# Studies of Global and Reactivity descriptors of Cyanuric acid tautomers in different solvents by using Chemometric Methods 

N.Surendra Babu ${ }^{1}$ and Jayaparkesh ${ }^{2}$<br>${ }^{1}$ Department of Chemistry, Hawassa University, Hawassa, Post Box No: 5, Ethiopia.<br>${ }^{2}$ Department of Chemistry, Acharya Nagarjuna University, Nagarjuna Nagar, Guntur, A. P., India.

## ARTICLE INFO

## Article history:

Received: 17 May 2015;
Received in revised form:
20 June 2015;
Accepted: 25 June 2015;

## Keywords

Chemometrics,
THF,
DMF,
Tautomers.


#### Abstract

Chemometrics is the application of statistical and mathematical methods, in particular multivariate methods, to handle chemical or process data. In this study, the global and reactivity descriptors with multivariate methods have been utilized as a potential tool for grouping of cyanuric acid. The global and reactivity descriptors of cyanuric acid tautomers were analyzed using Principal component analysis (PCA) and hierarchical cluster analysis (HCA) in different solvents. The tautomers constitute groups are similar in three solvents namely THF, DMF and water, but in ethanol solvent show different groups. In three solvents (THF,DMF and water), five tautomers, CA2, CA4, CA5, CA6 and CA7 constitute one group, but CA4, CA5, CA6 and CA7 unique well distinguished from the rest. The CA8, CA9 and CA1 tautomers constitute second group and CA10, CA11 and CA3 constitute a cluster of three tautomer types. In ethanol solvent, three tautomers, CA5, CA6 and CA7 constitute one group, the CA8, CA9 and CA1 tautomers constitute second group and CA10 and CA11 constitute a third group and CA2,CA3 and CA4 constitute a fourth group. The HCA results are very similar to those obtained with the PCA results.


© 2015 Elixir All rights reserved.

## Introduction

Chemometrics refers to the application of statistical and mathematical methods, in particular multivariate methods, to handle chemical or process data. The need for chemometrics methods originates from the massive amounts of data produced by modern measuring devices [1,2]. Chemometrics tends to deal with data tables or matrices consisting of several variables (columns of tables or matrices) and measurement targets (rows or tables or matrices) as a whole rather than as single variables or means or variations of single variables [3]. This multivariate approach enables finding the so-called latent variables or information of interrelated variables in the original data matrix which can then be extracted. The latent variable models are based on the assumption that the original data base dimensionality is not a full rank [4]. The new latent variables are projections of the original variables on multivariate space. Thus, even the 100 dimensional variable space can be reduced into a subspace consisting of a few latent variables that describes underlying phenomena [5] such that the originally 100 dimensional space can be visualized. There are several advantages of using multivariate methods over univariate techniques [5] such as robust modelling, noise removal, handling of interacting variables or overlapping spectral profiles, outlier or fault detection [6,4], variable reduction and understanding the reasons for similarity or dissimilarity of measurements (interpretation plus causality).

Chemometricians have adopted methods from other research fields such as econometrics and psychometrics where bilinear partial least squares and multi-way methods, respectively, have been applied and refined [1]. Chemometric methods have been widely applied in the food, biosciences, petroleum, oil and nowadays pharmaceutical industries, and it is continuing to diverge into new fields such as metabonomics.

## Principal component analysis (PCA)

Principal component analysis (PCA), is a linear projection method and used for reduction of dimensionality and multivariate data compression. The idea of PCA dates back in 19th century and was named by Hotelling in 1933 [7,8]. At that time, mathematicians explored multivariate data by fitting it onto lines and planes [7]. Today, PCA is one of the vast utilized multivariate method since its wide applicability for multivariate problems. PCA is deployed for data compression [9] and data exploring within different fields of science. PCA is also used for checking groupings of the $\mathbf{X}$ data, as well as grouping among the $\mathbf{Y}$ data matrix [10,11]. In process monitoring, PCA is used to detect trends, to find a correlation structure of variables and, in particular, to examine the changes in variable correlations [11, 12]. It should be noted that PCA is feasible for variable reduction if variables are correlated and thus contain a similar variance.

