Studies in Mathematical Sciences **ISSN 1923-8444 [Print]**
Vol. 1, No. 1, 2010, pp. 45-51 **ISSN 1923-8452 [Online]**
www.cscanada.org **www.cscanada.net**

Regression Analysis on the Chemical Descriptors of a Selected Class of DPP4 Inhibitors

Jose Isagani B. Janairo[1,4,*]
Frumencio F. Co[2]
Gerardo C. Janairo[3,4]
Derrick Ethelbhert C. Yu[3,4]

Abstract: The activity of a selected class of DPP4 inhibitors was assessed using quantum-chemical and physical descriptors. Using multiple linear regression model, it was found that ΔE, LUMO energy, dipole, area, volume, molecular weight and ΔH are the significant descriptors that can adequately assess the activity of the compounds. The model suggests that bulky and electrophilic inhibitors are desired. Furthermore a pair interaction between ΔE and dipole as well as for LUMO energy and dipole were determined. It is expected that the information derived herein will be beneficial for future design and development of DPP4 inhibitors.

Key Words: Multiple Linear Regression; Molecular Descriptors; 2D-QSAR; DPP4 Inhibitors

1. INTRODUCTION

The inhibition of dipeptidyl peptidase 4 (DPP4, EC 3.4.14.5) is a novel approach towards glycaemic control in type 2 diabetics [1] since this results into a normalized blood glucose level due to the prevention of degradation of its substrate, glucagon-like peptide 1 (GLP-1) [2]. GLP-1 is a gut hormone responsible for the stimulation of insulin secretion and biosynthesis, suppresses glucagon release and delays gastric emptying; therefore an increased half-life of GLP-1 in the system will positively contribute to an improved glucose metabolism [3]. The discovery of this viable target has prompted more research on the development and design of DPP4 inhibitors [4]. One of the approaches of drug design is structure-activity relationship (SAR) studies which can be done through synthetic means, theoretical studies, or a combination of both. The first method is usually conducted by synthesizing a library of compounds and thereafter measuring its activity. Theoretical investigation of activity on the other hand, is commonly called as quantitative structure-activity relationship (QSAR) study wherein a set of molecular descriptors are used to assess and predict the activity of the compounds through the derivation of an equation that relates all the selected parameters together [5].

[1]Physics Department, De La Salle University, 2401 Taft Avenue, Manila, Philippines.
[2]Mathematics Department, De La Salle University, 2401 Taft Avenue, Manila, Philippines.
[3]Chemistry Department, De La Salle University, 2401 Taft Avenue, Manila, Philippines.
[4]Materials Science and Nanotechnology Unit, Center for Natural Sciences and Environmental Research (CENSER), De La Salle University, 2401 Taft Avenue, Manila, Philippines.
[*]Corresponding author. E-mail address: jose.isagani.janairo@dlsu.edu.ph.
[†]Received 1 November 2010; accepted 20 November 2010.

Recently, a novel class of azalopyrimidine-based DPP4 inhibitors with high selectivity and efficacy was developed through SAR studies [6]. This paper herein describes the semi-empirical optimization of the reported azalopyrimidine derivatives and the subsequent calculation of their properties which shall be used as chemical descriptors in order to derive a multiple linear regression (MLR) model. The results that will be obtained are expected to be important in future development of DPP4 inhibitors since the significant molecular chemical descriptors as well as their interaction will be identified through a multiple linear regression model.

2. METHODOLOGY

The 11 azalopyrimidine derivatives were subjected to geometry optimization calculations employing AM1 semi-empirical method using Spartan 08 V.1.2.0 (Wavefun, Inc.). The resulting electronic and physical properties of the optimized structures were then utilized as parameters for the multiple linear regression model using Statistica V.9 (StatSoft). The reported binding affinity of the compounds with DPP4 was used as the dependent variable and the calculated molecular descriptors served as the independent variables.

Table 1: Reported binding affinity of the DPP4 inhibitors

Compound	Binding Affinity (nM)
1	64
2	67
3	137
4	93
5	73
6	50
7	18
8	50
9	31
10	29
11	106

The initial model was further refined by backward elimination in order to obtain significant descriptors which were then analyzed for descriptor interaction. The MLR equation equation was thereafter derived from the refined regression model. All statistical analyses used a significance level of 0.05.

