Comparitive study of calculated and expermintal Pka values for fluoro-acetoxy derivative

Abstract

Pka is the negative base -10 logarithm of the acid dissociation constant (Ka) of a solution. pKa = -log10Ka. The lower the pka value, the stronger the acid. It depends on the identity and chemical properties of the acid. Therefore, pKa was introduced as an index to express the acidity of weak acids. because of the importance of pka values for enzymatic reactions as an indicator to the difficulty of the uncatalyzed reaction, which is important for understanding the proficiency of the enzyme.

In this study we have calculated pKa value for some acetoxy group molecules using CBS-Q method which is one of the Complete Basis Set methods to find accurate energies. The acetoxy group molecules were also planned by Quantitative Structure Activity Relationship (QSAR) to study their effect on paraoxonase1 activity.

The results of this study shows strong relationship, (R^2=0.99) between the calculated and expermintal pka, also shows correlations between the activity of the enzyme and some of studied descriptors. Moreover, the results of the study reveal that by using the SPSS program that there is correlation between LUMO, Softness, Nucleofugality and Electrofugality as dependent variables and Cal. pKa as independent variable.

Keywords: Acetoxy, QSAR, pKa, HOMO, and LUMO

1-Introduction

pKa values known as acid dissociation constants, they are essential to understand many fundamental of chemical reactions in chemistry and biochemistry. It is one of the key physico-chemical parameters in the pharmaceutical industry for drug design and formulation. pKa plays a major role in drug absorbtion, and transport process, some parameters are dependent on pKa such as solubility and lipophilicity.

The pKa values (ionization constants) of the active-site residues in an enzyme are of importance to the functionality of the catalytic mechanism of the enzyme [1]. Predominantly the catalytic reaction is initiated by the transfer of a proton from a protein residue (the proton donor) to the substrate, and one of the steps in an enzymatic reaction mechanism is normally a nucleophilic attack on a substrate atom or the stabilization of a positively charged intermediate.

pKa is the negative base -10 logarithm of the acid dissociation constant (Ka) of a solution; It is expressed as : pKa = $-\log_{10}$ Ka.

Where dissociation constant (Ka) is a quantitative measure of the strength of an acid in solution. Ka is the equilibrium constant for a chemical reaction known as dissociation in the context of acid–base reactions; it is also known as acid-ionization constant, or acidity constant. Ka can be written representatively as:

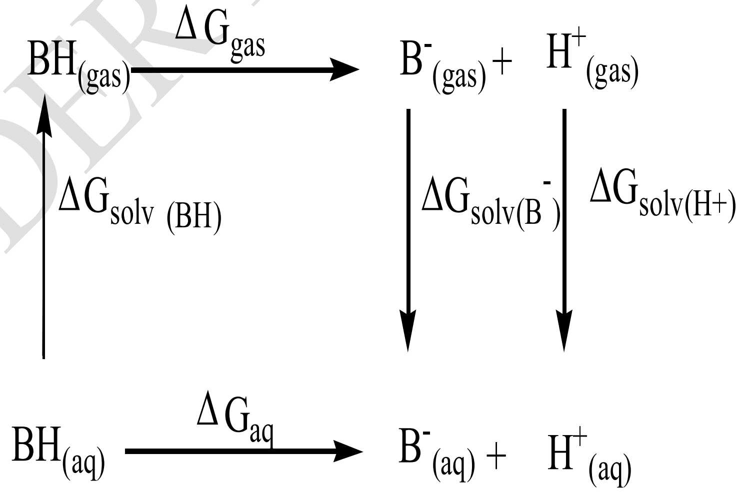
$$\mathrm{AH} \leftrightarrow \mathrm{A}^{-} + \mathrm{H}^{+}$$

Where AH and A^- undissociated and dissociated state of the weak acids respectively, and H^+ is the hydrogen ion as a result of weak acid dissociation.

$$\mathrm{Ka} = \frac{[\mathrm{A}^{-}][\mathrm{H}^{+}]}{[\mathrm{AH}]}$$

In this study we have calculated pKa value for acetoxy group using CBS-Q method which is one of the Complete Basis Set methods to find accurate energies [2-6] . We used the following thermodynamic cycles as [7].

1.1.Cycle 1 : The thermodynamic cycle 1 were given at Scheme 1 .



Scheme 1 Thermodynamic Cycle 1 Interrelationship between the gas phase and solution thermodynamic parameters.

Theoretical pKa values are commonly obtained by using the thermodynamic cycles. Experimental solvation free energy of H^+ were used to calculate pKa value in thermodynamic cycles.

pKa values are obtained by Eq.1

$$pKa = \Delta G_{aq} / 2.303RT \qquad (1)$$

$$\Delta G_{aq} = \Delta G_{gas} + \Delta\Delta G_{solv} \qquad (2)$$

In this cycle ΔGgas can be calculated as in Eq 3

$$\Delta G_{gas} = G_{gas}(H+) + G_{gas}(B-) - G_{gas}(BH) \qquad (3)$$

Since proton electronic energy is zero, Hgas(H+) is obtained by adding up the translational energy (E = 3/2RT) and PV = RT. This value is 1.48 kcal/mol at 298 K. Entropy, $S(H^+)$, is calculated by the Sackur–Tetrode equation for gas-phase monoatomic species, so TS = -7.76 kcal/mol at 298 K and 1 atm. Then $G_{gas}(H^+)$ equals -6.28 kcal/mol [8].

Since pKa determination employs the standard free deprotonation energy in solution 1 M, aqueous phase calculations also uses a reference state of 1 M and gas-phase free deprotonation energies are calculated to a reference state of 1 atm, gas-phase free energy difference, ΔG_{gas}, must be referred to 1 M by taking into account the factor RTln 24.46.

$$\Delta G_{gas}\ (1M) = \Delta\Delta G_{gas}\ (1atm) + RTln24.46 \qquad (4)$$

the calculation of $\Delta\Delta G_{solv}$ were obtained with the Eq.5

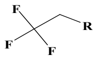
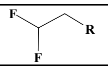
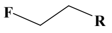
$$\Delta\Delta G_{solv} = \Delta G_{solv}(H^+) + \Delta G_{solv}(B^-) - \Delta G_{solv}(BH) \qquad (5)$$

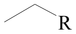
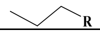
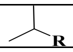
2- Materials and Methode

2.1 Molecules of study

Khersonsky, Tawfik had studied a group of molecules including these 7 molecules of study, published at [9].

