their residuals are larger specially with the saturated set (0.16→0.48) but not as much as in the unsaturated (0.00→0.10). These are due probably to the reduced size of the samples in each subset models ($n = 12$ for the saturated and $n = 7$ for the unsaturated). This leave-one-out method shows how the predictive quality of the regression equation will result in a satisfactory manner if only the number of samples is satisfied.

The independent variables, the pIC values of the combined set ($n = 19$), were randomized using the random generator of EXCEL. The randomized data replaced the inhibition data in the data set. When the pIC values were randomly scrambled out ten times, the largest r^2 and F values found are $r^2 = 0.40$ and $F = 3.3$ and the average $r^2 = 0.16$ and the average $F = 1.2$. When these data corresponding to a random statistics were compared to the model (Eqn. 1) $r^2 = 0.81$ and $F = 21$, these results clearly indicate that the model is significantly different from a random model, giving credence for the model based on the topological variables used in Eqn. 1. The same conclusion can also be derived from the randomization of the two subsets.

CONCLUSIONS

For the viral inhibition data set, excellent QSAR models are developed with two E-State structure descriptors and one molecular connectivity chi valence index for the combined set. To improve the correlation coefficient two subsets were introduced, the subset saturated FA and MG, where the QSAR is developed with one hydrogen E-State structure descriptor and three atom type E-State descriptors, and the subset with unsaturated FA and MG, developed with one hydrogen E-State descriptor and four atom type E-State descriptors. Furthermore, cross validation of the QSAR model signifies that the model may be used to make useful prediction for compounds not in the original data set. The structure information encoded in the descriptor is described, indicating significant and specific structure information that may be useful to aid compound design. This encoded structure information may be interpreted directly in terms of structural feature that can be helpful to synthetic chemists; skeletal branching and general molecular size $S^T(1X^v)$, presence and polar nature of oxygen atoms, $S^T(-OH)$, presence of non-polar hydrogen atoms, $HS^T(\text{other})$ and electron accessibility on oxygen atom in the carbonyl group $S^T(O=)$. These qualitative statements can be converted into numerical values of pIC by the QSAR model equations.

These findings on important functional groups that the antiviral (at least against lipid-coated virus) should possess for activity, recalls to mind recent developments on the use of antiviral essential oil components against the lipid-coated SARS-CoV-2 virus [Asif et al., 2020; Yadalam et al., 2021; Patne et al., 2020]. These essential oils contain monoterpenes [Denison et al., 1995; Astani et al., 2010] that are not far in structure make-up as the FA and MG and can be one of the targets for further study after this paper. Our model equations can be applied to recognize potential antiviral compounds from essential oil components or any other structurally related compounds. This does not prevent us in determining the antiviral activities of several fatty acid derivatives such as the fatty alcohols, fatty acid esters, fatty alcohol esters, and amides.

Our use of the E-State method is very appropriate in this kind of problem because in virus inhibition, as in toxicity or physico-chemical property studies, the dominant force is the non-covalent interaction. Electron accessibility attributes are very significant in this kind of interactions where both the electronic and steric forces are well encoded. In addition, the E-state method is based on the topological representation of molecular structure. E-State method combines valence state electronegativity with skeletal characterization of the molecule. Non-covalent interaction meanwhile involves intermolecular electron accessibility, which is the essence of molecular connectivity principle. Besides, topological method avoids the problems arising from assumption of a partitioning mechanism as well as the problems associated with finding a physico-chemical system to use as the basis for estimating, for example the log P values for each molecule [Rose and Hall, 2003].

ACKNOWLEDGEMENT

I would like to acknowledge the help of Dr. Junie Billones of the College of Math and Sciences, UP Manila, for providing the necessary journals needed in a particular aspect of problem met when writing this paper. Thanks are also due to Dr. Eloise Prieto-Valdez of the Molecular Biology and Biotechnology, College of Science, University of the Philippines, Diliman, Quezon City, for a critical review of the original problem before the paper was conceptualized.

