EUROPEAN JOURNAL OF BIOMEDICAL AND PHARMACEUTICAL SCIENCES
http://www.ejbps.com

ISSN 2349-8870
Volume: 4
Issue: 8
1128-1138
Year: 2017

CHEMOMETRICS STUDIES FOR GLOBAL AND LOCAL REACTIVITY DESCRIPTORS OF AMIDE-IMIDOL TAUTOMERS OF LACTAMS BY HF AND DFT METHODS AT DIFFERENT BASIS SETS IN GAS PHASE

Dr. Nambury Surendra Babu, Saidi Mohamedi Katundu and John Joseph Makangara

Department of Chemistry, College of Natural and Mathematical Sciences, The University of Dodoma, Dodoma, Tanzania.

***Corresponding Author: Dr. Nambury Surendra Babu**
Department of Chemistry, College of Natural and Mathematical Sciences, The University of Dodoma, Dodoma, Tanzania.

Article Received on 19/06/2017 **Article Revised on 10/07/2017** **Article Accepted on 31/07/2017**

ABSTRACT
In the present research work we have been calculated global and local DF reactivity descriptors for different ring size lactams tautomers at different basis sets by HF and DFT/B3LYP methods.The global reactivity descriptors such as ionization energy (I), electronaffinity (A) molecular hardness(η) and electrophilicity (ω),were calculated in gas phase. Calculated values lead to the conclusion that Global reactivity trend based on ω, is given by K1B > E2B > K1G > K1D > E1G > K1E > E1D > E1E > E1B > E2D > E2G > E2E and E2B > E1G > E1E > K1B > K1D > E1 > E1B > K1E > K1G > E2G > E2D > E2E for HF and DFT methods at 6-311G++(d,p) respectively.The Chemometricmethods PCA and HCA were employed to find the subset of variables that could correctly classify the compoundsaccording 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 lactam tautomers for the reactivity.

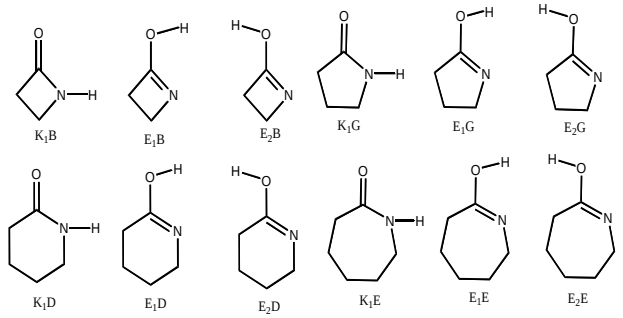
KEYWORDS:Lactams, HF and DFT , reactivity descriptors and Principal component analysis and Cluster analysis.

INTRODUCTION
The quantitative structure activity relationship (QSAR) is a mathematical representation of biological activity in terms of structural descriptors of a series of homologue molecules.[1–6] The main objective of QSAR is to look for new molecules with required properties using chemical intuition and experience transformed into a mathematically quantified and computerized form.[6] Once a correlation is established, the structure of any number of compounds with desired properties can be predicted. Recently the uses of quantum chemical descriptors in the development of QSAR have received attention due to reliability and versatility of prediction by these descriptors. In particular net atomic charges, HOMO–LUMO energies, frontier orbital electron densities and super delocalizabilities have been used to correlate with various biological activities.[6] Density functional theory based descriptors have found immense usefulness in the prediction of reactivity of atoms and molecules as well as site selectivity.[7–10] The resourcefulness of density functional descriptors in the development of QSAR has been recently reviewed by Chattaraj et al.[11] Chemical hardness (g), chemical potential (l), polarizability (a) and softness is known as global reactivity descriptors. Recently Parr et al.[12] have defined a new descriptor to quantify the global electrophilic power of the molecule as electrophilicity index (x), which defines a quantitative classification of the global electrophilic nature of a molecule within a relative scale. The earlier works of Maynard et al.[13] have formed the strong foundation for the electrophilicity index, which provided the direct relationship between the rates of reaction and the ability to identify the function or capacity of an electrophile and the electrophilic power of the inhibitors.

The present work reports the results of a systematic theoretical examination of different ring size lactams tautomers. Lactams can theoretically exist in three tautomeric structures, shown in seheme 1. In this work, HF and density functional theory (DFT/B3LYP) calculations have been carried out at different basis sets 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 and the tautomers of title compounds.

COMPUTATIONAL METHODS
The geometries of all compounds investigated (scheme 1) were completely optimized with the GAUSSIAN 09W

program,[14] employing HF and density functional theory DFT with the B3LYP functional[15–19] with different basis sets in the gas phase. In the present research paper we have been chosen the different ring size of lactams structures such β- lactams (four membered ring K1B,E1B and E2B), ϒ- lactams (five membered ring K1G,E1G and E2G) lactams, δ- lactams (six membered ring K1D,E1D and E2D) , ε- lactams (seven membered ring K1E,E1E and E2E) as shown in scheme 1. The correlation between the molecular properties calculated and the stability and reactivity studied was done by using the pattern recognition methods (PCA, HCA, and SDA) built into the statistical package Minitab (trial version) software with bivariate comparisons.



Scheme 1: Possible tautomeric forms of β- Lactams (K1B, E1B and E2B); γ-lactams (K1G,E1G and E2G) δ- lactams (K1D,E1D and E2D), ε - Lactams (K1E,E1E and E2E).

Theoretical Background

The chemical structures of the all ring size studied tautomers were optimized with HF and DFT /B3LYP method employing different basis sets in the 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[20] 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:

$$I = -E_{HOMO} \text{ and } A = -E_{LUMO} \quad (1)$$

When the values of I and A are known, one can determine through the following expressions[21] the values of the absolute electron negativity (χ), the absolute hardness (η),the chemical potential (μ) and the softness S[16] (the inverse of the hardness):

$$\chi = \frac{I+A}{2}; \ \eta = \frac{I-A}{2}; \quad \mu = -\frac{(I+A)}{2}; \quad s = \frac{1}{\eta} \quad (2)$$

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 has 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.[22]

$$\omega = \frac{\mu^2}{2\eta} \quad (3)$$

According to the definition, this index measures the propensity of chemical species to accept electrons. A good, more reactive, nucleophile is characterized by lower values 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 ΔNmax from the environment.[23]

$$\Delta N_{\max} = -\frac{\mu}{\eta} \quad (4)$$

The maximum charge transfer ΔNmax towards the electrophile was evaluated using Eq. (9). Thus, while the quantity defined by Eq.(8) describes the propensity of the system to acquire an additional electronic charge from the environment; the quantity defined in Eq. (9) describes the charge capacity of the molecule.

