Chemistry Research Journal, 2021, 6(3):46-59

Available online www.chemrj.org

Research Article

ISSN: 2455-8990
CODEN(USA): CRJHA5

On the compatibility between formal QSAR results and docking results: the relationship between electronic structure and H5N1 (A/goose/Guangdong/SH7/2013) neuraminidase inhibition by some Tamiflu derivatives as an example

Juan S. Gómez-Jeria*, Sebastián R. Crisóstomo-Cáceres, Andrés Robles-Navarro

Quantum Pharmacology Unit, Department of Chemistry, Faculty of Sciences, University of Chile. Las Palmeras 3425, Santiago 7800003, Chile
* facien03@uchile.cl

Abstract A study of the relationships between electronic structure and the inhibition of theN1 neuraminidase of a H5N1 strain was carried out with the Density Functional Theory at the B3lYP/6-31G(d,p) level. A statistically significant equation was obtained allowing us to suggest some of the molecule-site interactions. Also some chemical modifications to the molecules to obtain more active compounds are suggested. The most important result of the QSAR study indicates the possible existence of a C-H ... X hydrogen bond involving a specific carbon atom. As we expect that both, QSAR and docking results show some similitudes we docked the molecules to the N1 neuraminidase expecting to detect the C-H ... X hydrogen bond. AutoDock Vina software was used. We selected the most active molecule of the set for the analysis of the requirement of compatibility.None of the twelve binding modes obtained for the molecule showeda C-H ... X hydrogen bond. A literary memory suggested that perhaps this bond was not intermolecular but intramolecular. A full geometry optimization of the molecular geometry of the most active molecule in the presence of different solvents (water, benzene, chloroform, cyclohexane, heptane and dichloromethane) with the integral equation formalism model and the SMD solvation model produced no structures with a C-H ... X hydrogen bond. As a last resort we used MarvinView software and two different force fields (Dreiding and MMFF94) to find the ten lowest energy conformers of the molecule. Interestingly, several conformers with an intramolecular C-H…O were found. We suggest that the reinforcement or the"freezing" of this intramolecular hydrogen bond can lead to new families of molecules that can be tested for N1 neuraminidase inhibition. This suggestion will allow a better appreciation of the predictive power of the KPG model. On the other hand, the results of docking with AutoDock Vina should be taken with extreme caution because it seems that this program needs to be refined to take into account other situations like the one found here.

Keywords Klopman-Peradejordi-Gómez method, QSAR, local atomic reactivity indices, H5N1 virus, Tamiflu, N1 neuraminidase, AutoDock Vina, C-H…X bond

Introduction

For more than 40 years this Unit has been dedicated to expand and testa model that was finally able of generating formal quantitative relationships (FQSAR) between electronic structure and biological activity. The model has its

roots in the works of Gilles Klopman and Federico Peradejordi [1-4], being known as the Klopman-Peradejordi-Gómez (KPG) model.More than sixty articles have shown its usefulness for various biological activities and processes (receptor binding affinity, inhibitory activities, antiviral activity, etc.) [5-10]. In its current stage, this model, based on atom-atom interactions, has associated twenty local reactivity indices to each atom. Due to this number, and despite being a formal model [11], statistical tools have had to be used to obtain structure-activity relationships.

Almost all, but not all of the community of researchers working on finding QSAR (quantitative structure-activity relationships) and/or FQSAR equations, have placed a strong emphasis on the (supposed) predictive capacity of their resulting equations. Usually, and this fact is shown in several hundred or thousand papers, these equations are employed to predict the activity of one or more molecules without any mention of the shortcomings or restrictions the equations may have. During the year 2002 Stephen Johnson published a Letter entitled "The Trouble with QSAR (or How I Learned to Stop Worrying and Embrace Fallacy)" [12], stating that "a general feeling of disillusionment with QSAR has settled across the modeling community in recent years". The situation has not improved much today so it is necessary to provide more elements to understand what is happening. This is especially important in a world that is increasingly globalizing. In the fields of research in quantum chemistry and structure-activity relationships, the well-known threatening phrase "publish or perish" is contributing greatly to the fact that the literature begins to saturate with articles that have almost zero scientific contribution. This plague of papers has two main types of way of working. The first is: "*Take a molecule, mention some interesting properties of it (real or hypothetical) calculate everything you can (spectra, electrostatic potential, molecular orbitals and a long etc.) with some quantum chemistry software and publish*". The other is "*take some molecules and dock them to no matter what, but the pictures have to come out pretty*". There are hybrids of those two cases. What about the experimental biological activity? They just don't exist in those articles.

Now let's move on to the issue of the constraints of the QSAR equations. Let us consider a set $Q=\{q_1,q_2.....q_g\}$ comprising g variables that were employed to obtain an equation relating structure with activity. The final equation contains z variables with $z \subset Q$ and z<g. Each variable belonging to Q has a definite range of numerical values. Also the set of biological activities used to generate the QSAR equation has its own range of numerical values. These equations have some formal restrictions. The first limitation of a QSAR equation is that it cannot predict a biological activity value lying outside of the range of numerical values of the biological activities used to generate the equation. *Therefore it is strongly suggested to employ the widest possible range of biological activity values to generate a QSAR equation*. The second limitation is that the numerical value of each of the z variables of the molecule whose activity we want to predict must be within the range of the numerical values of the z variables that were originally used to generate the equation. The third limitation of a QSAR equation is that, in the molecule whose activity we want to predict, the numerical values of each of the variables of the molecule that are not in the equation but are in the set Q must be within the corresponding range of the original numerical values. And all these restrictions apply to every QSAR equation, regardless of the methodology used to generate it.

The opinion of one of us (J.S. G-J.) is that the equations allowing only general statements such as '*this part of the molecule must behydrophobic*' or '*this molecule must be more polar*' work precisely because of their generality-vagueness. In the case of QSAR equations produced with more exact methods, such as the one used here, if someone makes a correct activity prediction in a certain molecule without having checked that all the aforementioned restrictions are fulfilled, it is perfectly possible to speak of a happy coincidence. Without a doubt, predicting a biological activity that is restricted to a certain range already known in advance is a contribution to general knowledge, but something more important is the eventual ability of the equation to provide experimentalists with enough information for them to decide what modifications to make in the molecules in order to obtain molecules with activities well outside the original range of values. This is the exact goal of the QSAR equations obtained within the KPG method: to help to find one or more atomic sites that can be substituted to modulate activity.

A second aspect that deserves to be well analyzed is the relationship between the QSAR equations and the results of docking studies. In the ideal case where the two methods give perfect results, they should be completely compatible. In real work, partial compatibility can be accepted, but not some strong incompatibility of results.