Table 1: some of acetoxy derivative compounds [9].

Mol. No.	Name	Structure	Pka
1	Trifluoroethyl acetate		12.4
2	2,2-difluoroethyl acetate		13.3
3	2- fluoroethyl acetate		14.2

4	Ethyl acetate		16.1
5	Propyl acetate		16
6	Butyl acetate		16.1
7	Isopropyle acetate		17.1

CBS-Q was used for the calculation of pKa, these calculation have been used for gas and water phase calculations of acetoxy derivatives. all calculations were performed using Gaussian 09W program [10].

2.2 Calculation Methods

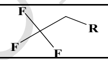
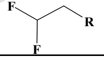
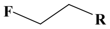
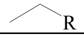
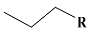
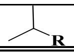
All calculations were performed on intel core-i7 Sony laptop Computer, using Gaussian 09W program [10]. The CBS-Q method has been used for all gas and water phase calculations for acetoxy molecules. All calculation results have no imaginary frequency at gas and water phase.

The CBS-Q method is one of the effective method of The Complete Basis Set Methods which were developed by Petersson and coworkers [11, 12, 6] . The CBS methods include some corrections for ab-inito calculation errors. These methods use relatively large basis sets for the structure calculation, medium sized basis sets for the second-order correlation correction, and small sized basis sets for higher order correlation corrections. Thus the CBS methods can compute energies for the molecules very accurately [13, 14, 11, 12].

3-Results

The pKa values for the 7 acetoxy derivatives were calculated, thermodynamic data for those molecules were calculated in the case of cis position and in trans position, these data are presented in Tables 2 and 3.

Table 2 calculated pKa to the experimental pka of some acetoxy derivatives compounds(cis-positions)

Mol. No.	Mol. Structure	CH2-g (kcal/mol)	gp-an (kcal/mol)	CH2-aq (kcal/mol)	aq-an (kcal/mol)	Cal. pKa	Exp. pKa
1		-604.71715	-604.179206	-604.724876	-604.269088	10.15	12.40
2		-505.531756	-504.597594	-505.540674	-505.083327	10.87	13.30
3		-406.353085	-405.807454	-406.360854	-405.901682	11.71	14.20
4		-307.193869	-306.638776	-307.200251	-306.737631	13.29	16.10
5		-346.42217	-345.867373	-346.428935	-345.966237	13.33	16.10
6		-385.650767	-385.096011	-385.657331	-385.19437	13.45	16.10
7		-346.426101	-345.870991	-346.432491	-345.968931	13.72	17.10

The calculated pka for the cis-position compounds were near of experimental pKa, and when test the correlation between them, we got good relationship, as presented by the Figure 1 followed.

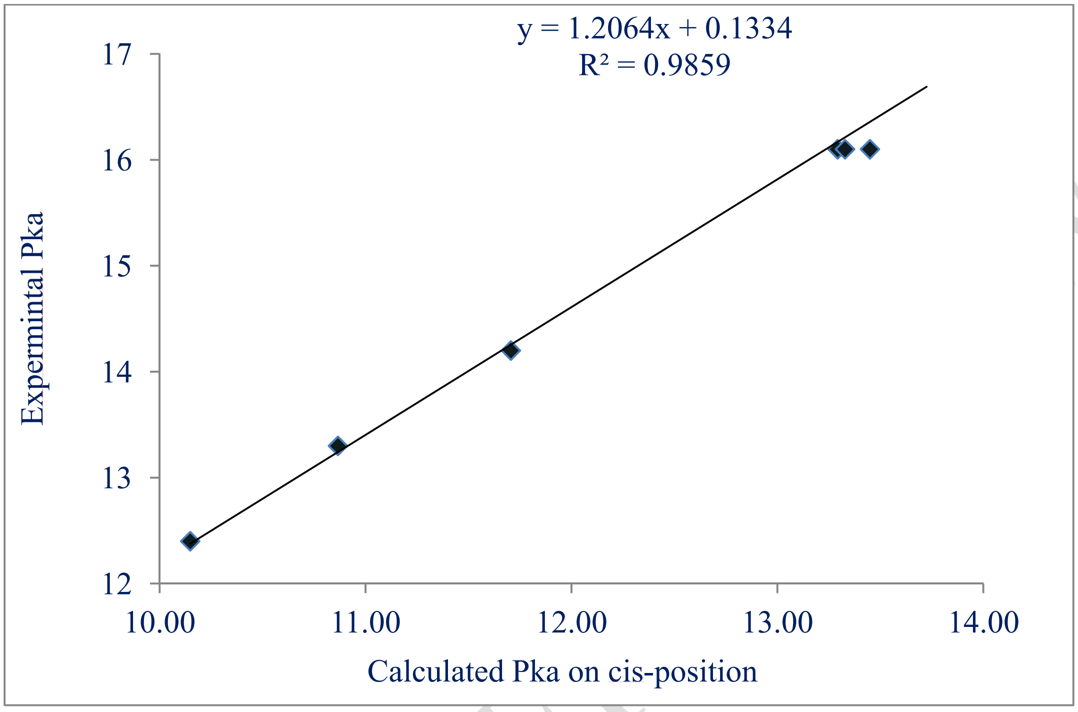


Figure 1: the relationship between the calculated and experimental pka in the cis-position

While in trans-position the calculated pka were little bit far from experimental one, but also it gave good relationship.