Very recently, Ayers and co-workers[24,25] have proposed two new reactivity indices to quantify nucleophilic and electrophilic capabilities of a leaving group, nucleofugality ΔE_n and electrofugality ΔE_e, defined as follows,

$$\Delta E_n = -A + \omega = \frac{(\mu+\eta)^2}{2\eta} \text{ and } \Delta E_e = I + \omega = \frac{(\mu-\eta)^2}{2\eta} \quad (5)$$

The global descriptors, chemical potential, chemical, hardness and chemical softness for all studied isomers are given in table 2 and 3. 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.

RESULTS AND DISCUSSION

HOMO-LUMO energies

The energies of the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) of the studied tautomers in the ground state.The HOMO and LUMO are calculated by using HF and DFT/B3LYP method with different basis sets and presented tn table 1.

Table 1: The theoretical electronic properties of HOMO and LUMO and energy gap (E_g) for studied tautomers by HF and DFt/B3LYP at different basis sets in the gas phase.

DFT method												
Tau tomer	6-311G			6-311G(d,p)			6-311G++			6-311G++(d,p)		
	HOMO	LUMO	ΔEg	HOMO	LUMO	ΔEg	HOMO	LUMO	ΔEg	HOMO	LUMO	ΔEg
K1B	-7.0799	0.4305	7.5104	-7.0434	0.7908	7.8342	-7.2834	-0.4512	6.8322	-7.2810	-0.4343	6.8467
E1B	-7.2747	0.2683	7.5430	-7.2241	0.6259	7.8500	-7.5033	-0.3578	7.1455	-7.4491	-0.3031	7.1460
E2B	-7.0750	0.2291	7.3041	-7.1231	0.4593	7.5824	-7.3144	-0.7682	6.5462	-7.3580	-0.6667	6.6913
K1G	-6.7043	0.4923	7.1966	-6.6967	0.7355	7.4322	-6.9068	-0.5083	6.3985	-6.9460	-0.4027	6.5433
E1G	-6.9716	0.3173	7.2889	-6.9424	0.4762	7.4186	-7.1683	-0.6985	6.4698	-7.1411	-0.6196	6.5215
E2G	-7.0701	0.4898	7.5599	-6.9857	0.7023	7.6880	-7.2486	-0.3396	6.9090	-7.1797	-0.3105	6.8692
K1D	-6.6105	0.3031	6.9136	-6.6377	0.4416	7.0793	-6.8107	-0.5538	6.2569	-6.8565	-0.5285	6.3280
E1D	-6.4793	0.2993	6.7786	-6.5822	0.4988	7.0810	-6.6584	-0.5970	6.0614	-6.7650	-0.5470	6.2180
E2D	-6.7609	0.3303	7.0912	-6.8263	0.5325	7.3588	-6.9348	-0.3880	6.5468	-6.9988	-0.3540	6.6448
K1E	-6.6129	0.3777	6.9906	-6.6363	0.6286	7.2649	-6.8004	-0.4966	6.3038	-6.8404	-0.4732	6.3672
E1E	-6.5196	0.2076	6.7272	-6.6431	0.3747	7.0178	-6.7079	-0.6656	6.0423	-6.8306	-0.6136	6.2170
E2E	-6.6663	0.2337	6.9000	-6.7577	0.4561	7.2138	-6.8423	-0.3540	6.4883	-6.9310	-0.3235	6.6075
HF method												
K1B	-10.7458	3.9182	14.6640	-10.8046	3.9851	14.7897	-10.8312	0.9113	11.7425	-10.9191	0.9391	11.8582
E1B	-10.6103	4.1285	14.7388	-10.5896	4.2406	14.8302	-10.7123	1.1293	11.8416	-10.7126	1.1535	11.8661
E2B	-10.5156	3.5375	14.0531	-10.5180	3.6515	14.1695	-10.6179	0.8381	11.4560	-10.6413	0.8838	11.5251
K1G	-10.4707	3.7424	14.2131	-10.5550	3.8480	14.4030	-10.5529	0.9086	11.4615	-10.6625	0.9410	11.6035
E1G	-10.3207	4.0101	14.3308	-10.2350	3.6784	13.9134	-10.4043	1.1197	11.5240	-10.3376	0.9265	11.2641
E2G	-10.2546	3.5631	13.8177	-10.3009	4.1304	14.4313	-10.3319	0.8925	11.2244	-10.3899	1.1366	11.5265
K1D	-10.3044	3.6836	13.9880	-10.4269	3.6599	14.0868	-10.3880	0.9037	11.2917	-10.5248	0.9385	11.4633
E1D	-10.2394	3.6654	13.9048	-10.2685	3.7737	14.0422	-10.3039	0.9045	11.2084	-10.3099	0.9350	11.2449
E2D	-10.3420	3.9299	14.2719	-10.3060	4.0333	14.3393	-10.4195	1.0727	11.4922	-10.4021	1.0991	11.5012
K1E	-10.2753	3.7976	14.0729	-10.3414	3.8765	14.2179	-10.3553	0.9385	11.2938	-10.4345	0.9671	11.4016
E1E	-10.1126	3.6722	13.7848	-10.0990	3.7590	13.8580	-10.1893	0.8961	11.0854	-10.1961	0.9238	11.1199
E2E	-10.2318	3.9642	14.1960	-10.1956	4.0665	14.2621	-10.3115	1.0735	11.3850	-10.2903	1.0977	11.3880

The HOMO and LUMO energies take part in chemical stability. The HOMO represents the ability to donate an electron, LUMO as an electron acceptor represent the ability to obtain an electron. The energy of the HOMO is directly related to the ionization potential and the energy of the LUMO is directly related to the electron affinity. The high value of HOMO energy is likely to indicate a tendency of the molecule to donate electrons to an appropriate acceptor molecule of low empty molecular orbital energy. The lower values of LUMO energy show more probable to accept electrons. So, the gap energy (Eg), i.e. the difference in energy between the HOMO and LUMO, is an important stability index. The smaller the LUMO and HOMO energy gaps, the easier it is for the HOMO electrons to be excited; the higher the HOMO energies, the easier it is for HOMO to donate electrons; the lower the LUMO energies, the easier it is for LUMO to accept electrons. The calculated HOMO level, LUMO level, and Eg are summarized in Table 1 and the HOMO and LUMO diagrams in ground state are represented by Fig (1).