## Properties of PCA

Principal components are so-called latent variables that are weighted linear combinations of the original data matrix. A special feature of a latent variable is that it cannot be measured directly; instead it consists of a linear combination of measurables, i.e. manifest variables [13]. The components are intended to capture the systematic structure of data and not to describe noise (non-systematic part). The principal components are based on the variance of original data matrix, and are extracted by different approaches, such as eigenvalue or singular value decomposition or in a sequential manner by using a noniterative partial least squares (NIPALS) algorithm. It has been proposed that NIPALS is preferable when the number of $x$ variables is large [14]. However, the commonality for all methods is that they find new sets of coordinate axis of the original data matrix $\mathbf{X}(\mathrm{I} \times \mathrm{J})$ with many objects (I) and variables (J) that are believed to be correlated and arranges them to
orthogonal directions where variance of the data is maximized. Thus, the PC space is the subspace of the original data space $\mathbf{X}$ and spans $\mathbf{X}$ in lower dimensions. The matrix notation for PCA is presented as

$$
\begin{equation*}
\mathbf{X}=\mathrm{TP}^{T}+\mathrm{E}_{F} \tag{1}
\end{equation*}
$$

where $\mathbf{T}(\mathrm{IxF})$ denotes score matrix, $\mathbf{P}(\mathrm{JxF})$ loadings matrix and $\mathbf{E}_{\mathrm{F}}(\mathrm{I} \times \mathrm{J})$ residual matrix after F components. Eq. 1 can be written as vector outer product, respectively

$$
\begin{equation*}
\mathrm{X}=t_{i} p_{i}^{T}+\ldots .+t_{F} p_{F}^{T}+E_{F}=\sum_{i=1}^{F} t_{i} p_{i}^{T}+E_{F} \tag{2}
\end{equation*}
$$

where $\mathrm{i}=1, \ldots, \mathrm{~F}$ and F is the number of latent components $(\mathrm{F} \leq$ I).

The first PC explains the largest part of the variance of the data corresponding to the largest eigenvalue of the eigenvector of the mean centered $\mathbf{X}^{\mathrm{T}} \mathbf{X}$ covariance matrix. The next component comprises the maximal variance of the residual data matrix of the first component that corresponds to the second largest eigenvalue, thus the direction of second largest variance. The variance explained by a subsequent principal component decreases with increasing order of PC. Since the basic concept of PCA is that data matrix with many variables is not a full rank and holds a latent structure that could be explained by a few latent variables, only a small number of the principal components is needed to explain the maximum variance of the original data. In the ideal case, the rest of the data contains redundant data, i.e., noise and error due to the measurement conditions.

## Clustering Methods

Clustering is a data analysis technique that, when applied to a set of heterogeneous items, identifies homogeneous subgroups as defined by a given model or measure of similarity. Of the many uses of clustering, a prime motivation for the increasing interest in clustering methods is their use in the selection and design of combinatorial libraries of chemical structures pertinent to pharmaceutical discovery.

Clustering methodology has been developed and used in a variety of areas including archaeology, astronomy, biology, computer science, electronics, engineering, information science, and medicine. Good, general introductory texts on the topic of clustering include those by Sneath and Sokal[15], Kaufmann and Rousseeuw[16], Everitt[17], and Gordon[18]. The main text that is devoted to clustering of chemical data sets is by Willett[19], with review articles by Bratchell[20], Barnard and Downs[21], and Downs and Willett[22].

The methods must be able to handle large data sets of highdimensional data. For small, low-dimensional data sets, most clustering methods are applicable, and descriptions in the standard texts and implementations available in standard statistical software packages [23,24[ suffice. Implementations designed for use on chemical data sets are available from most of the specialist software vendors, [25-30] the majority of which were reviewed by Warr [31].

The overall process of clustering involves the following steps:

1. Generate appropriate descriptors for each compound in the data set.
2. Select an appropriate similarity measure.
3. Use an appropriate clustering method to cluster the data set.
4. Analyze the results.

To address this problem, many numerical clustering techniques have been developed, and the techniques themselves
have been classified. For our purposes the methods considered belong to one of the following types.
(a) Hierarchical techniques in which the elements or objects are clustered to form new representative objects, with the process being repeated at different levels to produce a tree structure, the dendrogram.
(b) Methods employing optimization of the partitioning between clusters using some type of iterative algorithm, until some predefined minimum change in the groups is produced.
(c) Fuzzy cluster analysis in which objects are assigned a membership function indicating their degree of belonging to a particular group or set.

In this study, multivariate chemometric techniques have been applied in evaluating grouping operations in tautomers.

## Computational and Chemometric methods

Molecular geometries of tautomeric forms of Cyanuric acid were fully optimized by using the Gaussian quantum chemistry software package Gaussian 09 w [32] at DFT/B3LYP level of theory, using the $6-311++G(d, p)$ basis set. Following the geometry optimizations, analytical frequency calculations were preceded following the standard procedures, to obtain the thermo chemical properties. In addition the effects of solvents on the tautomeric structure properties were studied by means of the self-consistent reaction-field (SCRF) method based on PCM developed by Tomasi and coworkers [33] , it is one of the most widely used approaches. In this model, a solute is considered inside a cavity and the solvent as a structureless medium characterized by some parameters such as its dielectric constant, molar volume and polarizability. This consideration can substantially improve the simulation results for the electronic or vibrational spectroscopy of real molecular systems [34,35].The solvents chose for this studies are polar protic solvents namely water $(\varepsilon=74.80)$ and ethanol $(\varepsilon=24.55)$ and polar aprotic solvents like tetrahydrofuran(THF) $(\varepsilon=7.50)$ and dimethylformamide $(\varepsilon=38.00)$.

