Computational Studies and Multivariate Analysis of Global and Local Reactivity Descriptors of Five Membered Heterocycles Molecules by Density Functional Theory (DFT)

N. Surendra Babu
Department of Chemistry
Mahaveer Institute of Sciences and Technology
Hyderabad, India.

&

P. Prashanth Kumar
Department of Mathematics
Mahaveer Institute of Sciences and Technology
Hyderabad, India.

Abstract

In this work we have been calculated global and local DFT reactivity descriptors five membered heterocyclic molecules at B3LYP/6-311++G (d, p) level. Global reactivity descriptors such as ionization energy (I), electron affinity (A) molecular hardness (η) and electrophilicity (ω), were calculated to evaluate for the heterocyclic molecules reactivity in gas phase. Calculated values lead to the conclusion that the reactivity trend based on the energy gap between HOMO and LUMO is given by H27 < H10 < H23 < H12 < H11 < H28 < H5 < H3 < H7 < H4 < H2 < H25 < H15 < H24 < H16 < H1 < H17 < H26 < H14 < H13 < H6 < H22 < H8 < H21 < H20 < H19 < H18 < H9 respectively. The Chemometric methods PCA and HCA were employed to find the subset of variables that could correctly classify the compounds according to their reactivity. From the PCA and HCA results in this work, a classification model was built with the aim to be used in the search for the five membered heterocyclic molecules for the reactivity.

Keywords: N-(Pyridine-2-yl) acetamide, Density functional Theory, Reactivity descriptors and Principal component analysis (PCA).

1. Introduction

Heterocycles are an important class of compounds, making up more than half of all known organic compounds. Heterocycles are present in a wide variety of drugs, most vitamins, many natural products, biomolecules, and biologically active compounds, including antitumor, antibiotic, anti-inflammatory, antidepressant, antimalarial, anti-HIV, antimicrobial, antibacterial, antifungal, antiviral, antidiabetic, herbicidal, fungicidal, and insecticidal agents. Also, they have been frequently found as a key structural unit in synthetic pharmaceuticals and agrochemicals. Most of the

heterocycles possess important applications in materials science such as dyestuff, fluorescent sensor, brightening agents, information storage, plastics, and analytical reagents. In addition, they have applications in supra molecular and polymer chemistry, especially in conjugated polymers. Moreover, they act as organic conductors, semiconductors, molecular wires, photovoltaic cells, and organic light-emitting diodes (OLEDs), light harvesting systems, optical data carriers, chemically controllable switches, and liquid crystalline compounds.

Heterocyclic aromatic compounds are widely distributed pollutants in soil, air, sediments, surface water and groundwater, as well as in animal and plant tissues. They may be of natural origin (e.g. alkaloids), but high environmental concentrations mainly result from human activities. In particular, industrialized areas, such as creosote contaminated sites, represent important sources of tar oil pollutants. Creosote represents a complex mixture of over 10,000 single organic substances which are formed by thermal processes related to coal and fossil fuels. Beside technical and chemical processes that involve tar oil, heterocyclic compounds are also present in dyestuff, pesticides and pharmaceuticals.

A heterocyclic compound is one which possesses a cyclic structure with at least two different kinds of hetero atoms in the ring. Nitrogen, oxygen, and sulphur are the most common hetero atoms. Heterocyclic compounds are very widely distributed in nature and are essential to life in various ways. Heterocyclic compounds containing nitrogen, oxygen belonging to five/ six membered heterocyclic compounds has occupied enormous significance in the field of medicinal chemistry.

The present work reports the results of a systematic theoretical examination of five-membered heterocyclic systems, a set of one heteroatom and a set of heterocycles with more than one heteroatom (Scheme.1). In this work, density functional theory (DFT) calculations have been carried out at the B3LYP/6-311++G (d, p) level of theory to explore and calculate more representative descriptors. Furthermore, the multivariate methods, such as a principal component analysis (PCA) and hierarchical cluster analysis (HCA), have been employed with the aim of selecting the variables responsible for reactivity and to describe properly the relationship between the calculated descriptors for the five membered heterocyclic compounds.