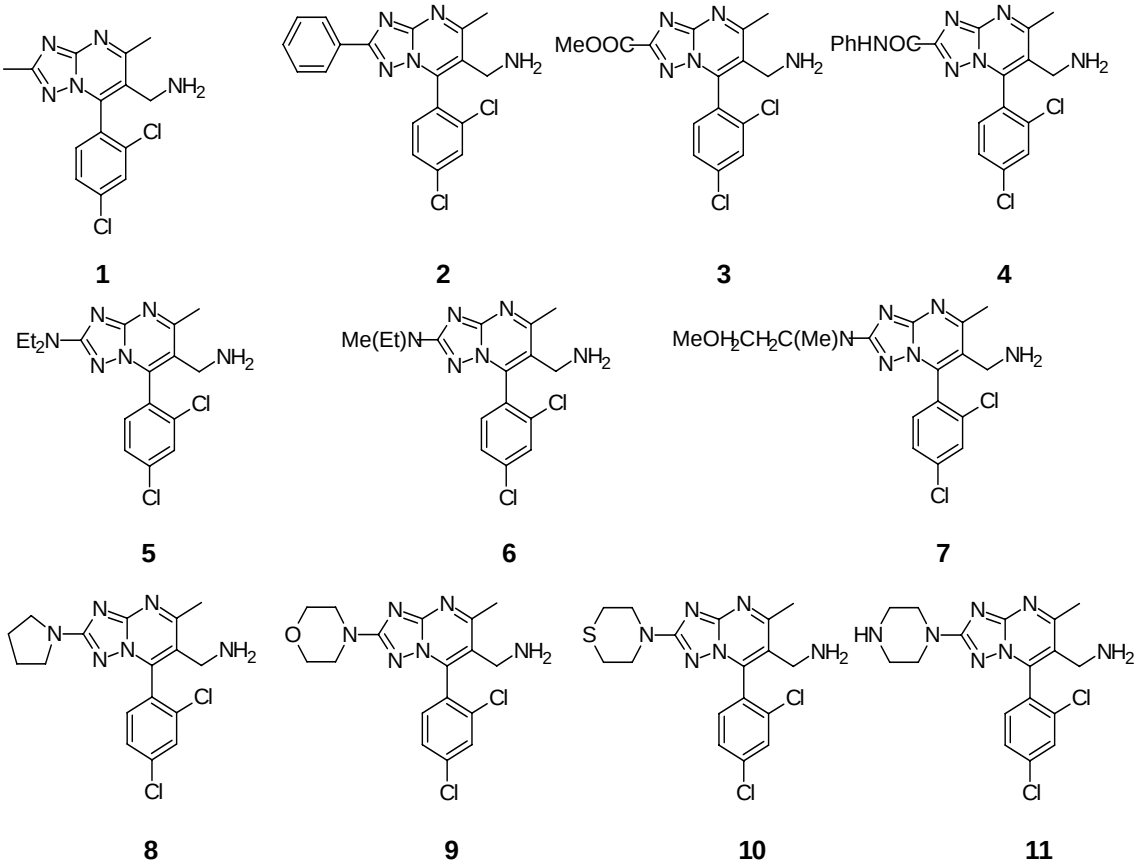

Figure 1: Structures of azalopyrimidine derivatives

3. RESULTS AND DISCUSSION

The geometry optimization and descriptor calculations were executed using the AM1 semi-empirical method since several comparative studies showed the surprising reliability of AM1 over ab initio (STO-3G) calculations when it comes to descriptor calculations [7]. The descriptors that were calculated included both quantum-chemical and physical descriptors in order to achieve a holistic depiction of the compounds. The quantum-chemical descriptors were limited to readily available parameters that did not require further calculations in order to highlight that simple descriptors are adequate to assess and predict the activity of the compounds. These quantum-chemical descriptors include energy (ΔE) & orbital energies (HOMO & LUMO), dipole, electron density centers (+ / - center) and enthalpy (ΔH). On the other hand, the physical descriptors comprise of the molecular weight (MW), polar surface area (PSA), area and volume of the molecule and hydrogen-bond donors and acceptors.