In our previous research paper [36] we calculated the HOMO and LUMO energies in order to determine, the usefulness of global reactivity descriptors namely, the electrophilicity, Chemical hardness ( $\eta$ ), chemical potential ( $\mu$ ), polarizability $(\alpha)$ electrophilicity index $(\omega)$, softness (S), nucleofugality, and electrofugality, values for the prediction of the reactivity of the cyanuric acid tautomers. In present work, we have been applied the chemometrics methods for the data of global reactivity descriptors. The correlation between the molecular properties calculated and the stability and reactivity studied was done by using the pattern recognition methods (PCA and HCA) built in the statistical package Ky-plot software with bivariate comparisons. P value below 0.05 was considered as statistically significant.

## Results and Discussion

Cyanuric acid shows keto-enol or more precisely aminoketo-iminoalcohol tautomerism. A prominent example is the isomerism between isocyanuric acid (1a) and cyanuric acid (1b) and may exist in several different tautomeric forms.


1a
Isocyanuric acid


1b
Cyanuric acid

Isocyanuric acid 1a has ten tautomeric forms: two mono hydroxy, six di-hydroxy and two tri-hydroxy isomers, these are differing in the mono proton, di proton and tri proton transfer and orientation of the hydroxyl groups. The isomers are labelled, CA1: Cyanuric acid (1,3,5-triazinane-2,4,6-trione), CA2 and CA3: 6-hydroxy-1,3,5- triazine-2,4(1H,3H)- dione , CA4, CA5, CA6, CA7, CA8 and CA9: 4,6-dihydroxy-1,3,5-triazin-2(1H)one and CA10 and CA11: 1,3,5-triazine-2,4,6-triol(scheme1).

The global and chemical reactivity descriptors [25], hardness ( $\eta$ ), chemical potential ( $\mu$ ), softness (S), electronegativity $(\chi)$ and electrophilicity index ( $\omega$ ) were calculated from HOMO and LUMO energies and incorporated in Table 1. in gas phase and in different solvents.


1,3,5-triazinane-2,46. trione(CAl)

24(1H3 H )-dions(CA2)

 $2(1 H)-\mathrm{ons}(\mathrm{CAS})$





Scheme 1. Diffrernt tautomeric forms of Cyanuric acid Scores and loadings

Principal components consist of scores and loadings as shown in Eqs. 1 and 2. Most commonly these vectors are plotted because score (Fig. 1) and loading (Fig. 2) plots visualize original observations (samples) and variables in new coordinate systems. The loading values depict how the original variables are weighted in order to comprise the new axis whereas the sample scores shows their position in a new coordinate system. These two plots (Figs. 1 and 2) are interactive, and thus reasoning for e.g. clustering of the samples or presence of outliers can be assessed.

We are now in a position to return to the complete set of global descriptors data in Table 1 and apply principal components analysis to the full data matrix. The techniques described and used in the above example to extract and determine the eigenvalues and eigenvectors for two variables can be extended to the more general, multivariate case but the procedure becomes increasingly difficult and arithmetically tedious with large matrices. Instead, the eigenvalues are usually found by matrix manipulation and iterative approximation methods using appropriate computer software. Before such an
analysis is undertaken, the question of whether to transform the original data should be considered.

Examination of Table 1 indicates that the variates considered have widely differing means and standard deviations. Rather than standardizing the data, since they are all recorded in the same units, one other useful transformation is to take logarithms of the values. Having performed the our data, the results of performing PCA on all 13 for the 11 tautomers are as given in Table 2.

According to the eigenvalue results present in Table 2, over $98 \%$ of the total variance in the original data can be accounted for by the first two principal components. The transformation of the 13 original variables to two new linear combinations represents considerable reduction of the data presented whilst retaining much of the original information. A scatter plot of the first two principal components scores is shown in Figure 14 and patterns to the samples according to the distribution of the tautomers in the data are evident.

Three tautomers, CA1, CA10 and CA11 constitute unique group, well distinguished from the rest. The CA8, CA4 and CA5 valves constitute a cluster of four tautomer types and CA2 and CA3 constitute a cluster of two tautomer types finally, there is a group of two tautomers CA7 and CA6 constitute a cluster in gas phase.