2. Materials and Methods

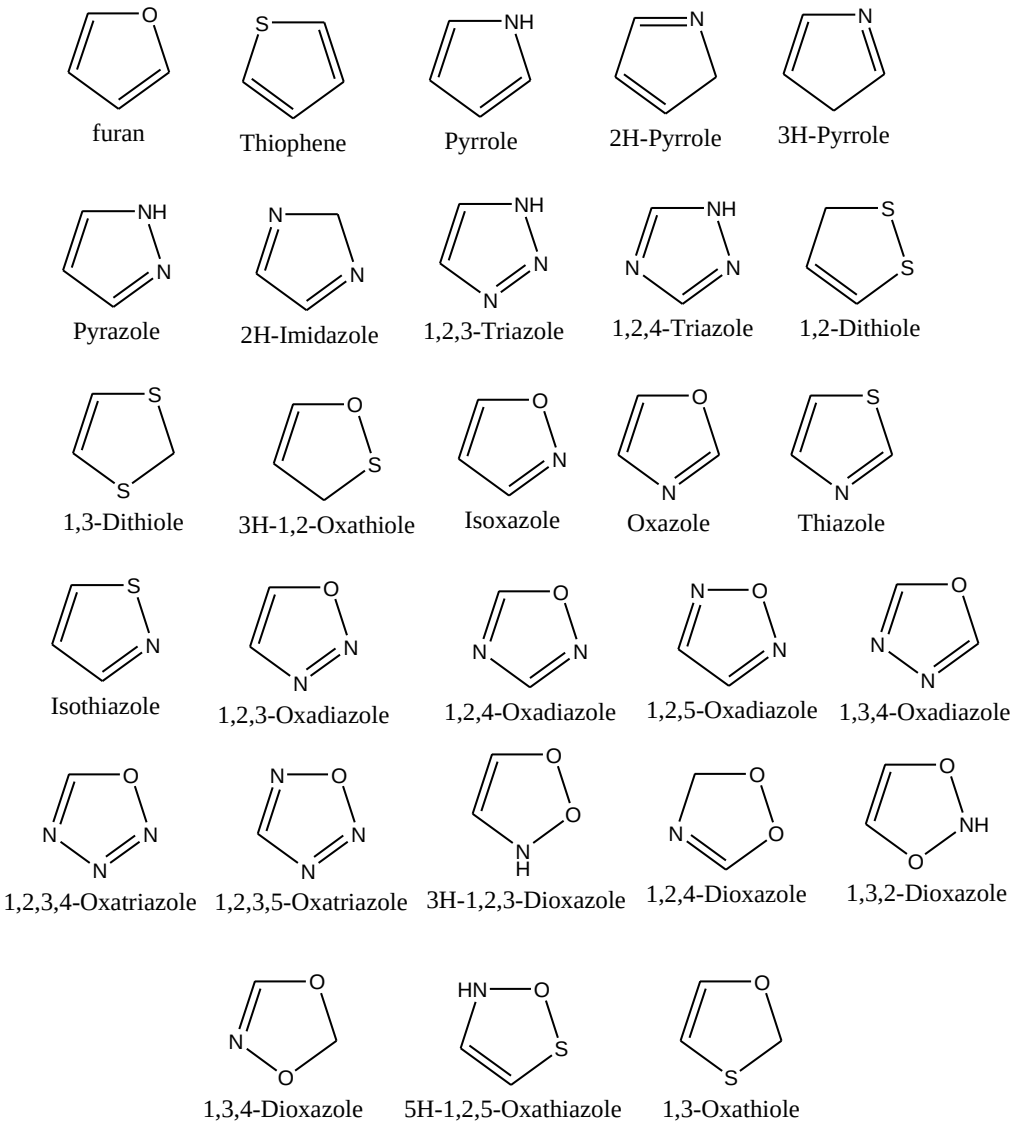
2.1 Computational Methods

The geometries of all heterocyclic compounds investigated (scheme 1) were completely optimized with employing the Becke3LYP functional [5] of the density functional theory [7] (DFT) with the polarized triple zeta split valence 6 311++G (d, p) basis set. The absence of imaginary frequencies confirmed that the structures are true minima on the potential energy surface. The entire calculations conducted in the present work were performed in the Gaussian 09W package [5] program. By combining the results of the Gauss view program [7] with symmetry considerations, vibrational frequency assignments were made with a high degree of accuracy. Density functional theory based descriptors have found immense usefulness in the prediction of reactivity of atoms and molecules as well as site selectivity [8]. The resourcefulness of density functional descriptors in the development of QSAR has been recently reviewed by Chattaraj et al [3]. Chemical hardness (η), chemical potential (μ), electrophilicity index (ω) and softness(s) are known as global reactivity descriptors. Recently Parr et al. [9] have defined a new descriptor to quantify the global electrophilic power of the molecule as electrophilicity index (ω), which defines a quantitative classification of the global electrophilic nature of a molecule within a relative scale.