In this article we present the results of a study of the relationship between electronic structure and N1(H5N1) neuraminidase inhibition by some oseltamivir derivatives. The selected experimental data was the activity against the, A/goose/Guangdong/SH7/2013 strain, because of the necessity of accumulating more information helping us to combat this disease [13-15]. The H5N1 strain of avian influenza virus was first detected in humans in Hong Kong in 1997. Subsequently to the re-emergence of human infections in Vietnam and China in 2003, the viruses have become rooted in poultry and continue evolving in some places of Asia, Africa and the Middle East. In early 2006, 184 people were confirmed with H5N1-derived infections. In February 2020 the running total of infected humans since year 2003 was 861. In our Unit we have carried out four FQSAR studies involving influenza viruses [16-19].

Also the molecules were docked to the active site of N1 neuraminidase to see if there was compatibility between the results of both, FQSAR and docking studies. This led to a really surprising result, with the possibility of generating one or more families of active molecules.

Model

During the last years, a formal linear relationship was developed to express the affinity constant in terms of several local atomic reactivity indices derived within the Hartree-Fock-Roothaan (HFR) framework [20-22]. Later this relationship was extended, with some constraints, to any kind of *in vivo* or *in vitro* biological activity (BA). The equation is the following:

$$\log(BA)=a+bM+c\log\left[\sigma/(ABC)^{1/2}\right]+\sum_{j=1}^{v}\left[e_jQ_j+f_jS_j^E+s_jS_j^N\right]+$$
$$+\sum_{j=1}^{v}\sum_{m=HOMO-2}^{HOMO}\left[h_j(m)F_j(m)^*+x_j(m)S_j^E(m)^*\right]+\sum_{j=1}^{v}\sum_{m'=LUMO}^{LUMO+2}\left[r_j(m')F_j(m')^*+t_j(m')S_j^N(m')^*\right]+$$
$$+\sum_{j=1}^{v}\left[g_j\mu_j^*+k_j\eta_j^*+o_j\omega_j^*+z_j\varsigma_j+w_jQ_j^{max*}\right] \quad (1)$$

where M is the drug's mass, σ its symmetry number and ABC the product of the drug's moments of inertia about the three principal axes of rotation. Q_j is the net charge of atom j, S_j^E and S_j^N are, respectively, the total atomic electrophilic and nucleophilic superdelocalizabilities of Fukui *et al.*, $F_{i,m}$ ($F_{i,m'}$) is the Fukui index of the occupied (empty) local MO m (m') located on atom i. $S_i^E(m)$ is the atomic electrophilic superdelocalizability of local MO m on atom i, etc. The total atomic electrophilic superdelocalizability of atom i corresponds to the sum over occupied MOs of all $S_i^E(m)$'s and the total atomic nucleophilic superdelocalizability of atom i corresponds to the sum over empty MOs of all $S_i^N(m')$'s.

The asterisks * refer to the so called *local* molecular orbitals. For atom p, the set of its local molecular orbitals corresponds to those MOs having a determinate electron population localized on p. In our work we considered only the three highest occupied and the three lowest empty local MOs. The influence of molecular orbitals that are not the frontier ones is well known.

The last bracket of the right side of Eq. 1 contains five new local atomic reactivity indices obtained by one of the authors (J.S.G-J). μ_j^*, η_j^*, ω_j^*, ς_j^* and Q_j^{max*} correspond, respectively, to the local atomic electronic chemical potential, the local atomic hardness, the local atomic electrophilicity the local atomic softness and the maximal amount of charge atom j may receive [23]. Note that these indices do not correspond at all to the ones derived within Density Functional Theory (DFT) theory. On the contrary, the units of the first four indices are eV while the units of the DFT indices are eV·e. The physical meanings of all the reactivity indices have been detailed before and we shall discus here only the ones appearing in the results [24-26].

The summation on v is over a set of atoms called the common skeleton (CS). This structure is defined as a set of atoms, common to all molecules analyzed, that accounts for approximately all the biological activity. In this model, it is considered that the action of the substituents resides in modifying the electronic structure of the CS and influencing the right alignment of the drug throughout the orientational parameters. It is important to stress that the hypothesis of the common skeleton implies that necessarily this structure must be oriented in the same way in all molecules.

If we consider only the atomic indices of Eq. 1 we can see that each atom of the common skeleton has 20 indices associated to it. The application of this technique, called the Klopman-Peradejordi-Gómez (KPG) method, produced excellent results for a variety of molecules and biological activities [5, 27-30].

Selection of molecules and biological activities

The selected molecules are a group of 7-amino-2-aryl/hetero-aryl-5-oxo-5,8-dihydro[1,2,4]triazolo[1,5-a] pyridine-6-carbonitriles selected from a recent study [31]. Their general formula and biological activity are displayed, respectively, in Fig. 1 and Table 1. The biological activity studied is the *in vitro* inhibitory activity of influenza neuraminidase N1(H5N1).

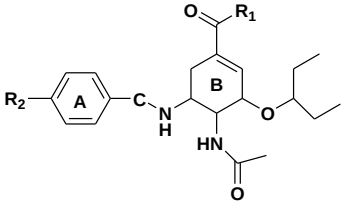

Figure 1: General formula of oseltamivir derivatives

Table 1: Oseltamivir derivatives and their H5N1 neuraminidase inhibition

Mol	Mol.	R_1	R_2	log(IC_{50})
6b	1	$NHCH_2pC_6H_5F$	1-pyrrolidinyl	1.30
6c	2	$NHCH_2pC_6H_5F$	SC_6H_5	0.92
6d	3	$NHCH_2pC_6H_5F$	C_6H_5	1.35
6e	4	$NHCH_2pC_6H_5F$	C_6H_5SCH3	1.54
10b	5	NHC(*i*Pr)C(O)OH	1-pyrrolidinyl	1.75
10c	6	NHC(*i*Pr)C(O)OH	SC_6H_5	0.97
10d	7	NHC(*i*Pr)C(O)OH	C_6H_5	1.85
10e	8	NHC(*i*Pr)C(O)OH	C_6H_5SCH3	-0.68
16b	9	$NHCH_2pC_6H_5Cl$	1-pyrrolidinyl	1.50
16e	10	$NHCH_2pC_6H_5Cl$	C_6H_5SCH3	1.44
18a	11	$NHCH_2pC_6H_5$ COOH	N(CH2CH3)2	1.61
18b	12	$NHCH_2pC_6H_5$ COOH	1-pyrrolidinyl	1.19
18c	13	$NHCH_2pC_6H_5$ COOH	SC_6H_5	0.49
18e	14	$NHCH_2pC_6H_5$ COOH	$C_6H_5SCH_3$	0.92

The common skeleton for oseltamivir derivatives is shown in Fig. 2.