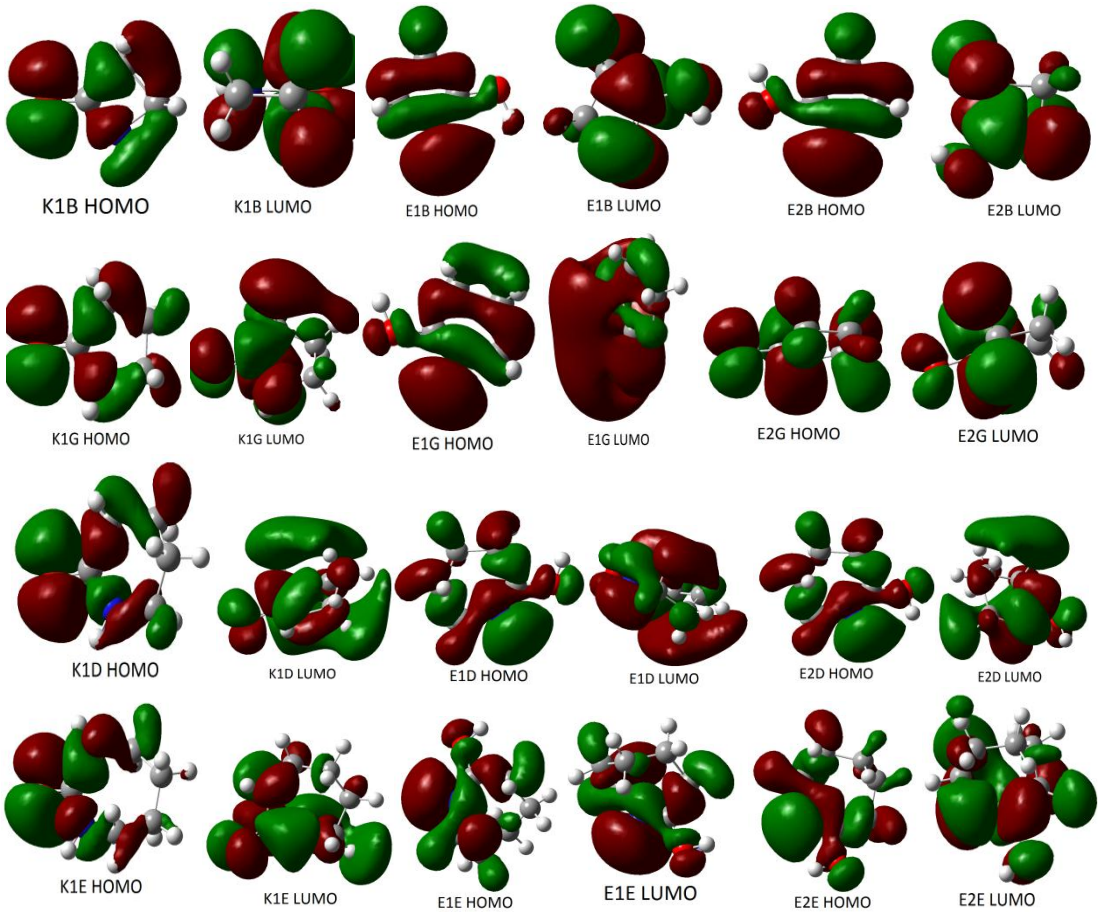


Figure 1: Optimized HOMO and LUMO structures of lactam tautomers by DFT method at B3LYP/6-311++G(d,p) basis set in gas phase.

The global and chemical reactivity descriptors, ionization potential (IP), electron affinity (EA), electronegativity (χ), hardness (η), softness (S), chemical potential (μ), electrophilicity index (ω), charge transfer (ΔNmax), nucleofugality (ΔEn) and electrofugality (ΔEe) of lactams tautomers calculated by HF and DFT/B3LYP at different basis sets in gas phase are listed in table 2 and 3. From table 2, the calculated values of global electrophilicity index ω show the nucleophility power of lactams tautomers . The obtained ω values for BIII exhibit a lower ω value comparing to others, one can expect better propensity of BIII to be involved in the reactions with electrophiles than for other tautomers. The values of the global descriptors (hardness, and electrophilicity) grow when going from acylamino form (A) to acylimino form and with the increase of stability for the three tautomeric forms. We can see in all the cases that the hardness value is higher for the amide form and lower in the case of imidol form.

Table 2: The theoretical reactive descriptors ionization potential (IP), electron affinity (EA), electronegativity (χ), hardness (η), softness (S), chemical potential (μ), electrophilicity index (ω), charge transfer (ΔNmax), nucleofugality (ΔEn) and electrofugality (ΔEe) of lactams tautomers calculated by HF at different basis sets in the gas phase.

6-311G												
	K1B	**E1B**	**E2B**	**K1G**	**E1G**	**E2G**	**K1D**	**E1D**	**E2D**	**K1E**	**E1E**	**E2E**
I	10.7458	10.6103	10.5156	10.4707	10.3207	10.2546	10.3044	10.2394	10.3420	10.2753	10.1126	10.2318
A	-3.9182	-4.1285	-3.5375	-3.7424	-4.0101	-3.5631	-3.6836	-3.6654	-3.9299	-3.7976	-3.6722	-3.9642
χ	3.4138	3.2409	3.4890	3.3641	3.1553	3.3458	3.3104	3.2870	3.2060	3.2388	3.2202	3.1338
η	7.3320	7.3694	7.0265	7.1065	7.1654	6.9088	6.9940	6.9524	7.1359	7.0365	6.8924	7.0980
μ	-3.4138	-3.2409	-3.4891	-3.3642	-3.1553	-3.3458	-3.3104	-3.2870	-3.2061	-3.2388	-3.2202	-3.1338
s	0.1364	0.1357	0.1423	0.1407	0.1396	0.1447	0.1430	0.1438	0.1401	0.1421	0.1451	0.1409
μ2	11.6541	10.5033	12.1734	11.3175	9.9559	11.1942	10.9588	10.8044	10.2787	10.4901	10.3697	9.8207
ω	0.7947	0.7126	0.8662	0.7963	0.6947	0.8101	0.7834	0.7770	0.7202	0.7454	0.7523	0.6918
Δ*N*	0.4656	0.4398	0.4966	0.4734	0.4403	0.4843	0.4733	0.4728	0.4493	0.4603	0.4672	0.4415
ΔEn	4.7129	4.8411	4.4037	4.5387	4.7049	4.3732	4.4670	4.4424	4.6501	4.5430	4.4244	4.6560
ΔEe	11.5405	11.3229	11.3818	11.2670	11.0155	11.0648	11.0879	11.0164	11.0622	11.0207	10.8648	10.9236
6-311G(d,p)												
	K1B	**E1B**	**E2B**	**K1G**	**E1G**	**E2G**	**K1D**	**E1D**	**E2D**	**K1E**	**E1E**	**E2E**
I	10.8046	10.5896	10.5180	10.5550	10.2350	10.3009	10.4269	10.2685	10.3061	10.3414	10.0990	10.1956
A	-3.9851	-4.2406	-3.6515	-3.8480	-3.6784	-4.1304	-3.6599	-3.7737	-4.0333	-3.8765	-3.7590	-4.0665
χ	3.4097	3.1745	3.4333	3.3535	3.2783	3.0852	3.3835	3.2474	3.1364	3.2324	3.1700	3.0645
η	7.3948	7.4151	7.0848	7.2015	6.9567	7.2156	7.0434	7.0211	7.1697	7.1090	6.9290	7.1310
μ	-3.4097	-3.1745	-3.4333	-3.3535	-3.2783	-3.0852	-3.3835	-3.2474	-3.1364	-3.2324	-3.1700	-3.0646
s	0.1352	0.1349	0.1411	0.1389	0.1437	0.1386	0.1420	0.1424	0.1395	0.1407	0.1443	0.1402
μ2	11.6262	10.0773	11.7873	11.2462	10.7472	9.5186	11.4478	10.5457	9.8369	10.4487	10.0489	9.3915
ω	0.7861	0.6795	0.8319	0.7808	0.7724	0.6596	0.8127	0.7510	0.6860	0.7349	0.7251	0.6585
Δ*N*	0.4611	0.4281	0.4846	0.4657	0.4712	0.4276	0.4804	0.4625	0.4375	0.4547	0.4575	0.4297
ΔEn	4.7712	4.9201	4.4834	4.6288	4.4509	4.7900	4.4726	4.5247	4.7193	4.6114	4.4841	4.7250
ΔEe	11.5907	11.2691	11.3499	11.3359	11.0075	10.9605	11.2395	11.0195	10.9921	11.0763	10.8241	10.8541
6-311G++												
	K1B	**E1B**	**E2B**	**K1G**	**E1G**	**E2G**	**K1D**	**E1D**	**E2D**	**K1E**	**E1E**	**E2E**
I	10.8312	10.7123	10.6179	10.5529	10.4043	10.3319	10.3880	10.3039	10.4195	10.3553	10.1893	10.3115
A	-0.9113	-1.1293	-0.8381	-0.9086	-1.1198	-0.8925	-0.9037	-0.9045	-1.0727	-0.9385	-0.8961	-1.0735
χ	4.9600	4.7915	4.8899	4.8221	4.6423	4.7197	4.7421	4.6997	4.6734	4.7084	4.6466	4.6190
η	5.8713	5.9208	5.7280	5.7307	5.7620	5.6122	5.6458	5.6042	5.7461	5.6469	5.5427	5.6925
μ	-4.9600	-4.7915	-4.8899	-4.8221	-4.6423	-4.7197	-4.7421	-4.6997	-4.6734	-4.7084	-4.6466	-4.6190
s	0.1703	0.1689	0.1746	0.1745	0.1736	0.1782	0.1771	0.1784	0.1740	0.1771	0.1804	0.1757
μ2	24.6012	22.9587	23.9110	23.2530	21.5506	22.2754	22.4878	22.0870	21.8409	22.1689	21.5911	21.3352