Table 3 calculated pka to the experimental pKa of some acetoxy derivatives compounds(trans-positions)

Mol. No.	CH2-g (kcal/mol)	gp-an (kcal/mol)	CH2-aq (kcal/mol)	aq-an (kcal/mol)	Cal. pka	Exp. pka
1	-604.7132	-604.105244	-604.720172	-604.2715	6.87	12.40
2	-505.528058	-504.999049	-505.534058	-505.08518	6.97	13.30
3	-406.346338	-405.811861	-406.354372	-405.90227	8.45	14.20
4	-307.189504	-306.642444	-307.196235	-306.73863	10.99	16.10
5	-346.416057	-345.871429	-346.423042	-345.96616	10.65	16.10
6	-385.645342	-385.10061	-385.652389	-385.19636	10.26	16.10
7	-346.470306	-345.877325	-346.429248	-345.97077	11.38	17.10

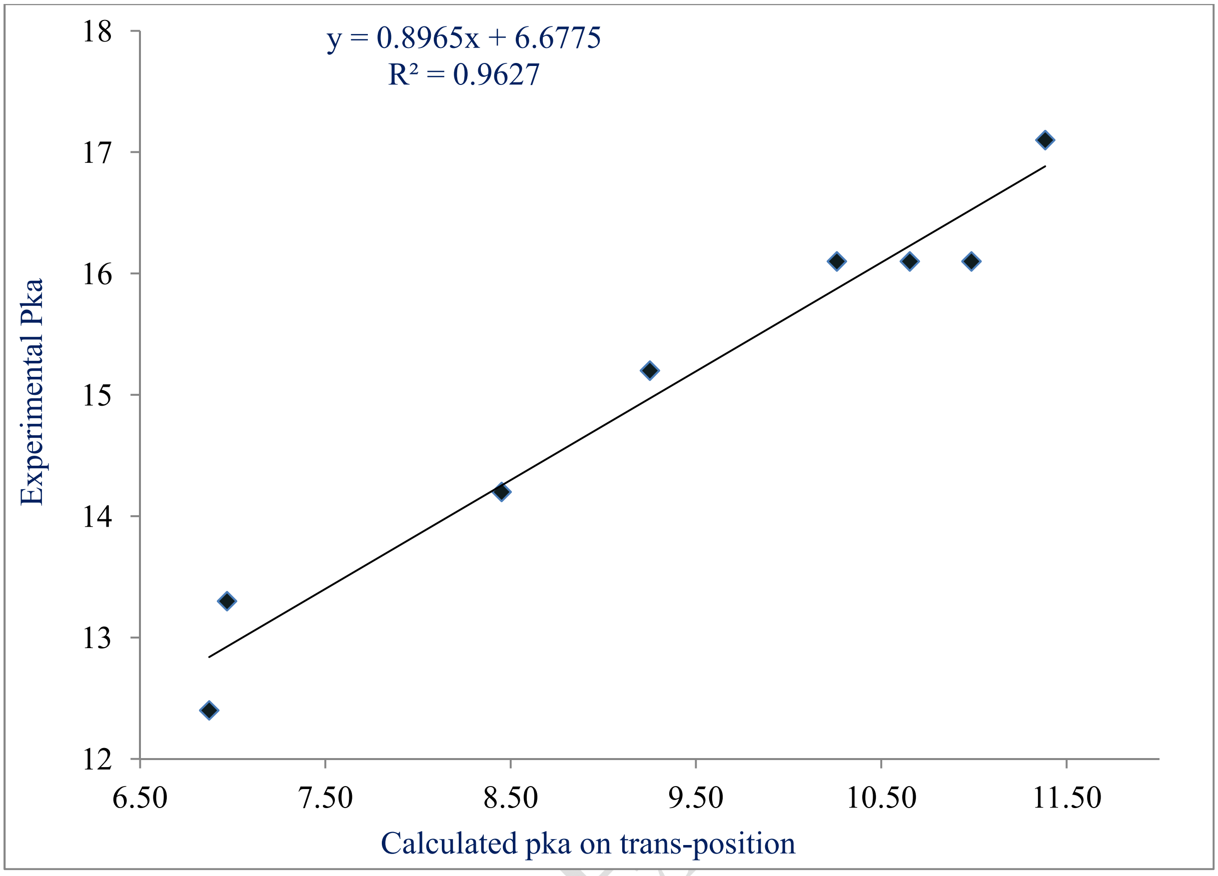


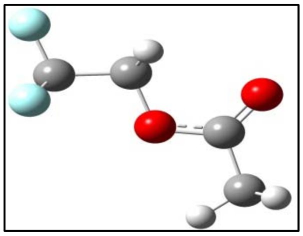
Figure 2: the relationship between the calculated and experimental pka in the trans-position

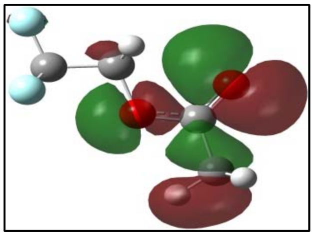
Moreover the quantum chemical calculations have been carried out at the CBS-Q level of theory using Gaussian-09 series of program package. Some descriptors for the same molecules like E_{HOMO}, E_{LUMO}, Energy gap, Hardness, Softness, Electronegativity, Chemical potential, Electrophilicity index, Electrofugality, and Nucleofugality were calculated, as shown in Table 4.