Fig 1. Score plots for the global descriptors of cyanuric acid tautomers in different solvents at B3LYP/6-311++G(d,p) level of theory
The tautomers constituting groups are similar in three solvents namely THF, DMF and water, but in ethanol solvent show different groups. In three solvents (THF,DMF and water), five tautomers, CA2, CA4, CA5, CA6 and CA7 constitute one group, but CA4, CA5, CA6 and CA7 unique well distinguished from the rest. The CA8, CA9 and CA1 tautomers constitute the second group and CA10, CA11 and CA3 constitute a third group . In ethanol solvent, three tautomers, CA5, CA6 and CA7 constitute one group, the CA8, CA9 and CA1 tautomers constitute group and CA10 and CA11 constitute a cluster of two tautomer types and CA2,CA3 and CA4 constitute a group.



Fig 2. Loading plots for the global descriptors of cyanuric acid tautomers in different solvents at B3LYP/6$311++G(d, p)$ level of theory.

## Hierarchical Techniques

When employing hierarchical clustering techniques, the original data are separated into a few general classes, each of which is further subdivided into still smaller groups until finally the individual objects themselves remain. Such methods may be agglomerative or divisive. By agglomerative clustering, small groups, starting with individual samples, are fused to produce larger groups. In contrast, divisive clustering starts with a single cluster, containing all samples, which is successively divided into smaller partitions. Hierarchical techniques are very popular, not least because their application leads to the production of a dendrogram which can provide a two-dimensional pictorial representation of the clustering process and the results. Agglomerative hierarchical clustering is very common and we will proceed with details of its application.

The entire process involved in undertaking agglomerative clustering using distance measures can be summarized by a fourstep algorithm.

Step 1. Calculation of the between-object distance matrix.
Step 2. Find the smallest elements in the distance matrix and join the corresponding objects into a single cluster.

Step 3. Calculate a new distance matrix, taking into account that clusters produced in the second step will have formed new objects and taken the place of original data points.

Table 1. The theoretical electronic properties (HOMO, LUMO) and energy gap (Eg) and reactive descriptors ionization potential (IP), electron affinity (EA), electronegativity ( $\alpha$ ), hardness ( $\boldsymbol{\eta}$ ), softness (S), chemical potential ( $\boldsymbol{\mu}$ ), electrophilicity index $(\omega)$, charge transfer ( $\Delta \mathbf{N}_{\text {max }}$ ), nucleofugality $\left(\Delta E_{n}\right)$ and electrofugality ( $\Delta \mathrm{E}_{\mathrm{e}}$ ) of cyanuric acid tautomers calculated by B3LYP/6-311++ G(d,p) in gas phase and different solvents