ω	2.0950	1.9388	2.0872	2.0288	1.8701	1.9845	1.9915	1.9706	1.9005	1.9629	1.9477	1.8740
Δ*N*	0.8448	0.8093	0.8537	0.8415	0.8057	0.8410	0.8399	0.8386	0.8133	0.8338	0.8383	0.8114
ΔEn	3.0064	3.0681	2.9253	2.9374	2.9898	2.8771	2.8952	2.8751	2.9732	2.9014	2.8438	2.9475
ΔEe	12.9263	12.6511	12.7051	12.5817	12.2743	12.3164	12.3795	12.2745	12.3200	12.3182	12.1370	12.1855
6-311G++(d,p)												
	K1B	**E1B**	**E2B**	**K1G**	**E1G**	**E2G**	**K1D**	**E1D**	**E2D**	**K1E**	**E1E**	**E2E**
I	10.9191	10.7126	10.6413	10.6625	10.3376	10.3899	10.5248	10.3099	10.4021	10.4345	10.1961	10.2903
A	-0.9391	-1.1535	-0.8838	-0.9410	-0.9266	-1.1366	-0.9385	-0.9350	-1.0991	-0.9671	-0.9238	-1.0977
χ	4.9900	4.7795	4.8787	4.8608	4.7055	4.6266	4.7932	4.6874	4.6515	4.7337	4.6361	4.5963
η	5.9291	5.9330	5.7626	5.8017	5.6321	5.7632	5.7317	5.6224	5.7506	5.7008	5.5600	5.6940
μ	-4.9900	-4.7796	-4.8787	-4.8608	-4.7055	-4.6266	-4.7932	-4.6874	-4.6515	-4.7337	-4.6361	-4.5963
s	0.1687	0.1685	0.1735	0.1724	0.1776	0.1735	0.1745	0.1779	0.1739	0.1754	0.1799	0.1756
μ2	24.9004	22.8441	23.8020	23.6271	22.1420	21.4056	22.9743	21.9721	21.6366	22.4079	21.4938	21.1258
ω	2.0998	1.9252	2.0652	2.0362	1.9657	1.8571	2.0042	1.9540	1.8813	1.9653	1.9329	1.8551
Δ*N*	0.8416	0.8056	0.8466	0.8378	0.8355	0.8028	0.8363	0.8337	0.8089	0.8304	0.8338	0.8072
ΔEn	3.0389	3.0786	2.9491	2.9772	2.8923	2.9937	2.9427	2.8890	2.9803	2.9324	2.8567	2.9528
ΔEe	13.0190	12.6377	12.7065	12.6987	12.3033	12.2469	12.5290	12.2638	12.2834	12.3998	12.1290	12.1454

Based on the obtained results the electrophilicity value is higher in the amide form and lower in the imidol form at large basis set. Global reactivity trend, based on ω, is given by K1B > E2B > K1G > K1D > E1G > K1E > E1D > E1E > E1B > E2D > E2G > E2E and E2B > E1G > E1E > K1B > K1D > E1 > E1B > K1E > K1G > E2G > E2D > E2E for HF and DFT methods at 6-311G++(d,p) respectively.

Table 3: The theoretical reactive descriptors ionization potential (IP), electron affinity (EA), electronegativity (χ), hardness (η), softness (S), chemical potential (μ), electrophilicity index (ω), charge transfer (ΔNmax), nucleofugality (ΔEn) and electrofugality (ΔEe) of lactams tautomers calculated by DFT/B3LYP at different basis sets in the gas phase.

6-311G												
	K1B	**E1B**	**E2B**	**K1G**	**E1G**	**E2G**	**K1D**	**E1D**	**E2D**	**K1E**	**E1E**	**E2E**
I	7.0799	7.2747	7.0750	6.7043	6.9716	7.0701	6.6105	6.4793	6.7609	6.6129	6.5196	6.6662
A	-0.4305	-0.2683	-0.2291	-0.4923	-0.3173	-0.4898	-0.3031	-0.2993	-0.3304	-0.3777	-0.2076	-0.2338
χ	3.3247	3.5032	3.4229	3.1060	3.3271	3.2901	3.1537	3.0900	3.2153	3.1176	3.1560	3.2163
η	3.7552	3.7715	3.6520	3.5983	3.6444	3.7799	3.4568	3.3893	3.5456	3.4953	3.3636	3.4500
μ	-3.3247	-3.5032	-3.4229	-3.1061	-3.3271	-3.2901	-3.1537	-3.0900	-3.2153	-3.1176	-3.1560	-3.2163
s	0.2663	0.2651	0.2738	0.2779	0.2744	0.2646	0.2893	0.2950	0.2820	0.2861	0.2973	0.2899
μ2	11.0536	12.2724	11.7164	9.6475	11.0699	10.8250	9.9456	9.5480	10.3382	9.7195	9.9602	10.3443
ω	1.4718	1.6270	1.6041	1.3406	1.5187	1.4319	1.4386	1.4086	1.4579	1.3904	1.4806	1.4992
Δ*N*	0.8854	0.9289	0.9373	0.8632	0.9129	0.8704	0.9123	0.9117	0.9068	0.8919	0.9383	0.9322
ΔEn	1.9023	1.8953	1.8332	1.8328	1.8360	1.9217	1.7417	1.7079	1.7882	1.7681	1.6882	1.7329
ΔEe	8.5516	8.9017	8.6791	8.0449	8.4903	8.5020	8.0490	7.8879	8.2188	8.0033	8.0002	8.1654
6-311G(d,p)												
	K1B	**E1B**	**E2B**	**K1G**	**E1G**	**E2G**	**K1D**	**E1D**	**E2D**	**K1E**	**E1E**	**E2E**