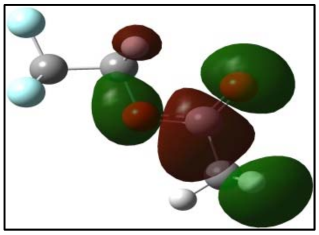
Table 4: calculated descriptors for the acetoxy derivative compounds (cis-position)

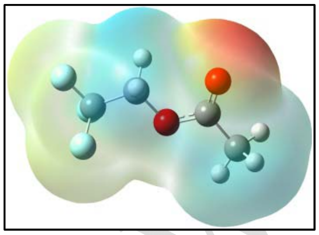
Mol. No.	**HOMO (eV)**	**LUMO (eV)**	**Energy Gap(eV)**	**Hardness (eV)**	**Softness (eV^{-1})**	**Electro-negativity (eV)**	**Chemical potential (eV)**	**electrophili-city index(ω)**	**Nucleo-fugality**	**Electr-fugality**	**Cal. pka**	**exp. pka**
1	-12.5081	4.428144	16.93623	8.468114	0.059045	4.03997	-4.03997	0.963695	1.157782	9.237722	10.2	12.4
2	-12.3201	4.594951	16.915	8.457501	0.059119	3.862551	-3.86255	0.882016	1.248216	8.973317	10.9	13.3
3	-12.1377	4.761214	16.89895	8.449474	0.059175	3.68826	-3.68826	0.804977	1.341454	8.717974	11.7	14.2
4	-11.7905	5.11959	16.9101	8.455052	0.059136	3.335462	-3.33546	0.657909	1.549973	8.220897	13.3	16.1
5	-11.757	5.079589	16.83663	8.418317	0.059394	3.338727	-3.33873	0.662074	1.532505	8.20996	13.3	16.1
6	-11.7252	5.092379	16.81759	8.408793	0.059462	3.316414	-3.31641	0.653994	1.541977	8.174804	13.4	16.1
7	-11.7472	5.137006	16.88425	8.442127	0.059227	3.305121	-3.30512	0.646983	1.562925	8.173167	13.7	17.1

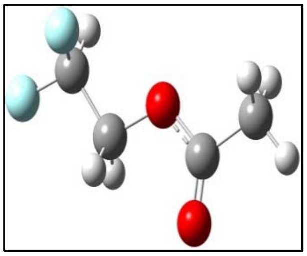
E_{HOMO} is associated with the ability of molecule to donate electron, the high E_{HOMO} value indicates the tendency of the molecule to donate electrons to an appropriate acceptor molecule with lower energy MO [15-17]. HOMO and LUMO orbitals of the 7 molecules were obtained from the quantum chemical calculation by the DFT using cbs-q bases sets as shown in Figure 1:

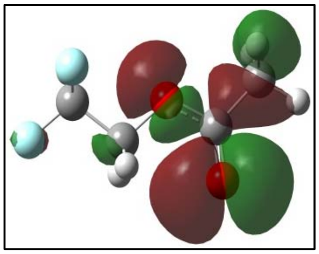

1-OPT

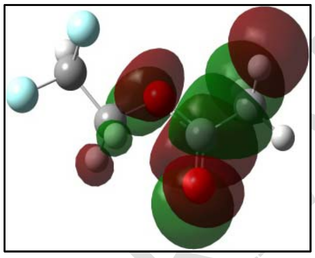

1-HOMO

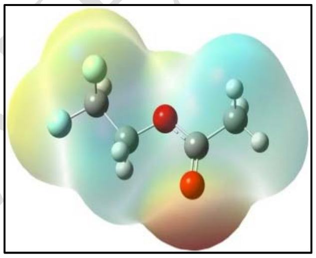

1-LUMO

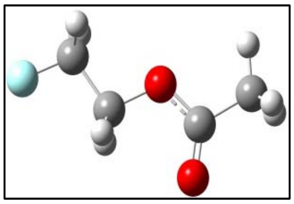

1-ESP

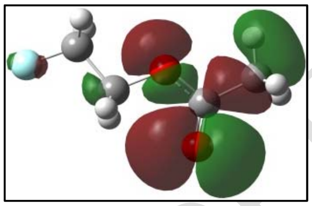

2-OPT

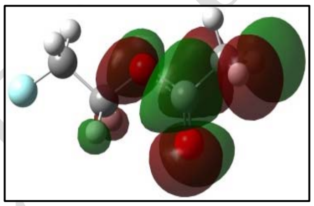

2-HOMO

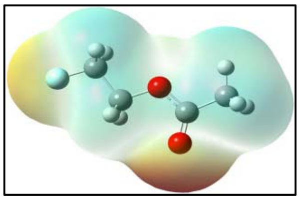

2-LUMO

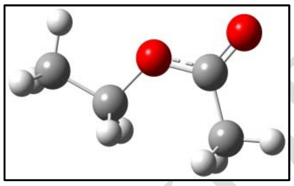

2-ESP

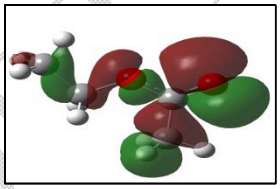

3-OPT

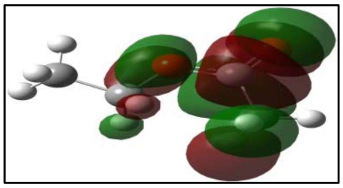

3-HOMO

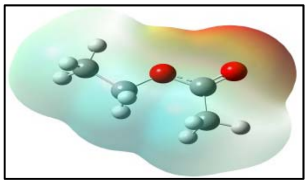

3-LUMO

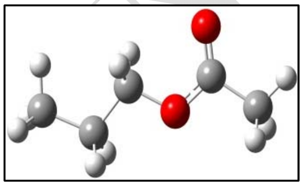

3-ESP

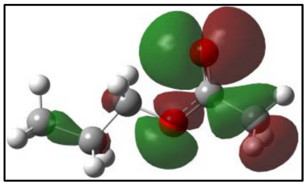

4-OPT

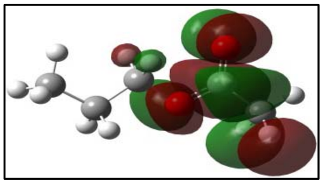

4-HOMO

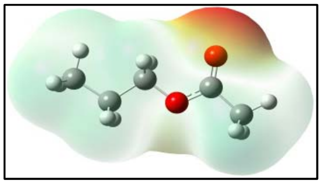

4-LUMO

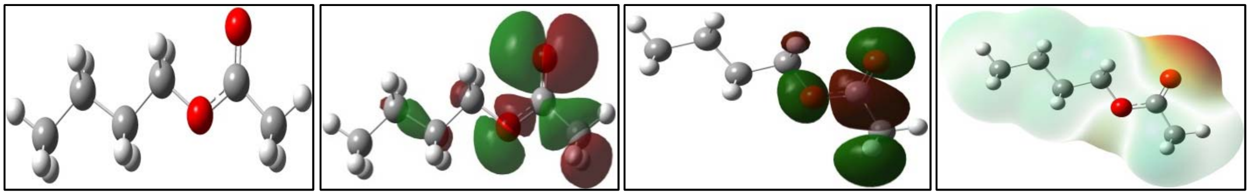

4-ESP


5-OPT

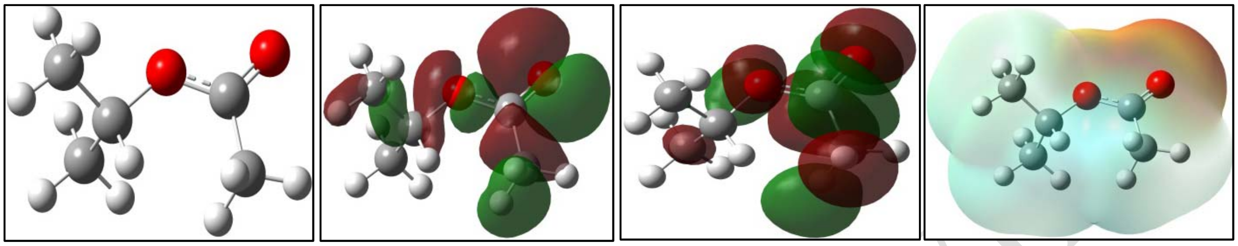

5-HOMO


5-LUMO


5-ESP



6-OPT 6-HOMO 6-LUMO 6-ESP



7-OPT 7-HOMO 7-LUMO 7-ESP

Figure 3: Molecular structure, HOMO, LUMO & ESP of the 8 molecules

The highest E_{HOMO} among the 7 cis-molecules were: -11.7252 eV, -11.7472 eV, -11.757 eV, -11.7905 eV, -12.1377 eV, -12.3201 eV and -12.5081 eV, recorded with molecules: 6, 7, 5, 4, 3, 2 and 1. The E_{LUMO} of these molecules were: 5.137006 eV, 5.11959 eV, 5.092379 eV, 5.079589 eV, 4.761214 eV, 4.594951 eV and 4.428144 eV, these values for 7, 4, 6, 5, 3, 2 and 1 molecules. The Energy Gap among these molecules were: 16.93623 eV, 16.9150 eV, 16.9101 eV, 16.89895 eV, 16.88425 eV, 16.83663 eV and 16.81759 eV, these for molecules 1, 2, 4, 3, 7, 5 and 6. As the molecules with smaller E_{HOMO}-E_{LUMO} energy gap lead to lower kinetic stability and higher chemical reactivity, so the molecules that have high activity are 1, 2, 4, 3, 7, 5 and 6 respectively, [15].