2.2 Chemometric Methods

Principal component analysis (PCA) and Hierarchical cluster analysis (HCA) are two important techniques in multivariate analysis to analyze data that corresponds to more than one variable. The main objective of PCA [2], [4] and HCA are to study how the variables are related to one another, and how they work in combination to distinguish between multiple cases of observations.

PCA uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables. The number of PCs is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to (i.e., uncorrelated with) the preceding components. The principal components are orthogonal because they are the eigenvectors of the covariance matrix, which is symmetric. PCA is sensitive to the relative scaling of the original variables. Because of the lack of optimal data distribution (different units and variances), some pre processing operation is required, as auto scaling



Scheme.1: Five-membered heterocyclic systems, a set of one heteroatom and a set of heterocycles with more than one heteroatom.

(the scaled variables have zero mean and unity variance). Once the redundancy is removed, only the first few PCs are required to describe most of the information contained in the original data.

HCA [2] is one of the most straightforward methods. It displays the data in 2D space, qualitatively, in a form of dendrograms with similarities among samples or variables. The distances between samples or variables are calculated, transformed into a similarity matrix **S**, and compared. For any two samples k and l, the similarity index is

$$S_{kl} = 1.000 - \frac{d_{kl}}{d_{\max}} \quad (1)$$

Where S_{kl} is an element of **S**, d_{max} is the largest distance among each pair of samples in the data, and d_{kl} is the Euclidean distance among samples k and l. All chemometric methods were performed using software Mini tab 17 version (Trial version).

3. Theoretical Background

Global and local reactivity descriptors

The chemical structures of the five membered heterocyclic molecules were optimized with B3LYP method employing 6- 311++ G (d, p) basis set in gas phase. The calculated descriptors can be classified into four different electronic categories including: local charges, dipoles, orbital energies and the quantum chemical indices.

According to, the Koopmans' theorem [4] for closed-shell molecules, ionization potential (I) and electron affinity (A) can be expressed as follows in terms of E_{HOMO} and E_{LUMO} the highest occupied molecular orbital energy, and the lowest unoccupied molecular orbital energy, respectively:

$$I = -E_{HOMO} \quad \text{and} \quad A = -E_{LUMO} \quad (2)$$

When the values of I and A are known, one can determine through the following expressions [2] the values of the absolute electron negativity (χ), the absolute hardness (η),the chemical potential (μ) and the softness S (the inverse of the hardness):

$$\chi = \frac{I + A}{2}; \quad \eta = \frac{I - A}{2}; \quad \mu = \frac{(I + A)}{2} \quad \text{and} \quad S = \frac{1}{\eta} \quad (3)$$

The electrophilicity is a descriptor of reactivity that allows a quantitative classification of the global electrophilic nature of a molecule within a relative scale. Parr have proposed electrophilicity index as a measure of energy lowering due to maximal electron flow between donor and acceptor and defined electrophilicity index (ω) as follows [6].

$$\omega = \frac{\mu^2}{2\eta} \quad (4)$$

According to the definition, this index measures the propensity of chemical species to accept electrons. A good, more reactive, nucleophile is characterized by lower value of μ, ω, and conversely a good electrophile is characterized by a high value of μ, ω. This new reactivity index measures the stabilization in energy when the system acquires an additional electronic charge ΔN_{max} from the environment [5].

$$\Delta N_{\max} = -\frac{\mu}{\eta} \quad (5)$$

The maximum charge transfer ΔN_{max} towards the electrophile was evaluated using Eq. (5). Thus, while the quantity defined by Eq. (3) describes the propensity of the system to acquire additional electronic charge from the environment; the quantity defined in Eq. (5) describes the charge capacity of the molecule. Very recently, Ayers and co-workers [1], [11] have proposed two new reactivity

indices to quantify nucleophilic and electrophilic capabilities of a leaving group, nucleofugality (ΔEn) and electrofugality (ΔEe), defined as follows,

$$\Delta E_n = -A + \omega = \frac{(\mu+\eta)^2}{2\eta} \text{ and } \qquad \Delta E_e = I + \omega = \frac{(\mu-\eta)^2}{2\eta} \tag{6}$$

4. Results and Discussion

HOMO and LUMO Energies

The equilibrium geometry optimization for the isomers has been achieved by energy minimization, using DFT at the B3LYP level, employing the basis set 6-311++G(d, p). The optimized geometry of the five membered heterocyclic molecules under study are confirmed to be located at the local true minima on potential energy surface, as the calculated vibrational spectra has no imaginary frequency.