REFERENCES

- Abou-Shaaban RR, Al-Khamees HA, Abou-Auda HS, Simonelli AP. Atom level electrotopological state indexes in QSAR: Designing and testing antithyroid agents. *Pharm Res.* 1996 Jan; 13:129-136. <https://doi.org/10.1023/A:1016049921842>
- Asif M, Saleem M, Saadullah M, Yaseen HS, Al Zarzour R. COVID-19 and therapy with essential oils having antiviral anti-inflammatory, and immunomodulatory properties. *Inflammo-pharmacol.* 2020 Aug; 28(5):1153-1161. <https://doi.org/10.1007/s10787-020-00744-0>
- Astani A, Reichling J, Schnitzler P. Comparative study on the antiviral activity of selected monoterpenes derived from essential oils. *Phytother Res.* 2010 Aug; 24(5): 673-679. <https://doi.org/10.1002/ptr.2955>
- Carrouel F, Gonçalves LS, Conte MP, Campus G, Fisher J, Fraticelli L, Gadea-Deschamps E, Ottolenghi L, Bourgeois D. Antiviral activity of reagents in mouth rinses against SAR-CoV-2. *J Dent Res.* 2021 Feb; 100(2):124-132. <https://doi.org/10.1177/0022034520967933>
- Contrera JF, MacLaughlin P, Hall LH, Kier LB. QSAR modeling of carcinogenic risk using discriminant analysis and topological molecular descriptors. *Curr Drug Disc Technol.* 2005; 2(2):55-67. <https://doi.org/10.2174/1570163054064684>
- Denison DK, Meredith GM, Shillitoe EJ, Caffesse RG. The oral antibacterial spectrum of Listerine antiseptic. *Oral Surg. Oral med. Oral pathol. Oral Radiol. Endod.* 1995 April; 79(4):442-448. [https://doi.org/10.1016/S1079-2104\(05\)80124-6](https://doi.org/10.1016/S1079-2104(05)80124-6)
- Hall LH, Mohny B, Kier LB. The electrotopological state: structure information at the atomic level for molecular graphs. *J Chem Inf Comput Sci.* 1991 Feb; 31:76-82. <https://doi.org/10.1021/ci00001a012>
- Hall LH, Story CT. Boiling point and critical temperature of heterogeneous data set: QSAR with atom type electrotopological state using neural networks. *J Chem Inf Comput Sci.* 1996 Sept; 36:1004-1014. <https://doi.org/10.1021/ci960375x>
- Hall LH, Story CT. Boiling point of a set of alkanes, alcohols and chloroalkanes: QSAR with atom-type electrotopological state indices using artificial neural networks. *SAR-QSAR Environ Res.* 1997; 6: 139 -161.
- Hall LH, Vaughn TA. QSAR of phenol toxicity using electrotopological kappa shape indices. *Med. Chem. Res.* 1997; 7:407-416.
- Hall LH, Kier LB, Hall LM. Topological quantitative structure-activity relationship applications: structure information representation in drug discovery. In: *Comprehensive Medical Chemistry II.* 2007; 4: p 537-574. Elsevier Ltd. <https://doi.org/10.1016/B0-08-045044-X/00265-0>
- Hilmarsson H, Kristmundsdottir T, Gunnarson E, Thormar H. Intranasal delivery of formulations containing virucidal lipid for treatment of respiratory syncytial virus (RSV) infection in rats. *J Drug Deliv Sci Technol.* 2013 Dec; 23(05):455-457. [https://doi.org/10.1016/S1773-2247\(13\)50065-7](https://doi.org/10.1016/S1773-2247(13)50065-7)
- Huuskonen J. Estimation of aqueous solubility for a diverse set of organic compounds based on molecular topology. *J Chem Inf Comput Sci.* 2001; 40:773-777. <https://doi.org/10.1021/ci9901338>

Maw HH, Hall LH. E-State modeling of dopamine transporter binding. Validation of the model for a small data set. *J Chem Inf Comput Sci.* 2000 Aug; 40:1270–1275. <https://doi.org/10.1021/ci000023x>

Molecular descriptor's guide. Description of the molecular descriptors appearing in the toxicity estimation software tool. Version 1.0.2. US Environmental Protection Agency. 2008.