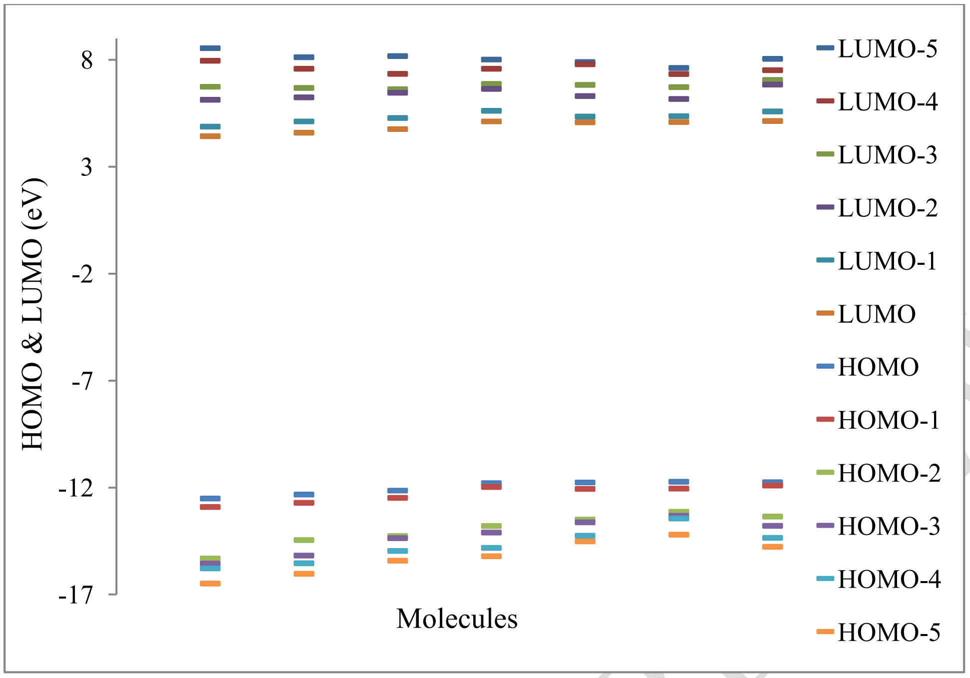


Figure 4: HOMO & LUMO for cis-molecules

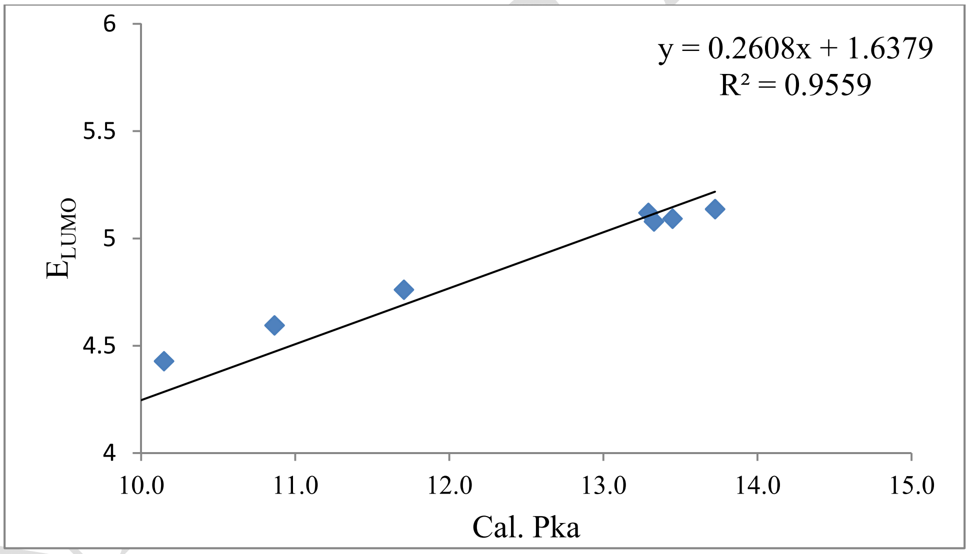


Figure 5:the correlation between calculated pka and LUMO in case of cis-position

Nucleofugality is: defined as the propensity of an atom or group of them to depart bearing the bonding electron pair in a heterolytic cleavage process, [18,19], the highest nucleofugality of the 7 molecules were: 1.562925, 1.549973, , 1.541977, 1.532505, 1.341454, 1.248216 and 1.157782. these results for molecules: 7, 4, , 6, 5, 3, 2 and 1. According to these values, those molecules have the activation activity to PON1. Figure 6 represents the relationship between the calculated Pka and nucleogugality, which confirmed by the statistical analysis.

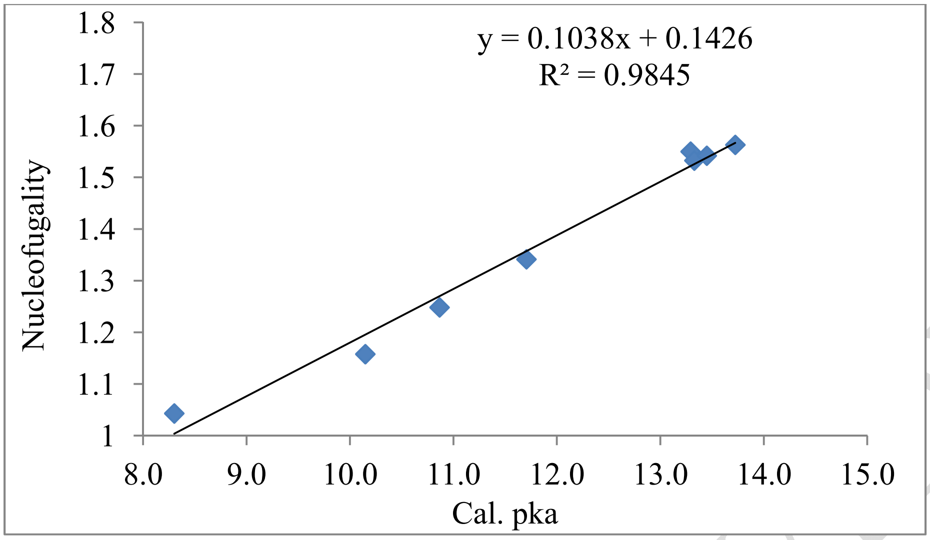


Figure 6:The correlation between calculated pka and Nucleofugality in case of cis-position

Table 5: calculated discriptors for the acetoxy derivative compounds (trans-position)

Mol. No.	**HOMO (eV)**	**LUMO (eV)**	**Energy Gap(eV)**	**Hardness (eV)**	**Softness (eV^{-1})**	**Electro-negativity (eV)**	**Chemical potential (eV)**	**electrophili-city index(ω)**	**Nucleo-fugality**	**Electr-fugality**	**Cal. pka**	**exp. pka**
1	-12.6836	4.4488	17.132423	8.566212	0.058369	4.117387	-4.117387	0.98952	1.155239	9.390013	6.87	12.4
2	-12.5206	4.6831	17.203718	8.601859	0.058127	3.918743	-3.918743	0.892629	1.274816	9.112301	6.97	13.3
3	-12.3105	4.7901	17.100586	8.550293	0.058478	3.760235	-3.760235	0.826835	1.341747	8.862217	8.45	14.2
4	-11.8975	5.0527	16.950106	8.475053	0.058997	3.422403	-3.422403	0.691019	1.506142	8.350948	10.99	16.1
5	-11.9315	5.1596	17.091062	8.545531	0.05851	3.385939	-3.385939	0.670794	1.55762	8.329499	10.65	16.1
6	-11.9029	5.1724	17.075279	8.53764	0.058564	3.365259	-3.365259	0.663237	1.566799	8.297316	10.26	16.1
7	-11.7818	5.0573	16.839082	8.419541	0.059386	3.362265	-3.362265	0.671345	1.51885	8.243381	11.38	17.1

In the case of trans position, the calculated discriptors were presented in Table 5, the values were not far of the cis-position case, but there was a rearrangement of the molecules, especially for the HOMO and LUMO, Energy Gap and nucleofugality as shown in Table 6, as well as Figures 7 and 8.

Table 6: comparison between trans and cis-positions highest E_{HOMO}, E_{LUMO} and Energy Gap.