I	7.0434	7.2241	7.1231	6.6967	6.9424	6.9857	6.6377	6.5822	6.8263	6.6363	6.6431	6.7577
A	-0.7908	-0.6259	-0.4593	-0.7355	-0.4762	-0.7023	-0.4416	-0.4988	-0.5325	-0.6286	-0.3747	-0.4561
χ	3.1263	3.2991	3.3319	2.9806	3.2331	3.1417	3.0980	3.0417	3.1469	3.0039	3.1342	3.1508
η	3.9171	3.9250	3.7912	3.7161	3.7093	3.8440	3.5397	3.5405	3.6794	3.6325	3.5089	3.6069
μ	-3.1263	-3.2991	-3.3319	-2.9806	-3.2331	-3.1417	-3.0980	-3.0417	-3.1469	-3.0039	-3.1342	-3.1508
s	0.2553	0.2548	0.2638	0.2691	0.2696	0.2601	0.2825	0.2824	0.2718	0.2753	0.2850	0.2772
μ2	9.7739	10.8841	11.1016	8.8840	10.4531	9.8702	9.5977	9.2519	9.9027	9.0232	9.8233	9.9276
ω	1.2476	1.3865	1.4641	1.1953	1.4090	1.2838	1.3557	1.3066	1.3457	1.2420	1.3998	1.3762
Δ*N*	0.7981	0.8405	0.8788	0.8021	0.8716	0.8173	0.8752	0.8591	0.8553	0.8270	0.8932	0.8736
ΔEn	2.0384	2.0124	1.9234	1.9309	1.8852	1.9862	1.7974	1.8054	1.8782	1.8706	1.7745	1.8323
ΔEe	8.2910	8.6106	8.5872	7.8921	8.3515	8.2696	7.9934	7.8888	8.1720	7.8783	8.0429	8.1339
6-311G++												
	K1B	**E1B**	**E2B**	**K1G**	**E1G**	**E2G**	**K1D**	**E1D**	**E2D**	**K1E**	**E1E**	**E2E**
I	7.2834	7.5033	7.3144	6.9068	7.1683	7.2486	6.8107	6.6584	6.9348	6.8004	6.7079	6.8423
A	0.4512	0.3578	0.7682	0.5083	0.6985	0.3396	0.5538	0.5970	0.3880	0.4966	0.6656	0.3540
χ	3.8673	3.9306	4.0413	3.7076	3.9334	3.7941	3.6822	3.6277	3.6614	3.6485	3.6867	3.5982
η	3.4161	3.5727	3.2731	3.1992	3.2349	3.4545	3.1285	3.0307	3.2734	3.1519	3.0211	3.2441
μ	-3.8673	-3.9306	-4.0413	-3.7076	-3.9334	-3.7941	-3.6823	-3.6277	-3.6614	-3.6485	-3.6867	-3.5982
s	0.2927	0.2799	0.3055	0.3126	0.3091	0.2895	0.3196	0.3300	0.3055	0.3173	0.3310	0.3082
μ2	14.9559	15.4492	16.3321	13.7460	15.4717	14.3951	13.5589	13.1601	13.4061	13.3116	13.5920	12.9468
ω	2.1890	2.1621	2.4949	2.1483	2.3914	2.0835	2.1670	2.1712	2.0477	2.1117	2.2495	1.9954
Δ*N*	1.1321	1.1002	1.2347	1.1589	1.2159	1.0983	1.1770	1.1970	1.1185	1.1576	1.2203	1.1091
ΔEn	1.7379	1.8043	1.7267	1.6400	1.6929	1.7439	1.6133	1.5741	1.6597	1.6151	1.5839	1.6414
ΔEe	9.4724	9.6654	9.8093	9.0551	9.5597	9.3321	8.9777	8.8295	8.9826	8.9121	8.9574	8.8377
6-311G++(d,p)												
	K1B	**E1B**	**E2B**	**K1G**	**E1G**	**E2G**	**K1D**	**E1D**	**E2D**	**K1E**	**E1E**	**E2E**
I	7.2810	7.4491	7.3580	6.9460	7.1411	7.1797	6.8565	6.7650	6.9988	6.8404	6.8306	6.9310
A	0.4343	0.3031	0.6667	0.4027	0.6196	0.3105	0.5284	0.5469	0.3540	0.4732	0.6136	0.3235
χ	3.8576	3.8761	4.0123	3.6744	3.8803	3.7451	3.6925	3.6560	3.6764	3.6568	3.7221	3.6273
η	3.4233	3.5730	3.3456	3.2716	3.2607	3.4346	3.1640	3.1090	3.3224	3.1836	3.1085	3.3037
μ	-3.8576	-3.8761	-4.0123	-3.6744	-3.8804	-3.7451	-3.6925	-3.6560	-3.6764	-3.6568	-3.7221	-3.6273
s	0.2921	0.2799	0.2989	0.3057	0.3067	0.2912	0.3161	0.3216	0.3010	0.3141	0.3217	0.3027
μ2	14.8813	15.0244	16.0987	13.5009	15.0571	14.0258	13.6342	13.3663	13.5159	13.3722	13.8541	13.1572
ω	2.1735	2.1025	2.4059	2.0633	2.3088	2.0418	2.1546	2.1496	2.0341	2.1002	2.2284	1.9913
Δ*N*	1.1269	1.0848	1.1993	1.1231	1.1900	1.0904	1.1670	1.1759	1.1066	1.1486	1.1974	1.0979
ΔEn	1.7392	1.7994	1.7392	1.6606	1.6892	1.7313	1.6261	1.6026	1.6801	1.6270	1.6148	1.6677
ΔEe	9.4545	9.5516	9.7639	9.0093	9.4499	9.2216	9.0110	8.9146	9.0328	8.9406	9.0590	8.9223

Chemometrics Studies

Chemometrics is a scientific discipline between measurement oriented chemistry and applied statistics. Analytical chemistry is the most significant discipline where chemometrics plays an important role[26,27]. Chemometrics uses mathematical and statistical methods to choose optimal procedures and experiments and to provide chemical information by analyzing chemical data.[28,29] In many fields of chemistry there is a need for solving practical oriented problems. Chemists usually can not measure directly the parameters they want to determine, but they are measuring signals on the instruments which then should be calculated into individual parameters, like concentration of components in samples, or different physical parameters.