Table 1: The HOMO and LUMO energies and the energy gap between HOMO and LUMO (Δ Eg), in eV units at DFT/6-311++G(d, p) level in gas phase

S.No.	Molecule	Code	HOMO	LUMO	ΔE_g
1	Furan	H1	-6.53088	-0.16871	-6.36217
2	Thiophene	H2	-6.68925	-0.69853	-5.99072
3	Pyrrole	H3	-5.94609	-0.33580	-5.61030
4	2H-Pyrrole	H4	-7.23295	-1.35978	-5.87317
5	3H-Pyrrole	H5	-6.86259	-1.27869	-5.58390
6	Pyrazole	H6	-7.05526	-0.33471	-6.72055
7	2H-Imidazole	H7	-7.88386	-2.02376	-5.86010
8	1,2,3-Triazole	H8	-7.55677	-0.60356	-6.95321
9	1,2,4-Triazole	H9	-10.5229	-0.87840	-9.64448
10	1,2-Dithiole	H10	-5.57247	-1.73667	-3.83580
11	1,3-Dithiole	H11	-5.47968	-0.77391	-4.70577
12	3H-1,2-Oxathiole	H12	-5.89847	-1.43380	-4.46467
13	Isoxazole	H13	-7.70100	-0.98126	-6.71973
14	Oxazole	H14	-7.25798	-0.61853	-6.63946
15	Thiazole	H15	-7.19268	-1.16277	-6.02991
16	Isothiazole	H16	-7.43758	-1.33148	-6.10610
17	1,2,3-Oxadiazole	H17	-8.17176	-1.75871	-6.41305
18	1,2,4-Oxadiazole	H18	-8.76444	-1.47489	-7.28955
19	1,2,5-Oxadiazole	H19	-9.18677	-1.94103	-7.24574
20	1,3,4-Oxadiazole	H20	-8.26973	-1.03923	-7.23050
21	1,2,3,4-Oxatriazole	H21	-9.45318	-2.27601	-7.17717
22	1,2,3,5-Oxatriazole	H22	-9.60284	-2.73100	-6.87185

23	3H-1,2,3-Dioxazole	H23	-7.31132	-3.03060	-4.28072
24	1,2,4-Dioxazole	H24	-7.18751	-1.09909	-6.08841
25	1,3,2-Dioxazole	H25	-6.30910	-0.31185	-5.99725
26	1,3,4-Dioxazole	H26	-6.84817	-0.36464	-6.48353
27	5H-1,2,5-Oxathiazole	H27	-5.40757	-1.60877	-3.79880
28	1,3-Oxathiole	H28	-5.51968	-0.36981	-5.14987

The most important orbitals in a molecules are the frontier molecular orbitals, called highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO). These orbitals determine the way the molecule interacts with other species.

The frontier orbital energy gap helps characterize the chemical reactivity and kinetic stability of the molecule. A molecule with a small frontier orbital gap is more polarizable and is generally associated with a high chemical reactivity, low kinetic stability and is also termed as soft molecule. The HOMO is the orbital that primarily acts as an electron donor and the LUMO is the orbital that largely acts as the electron acceptor. Table 1 summarizes the highest occupied molecular (HOMO), the lowest unoccupied molecular orbital (LUMO) and HOMO and LUMO energy gaps (ΔEg) for studied molecules calculated at DFT level in the 6-311++G(d, p) basis set. The 3D plots of the frontier orbitals HOMO and LUMO figures for all isomers are shown in Fig. 1and 2. From the resulting data shown in Table 1, based on the frontier orbital energy gap, the reactivity order of the isomers are H27 < H10 < H23 < H12 < H11 < H28 < H5 < H3 < H7 < H4 < H2 < H25 < H15 < H24 < H16 < H1 < H17 < H26 < H14 < H13 < H6 < H22 < H8 < H21 < H20 < H19 < H18 < H9 respectively.