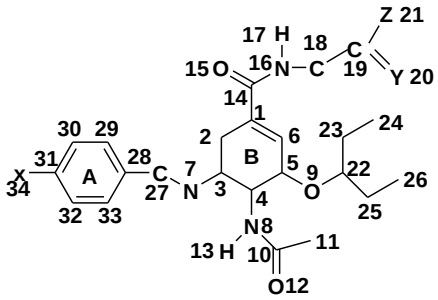

Figure 2: The common skeleton for oseltamivir derivatives of Table 1

Theoretical approach

A preliminary geometry optimization was performed with Molecular Mechanics with the OPLS (optimized potentials for liquid simulations) force field using the Hyperchem software [32]. This is the starting point for the nexttheoretical approach employed here. Subsequentgas phase full geometry optimization without symmetry constraints was carried out at the B3LYP/6-31G(d,p) level of theory withthe Gaussian16 software [33]. The same software was used for all electronic structure calculations.

The information to calculate numerical values for the local atomic reactivity indices was obtained from the Gaussian results file with the D-Cent-QSAR software [34]. All electron populations smaller than or equal to 0.01e were considered as being zero [35]. Negative electron populations coming from Mulliken Population Analysis were corrected. The resolution of the system of linear equations 1 is not possible because we have not enough molecules. Therefore, we made use of Linear Multiple Regression Analysis (LMRA) techniques to find the best solution. A matrix containing the dependent variable (the biological activity) and the local atomic reactivity indices of the atoms of the common skeleton as independent variables was built. The Statistica software was used for LMRA [36]. It is important to note that the LMRA resulting equation shows the relationship between the variation of the biological activity in terms of the variation of the numerical values of a set of local atomic reactivity indices.

Docking

We downloaded from the Protein Data Bank the N1 neuraminidase in complex with oseltamivir (PDB code: 2HU0) [37]. It contains two biological assemblies consisting of tetramers of the protein. The active site is located within the B chain of the tetramer, where the residues Glu119, Val149, Asp151 and Glu276 are located; therefore, for docking purposes the extra chains were eliminated. Using Discovery Studio Visualizer [38], hydrogen atoms were added to the receptor structure and the structure was prepared with AutodockTools 1.5.7 for docking with Autodock Vina [39]. For the docking procedure we employed the flexible residues approach. In this case the list of flexible residues is: Arg118, Glu119, Asp151, Arg152, Arg156, His274, Glu276, Glu277, Arg292, Try347, Arg371 and Tyr406. The grid box used for all dockings has 50 points on each axis with a spacing of 0.375 Å and is centered on the position (3.054, 19.311, 107.464) which includes the space in which the flexible residues are situated. The exhaustiveness of the calculation was set at 64. Twelve binding modes were obtained for each molecule.

Results

The best equation obtained was:

$$\log(IC_{50}) = 1.75 + 0.19\mathbf{S_{27}}^{\mathbf{N}}(\mathbf{LUMO+1})^{*} + 0.24\mathbf{S_{13}}^{\mathbf{N}}(\mathbf{LUMO})^{*} - 5.88\mathbf{S_{23}}^{\mathbf{N}}(\mathbf{LUMO})^{*} - \\ -0.003\mathbf{S_{19}}^{\mathbf{N}}(\mathbf{LUMO})^{*} \quad (2)$$

with n=14, R=0.98, R^2=0.97, adj-R^2=0.95, F(4,9)=67.36 (p<0.00001) and a standard deviation of 0.14. No outliers were detected and no residuals fall outside the ±2σ limits. Here,S_{27}^{N}(LUMO+1)* is the nucleophilic superdelocalizability of the second lowest empty MO localized on atom 27, S_{13}^{N}(LUMO)* is the nucleophilic superdelocalizability of the lowest empty MO localized on atom 13, S_{23}^{N}(LUMO)*)* is the nucleophilic superdelocalizability of the lowest empty MO localized on atom 23 and S_{19}^{N}(LUMO)* is the nucleophilic superdelocalizability of the lowest empty MO localized on atom 19.

Tables 2 and 3 show the beta coefficients, the results of the t-test for significance of coefficients and the matrix of squared correlation coefficients for the variables of Eq. 2.There are no significant internal correlations between independent variables (Table 3). Figure 3 displays the plot of observed *vs.* calculated log(IC_{50}) values.

Table 2: Beta coefficients and t-test for significance of coefficients in Eq. 2

	Beta	**t(9)**	**p-level**
S_{27}^{N}(LUMO+1)*	0.60	8.78	<0.00001
S_{13}^{N}(LUMO)*	0.20	3.04	<0.01
S_{23}^{N}(LUMO)*	-0.37	-5.81	<0.0003
S_{19}^{N}(LUMO)*	-0.30	-4.79	<0.001

Table 3: Matrix of squared correlation coefficients for the variables in Eq. 2

	S_{27}^{N}(LUMO+1)*	S_{13}^{N}(LUMO)*	S_{23}^{N}(LUMO)*
S_{27}^{N}(LUMO+1)*	1.00		
S_{13}^{N}(LUMO)*	0.17	1.00	
S_{23}^{N}(LUMO)*	0.08	0.06	1.00
S_{19}^{N}(LUMO)*	0.03	0.00	0.01

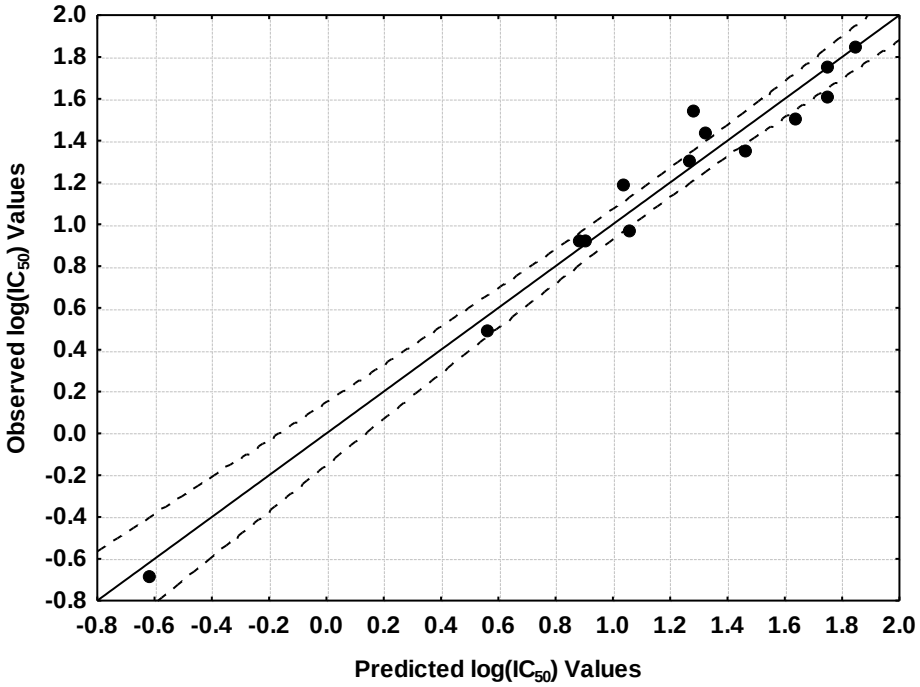

Figure 3: Plot of predicted vs. observed log(IC_{50}) values (Eq. 2). Dashed lines denote the 95% confidence interval.