Cis- position						**Trans-position**					
Mol. No.	E_{HOMO}	Mol. No.	E_{LUMO}	Mol. No.	Energy Gap	Mol. No.	E_{HOMO}	Mol. No.	E_{LUMO}	Mol. No.	Energy Gap
6	-11.7252	7	5.137006	1	16.93623	7	-11.7818	6	5.1724	2	17.20372
7	-11.7472	4	5.11959	2	16.9150	4	-11.8975	5	5.1596	1	17.13242
5	-11.757	6	5.092379	4	16.9101	6	-11.9029	7	5.0573	3	17.10059
4	-11.7905	5	5.079589	3	16.89895	5	-11.9315	4	5.0527	5	17.09106
3	-12.1377	3	4.761214	7	16.88425	3	-12.3105	3	4.7901	6	17.07528
2	-12.3201	2	4.594951	5	16.83663	2	-12.5206	2	4.6831	4	16.95011
1	-12.5081	1	4.428144	6	16.81759	1	-12.6836	1	4.4488	7	16.83908

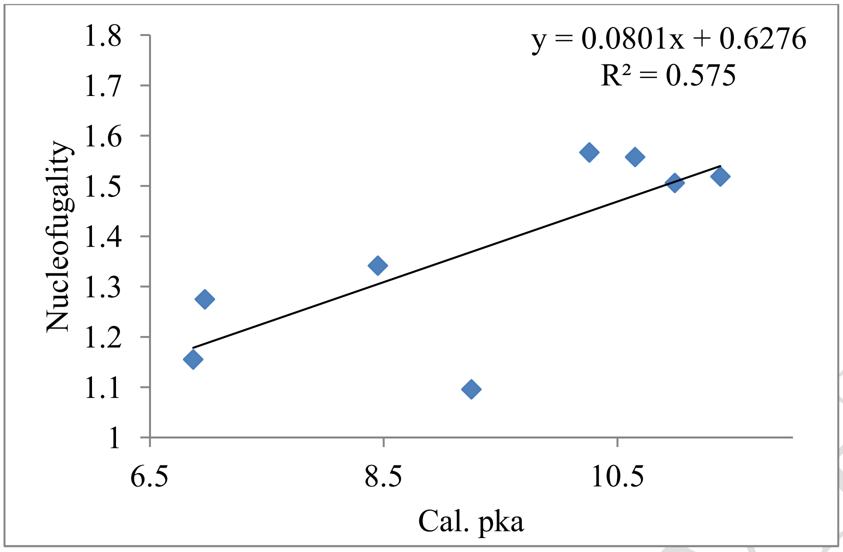


Figure 7:the correlation between calculated pka and Nucleofugality in case of tras-position

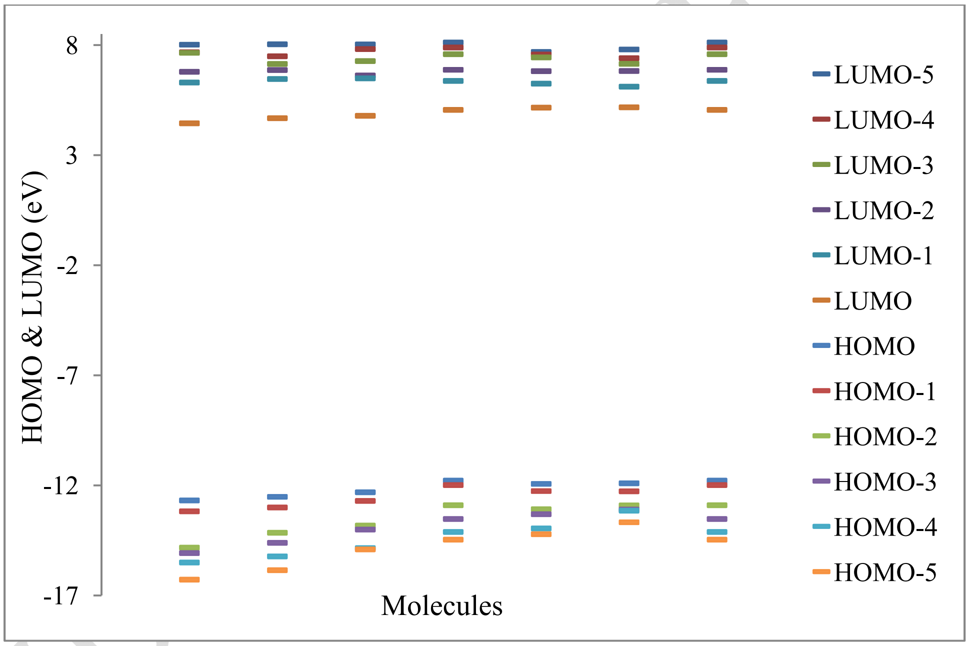


Figure 8: Frontier orbital energies near HOMO and LUMO near for trans-molecules

Statistical analysis

To detect whether there is a relationship or effect between the independent variable (Calculated pka) and dependent variables (E_{HOMO}, E_{LOMO}, Hardness, Energy Gap, Softness, Electronegativity, Chemical potential, electrophilicity index(ω), Nucleofugality and Electrfugality), and, which of these variables is more effective, we used multiple regression analysis, because it is the standard method, it is used to enter all independent variables, in which not excludes any variables.

Given the correlation matrix between all independent and dependent variables, it is shown that the variables (E_{HOMO}, E_{LOMO}, Energy Gap, Softness Hardness, Electronegativity and Chemical potential) have no correlation between them, so these variables will be excluded because they are not effective. While the remaining independent variables (electrophilicity index (ω), Nucleofugality, Electrfugality) are associated with the dependent variable, this is in case of cis-position molecules.

The value of the correlation coefficient between the dependent variable and the independent variables under study is R= 0.999 intermediate value and indicates a relationship between these variables. The coefficient of determination $R^2 = 0.998$, this means that the independent variables were able to explain 100% of the differences and changes in (Cal. pka).

Table 7: the model summary of the statistical analysis in cis-position

Model	R	R^2	Adjusted R^2	Std. Error of the Estimate
1	.999[a]	.998	.996	.1205

a. Predictors: (Constant), Electrfugality, Nucleofugality, electrophilicity index(ω).

Anova test shows that there is a very strong relationship between the independent variable and dependent variables. thus confirming the high explanatory power of the statistically multiple linear regression model. From the coefficients table we conclude that the statistically independent variables and T test at the significant level ($P \leq 0.05$) had no significant effect on the multiple regression model, although there is a correlation between these variables and the independent variable.

Regression equation was obtained using non-standard beta (fixed limit) as follows:

Cal. pka= -3.010+2.720 electrophilicity index(ω)+ 10.259 Nucleofugality-.145 Electrfugality.