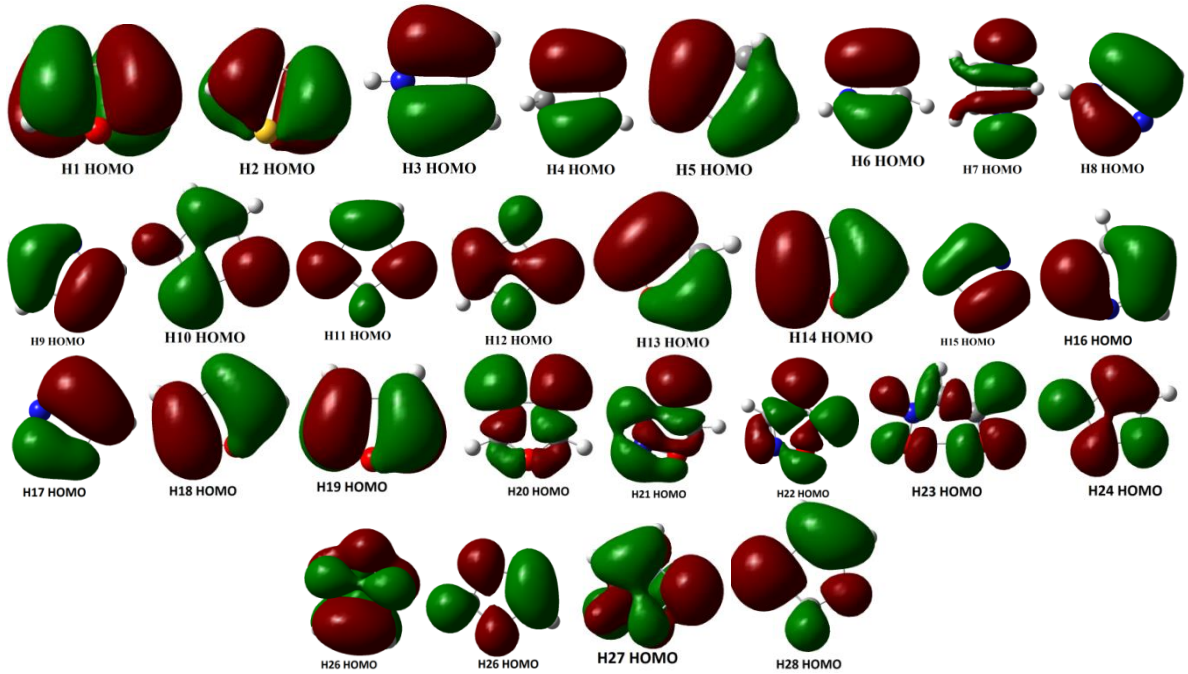


Figure: 1.The HOMO diagrams of heterocyclic molecules at B3LYP/6-311++G (d, p) level in gas phase.

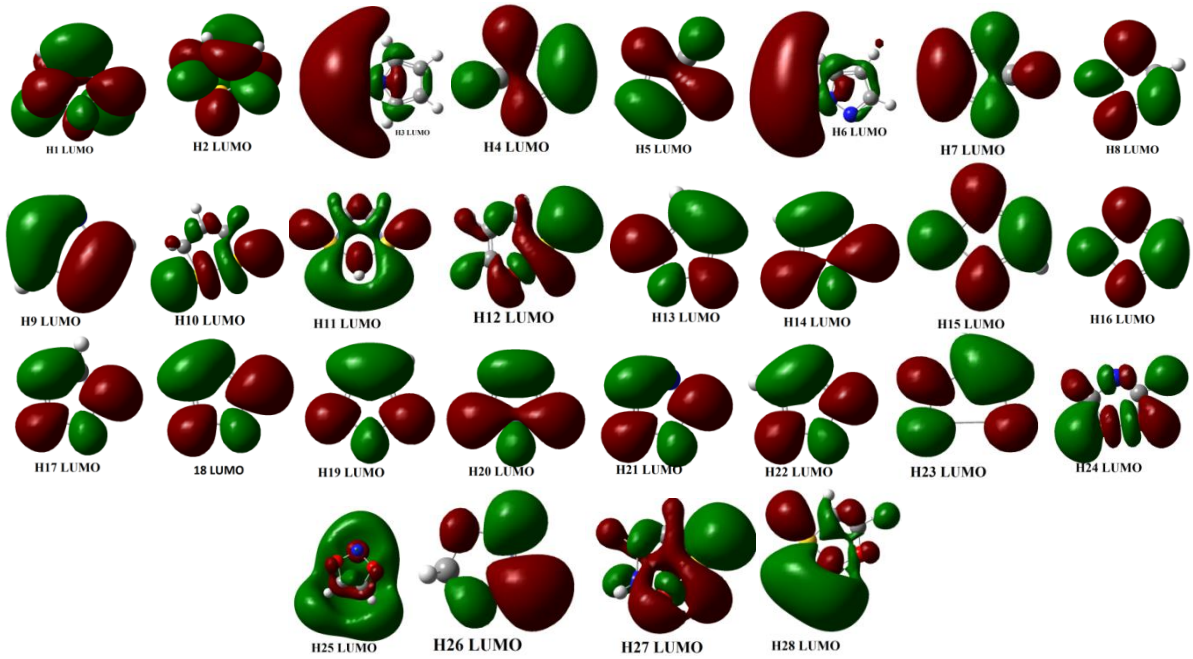


Figure: 2.The LUMO diagrams of heterocyclic molecules at B3LYP/6-311++G (d, p) level in gas phase.