The associated statistical parameters of Eq. 2 indicate that this equation is statistically significant and that the variation of the numerical values of a group of four local atomic reactivity indices of some atoms of the common skeleton explains about 95% of the variation of log(IC_{50}). Figure 3, spanning about 2.6 orders of magnitude, shows that there is a good correlation of observed *versus* calculated values.

It is important point to understand that when a local atomic reactivity index of an inner occupied MO or of a higher vacant MO appears in any equation, this means that the remaining of the upper occupied MOs (for example, if HOMO-2 appears, upper means HOMO-1 and HOMO) or the remaining of the empty MOs (for example, if LUMO+1 appears, lower means the LUMO) contribute to the interaction. Their absence in the equation only means that the variation of their numerical values does not account for the variation of the numerical value of the biological property.

Local Molecular Orbitals

Table 4 shows the local MO structure of all atoms appearing in Eq. 2 (see Fig. 2). Nomenclature: Molecule (HOMO) / (HOMO-2)* (HOMO-1)* (HOMO)* - (LUMO)* (LUMO+1)* (LUMO+2)*.

Table 4: Local Molecular Orbitals of atoms appearing in Eq. 2

Mol	Mol.	Atom 13	Atom 19	Atom 23	Atom 27
6b	1 (148)	110σ114σ116σ-155σ156σ157σ	143π145π146π-149π150π151π	136σ138σ139σ-164σ168σ170σ	137σ147σ148σ-152σ153σ162σ
6c	2 (157)	94σ119σ121σ-168σ169σ170σ	149π153π155π-160π163π166π	147σ151σ154σ-172σ177σ178σ	143σ144σ156σ-158σ167σ169σ
6d	3 (149)	113σ115σ126σ-	145π146π147π-	135σ136σ139σ-	138σ148σ149σ-

		158σ159σ160σ	151π152π154π	164σ167σ170σ	150σ157σ166σ
6e	4 (141)	84σ107σ120σ- 149σ150σ152σ	137π138π139π- 142π143π145π	129σ131σ132σ- 155σ158σ161σ	130σ140σ141σ- 144σ157σ159σ
10b	5 (147)	108σ111σ117σ- 153σ154σ155σ	131π137π143π- 149π153σ154π	141σ142σ144- 156σ163σ168σ	141σ146σ147σ- 150σ151σ156σ
10c	6 (156)	120σ121σ124σ- 165σ166σ168σ	151σ152σ153π- 157π161π165σ	148σ150σ154σ- 172σ173σ177σ	154σ155σ156σ- 158σ167σ174σ
10d	7 (148)	110σ113σ115σ- 156σ158σ159σ	133π137π142π- 152π156σ157σ	140σ141σ146σ- 160σ164σ167σ	145σ147σ148σ- 150σ151σ153σ
10e	8 (140)	105σ108σ109σ- 147σ148σ149σ	133σ135σ136π- 141π144π147π	132σ133σ138σ- 153σ154σ160σ	131σ139σ140σ- 142σ143σ152σ
16b	9 (152)	90σ113σ119σ- 160σ161σ162σ	147π149π150π- 153π154π155π	138σ140σ144σ- 168σ169σ173σ	141σ151σ152σ- 156σ157σ166σ
16e	10 (145)	110σ111σ114σ- 155σ156σ157σ	138π141π143π- 146π147π148π	137σ139σ142σ- 158σ162σ164σ	134σ144σ145σ- 149σ158σ161σ
18a	11 (156)	92σ116σ118σ- 165σ166σ167σ	151π152π154π- 157π158σ159π	150σ151σ154σ- 172σ177σ183σ	145σ155σ156σ- 160σ161σ166σ
18b	12 (155)	115σ116σ121σ- 162σ163σ164σ	150π151π152π- 156π158π162π	141σ142σ152σ- 173σ175σ178σ	145σ154σ155σ- 159σ160σ170σ
18c	13 (164)	127σ130σ144σ- 173σ177σ178σ	160π161σ162π- 165π170π173π	152σ154σ162σ- 182σ184σ185σ	149σ150σ163σ- 167σ174σ176σ
18e	14 (148)	113σ116σ125σ- 156σ157σ158σ	142π144π145π- 149π151π156π	131σ133σ135σ- 167σ169σ170σ	137σ147σ148σ- 152σ166σ168σ

Discussion

Table 2 shows that the importance of variables in Eq. 2 is
S_{27}^{N}(LUMO+1)*>>S_{23}^{N}(LUMO)*>S_{19}^{N}(LUMO)*>S_{13}^{N}(LUMO)*.

The variable-by-variable analysis of Eq. 2 shows that a high H5N1 neuraminidase inhibitory activity is associated with at least low numerical values for S_{27}^{N}(LUMO+1)* and S_{13}^{N}(LUMO)*, and with high numerical values for S_{23}^{N}(LUMO)* and S_{19}^{N}(LUMO)*.We need to comment about the appearance in the resulting equation of local atomic indices associated to MOs that are not the frontier ones. When this situation occurs and the equivalent LARI corresponding to the frontier MO is not present we have we have chosen to assign the same interpretation to both.

Atom 27 is a saturated carbon atom in the chain linking rings A and B (Fig. 2). All local MOs have a σ nature (Table 4). A high neuraminidase inhibitory activity is associated with low numerical values for $S_{27}^{N}(LUMO+1)_{27}$*. These values are obtained by lowering the electron population of this MO on atom 27 or by increasing the $(LUMO+1)_{27}$* energy. One of the ideal situations is when $(LUMO+1)_{27}$* is not more localized on atom 7 (i.e., $(LUMO+1)_{27}$* is replaced by a higher energy empty MO).The other is when $(LUMO+1)_{27}$* remains localized on atom C27 but its energy increased. On the other hand, Table 4 shows that $(HOMO)_{27}$* coincides with the molecule's HOMO or (HOMO-1). A direct interaction of C27 with another center is precluded because it forms part of a saturated CH_2 group. Therefore, the most probable interaction is anon-classical CH…X hydrogen bond (X= O, N) [40-43]. The experimentalist should explore the substitution of the hydrogen atoms attached to atom C27 by methyl or ethyl groups.