In case of trans-position the independent variables (LOMO, Softness, Nucleofugality, Electrfugality) are associated with the dependent variable. The value of the correlation coefficient between the dependent variable and the independent variables under study is R= 0.682 intermediate value and indicates a relationship between these variables. The coefficient of determination $R^2 = 0.465$ this means that the independent variables were able to explain 47% of the differences and changes in (Cal. pKa).

Table 8: the model summary of the statistical analysis in trans-position

Model	R	R^2	Adjusted R^2	Std. Error of the Estimate
1	.682[a]	.465	.363	2.55

a. Predictors: (Constant), Electrfugality, LOMO, Nucleofugality, Softness)

Anova test show that there is relationship between the independent variable and dependent variables. thus confirming the high explanatory power of the statistically multiple linear regression model. From the coefficients table we conclude that the statistically independent variables and T test at the significant level ($P \leq 0.05$) had no significant effect on the multiple regression model, although there is a correlation between these variables and the independent variable.

Conclusion

This paper aims to study the relation of calculated pKa with the experimental for 7 molecules, and to study the effect of some discriptors on above mentioned molecles and their correlation with calculated pKa. The values of calculted pKa obtained reveal that there is a strong relationships between the calculated and expermintal pKa. The calculated values were nearby the expermintal values of cis-position moleculs, calculation of pKa using other methods could be more close to the expermintal.

References:

1. Nielsen J. E. & Mccammon J. A. Calculating pKa values in enzyme active sites. Protein Science. 2003;(12):1894–1901.
2. Nyden, M. R.; Petersson, G. A. Complete basis set correlation energies. I. The asymptotic convergence of pair natural orbital expansions. J Chem Phys 1981;(75):1843–1862.
3. Petersson, G. A.; Bennett, A.; Tensfeldt, T. G.; Al-Laham, M. A.; Shirley, W. A.; Mantzaris, J. A complete basis set model chemistry. I. The total energies of closed□shell atoms and hydrides of the first□row elements. J Chem Phys. 1988;(89):2193–2218.
4. a- Petersson, G. A.; Al-Laham, M. A. A complete basis set model chemistry. II. Open□shell systems and the total energies of the first□row atoms. J Chem Phys 1991;(94):6081–6090.
5. b- Petersson, G. A.; Tenfeldt, T. G.; Mongomery, J. A. A complete basis set model chemistry. III. The complete basis set□quadratic configuration interaction family of methods. J Chem Phys. 1991;(94): 6091–6101.
6. Petersson G. A., Malick D. K., Wilson W. G, Ochterski J. W., Montgomery J. A. and Frisch J. M. Calibration and comparison of the Gaussian-2, complete basis set, and density functional methods for computational thermochemistry. J. Chem. Phys. 1998;(109): 10570-10579.

7. Casasnovas, R.; Frau J.; Ortega-Castro J.; Salvà A.; Donoso J.; Muñoz F. Absolute and relative pKa calculations of mono and diprotic pyridines by quantum methods, Journal of Molecular Structure: THEO. CHEM. 2009, 912: 5–12.

8. Casasnovas, R.; Fernandez, D.; Ortega-Castro, J.; Frau, J.; Donoso J.; Muñoz F. (). Avoiding gas-phase calculations in theoretical pKa predictions, Theor Chem Acc. 2011 130: 1–13.

9. Khersonsky O; Tawfik DS. Structure-reactivity studies of serum paraoxonase PON1 suggest that its native activity is lactonase. Biochemistry 2005;(44): 6371-6382.

10. Gao, D.Q.; Svoronos, P.; Wong, P.K.; Maddalena, D.; Hwang, J.; Walker, H. pK(a) of acetate in water: A computational study. J. Phys. Chem. 2005;109(47):10776—85.

11. Mongomery, J. A.; Ochterski, J. W.; Petersson, G. A. A complete basis set model chemistry. IV. An improved atomic pair natural orbital method. J Chem Phys 1994;(101): 5900–5909.

12. Montgomery J. A., Frisch M. J., Ochterski J. W., Petersson G. A. A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. J. Chem. Phys. 1999;(110): 2822-2827.

13. Ochterski J. W., Petersson G. Montgomery, J. A. A complete basis set model chemistry. V. Extensions to six or more heavy atoms. J. Chem.Phys. 1998;(104):2598-2619.

14. Liptak M. D. and Shields G. C. Accurate pKa Calculations for Carboxylic Acids Using Complete Basis Set and Gaussian-n Models Combined with CPCM Continuum Solvation Methods, J. Am. Chem. Soc. 2001;(123): 7314-7319.

15. Liptak M.D., Gross K.C., Seybold P.G., Feldgus S. and Shields G.C. Absolute pKa Determinations for Substituted Phenols. J. Am. Chem. Soc. 2002;(124):6421-6427.

16. Aihara, J.; “Reduced HOMO&LUMO Gap as an Index of Kinetic Stability for Polycyclic Aromatic Hydrocarbons”. J. Phys. Chem. 1999;(103): 7487–7495.

17. Kim KH, Han YK, Jung J. Basis set effects on relative energies and HOMO–LUMO energy gaps of fullerene C36. Theoretical Chemistry Accounts. 2005;(113): 233–237.

18. Osman ODFT Study of the Structure, Reactivity, Natural Bond Orbital and Hyperpolarizability of Thiazole Azo Dyes. Int J Mol Sci 2017;18.

19. Campod´onico P.R.; Ormaz´abal-T, R.; Aizman A.; Contreras R. “Permanent group effect on nucleofugality in aryl benzoates,” Chemical Physics Letters, 2010;(498): 221–225.

20. Mishra V., Pandeya S. N., Pannecouque C., Witvrouw M. and De Clercq E. Anti-HIV Activity of Thiosemicarbazone and Semicarbazone Derivatives of (±)-3-Menthone. Arch. Pharm. 2002;335(5):183-186.