| Gas |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Tautomers | HOMO | LUMO | $\Delta \mathrm{E}_{\mathrm{g}}$ | I | A | $\chi$ | $\eta$ | S | $\mu$ | $\omega$ | $\Delta \mathrm{N}_{\text {max }}$ | $\Delta \mathrm{E}_{\mathrm{n}}$ | $\Delta \mathrm{E}_{\mathrm{e}}$ |
| CA1 | -8.3242 | -1.1021 | 7.2221 | 8.3228 | 1.1029 | 4.7128 | 3.6099 | 0.2770 | -4.7128 | 3.0764 | 1.3055 | 1.9735 | 11.3991 |
| CA2 | -8.0738 | -1.5184 | 6.5554 | 8.0749 | 1.5182 | 4.7965 | 3.2784 | 0.3050 | -4.7965 | 3.5089 | 1.4631 | 1.9907 | 11.5837 |
| CA3 | -8.0874 | -1.4694 | 6.6180 | 8.0860 | 1.4703 | 4.7782 | 3.3079 | 0.3023 | -4.7782 | 3.4510 | 1.4445 | 1.9807 | 11.5370 |
| CA4 | -7.8262 | -1.6354 | 6.1907 | 7.8248 | 1.6365 | 4.7307 | 3.0941 | 0.3232 | -4.7307 | 3.6164 | 1.5289 | 1.9799 | 11.4412 |
| CA5 | -7.7990 | -1.6218 | 6.1771 | 7.7990 | 1.6207 | 4.7099 | 3.0891 | 0.3237 | -4.7099 | 3.5905 | 1.5247 | 1.9697 | 11.3894 |
| CA6 | -7.8235 | -1.5647 | 6.2588 | 7.8235 | 1.5658 | 4.6946 | 3.1288 | 0.3196 | -4.6946 | 3.5220 | 1.5004 | 1.9562 | 11.3454 |
| CA7 | -7.8017 | -1.5565 | 6.2452 | 7.8022 | 1.5571 | 4.6796 | 3.1226 | 0.3202 | -4.6796 | 3.5066 | 1.4986 | 1.9495 | 11.3088 |
| CA8 | -7.5704 | -1.8069 | 5.7635 | 7.5696 | 1.8055 | 4.6875 | 2.8820 | 0.3470 | -4.6875 | 3.8121 | 1.6265 | 2.0066 | 11.3816 |
| CA9 | -7.4670 | -1.5456 | 5.9213 | 7.4659 | 1.5443 | 4.5051 | 2.9608 | 0.3377 | -4.5051 | 3.4274 | 1.5216 | 1.8831 | 10.8933 |
| CA10 | -8.2561 | -1.1456 | 7.1105 | 8.2561 | 1.1467 | 4.7014 | 3.5547 | 0.2813 | -4.7014 | 3.1090 | 1.3226 | 1.9623 | 11.3651 |
| CA11 | -8.3976 | -1.1048 | 7.2928 | 8.3 | 1.1056 | 4.7519 | 3.6463 | 0.2743 | -4.7519 | 3.0964 | 1.3032 | 1.9908 | 11.4946 |
| THF |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CA1 | -8.2126 | -0.9470 | 7.2656 | 8.2123 | 0.9475 | 4.5799 | 3.6324 | 0.2753 | -4.5799 | 2.8873 | 1.2609 | 1.9398 | 11.0996 |
| CA2 | -8.1554 | -1.4014 | 6.7540 | 8.1565 | 1.4022 | 4.7794 | 3.3771 | 0.2961 | -4.7794 | 3.3819 | 1.4152 | 1.9797 | 11.5384 |
| CA3 | -8.1636 | -1.3633 | 6.8003 | 8.1628 | 1.3622 | 4.7625 | 3.4003 | 0.2941 | -4.7625 | 3.3352 | 1.4006 | 1.9730 | 11.4980 |
| CA4 | -7.8479 | -1.5239 | 6.3241 | 7.8466 | 1.5236 | 4.6851 | 3.1615 | 0.3163 | -4.6851 | 3.4715 | 1.4819 | 1.9479 | 11.3181 |
| CA5 | -7.8289 | -1.5212 | 6.3077 | 7.8297 | 1.5206 | 4.6752 | 3.1546 | 0.3170 | -4.6752 | 3.4644 | 1.4820 | 1.9438 | 11.2941 |
| CA6 | -7.8588 | -1.4858 | 6.3731 | 7.8596 | 1.4863 | 4.6730 | 3.1867 | 0.3138 | -4.6730 | 3.4263 | 1.4664 | 1.9400 | 11.2859 |
| CA7 | -7.8425 | -1.4858 | 6.3567 | 7.8422 | 1.4860 | 4.6641 | 3.1781 | 0.3147 | -4.6641 | 3.4225 | 1.4676 | 1.9365 | 11.2647 |
| CA8 | -7.7690 | -1.0830 | 6.6860 | 7.7701 | 1.0828 | 4.4264 | 3.3437 | 0.2991 | -4.4264 | 2.9299 | 1.3238 | 1.8472 | 10.7000 |
| CA9 | -7.7772 | -1.0585 | 6.7186 | 7.7775 | 1.0594 | 4.4184 | 3.3590 | 0.2977 | -4.4184 | 2.9059 | 1.3154 | 1.8466 | 10.6834 |
| CA10 | -8.3813 | -1.1211 | 7.2602 | 8.3821 | 1.1211 | 4.7516 | 3.6305 | 0.2754 | -4.7516 | 3.1095 | 1.3088 | 1.9884 | 11.4916 |
| CA11 | -8.4031 | -1.1048 | 7.29 | 8.4 | 1.1056 | 4.7543 | 3.6487 | 0.2741 | -4.7543 | 3.0975 | 1.3030 | 1.9919 | 11.5006 |
| Ethanol |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CA1 | -8.1963 | -0.9225 | 7.2738 | 8.1968 | 0.9225 | 4.5596 | 3.6372 | 0.2749 | -4.5596 | 2.8580 | 1.2536 | 1.9356 | 11.0548 |
| CA2 | -8.1690 | -1.3878 | 6.7812 | 8.1701 | 1.3881 | 4.7791 | 3.3910 | 0.2949 | -4.7791 | 3.3677 | 1.4093 | 1.9796 | 11.5378 |
| CA3 | -8.1826 | -1.3443 | 6.8384 | 8.1832 | 1.3448 | 4.7640 | 3.4192 | 0.2925 | $-4.7640$ | 3.3189 | 1.3933 | 1.9741 | 11.5021 |
| CA4 | -8.2207 | -1.5075 | 6.7132 | 8.2197 | 1.5075 | 4.8636 | 3.3561 | 0.2980 | -4.8636 | 3.5242 | 1.4492 | 2.0166 | 11.7438 |