The global descriptors, chemical potential, chemical, hardness and chemical softness for all studied isomers are given in Table 2. Ionization energy is a fundamental descriptor of the chemical reactivity of atoms and molecules. High ionization energy indicates high stability and chemical inertness and small ionization energy indicates high reactivity of the atoms and molecules. Absolute hardness and softness are important properties to measure the molecular stability and reactivity. It is apparent that the chemical hardness fundamentally signifies the resistance towards the deformation or polarization of the electron cloud of the atoms, ions or molecules under small perturbation of chemical reaction.

Table 2: Calculated Global Quantities ionization potential (I), electron affinity (A) ,Chemical potential (μ), Electronegativity (χ), Hardness (η), Softness (s), and Electrophilicity (ω) of different five membered heterocyclic in eV at DFT/6-311++G(d, p) level in gas phase.

Mol.	**I**	**A**	**χ**	**η**	**S**	**μ**	**ω**	**ΔN_{max}**	**ΔE_n**	**ΔE_e**
H1	6.5309	0.1687	3.3498	3.1811	0.3144	-3.3498	1.7637	1.0530	1.5950	8.2946
H2	6.6893	0.6985	3.6939	2.9954	0.3338	-3.6939	2.2777	1.2332	1.5791	8.9669
H3	5.9461	0.3358	3.1409	2.8051	0.3565	-3.1409	1.7585	1.1197	1.4227	7.7046
H4	7.2329	1.3598	4.2964	2.9366	0.3405	-4.2964	3.1429	1.4630	1.7831	10.3758
H5	6.8626	1.2787	4.0706	2.7920	0.3582	-4.0706	2.9675	1.4580	1.6888	9.8301
H6	7.0553	0.3347	3.6950	3.3603	0.2976	-3.6950	2.0315	1.0996	1.6968	9.0868
H7	7.8839	2.0238	4.9538	2.9301	0.3413	-4.9538	4.1877	1.6907	2.1639	12.0715
H8	7.5568	0.6036	4.0802	3.4766	0.2876	-4.0802	2.3943	1.1736	1.7907	9.9510
H9	10.5229	0.8784	5.7006	4.8222	0.2074	-5.7006	3.3695	1.1822	2.4911	13.8924
H10	5.5725	1.7367	3.6546	1.9179	0.5214	-3.6546	3.4819	1.9055	1.7452	9.0544
H11	5.4797	0.7739	3.1268	2.3529	0.4250	-3.1268	2.0776	1.3289	1.3037	7.5573

H12	5.8985	1.4338	3.6661	2.2323	0.4480	-3.6661	3.0104	1.6423	1.5766	8.9089
H13	7.7010	0.9813	4.3411	3.3599	0.2976	-4.3411	2.8045	1.2921	1.8232	10.5055
H14	7.2598	0.6185	3.9391	3.3206	0.3011	-3.9391	2.3365	1.1862	1.779	9.5962
H15	7.1927	1.1628	4.1777	3.0150	0.3317	-4.1777	2.8945	1.3857	1.7317	10.0871
H16	7.4376	1.3315	4.3845	3.0531	0.3275	-4.3845	3.1483	1.4361	1.8169	10.5859
H17	8.1718	1.7587	4.9652	3.2065	0.3119	-4.9652	3.8443	1.5485	2.0856	12.0160
H18	8.7644	1.4749	5.1197	3.6448	0.2744	-5.1197	3.5957	1.4047	2.1208	12.3601
H19	9.1868	1.9410	5.5639	3.6229	0.2760	-5.5639	4.2724	1.5358	2.3314	13.4592
H20	8.2697	1.0392	4.6545	3.6153	0.2766	-4.6545	2.9962	1.2875	1.9570	11.2659
H21	9.4532	2.2760	5.8646	3.5886	0.2787	-5.8646	4.7921	1.6342	2.5161	14.2452
H22	9.6028	2.7310	6.1669	3.4359	0.2910	-6.1669	5.5343	1.7948	2.8033	15.1371
H23	7.3113	3.0306	5.1710	2.1404	0.4672	-5.1710	6.2463	2.4159	3.2157	13.5577
H24	7.1875	1.0991	4.1433	3.0442	0.3285	-4.1433	2.8196	1.3610	1.7205	10.0071
H25	6.3091	0.3118	3.3105	2.9986	0.3335	-3.3105	1.8274	1.1040	1.5155	8.1365
H26	6.8482	0.3646	3.6064	3.2418	0.3085	-3.6064	2.0060	1.1125	1.6414	8.8542
H27	5.4076	1.6088	3.5082	1.8994	0.5265	-3.5082	3.2398	1.8470	1.6310	8.6473
H28	5.5197	0.3698	2.9447	2.5749	0.3884	-2.9447	1.6838	1.1436	1.3140	7.2035

5. Principal Component Analysis (PCA)

In the present research work, we auto scaled all calculated variables so that they can be compared in the same scale. Afterwards, PCA was used to reduce the number of variables and select the most relevant ones, i.e. those responsible for the tautomers reactivity. Several tests were performed and we obtained a good separation between more active and less active Tautomers compounds using eleven variables: I, A, χ , η, s, μ, ω, ΔN_{max}, E_n, E_e, (see Table 2).