There is also possibility to use chemometrics for determining different properties of individual samples in biology, food chemistry, forensic chemistry, geochemistry, archaeology etc.[30-32] It has to be stressed that measurements should be of good quality, accurate and reliable, as this is necessary for obtaining high quality information from multivariate data. There is also possibility to use chemometrics for determining different properties of individual samples in biology, food chemistry, forensic chemistry, geochemistry, archaeology etc. It has to be stressed that measurements should be of good quality, accurate and reliable, as this is necessary for obtaining high quality information from multivariate data.

Initial analysis

From table 1 the energy gap between HOMO and LUMO (ΔEg) at different basis sets 6-311G, 6-311G (d,p), 6-311G ++ and 6-311G ++ (d,p) for 12studied tautomers of lactams. Since we have more than one variable, it is possible to calculate a product–moment (Pearson) correlation coefficient for each pair of variables. These are summarized in the correlation matrix in Table 4, obtained using Minitab. The results show that, the correlation coefficient for the ΔEg at 6-311G and 6-331G ++ is 0.972 for HF method. The correlation coefficient for the ΔEg at 6-311G ++ and 6-311G ++ (d,p) is 0.989. The relationships between pairs of variables can be illustrated by a draftsman plot as shown in Figure 2 and 3. This gives scatter diagrams for each pair of variables. Both the correlation matrix and the scatter diagrams indicate that there is some correlation between some of the pairs of variables.

Table 4: The correlation matrix for the energy gap between HOMO and LUMO (ΔEg) at different basis sets.

HF method				DFT method			
	6-311G	6-311G(d,p)	6-311G++		6-311G	6-311G(d,p)	6-311G++
6-311G(d,p)	0.731			6-311G(d,p)	0.971		
6-311G++	0.972	0.758		6-311G++	0.913	0.929	
6-311G++(d,p)	0.757	0.951	0.822	6-311G++(d,p)	0.892	0.929	0.989

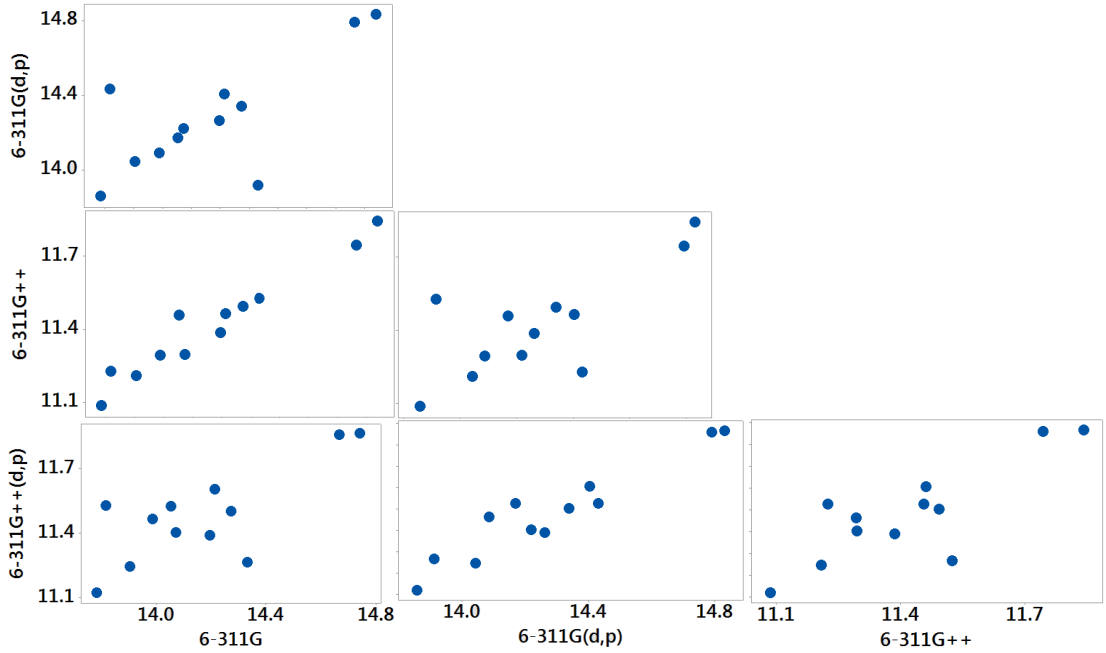
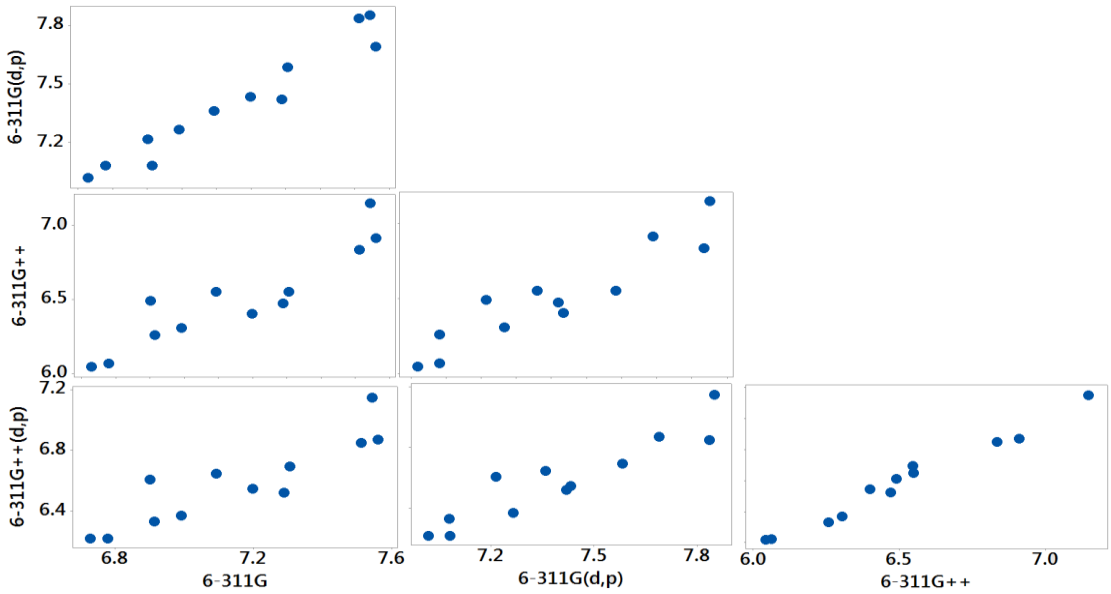


Figure 2: Draftsman plot for the energy gap between HOMO and LUMO (ΔEg) at different basis sets by HF method.