| CA5 | -7.8398 | -1.5075 | 6.3322 | 7.8406 | 1.5075 | 4.6741 | 3.1665 | 0.3158 | -4.6741 | 3.4497 | 1.4761 | 1.9421 | 11.2903 |
| CA6 | -7.8697 | -1.4749 | 6.3948 | 7.8683 | 1.4738 | 4.6711 | 3.1973 | 0.3128 | -4.6711 | 3.4121 | 1.4610 | 1.9383 | 11.2805 |
| CA7 | -7.8561 | -1.4776 | 6.3785 | 7.8556 | 1.4771 | 4.6663 | 3.1892 | 0.3136 | -4.6663 | 3.4137 | 1.4631 | 1.9367 | 11.2693 |
| CA8 | -7.8507 | -1.0912 | 6.7595 | 7.8496 | 1.0901 | 4.4698 | 3.3797 | 0.2959 | -4.4698 | 2.9558 | 1.3225 | 1.8657 | 10.8054 |
| CA9 | -7.8534 | -1.0749 | 6.7785 | 7.8528 | 1.0741 | 4.4634 | 3.3894 | 0.2950 | -4.4634 | 2.9389 | 1.3169 | 1.8649 | 10.7918 |
| CA10 | -8.3949 | -1.1184 | 7.2765 | 8.3941 | 1.1184 | 4.7562 | 3.6378 | 0.2749 | -4.7562 | 3.1093 | 1.3074 | 1.9908 | 11.5033 |
| CA11 | -8.4085 | -1.1102 | 7.29 | 8. | 1.1 | 4.7587 | 3.6493 | 0.2740 | -4.7587 | 3.1027 | 1.3040 | 1.9933 | 11.5107 |
| DMF |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CA1 | -8.1935 | -0.9170 | 7.2765 | 8.1933 | 0.9165 | 4.5549 | 3.6384 | 0.2748 | -4.5549 | 2.8511 | 1.2519 | 1.9346 | 11.0444 |
| CA2 | -8.1690 | -1.3851 | 6.7840 | 8.1677 | 1.3851 | 4.7764 | 3.3913 | 0.2949 | -4.7764 | 3.3636 | 1.4084 | 1.9785 | 11.5313 |
| CA3 | -8.1881 | -1.3416 | 6.8465 | 8.1873 | 1.3410 | 4.7641 | 3.4231 | 0.2921 | -4.7641 | 3.3152 | 1.3917 | 1.9742 | 11.5025 |
| CA4 | -7.8588 | -1.5048 | 6.3540 | 7.8575 | 1.5043 | 4.6809 | 3.1766 | 0.3148 | -4.6809 | 3.4488 | 1.4736 | 1.9445 | 11.3062 |
| CA5 | -7.8425 | -1.5048 | 6.3377 | 7.8428 | 1.5051 | 4.6739 | 3.1688 | 0.3156 | -4.6739 | 3.4470 | 1.4750 | 1.9419 | 11.2897 |
| CA6 | -7.8697 | -1.4722 | 6.3975 | 7.8703 | 1.4711 | 4.6707 | 3.1996 | 0.3125 | -4.6707 | 3.4091 | 1.4598 | 1.9380 | 11.2793 |
| CA7 | -7.8588 | -1.4749 | 6.3839 | 7.8583 | 1.4754 | 4.6669 | 3.1914 | 0.3133 | -4.6669 | 3.4122 | 1.4623 | 1.9368 | 11.2705 |
| CA8 | -7.8670 | -1.0912 | 6.7758 | 7.8681 | 1.0917 | 4.4799 | 3.3882 | 0.2951 | -4.4799 | 2.9617 | 1.3222 | 1.8700 | 10.8298 |
| CA9 | -7.8697 | -1.0776 | 6.7921 | 7.8700 | 1.0773 | 4.4737 | 3.3963 | 0.2944 | -4.4737 | 2.9464 | 1.3172 | 1.8690 | 10.8163 |
| CA10 | -8.3949 | -1.1184 | 7.2765 | 8.3952 | 1.1179 | 4.7565 | 3.6387 | 0.2748 | -4.7565 | 3.1089 | 1.3072 | 1.9910 | 11.5041 |
| CA11 | -8.4085 | -1.1102 | 7.298 | 8.40 | 1.1100 | 4.7594 | 3.6494 | 0.2740 | -4.7594 | 3.1035 | 1.3042 | 1.9935 | 11.5123 |
| Water |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CA1 | -8.1908 | -0.9116 | 7.2792 | 8.1900 | 0.9121 | 4.5511 | 3.6389 | 0.2748 | -4.5511 | 2.8459 | 1.2507 | 1.9338 | 11.0359 |
| CA2 | -8.1690 | -1.3878 | 6.7812 | 8.1696 | 1.3881 | 4.7788 | 3.3908 | 0.2949 | -4.7788 | 3.3676 | 1.4094 | 1.9795 | 11.5372 |
| CA3 | -8.1881 | -1.3388 | 6.8493 | 8.1892 | 1.3394 | 4.7643 | 3.4249 | 0.2920 | -4.7643 | 3.3137 | 1.3911 | 1.9743 | 11.5029 |
| CA4 | -7.8588 | -1.5021 | 6.3567 | 7.8583 | 1.5026 | 4.6805 | 3.1778 | 0.3147 | -4.6805 | 3.4468 | 1.4729 | 1.9442 | 11.3051 |
| CA5 | -7.8452 | -1.5048 | 6.3404 | 7.8439 | 1.5040 | 4.6739 | 3.1699 | 0.3155 | -4.6739 | 3.4458 | 1.4745 | 1.9418 | 11.2896 |
| CA6 | -7.8724 | -1.4694 | 6.4030 | 7.8713 | 1.4697 | 4.6705 | 3.2008 | 0.3124 | -4.6705 | 3.4076 | 1.4592 | 1.9378 | 11.2789 |
| CA7 | -7.8588 | -1.4749 | 6.3839 | 7.8599 | 1.4746 | 4.6673 | 3.1926 | 0.3132 | -4.6673 | 3.4115 | 1.4619 | 1.9369 | 11.2714 |
| CA8 | -7.8779 | -1.0939 | 6.7840 | 7.8765 | 1.0928 | 4.4847 | 3.3918 | 0.2948 | -4.4847 | 2.9648 | 1.3222 | 1.8720 | 10.8413 |
| CA9 | -7.8779 | -1.0803 | 6.7976 | 7.8781 | 1.0800 | 4.4791 | 3.3991 | 0.2942 | -4.4791 | 2.9512 | 1.3177 | 1.8711 | 10.8293 |
| CA10 | -8.3949 | -1.1184 | 7.2765 | 8.3957 | 1.1176 | 4.7567 | 3.6391 | 0.2748 | -4.7567 | 3.1087 | 1.3071 | 1.9911 | 11.5045 |
| CA11 | -8.4058 | -1.1075 | 7.2983 | 8.4069 | 1.1086 | 4.7577 | 3.6491 | 0.2740 | -4.7577 | 3.1016 | 1.3038 | 1.9930 | 11.5085 |