From PCA results, we can observe that the first three principal components (PC1, PC2 and PC3) describe 99.41% of the overall variance as follows: PC1 = 66.95%, PC2 = 31.21% and PC3 = 1.22%. Since almost all of the variance is explained by the first two PCs, their score plot is a reliable representation of the spatial distribution of the points for the data set studied. The most informative score plot is presented in Figure 2 (PC1 versus PC2) and we can see that PC1 alone is responsible for the separation between more active and less active compound tautomers. Looking at Figure 3, we can see that the 28 compounds studied were separated into two groups: A (more active compounds - compounds H1,H2,H3,H4,H5, H6,H8,H10,H11,H12,H14,H15,H24,H25,H26,H27,H28 in Table 1) and B (less active compounds - compounds H7,H9,H13,H16,H17,H18,H19,H20,H21,H22,H23 in Table 1) where PC1>0 for the more active compounds and PC1<0 for the less active ones.

The loading vectors for the first two principal components (PC1 and PC2) are displayed in figure 2. According to figure 2 PC1 can be expressed through the following equation.

$$PC1 = 0.337\,[\,I\,] + 0.307\,[\,A\,] + 0.178\,[\,\eta\,] - 0.382\,[\,\mu\,] + 0.345\,[\,\omega\,] + 0.209\,[\Delta N_{max}] - 0.145\,[\,S\,] + 0.382\,[\,\chi\,] + 0.365\,[\,E_n\,] + 0.385\,[\,E_e\,] \quad (7)$$

From equation (7) we can see that more active tautomers compounds (PC1 > 0) can be obtained when we have higher A, χ , s, ω, ΔN_{max}, En, Ee, values (notice that I,A, η, ω, ΔN_{max}, s, χ , E_n, E_e, have positive coefficients in PC1 equation) combined with negative μ va and S values than those obtained for the less active tautomers compounds. In this way, some important features on the more active tautomers can be observed:

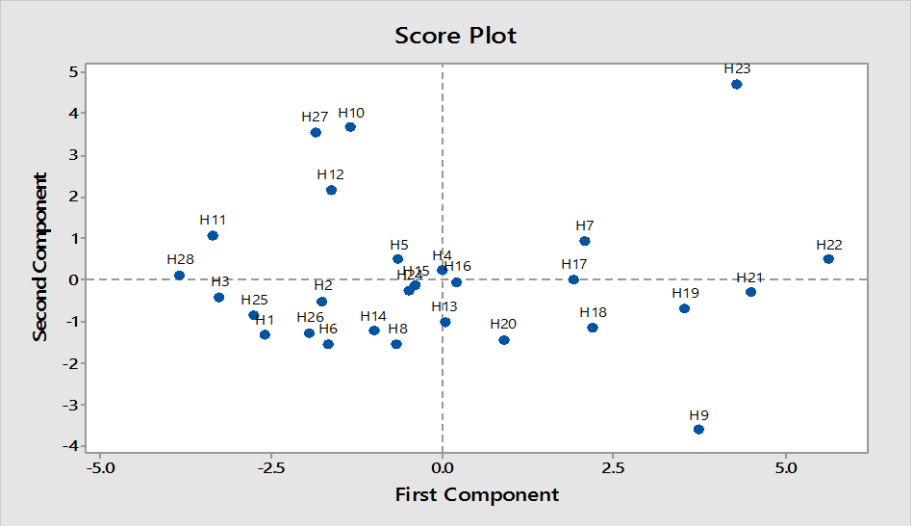


Figure 3. Score plot for the five membered heterocyclic molecules in gas phase.