Atom 13 is a hydrogen atom bonded to a nitrogen atom (N8, Fig. 2). All local MOs have a σ nature (Table 4). $(HOMO)_{13}$* corresponds to an occupied molecular MO with an energy very far from the HOMO energy. $(LUMO)_{13}$* is relatively closer in energy to molecule's LUMO. These conditions suggest that the nitrogen N8 atom can share hydrogen atom H13 with another atom such as oxygen or nitrogen, forming a hydrogen bond.The fact that a high H5N1neuraminidase inhibitory activity berelated with low numerical values for S_{13}^{N}(LUMO)* suggests that if we modify the localization of the actual $(LUMO)_{13}$* in such a way to replace it by a an upper empty MO we are obtaining an atom that is a bad electron acceptor and a bad electron donor and susceptible to electrostatic

interactions. Given the special position of H13 inside the molecule, we may only suggest to substitute it with a fluorine atom and/or substitute C11 with an ethyl group.

Atom 23 is a saturated carbon atom in a side chain (Fig. 2). Table 4 shows that all MOs have a σ nature. $(LUMO)_{23}$* coincides with an empty molecular MO with an energy very far from the LUMO energy, i.e. this atom behaves as bad electron acceptor. The same fact holds for $(HOMO)_{27}$*, the only difference being that this local MO is closer to the molecular HOMO. Eq. 2 shows that a high H5N1 neuraminidase inhibitory activity is associated with high numerical values for S_{23}^{N}(LUMO)*. These values are obtained by lowering the MO energy, making atom C23 more prone to interact with electron-rich systems. The interactions can be of σ-π, alkyl and/or alkyl-π nature. A possible way to increase the sigma electron density on atom C23 (and atom C25) could be to replace the $CH(Et)_2$ group by a $CH(Me)_2$ group.

Atom 19 is a carbon atom that can be a fragment of a ring or of a carboxylate group (Fig. 2). Table 4 shows that all frontier local MOs of atom 19 have a π nature. Note that a high H5N1neuraminidase inhibitory activity is associated with high numerical values for S_{19}^{N}(LUMO)*. High numerical values for this reactivity index are obtained by increasing the localization of this MO on atom C19 (2.0 is the maximum value) and/or by lowering the $(LUMO)_{19}$* energy. This strongly suggests that atom C19 is interacting with an electron-rich center, alone or as part of a larger π system. This interaction could be π-π stacked, π-π T-shaped, π-amide stacked, π-σ or π-alkyl. We can improve the electron-accepting capacity of C19 by changing C18 by O18, trying C19-O21-F in the case of a carboxylate group or substituting the phenyl ring in the case of a phenyl ring (molecules 1-4 and 9-10 or Table 1). All the suggestions are displayed in Fig. 4.

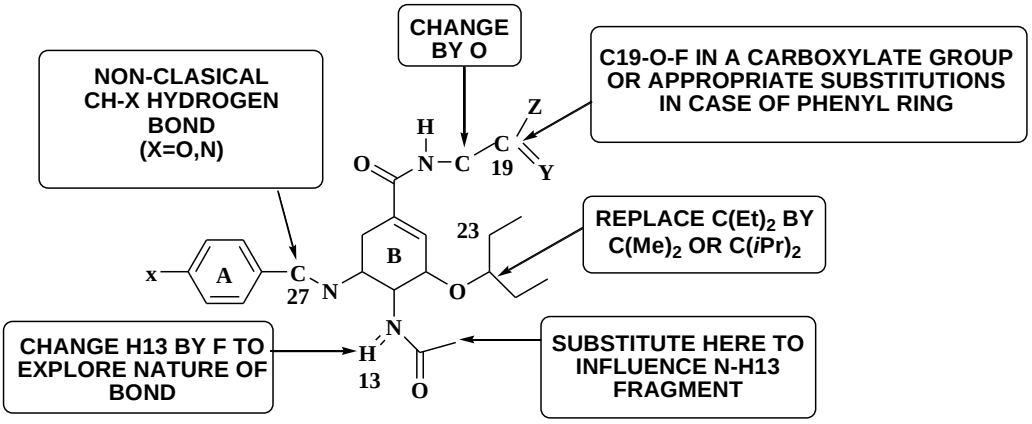

Figure 4: Some possible suggestions to improve biological activity

Figure 5 shows the possible interactions of atoms appearing in Eq. 2.

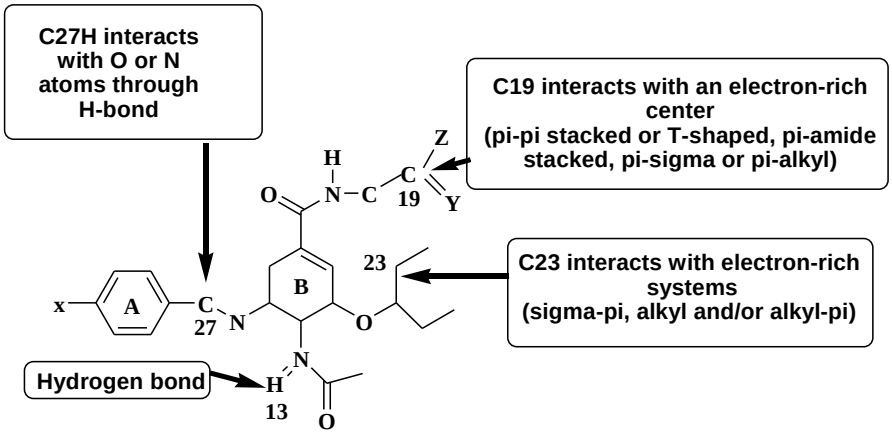

Figure 5: Possible intermolecular interactions of atoms appearing in Eq. 2

Eq. 2 contains only some local atomic reactivity indices involved in the inhibitory process. This equation was obtained by using a common skeleton for all them and it is statistically significant. This allows holding the hypothesis that this common skeleton has the same orientation in all molecules. But we cannot discard the possibility that one (or more) molecules acts in a different orientation than the others. In some previous FQSAR studies using the KPG method the multiple regression analysis has discarded sometimes one or two molecules. This fact opens the door for future research.

Now, what we expect from the Autodock Vina docking results is to find one or more interactions obtained from Eq. 2.The Webpage of Autodock Vina contains these two important phrases: *by design, the results should not have a statistical bias related to the conformation of the input structure*, and *Vina avoids imposing artificial restrictions, such as the number of atoms in the input, the number of torsions, the size of the search space, the exhaustiveness of the search, etc*.

For all that follows we selected molecule 8 because it is the most active one. Therefore we expect that the local atomic reactivity indices appear in one or more of the binding modes. Figure 6 shows the first binding mode.