From the obtained data, a multiple linear regression model was set-up having a pearson's coefficient (r) of 0.999 which translates that the selected descriptors possess a linear relationship with the binding affinity. Despite exhibiting linearity, the number of descriptors was eliminated in order to identify which among th-em were significant. The backward elimination procedure was carried out on the basis of removing descri-ptors having a p-value greater than the significance level. Thus, the refined model consisted of energy, LU MO eigenvalue, dipole, area, volume, molecular weight and enthalpy. The refinement is justified since *r*=

0.998 which means that linearity is still maintained despite the removal of descriptors which were deemed to be insignificant in assessing and predicting the activity.

Table 2: Summary of calculated descriptors for the 11 DPP4 inhibitors

	ΔE (kj/ mol)	HOMO (eV)	LUMO (eV)	Dipole (debye)	Area (A^2)	Volume (A^3)	PSA (A^2)	Mol. Mass
1	242.27	-10.03	-1.22	3.54	328.9	309.75	50.26	351.193
2	363.02	-9.26	-1.29	3.69	394.63	375.92	46.83	413.26
3	256.43	-10.12	-1.32	6.41	343.6	320.4	73.1	366.21
4	274.88	-8.91	-1.54	2.14	427.96	408.96	73.52	456.29
5	264.66	-8.93	-1.02	3.65	401.37	379.11	50.54	408.29
6	288.02	-9.02	-1.06	3.39	383.44	360.56	50.97	394.26
7	114.41	-9.05	-1.13	3.04	414.65	388.54	58.43	424.29
8	292.53	-8.93	-1.03	3.67	386.66	367.55	51.3	406.27
9	140.67	-9.14	-1.14	3.45	394.34	375.53	58.99	422.72
10	331.1	-8.74	-1.16	3.56	401.87	384.18	50.73	438.34
11	328.78	-9.05	-1.1	4.6	398.11	378.58	63.63	421.29

	H-Bond Donor	H-Bond Acceptor	+ Center	- Center	ΔH (kJ/mol)	H-Bond Donor	H-Bond Acceptor
1	0	4	1	0	984.24	0	4
2	0	4	1	0	1259.04	0	4
3	1	5	2	0	1050.67	1	5
4	0	6	2	0	1252.8	0	6
5	0	4	2	0	1289.66	0	4
6	0	4	2	0	1233.67	0	4
7	0	5	2	0	1156.58	0	5
8	0	4	2	0	1261.3	0	4
9	0	5	2	0	1126.88	0	5
10	0	5	2	0	1304.53	0	5
11	0	5	3	0	1347.36	0	5

Table 3: Molecular descriptors and their corresponding p-values obtained from the regression model

Descriptor	P – value
ΔE	0.0524
HOMO	0.110
LUMO	0.0479
Dipole	0.0527
Area	0.0344
Volume	0.0607
PSA	0.147
MW	0.0278
ΔH	0.0453

The MLR equation is thus derived from refined regression model which is mathematically expressed as

Binding Affinity = -80.647 - 1.051(ΔE) - 480.758(LUMO) + 10.367(Dipole) - 5.803(Area) + 4.122(Volume) - 2.515(MW) + 1.313(ΔH).

The equation relates that binding affinity is a function of the determined significant descriptors. Since it is desirable for binding affinity to possess a low value it implies that qualitatively, a high positive ΔE, LUMO and Area are needed. Conversely, a negative or a small value for Dipole, Volume and MW are required as well.

Table 4: Refined multiple linear regression model

Descriptor	P- value	Slope (Intercept = -80.647)
ΔE	0.00121	-1.051
LUMO	0.00035	-480.758
Dipole	0.02194	10.367
Area	0.00756	-5.803
Volume	0.03466	4.122
MW	0.00319	-2.515
ΔH	0.00093	1.313