Figutre 3: Draftsman plot for the energy gap between HOMO and LUMO (ΔEg) at different basis sets by DFT method.

Principal component analysis

Principal component analysis PCA[33-34] is by far the most important method in multivariate data analysis. Principal component analysis PCA is also the most frequently used concept for defining a latent variable. It tries to represent the multidimensional data structure optimally. Direction which best describes the relative distance between the objects is the direction with maximum variance. This direction is called first principal component PC1. The second principal component (PC2) is orthogonal to PC1 and has the second maximum possible variance. PCA can transform data from multidimensional space into two dimensions without losing considerable amount of information. PCA estimates the correlation structure of the variables and hence the possible structure of latent variables.

In PCA the numerical data should be organized into matrix X (measured/ calculated data) with *n* rows (attributed to different compounds or systems, "samples") and *m* columns (referred to different experiments/calculations, or different properties; each column representing one "variable" or "parameter"). Because of the lack of optimal data distribution (different units and variances) in X, some preprocessing operation is required, as auto scaling (the scaled variables have zero mean and unity variance). Preprocessed X is decomposed in PCA as:

$$X = TV^T$$

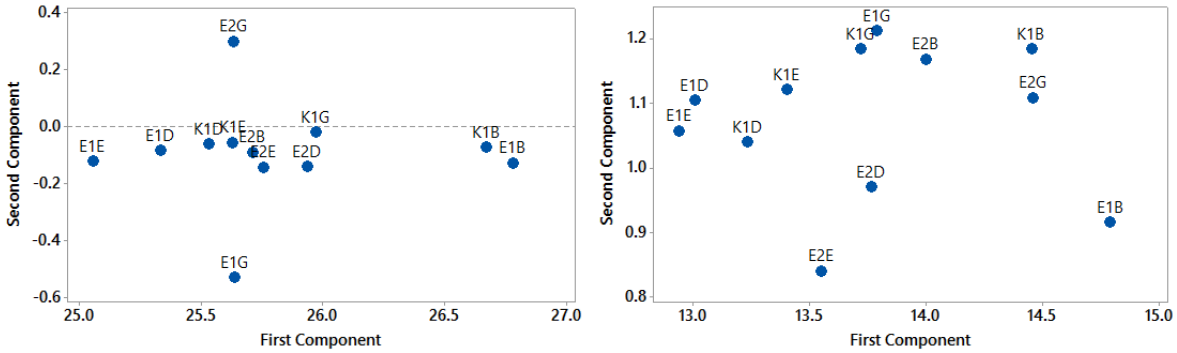
The matrix T ("scores" matrix) represents the position of samples in the new orthogonal coordinate system with the principal components (PCs) as the axes. The matrix V ("loadings" matrix) has columns that describe how the PCs are built from the old axes. Thus, a new set of completely uncorrelated variables, PCs, is generated. The first principal component, PC1, is defined in the direction of maximum variance in the data set, and PC2 is in the direction of next greatest variation, and so on. It is often found that PC1 and PC2 then account between them for most of the variation in the data set. As a result the data can be represented in only two dimensions instead of the original n. The principal components are obtained from the covariance matrix. In mathematical terms the principal components are the eigenvectors of the covariance matrix and the technique for finding these eigenvectors is called eigenanalysis. Corresponding to each principal component (i.e. eigenvector) is an eigenvalue which gives the amount of variance in the data set which is explained by that principal component.

Table 5: The Eigenanalysis of the Covariance Matrix for HOMO and LUMO (ΔEg) at different basis sets by DFT method.

	HF method				DFT method			
Eigenvalue	0.24631	0.03278	0.00387	0.00067	0.34270	0.01347	0.00268	0.00057
Proportion	0.868	0.116	0.014	0.002	0.953	0.037	0.007	0.002
Cumulative	0.868	0.984	0.998	1.000	0.953	0.991	0.998	1.000
Eigenvectors								
Variable	**PC1**	**PC2**	**PC3**	**PC4**	**PC1**	**PC2**	**PC3**	**PC4**
6-311G	0.564	-0.615	-0.334	0.438	0.488	0.617	-0.535	-0.308
6-311G(d,p)	0.570	0.611	-0.488	-0.251	0.482	0.398	0.709	0.327
6-311g++	0.417	-0.359	0.412	-0.726	0.557	-0.479	-0.372	0.568
6-311G++(d,p)	0.427	0.347	0.693	0.467	0.469	-0.481	0.269	-0.690

The sum of the variances of the original variables can be calculated from the covariance matrix. It is equal to 0.24631 + 0.03278 + 0.00387 + 0.00067 = 0.28363 at HF and 0.34270 + 0.01347 + 0.00268 + 0.00057 = 0.35942. It can be seen that the variances of the variables are different. The first line of the table shows how the total variance is shared between the four principal components and the proportion shows each variance as a proportion of the total. Thus PC1 has a variance of 0.24631, which is 86.8% of the total in HF and PC1 has a variance of 0.34270, which is 95.3% of the total in DFT method. This is a much greater proportion than any of the original variables. PC2 accounts for 11.6% of the total variance in HF and 3.7% in DFT method. The cumulative proportion shows, that between them, PC1 and PC2 account for 98.4% and 99.1% of the variation in HF and DFT methods respectively.

The eigenvectors give the coefficients of the principal components. The first principal component is Z1 = 0.564X1 + 0.570X2 + 0.417X3 – 0.427 X4, where X1, X2, X3 and X4 are the basis sets at 6-311G, 6-311G (d,p), 6-311G ++ and 6-311G ++ (D,P) respectively for HF method and similarly to calculate for DFT method. The values that the principal components take for each of the compounds can be calculated by substituting the relevant values of X1, X2, X3 and X4 into this formula. This value is sometimes referred to as the 'score' of PC1 for individual tautomers. Figure 4 plots the scores of the first two principal components, calculated in this way, for the all tautomers. This diagram reveals that the compounds fall into two distinct groups, a fact which is not readily apparent from the original data.



Figurer 4: The scores of the first two principal components for studying molecular descriptors at 6-311G++ (d, p) basis set. Left side figure is HF and a right side figure is DFT methods.