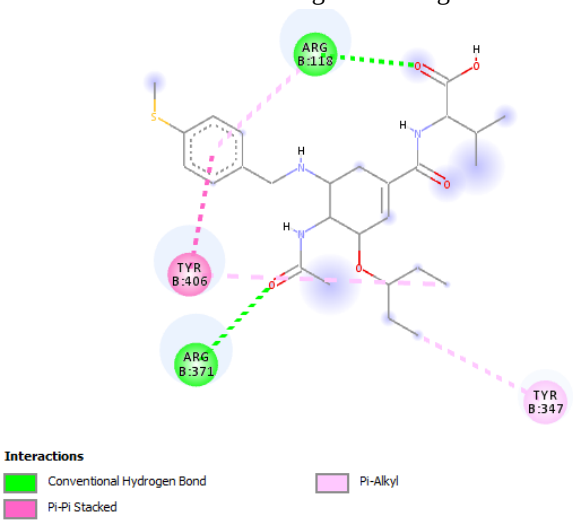

Figure 6: The first binding mode of molecule 8 (Autodock Vina result)

We can notice that the only interaction that could be tentatively assigned to Eq. 2 is the π-alkyl interaction between Tyr-347 and one methyl group of the O(CH)-Me_2 moiety. We need to stress that the other interactions could exist or not during the real drug-site interaction. Following Table 2, we should expect at least a binding mode involving atoms 27 and 23. If an interaction involving atom 19 also appears probably this mode is closer to reality than the remaining ones. Binding 2 mode does not show C-H hydrogen bonds.In the third binding mode (Figure 7) a C-H hydrogen bond is formed between a CHgroup from Arg371 and O12. We expect something similar but involving C27-H.

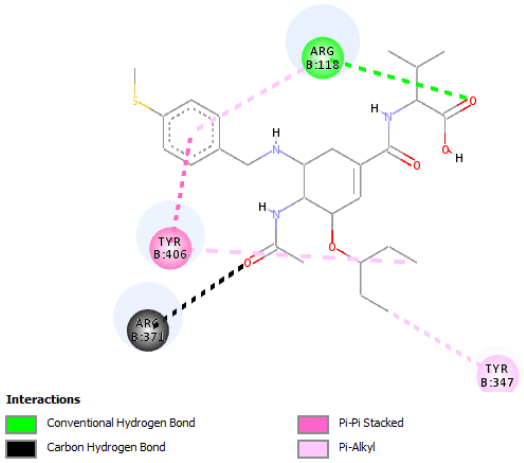

Figure 7: The third binding mode of molecule 8 (Autodock Vina result)

None of the twelve binding modes show an interaction involving C27-H and any other atom or group.

At this stage of the research, we can only tentatively conclude that it seems necessary to modify the Autodock Vina to account for the C23-H interaction. Thinking about this problem one of us (J.S. G.-J.) remembered a literary counsel given in *The Sign of the Four*: eliminate all other factors, and the one which remains must be the truth. Following this counsel we hypothesized that perhaps the C-H hydrogen bond was not intermolecular but intramolecular.

Therefore we began to study this hypothesis proceeding in the following way. The geometry of molecule 8 was fully optimized again at the B3LYP/6-31G(d,p)in the presence of a solvent and using two methods. In the first one we employed the reaction field calculation using the integral equation formalism model (IEFPCM, with water, benzene, chloroform, cyclohexane, and heptane as solvents). The second one was an IEFPCM calculation with radii and non-electrostatic terms for Truhlar and coworkers' SMD solvation model (SMD method hereafter with water, benzene, chloroform, cyclohexane, dichloromethane, and heptane as solvents). The final geometries were displayed analyzed with the Discovery Studio software. Also, an *in vacuo* calculation was also carried out. The result of the *in vacuo* geometry optimization is show in Fig. 8.

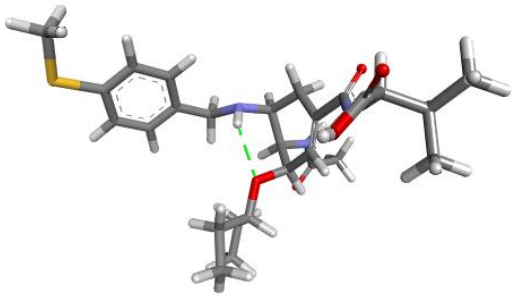

Figure 8: In vacuo full geometry optimization of molecule 8

We can observe that there is a conventional hydrogen bond (green line) between N7 and O9 (see Fig. 2). This hydrogen bond is also present in the IEFPCM results with all solvents. No other intramolecular bonds were found. In the case of SMD results, the same intramolecular bond was observed for benzene, chloroform, cyclohexane, dichloromethane and heptane. In the case of water result, no intramolecular bonds existed.

As the last possibility we hypothesized that there was the possibility that a conformer of molecule 8 could have an intramolecular C-H bond involving C27. Using MarvinView software [44] we carried out a search of the ten first conformers using the Dreiding and MMFF94 force fields (diversity limit of 0.5 kcal/mol). Surprisingly, we found some conformers with an intramolecular C27-H…O9 bond. For example, figure 9 shows the fourth conformer of the

list calculated with the Dreiding force field (the differences between these two isomers lie in the conformation of the rest of the molecule.).

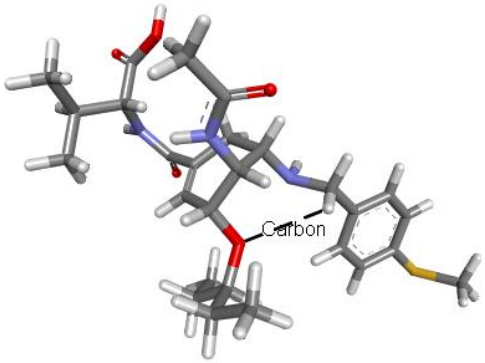

Figure 9: Fourth conformer obtained with the Dreiding force field showing C27-H…O9 intramolecular bond.

This is the only intramolecular bond in this conformer. The same bond is observed in the ninth conformer. The first three conformers have no intramolecular bonds.

The MMFF94 force field calculations show this same king of bond in the isomers number 6, 7, 8, 9 and 10. Figure 10 shows the C27-H…O9 hydrogen bond of the sixth conformer.

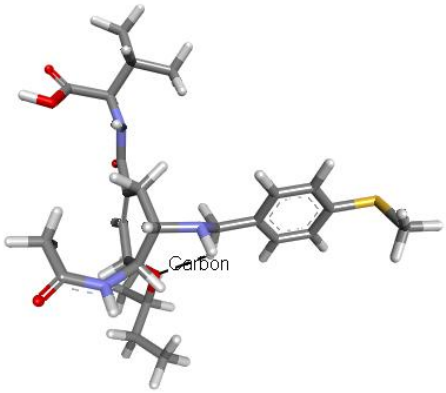

Figure 10: Fourth conformer obtained with the MMFF94 force field

This is the only intramolecular bond that this isomer has. In the case of the first five conformers, three of them have no intramolecular bonding and two have C-H hydrogen bonds between other atoms.