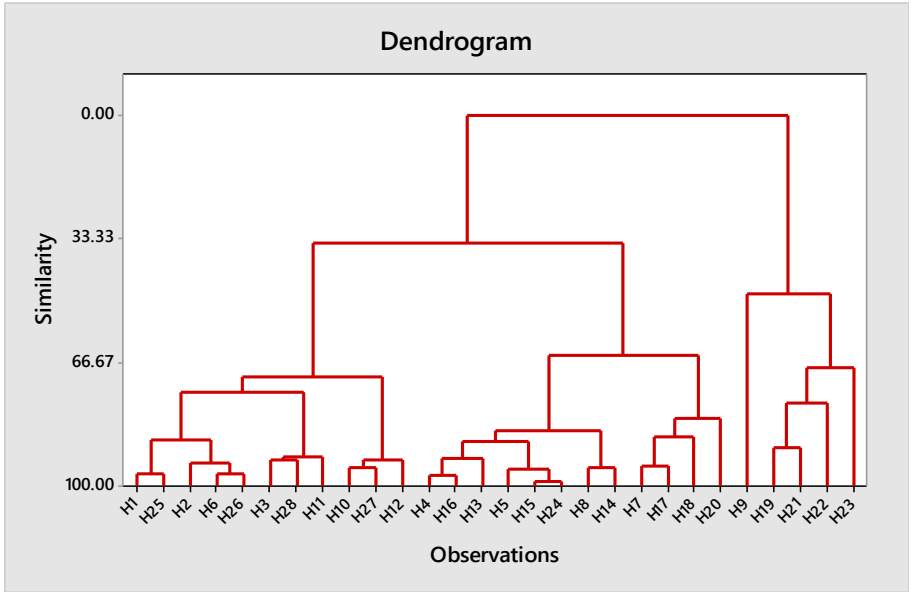


Figure 4. Dendrogram obtained for the heterocyclic compounds studied

6. Hierarchical Cluster Analysis (HCA)

Figure 3 shows HCA analysis of the current study. The horizontal lines represent the compounds and the vertical lines the similarity values between pairs of compounds, a compound and a group of compounds and among groups of compounds. The similarity value between the two classes of compounds was 0.0 and this means these two classes are distinct. From Figure 4, we can see that the HCA results are very similar to those obtained with the PCA analysis, i.e. the compounds studied were grouped into four categories: more actives and less active and in between active and inactive.

7. Conclusion

The reactivity of five membered heterocyclic molecules were theoretically investigated with the density functional theory employing the 6-311+G (d, p) basis sets. Based on the frontier orbital energy gap, the reactivity order of the isomers are H27 < H10 < H23 < H12 < H11 < H28 < H5 < H3 < H7 < H4 < H2 < H25 < H15 < H24 < H16 < H1 < H17 < H26 < H14 < H13 < H6 < H22 < H8 < H21 < H20 < H19 < H18 < H9 respectively. From PCA results, we can see that PC1 alone is responsible for the separation between more active and less active molecules. PC1>0 is more reactive molecules, and PC1<0 for the less reactive molecules respectively. From the HCA results are very similar to those obtained with the PCA analysis, i.e. the molecules studied were grouped into two categories: more reactive and less reactive tautomers.

8. Acknowledgement

The authors thankful to the Chairman and secretary of the Mahaveer institute of Science & Technology, to provide the facilities for carrying this research work and also thankful to principal to his cooperation.

References

[1]. Ayers, P. W., Anderson, J. S. M. and Bartolotti, L. J. 2005; *Int. J. Quantum Chem*, 101: 520.

[2]. Becke, A. D. 1993; *J. Chem. Phys.*, 98: 5648.

[3]. Chattaraj, P. K., Nath, S. and Maiti, B. 2003; Reactivity Descriptors In Computational Medicinal Chemistry for Drug Discovery, *Eds. Marcel Dekker*, New York, Chap. 11: 295–322.

[4]. Koopmans, T. 1993; *Physica*., 1: 104.

[5]. Lee, C., Yang, W. and Paar, R. G. 1988; *Phys. Rev.*, B37: 785.

[6]. Lesar, A. and Milosev, I. 2009; *Chem. Phys. Lett.*, 483 (4-6): 198.

[7]. Parr, R. G. and Wang, W.1994; Density-Functional Theory of Atoms and Molecules, *Oxford University Press*: New York.

[8]. Parr, R. G. and Yang, W. 1989; Density Functional Theory of Atoms and Molecules, *Oxford University Press*, UK.

[9]. Parr, R. G., Szentpaly, L. V. and Liu, S. J. 1999; *Am. Chem. Soc*.: 121.

[10]. Parr, R. G., Szentpaly, L. and Liu, S. 1999; *J. Am. Chem. Soc*.:121.

[11]. Roos, G., Loverix, S., Brosens, E. Belle, K., Wyns, Van L., Geerlings, P. and Messens, J.; 2006; *Chem BioChem.*, 7: 981.