Cluster analysis (CA)

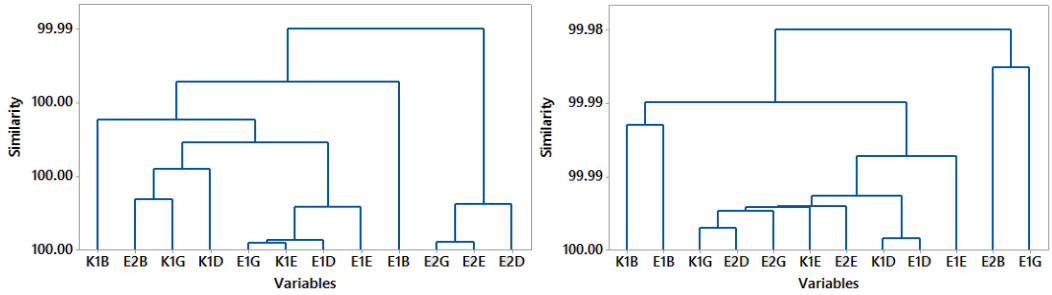
Clustering techniques have been applied to a wide variety of research problems, usually for classification or characterization. Cluster analysis belongs to exploratory data analysis. It allows grouping of samples on the basis of their similarities or differences. Cluster analysis uses all the variance or information contained in original data set. The starting point is a distant matrix containing the distances between all possible pairs of objects. The pairs with the smallest distance are merged and thus number of object is reduced. Result is a hierarchical tree, named "dendrogram".It displays the data in 2D space, qualitatively, in a form of dendograms 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

$$S_{kl} = 1.000 - \frac{d_{kl}}{d_{\max}}$$

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*. The similarity scale ranges from 0 to 1.

HCA RESULTS

Hierarchical cluster analysis (HCA) was used in this work as it groups the compounds based on their similarity degree. In this technique, each compound is initially assumed to be alone cluster and one similarity matrix is built, generally calculating the Euclidean distance between all of the objects. Then, the compounds are clustered together and treated as a single cluster and successive iterations lead to the total clustering of all compounds according to their similarity level, generating a dendrogram Figure 5 shows our results obtained with the HCA analysis. 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 5, we can see that the HCA results are very similar to those obtained with the PCA analysis. In Figure 5, this analysis suggests that the tautomers fall into two distinct groups. Not surprisingly, the groups contain the same members as they did with PCA.



Figurer 5: Dendrogram for studying molecular descriptors at 6-311G++ (d, p) basis set. Left side figure is HF and a right side figure is DFT methods.

CONCLUSION

The tautomerism, of different ring size lactams were theoretically investigated with HF and DFT/B3LYP methods employing at different basis sets. The energies of the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) of the studied tautomers in the ground state. Global reactivity trend, based on ω, is given by K1B > E2B > K1G > K1D > E1G > K1E > E1D > E1E > E1B > E2D > E2G > E2E and E2B > E1G > E1E > K1B > K1D > E1 > E1B > K1E > K1G > E2G > E2D > E2E for HF and DFT methods at 6-311G++(d,p) respectively.From PCA results, the tautomers fall into two distinct groups, a fact which is not readily apparent from the original data. From the HCA results are very similar to those obtained by the PCA analysis, i.e. the compounds studied were grouped into two categories.

REFERENCES

1. Hansch, C., Leo, A., Taft and R.W. Chem. Rev, 1991; 91: 165.
2. Gao,H., Katzenellenbogen,J.A., Garg,V.and Hansch, C. Chem. Rev, 1999; 99: 723.
3. Franke, R. Theoretical Drug Design Methods; Elsevier: Amsterdam, New York, 1984: 115.
4. Gupta, S.P., Singh, P. and Bindal, M.C. Chem. Rev, 1983; 83: 633.
5. Gupta, S.P. Chem. Rev, 1991; 91: 1109.
6. Karelson, M. Lobanov, V.S. and Katritzky, A. R. Chem. Rev, 1996; 96: 1027.
7. Parr, R.G,. Annu. Rev. Phys. Chem, 1983; 34: 631.
8. Parr, R.G.and Yang, W. Density Functional Theory of Atoms and Molecules; Oxford University Press: Oxford, UK, 1989.
9. Chermette, H.J. Comput. Chem, 1999; 20: 129.
10. Geerlings, P., De Proft, F.and Langenaeker, W. Chem. Rev, 2003; 103: 1793.
11. Chattaraj, P.K., Nath, S. and Maiti, B. Reactivity Descriptors. In Computational Medicinal Chemistry for Drug Discovery; J Tollenaere; P Bultinck; HD Winter; W Langenaeker. Eds.; MarcelDekker: New York, 2003; 295–322.
12. Parr, R.G., Szentpaly, L.V. and Liu, S.J. Am. Chem. Soc, 1999; 121: 1922.
13. Maynard, A.T., Huang, M., Rice, W.G. and Covell, D.G. Proc. Natl. Acad. Sci. U.S.A, 1998; 95: 11578.
14. Frisch, M.J, Trucks, G.W., Schlegel, H.B., Gill, P.M.W., Johnson, B.G. and Robb, M.A. et al., GAUSSIAN 09, Revision, Gaussian, Inc., Wallingford CT, 2009.
15. Becke, A.D. Phys. Rev, 1988; A38: 3098.
16. Becke, A.D. J. Chem. Phys, 1993; 98: 5648.
17. Lee, C., Yang, W. and Paar, R.G. Phys. Rev, 1988; B3: 785.
18. Parr, R.G. and Wang, W. Density-Functional Theory of Atoms and Molecules; Oxford University Press: New York, 1994.
19. Neumann, R., Nobes, R.H. and Handy, N.C. Mol. Phys, 1996; 87: 1.
20. Koopmans, T. Ordering of wave functions and eigenergies to the individual electrons of an atom. Physica, 1993; 1: 104.
21. Pearson, R.G. J.Am.Chem. Soc, 1963; 85(22): 3533.
22. Lesar, A. and Milosev, I. Chem. Phys. Lett, 2009; 483(4-6): 198.
23. Parr, R.G., Szentpaly, L., Liu, S. J. Am. Chem. Soc, 1999; 121: 1922.
24. Ayers, P. W., Anderson, J. S. M. and Bartolotti, L. J. Int. J. Quantum Chem, 2005; 101: 520.
25. Roos, G., Loverix, S., Brosens, E., Van Belle, K., Wyns, L., Geerlings, P., Messens, J. ChemBioChem, 2006; 7: 981.
26. Adams, M.J. Chemometrics in Analytical Spectroscopy. The Royal Society of Chemistry, Cambridge, 1990.
27. Wilkins, C.L. Computer- Enhanced Analytical Spectroscopy. The Royal society of chemistry, Cambridge, 1990.
28. Willet, P. Similarity and Clustering in Chemical Information. Wiley, New York, 1987.
29. Nilsson, N.J. Linear Learning Machines. Mc.Graw-Hill, NewYork, 1965.
30. Lankmayr, E., Mocak, J., Šnuderl, K., Balla, B., Wenzl, T., Bandoinene, D., Gfrerer, M.and Wagner, S. J. Biochem. Biophys.Methods, 2004; 61: 95.
31. Šnuderl, K., Netriová, J., Mocak, J., Lehotay, J. and Brodnjakvončina, D.Fac. Chem.Technol., 2005; 11: 315.
32. Brodnjak-Vončina, D., Cencič-Kodba, Z.; Novič, M. and Chemometr. Intell. Lab. Syst., 2005; 75: 31.
33. Malinowski, E.R. Factor Analysis in Chemistry, 2nd edition; Wiley: New York, 1991.
34. Beebe, K.R., Pell, R.J., Seasholtz, M.B. Chemometrics: A Practical Guide; Wiley: New York, 1998.