In light of these results, and considering that we have used AutoDock Vina in several previous studies, we can only conclude that this software is not correctly representing some situations. We have given preference to the results of the KPG method simply because its predictive capacity has been quantitatively proven in at least one case. A good rationale for preferring KPG results is that the method has been developed from physically rigorous equations. Another argument is that a more primitive version of the KPG model was able to predict the human hallucinogenic dose of (±)-2,5-dimethoxy-4-nitro-amphetamine (DON) [45, 46].

In any case, there is an experimental way to discern between the two models. The KPG results indicate, on the one hand, that there is a C27-H hydrogen bond and the results of docking and search for conformers, on the other hand, indicate that this C27-H hydrogen bond appears to be intramolecular. If both conclusions are correct, then it would be necessary to synthesize compounds in which the C27-H ... O9 bond is reinforced or modified. For example, one way to do this would be to replace C27-H ... O9 with a covalent structure of the type C27-X-N. If this suggestion is correct and leads to a whole family of new molecules therefore the KPG model performs well and it fulfills its objective of delivering elements that allow the synthesis of more active molecules.

References

[1]. Peradejordi, F.; Martin, A. N.; Cammarata, A. Quantum chemical approach to structure-activity relationships of tetracycline antibiotics. *Journal of Pharmaceutical Sciences* 1971, 60, 576-582.

[2]. Klopman, G. Chemical reactivity and the concept of charge- and frontier-controlled reactions. *Journal of the American Chemical Society* 1968, 90, 223-234.

[3]. Klopman, G.; Hudson, R. F. Polyelectronic perturbation treatment of chemical reactivity. *Theoretica chimica acta* 1967, 8, 165-174.

[4]. Hudson, R. F.; Klopman, G. A general perturbation treatment of chemical reactivity. *Tetrahedron Letters* 1967, 8, 1103-1108.

[5]. Kpotin, G. A.; Bédé, A. L.; Houngue-Kpota, A.; Anatovi, W.; Kuevi, U. A.; Atohoun, G. S.; Mensah, J.-B.; Gómez-Jeria, J. S.; Badawi, M. Relationship between electronic structures and antiplasmodial activities of xanthone derivatives: a 2D-QSAR approach. *Structural Chemistry* 2019, 30, 2301-2310.

[6]. Gómez-Jeria, J. S.; Soloaga Ardiles, C. E.; Kpotin, G. A. A DFT Analysis of the Relationships between Electronic Structure and Human κ, δ and μ Opioid Receptor Binding Affinity in a series of Diphenethylamines. *Chemistry Research Journal* 2020, 5, 32-46.

[7]. Gómez-Jeria, J. S.; Lagos-Arancibia, L. Quantum-chemical structure-affinity studies on kynurenic acid derivatives as Gly/NMDA receptor ligands. *International Journal of Quantum Chemistry* 1999, 71, 505-511.

[8]. Gómez-Jeria, J. S.; Ojeda-Vergara, M.; Donoso-Espinoza, C. Quantum-chemical Structure-Activity Relationships in carbamate insecticides. *Molecular Engineering* 1995, 5, 391-401.

[9]. Gómez-Jeria, J. S.; Morales-Lagos, D. R. Quantum chemical approach to the relationship between molecular structure and serotonin receptor binding affinity. *Journal of Pharmaceutical Sciences* 1984, 73, 1725-1728.

[10]. Gómez-Jeria, J. S.; Morales-Lagos, D. The mode of binding of phenylalkylamines to the Serotonergic Receptor. In *QSAR in design of Bioactive Drugs*, Kuchar, M., Ed. Prous, J.R.: Barcelona, Spain, 1984; pp 145-173.

[11]. Martin, Y. C. *Quantitative drug design: a critical introduction*. M. Dekker: New York, 1978; p x, 425 p.

[12]. Johnson, S. R. The Trouble with QSAR (or How I Learned To Stop Worrying and Embrace Fallacy). *Journal of Chemical Information and Modeling* 2007, 48, 25-26.

[13]. Chen, X.; Wang, W.; Wang, Y.; Lai, S.; Yang, J.; Cowling, B. J.; Horby, P. W.; Uyeki, T. M.; Yu, H. Serological evidence of human infections with highly pathogenic avian influenza A (H5N1) virus: a systematic review and meta-analysis. *BMC medicine* 2020, 18, 1-16.

[14]. Dolan, B. It wasn't supposed to be a coronavirus: The quest for an influenza A(H5N1)-derived vaccine and the limits of pandemic preparedness. *Centaurus* 2020, 62, 331-343.

[15]. Qi, Y.; Ni, H. B.; Chen, X.; Li, S. Seroprevalence of highly pathogenic avian influenza (H5N1) virus infection among humans in mainland China: A systematic review and meta-analysis. *Transboundary and emerging diseases* 2020, 67, 1861-1871.

[16]. Gómez-Jeria, J. S.; Castro-Latorre, P.; Kpotin, G. Quantum Chemical Study of the Relationships between Electronic Structure and Antiviral Activities against Influenza A H1N1, Enterovirus 71 and Coxsackie B3 viruses of some Pyrazine-1,3-thiazine Hybrid Analogues. *International Journal of Research in Applied, Natural and Social Sciences* 2017, 5, 49-64.

[17]. Muñoz-Gacitúa, D.; Gómez-Jeria, J. S. Quantum-chemical study of the relationships between electronic structure and anti influenza activity. 2. The inhibition by 1H-1,2,3-triazole-4-carboxamide derivatives of the cytopathic effects produced by the influenza A/WSN/33 (H1N1) and A/HK/8/68 (H3N2) strains in MDCK cells. *Journal of Computational Methods in Molecular Design* 2014, 4, 48-63.

[18]. Muñoz-Gacitúa, D.; Gómez-Jeria, J. S. Quantum-chemical study of the relationships between electronic structure and anti influenza activity. 1. The inhibition of cytophatic effects produced by the influenza A/Guangdong Luohu/219/2006 (H1N1) strain in MDCK cells by substituted bisaryl amide compounds. *Journal of Computational Methods in Molecular Design* 2014, 4, 33-47.

[19]. Alarcón, D. A.; Gatica-Díaz, F.; Gómez-Jeria, J. S. Modeling the relationships between molecular structure and inhibition of virus-induced cytopathic efects. Anti-HIV and anti-H1N1 (Influenza) activities as examples. *Journal of the Chilean Chemical Society* 2013, 58, 1651-1659.