Table 2. Results of principal components analysis on the global descriptors data

| Eigenvalues |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Gas phase |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | PC 1 | PC 2 | PC 3 | PC 4 | PC 5 | PC 6 | PC 7 | PC 8 | PC 9 | PC 10 | PC 11 | PC 12 | PC 13 |
| Variance | 9.3282 | 3.5010 | 0.1705 | 0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | -0.0000 | -0.0000 |
| \% Variance | 71.8 | 26.9 | 13.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Cumulative | 71.8 | 98.7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| THF solvent |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | PC 1 | PC 2 | PC 3 | PC 4 | PC 5 | PC 6 | PC 7 | PC 8 | PC 9 | PC 10 | PC 11 | PC 12 | PC 13 |
| Variance | 7.2636 | 5.7186 | 0.0175 | 0.0003 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | -0.0000 |
| \% Variance | 55.9 | 44.0 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Cumulative | 55.9 | 99.9 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Ethanol solvent |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | PC 1 | PC 2 | PC 3 | PC 4 | PC 5 | PC 6 | PC 7 | PC 8 | PC 9 | PC 10 | PC 11 | PC 12 | PC 13 |
| Variance | 7.1224 | 5.8589 | 0.0178 | 0.0009 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| \% Variance | 54.8 | 45.1 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Cumulative | 54.8 | 99.9 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| DMF solvent |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | PC 1 | PC 2 | PC 3 | PC 4 | PC 5 | PC 6 | PC 7 | PC 8 | PC 9 | PC 10 | PC 11 | PC 12 | PC 13 |
| Variance | 7.5657 | 5.4086 | 0.0254 | 0.0003 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| \% Variance | 58.2 | 41.6 | 2.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Cumulative | 58.2 | 99.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Water |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | PC 1 | PC 2 | PC 3 | PC 4 | PC 5 | PC 6 | PC 7 | PC 8 | PC 9 | PC 10 | PC 11 | PC 12 | PC 13 |
| Variance | 7.5953 | 5.3778 | 0.0265 | 0.0004 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | -0.0000 | -0.0000 | -0.0000 |
| \% Variance | 58.4 | 41.4 | 2.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | -0.0 | -0.0 | -0.0 |
| Cumulative | 58.4 | 99.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

Step 4. Return to Step 2 or stop if the final two clusters have been fused into the final, single cluster.