Meiller TF, Silva A., Ferreira SM, Jabra-Rizk MA, Kelly JI, DePaola LG. Efficacy of Listerine antiseptic in reducing viral contamination of saliva. *J Clin Periodontol.* 2005 April; 32(4):341-346. <https://doi.org/10.1111/j.1600-051X.2005.00673.x>

O'Donnell VB, Thomas DW, Stanton R, Maillard J-Y, Murphy RC, Jones SA, Humpreys I, Wakelam MJO, Fegan C, Wise MP, Bosch A, Sattar SA. Potential role of oral rinses targeting the viral lipid envelope in SARS-CoV-2 infection. *Function.* 2020 June; 1(1):2qaa002, <https://doi.org/10.1093/function/zqaa002>

Patne T, Mahore J, Tokmurke P. Inhalation of Essential oils could be adjuvant therapeutic strategy for Covid-19. *Int J Pharm Sci Res.* 2020 Sept; 11(9):4095-4103. [https://doi.org/10.13640/IJPSR.0975-8232.11\(9\).4095-03](https://doi.org/10.13640/IJPSR.0975-8232.11(9).4095-03)

Pfaender S, Heyden J, Friesland M, Ciesek S, Ejaz A, Steinmann Joerg, Steinmann Jechen, Malarski A, Stoiber H, Tsiavaliaris G, Bader W, Jahreis G, Pietschmann T, Steinmann E. Inactivation of hepatitis C virus infectivity by human breast milk. *J Infect Dis.* 2013 Dec; 208(12):1943–1952. <https://doi.org/10.1093/infdis/jit519>

Rose K, Hall LH. E-State modeling of fish toxicity independent of 3D structure information. SAR and QSAR in Environmental Research. 2003; 14(2): 113–129. <https://doi.org/10.1080/1062936031000073144>

Sands J, Auperin D, Snipes W. Extreme sensitivity of enveloped viruses, including *Herpes simplex*, to long chain unsaturated monoglycerides and alcohols. *Antimicrob Agents Chemother.* 1979 Jan; 15(1) 67-73. <https://doi.org/10.1128/aac.15.1.67>

Schroder U, Svenson S. Vaccine formulation comprising of monoglyceride and fatty acid as adjuvant through the intranasal route. European Patent Office EP1150713B1; World wide application: EP00903678A events. 1999.

Schroder U, Svenson S. Nasal and parenteral immunization with diphtheria toxoid using monoglyceride/fatty acid suspension as adjuvants. *Vaccine.* 1999; 17(15): 2096 -2103. [https://doi.org/10.1016/s0264-410x\(98\)00408-3](https://doi.org/10.1016/s0264-410x(98)00408-3)

Thormar H, Isaacs CE, Brown HR, Barshatzky MR, Pessolano T. Inactivation of enveloped viruses and killing of cells by fatty acids and monoglycerides. *Antimicrob Agents Chemother.* 1987 Jan; 31(1):27-31. <https://doi.org/10.1128/AAC.31.1.27>

Votano JR, Parham M, Hall LH, Kier LB. New predictors for several ADME/Tox properties: Aqueous solubility, human oral absorption, and Ames genotoxicity using topological descriptors. *J. Mol. Divers.* 2004 Dec; 8:379-391. <https://doi.org/10.1023/B:MODI.0000047512.82293.75>

Yadalam PK, Varatharajan K, Rajapandian K, Chopra P, Arumunaganainar D, Nagarathnam T, Sohn H, Madhavan T. Antiviral essential oil components against SARS-CoV-2 in pre-procedural mouth rinses for dental settings during COVID-19. A computational study. *Front Chem.* 2021 March; 9: 642026. <https://doi.org/10.3389/fchem.2021642026>