[20]. Gómez-Jeria, J. S. On some problems in quantum pharmacology I. The partition functions. *International Journal of Quantum Chemistry* 1983, 23, 1969-1972.

[21]. Gómez-Jeria, J. S. Modeling the Drug-Receptor Interaction in Quantum Pharmacology. In *Molecules in Physics, Chemistry, and Biology*, Maruani, J., Ed. Springer Netherlands: 1989; Vol. 4, pp 215-231.
[22]. Gómez-Jeria, J. S.; Ojeda-Vergara, M. Parametrization of the orientational effects in the drug-receptor interaction. *Journal of the Chilean Chemical Society* 2003, 48, 119-124.
[23]. Gómez-Jeria, J. S. A New Set of Local Reactivity Indices within the Hartree-Fock-Roothaan and Density Functional Theory Frameworks. *Canadian Chemical Transactions* 2013, 1, 25-55.
[24]. Gómez-Jeria, J. S. 45 Years of the KPG Method: A Tribute to Federico Peradejordi. *Journal of Computational Methods in Molecular Design* 2017, 7, 17-37.
[25]. Gómez-Jeria, J. S.; Kpotin, G. Some Remarks on The Interpretation of The Local Atomic Reactivity Indices Within the Klopman-Peradejordi-Gómez (KPG) Method. I. Theoretical Analysis. *Research Journal of Pharmaceutical, Biological and Chemical Sciences* 2018, 9, 550-561.
[26]. Gómez-Jeria, J. S.; Robles-Navarro, A.; Kpotin, G.; Garrido-Sáez, N.; Gatica-Díaz, N. Some remarks about the relationships between the common skeleton concept within the Klopman-Peradejordi-Gómez QSAR method and the weak molecule-site interactions. *Chemistry Research Journal* 2020, 5, 32-52.
[27]. Gómez-Jeria, J. S.; Kpotin, G. A Density Functional Theory Analysis of the relationships between electronic structure and KCNQ2 potassium channels inhibition by a series of retigabine derivatives. *Chemistry Research Journal* 2019, 4, 68-79.
[28]. Abdussalam, A.; Gómez-Jeria, J. S. Quantum Chemical Study of the Relationships between Electronic Structure and Corticotropin-Releasing Factor 1 Receptor Binding Inhibition by a Group of Benzazole Derivatives. *Journal of Pharmaceutical and Applied Chemistry* 2019, 5, 1-9.
[29]. Gómez-Jeria, J. S.; Surco-Luque, J. C. A Quantum Chemical Analysis of the Relationships between Electronic Structure and the inhibition of Botulinum Neurotoxin serotype A by a series of Derivatives possessing an 8-hydroxyquinoline core. *Chemistry Research Journal* 2017, 2, 1-11.
[30]. Gómez-Jeria, J. S.; Moreno-Rojas, C. Dissecting the drug-receptor interaction with the Klopman-Peradejordi-Gómez (KPG) method. I. The interaction of 2,5-dimethoxyphenethylamines and their N-2-methoxybenzyl-substituted analogs with 5-HT_{1A} serotonine receptors. *Chemistry Research Journal* 2017, 2, 27-41.
[31]. Jia, R.; Zhang, J.; Ai, W.; Ding, X.; Desta, S.; Sun, L.; Sun, Z.; Ma, X.; Li, Z.; Wang, D. Design, synthesis and biological evaluation of "Multi-Site"-binding influenza virus neuraminidase inhibitors. *European journal of medicinal chemistry* 2019, 178, 64-80.
[32]. Hypercube. *Hyperchem 7.01*, 7.01; 419 Phillip St., Waterloo, Ontario, Canada, 2002.
[33]. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P. *Gaussian 16 16Rev. A.03*, Gaussian: Pittsburgh, PA, USA, 2016.
[34]. Gómez-Jeria, J. S. *D-Cent-QSAR: A program to generate Local Atomic Reactivity Indices from Gaussian16 log files*, v. 1.0; Santiago, Chile, 2020.
[35]. Gómez-Jeria, J. S. An empirical way to correct some drawbacks of Mulliken Population Analysis (Erratum in: J. Chil. Chem. Soc., 55, 4, IX, 2010). *Journal of the Chilean Chemical Society* 2009, 54, 482-485.
[36]. Statsoft. *Statistica v. 8.0*, 2300 East 14 th St. Tulsa, OK 74104, USA, 1984-2007.
[37]. Russell, R. J.; Haire, L. F.; Stevens, D. J.; Collins, P. J.; Lin, Y. P.; Blackburn, G. M.; Hay, A. J.; Gamblin, S. J.; Skehel, J. J. The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature* 2006, 443, 45-49.
[38]. Dassault. *Dassault Systèmes Biovia Corp. Discovery Studio Visualizer v. 17.2.0.16349*, San Diego, USA, 2016.
[39]. Trott, O.; Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* 2010, 31, 455-461.
[40]. Frey, R.; Hayashi, T.; Buller, R. M. Directed evolution of carbon–hydrogen bond activating enzymes. *Current opinion in biotechnology* 2019, 60, 29-38.
[41]. Tresca, B. W.; Hansen, R. J.; Chau, C. V.; Hay, B. P.; Zakharov, L. N.; Haley, M. M.; Johnson, D. W. Substituent effects in CH hydrogen bond interactions: linear free energy relationships and influence of anions. *Journal of the American Chemical Society* 2015, 137, 14959-14967.

[42]. Pierce, A. C.; Sandretto, K. L.; Bemis, G. W. Kinase inhibitors and the case for CH… O hydrogen bonds in protein–ligand binding. *Proteins: Structure, Function, and Bioinformatics* 2002, 49, 567-576.
[43]. Popelier, P.; Bader, R. The existence of an intramolecular C-H-O hydrogen bond in creatine and carbamoyl sarcosine. *Chemical Physics Letters* 1992, 189, 542-548.
[44]. Chemaxon. *MarvinView*, 21.3.0; www.chemaxon.com: USA, 2021.
[45]. Gómez-Jeria, J. S.; Cassels, B. K.; Saavedra-Aguilar, J. C. A quantum-chemical and experimental study of the hallucinogen (±)-1-(2,5-dimethoxy-4-nitrophenyl)-2-aminopropane (DON). *European Journal of Medicinal Chemistry* 1987, 22, 433-437.
[46]. Gómez-Jeria, J. S.; Cassels, B. K.; Clavijo, R. E.; Vargas, V.; Quintana, R.; Saavedra-Aguilar, J. C. Spectroscopic characterization of a new hallucinogen: 1-(2,5-dimethoxy-4-nitrophenyl)-2-aminipropane (DON). *Microgram (DEA)* 1986, 19, 153-162.