According to the refined multiple linear regression model, the LUMO energy, enthalpy and energy are the most important descriptors since they are the ones who possess the lowest p-values. Since ΔE is related to the structure of the molecule since it is based on the individual atomic interactions, it can be assumed that a bulky compound is desired since such a molecule possesses a high positive value for ΔE. Such an assumption is logical since the lock-and-key model is invoked therefore a bulky inhibitor can effectively occupy the active site thus preventing the natural substrate from occupying it. The relatively smaller magnitude of the LUMO energy suggests that the compounds exhibit more electron accepting properties or they are more electrophilic than nucleophilic [8]. The large value of the slope of the LUMO term indicates that a smaller value of the LUMO energy is needed. The enthalpy on the other hand, requires a small magnitude or a negative value. Enthalpy is usually associated in docking calculations wherein, ideally a large positive value is needed since enthalpy-based docking calculations are mathematically expressed as:

$$\Delta H_{Bind} = \Delta H_{system} - (\Delta H_{ligand} + \Delta H_{receptor}).$$

Despite this contradiction, it must be recalled that the descriptors exhibit additive interaction. Therefore any non-ideal value that will be obtained can be off-set by the other parameters.

In order to determine which among the descriptors exhibit pair interaction, a descriptor interaction matrix was constructed and the p-value was obtained for every pair of descriptor. Similar to what has been done previously, a descriptor possessing a p-value that is smaller than the significance level is deemed significant. From the matrix, it was determined that significant interaction exists between ΔE and Dipole descriptors as well as for the LUMO energy – Dipole descriptors. This means that aside from the non-specific additive interaction existing among the descriptors, a pair multiplicative interaction also exists for the aforementioned paired descriptors.

Table 5: Descriptor interaction matrix

	ΔE	LUMO	Dipole	Area	Volume	MW	ΔH
ΔE	*	0.112	0.028	0.350	0.369	0.355	0.380
LUMO		*	0.012	0.549	0.579	0.533	0.328
Dipole			*	0.060	0.062	0.059	0.057
Area				*	0.346	0.365	0.595
Volume					*	0.370	0.592
MW						*	0.621
ΔH							*

4. CONCLUSION

An MLR equation has been derived using a multiple linear regression model from significant quantum-chemical and physical descriptors. It was determined that linearity has not been compromised despite the removal of other descriptors. The significant descriptors were found to be ΔE, LUMO energy, dipole, area, volume, molecular weight and ΔH. Furthermore pair interaction between ΔE and dipole as well as for LUMO energy and dipole were determined. The derived MLR equation implies that favourable inhibitors should be bulky by virtue of their high energy and enthalpy; and the inhibitors should be electrophilic. Moreover, it has been demonstrated that simple descriptors can adequately assess and predict the activity of the selected class of DPP4 inhibitors. Results obtained are useful for virtual screening process in which the MLR equation can be utilized to rapidly screen molecules with the desired properties.

REFERENCES

[1] Gadsby, R. (2007). New treatment for type 2 diabetes – the DPP4 inhibitors. *Primary Care Diabetes I*. 209 -211.

[2] Havale, H., & Pal, M. (2009). Medicinal chemistry approaches to the inhibition of DPP4 for the treatment of type 2 diabetes. *Bioorg. Med. Chem. 17*, 1783-1802.

[3] Buteau, J. (2008). GLP-1 receptor signalling: effects on pancreatic β-cell proliferation and survival. *Diabetes & Metabolism. 34*, 73-77.

[4] Pratley, R., & Salsali, A. (2007). Inhibition of DPP-4: a new therapeutic approach for the treatment of type 2 diabetes. *Current Medical Research and Opinion. 23*, 919-931.

[5] Young, D. (2009). *Computational drug design*. New Jersey: John Wiley & Sons.

[6] Brigance, R., Meng, W., Fura, A., Harrity, T., Wang, A., Zahler, R., Kirby, M., & Hamann, L. (2010). Synthesis and SAR of azalopyrimidines as potent and selective DPP4 inhibitors for type 2 diabetes. *Bioorganic and Medicinal Chemistry Letters. 20*, 4395-4398.

[7] Kelson, M., & Lobanov, V. (1996). Quantum – chemical descriptors in QSAR/QSPR studies. *Chemical Reviews. 96*, 1027-1043.

[8] Bhattacharjee, A., Gupta, R., Ma, D., & Karle, J. (2000). Molecular similarity analysis between insect juvenile hormone and N, N-diethyl-*m*-toluamide (DEET) analogs may aid design of novel insect repellents. *Journal of Molecular Recognition. 13*, 213-220.