The wide range of agglomerative methods available differ principally in the implementation of Step 3 and the calculation of the distance between two clusters. The different betweengroup distance measures can be defined in terms of the general formula

$$
\begin{equation*}
d_{k(i, j)}=\alpha_{i} d_{k, i}+\alpha_{j} d_{k, j}+\beta d_{i, j}+\gamma\left|d_{k, i}-d_{k, j}\right| \tag{3}
\end{equation*}
$$

where $d_{i, j}$, is the distance between objects i and j and $d_{k(i, j)}$ is the distance between group k and a new group $(i, j)$ formed by the fusion of groups $i$ and $j$. The values of coefficients $\alpha, \beta$, and $\gamma$ are chosen to select the specific between-group metric to be used.

The complete dendrogram is shown in Figure 3 for table. 1 in different solvents by using the hierarchical clustering technique. The horizontal lines represent the methods and vertical lines the similarity values between pairs of methods, a method and a group of methods and among groups of methods. From Fig.3, the HCA results are very similar to those obtained with the PCA results.




Fig 3. Dendrogram obtained for the global descriptors for cyanuric acid tautomers in different solvents.

## Conclusion

Molecular geometries of tautomeric forms of Cyanuric acid were fully optimized by using the Gaussian quantum chemistry software package Gaussian 09 w at DFT/B3LYP level of theory, using the $6-311++G(d, p)$ basis set. we were calculated the HOMO and LUMO energies in order to determine, the usefulness of global reactivity descriptors namely, the electrophilicity, Chemical hardness ( $\eta$ ), chemical potential ( $\mu$ ), polarizability $(\alpha)$ electrophilicity index $(\omega)$, softness (S), nucleofugality, and electrofugality, values for the prediction of the reactivity of the cyanuric acid tautomers. The global and reactivity descriptors of cyanuric acid tautomers were analyzed using Principal component analysis (PCA) and hierarchical cluster analysis (HCA) in different solvents. Three tautomers, CA1, CA10 and CA11 constitute unique group, well distinguished from the rest. The CA8, CA4 and CA5 valves constitute a cluster of four tautomer types and CA2 and CA3 constitute a cluster of two tautomer types finally, there is a group of two tautomers CA7 and CA6 constitute a cluster in gas phase. The tautomers constitute groups are similar in three solvents namely THF, DMF and water, but in ethanol solvent show different groups. The HCA results are very similar to those obtained with the PCA results.

## References

1.P.Geladi, K. Esbensen, J.Chemometr. 4 (1990) 337-354.
2. K. Esbensen, P. Geladi, J. Chemometr. 4(1990) 389-412.
3. J.Workman, Chemometr Intell Lab. 60(2002) 13 - 23.
4. T. Kourti, Crit Rev Anal Chem 36(2006) 257-278.
5. R. Bro, Anal Chim Acta. 500(2003) 185-194.
6. T. Kourti, P. Nomikos , M. JF, J. Proc Cont 5(1995) 277-284.
7. A.Smilde, R. Bro, R, P G: Multi-way analysis with applications in the chemical sciences. John Wiley \& Sons Ltd, 2005
8. R.G.Brereton, Chemometrics: Data analysis for the laboratory and chemical plant. John Wiley \& Sons Ltd, 2003
9. G. Reich, Adv Drug Deliver Rev. 57(2005) 1109-1143.
10. S.Garca-Mu noz, T.Kourti, J.MacGregor, A. Mateos, G. Murphy, Ind Eng Chem Res. 42(2003) 3592-3601.
11.L.H. Chiang , L.F. Colegrove, Chemometr Intell Lab. 88(2007) 143-153.
12.B. Wise, N.Gallagher, J Process Contr. 6(1996) 329-348.
13.H.Martens, M. Martens, Multivariate Analysis of Quality: An Introduction. John Wiley \& Sons, Ltd, 2001
14.T. Kourti, IEEE Contr Syst Mag. 22(2002) 10-25.
15. P. H. A. Sneath, R. R. Sokal, Numerical Taxonomy, W. H. Freeman, San Francisco, CA, 1973.
16. L. Kaufman,P. J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, Wiley-Interscience, New York, 1990.
17. B. S. Everitt, Cluster Analysis, 3rd ed, Edward Arnold, London, 1993.
18. A. D. Gordon, Classification, 2nd ed, Chapman and Hall, London, 1999.
19. P. Willett, Similarity and Clustering in Chemical Information Systems, Research Studies Press, Letchworth, UK, 1987.
20. N. Bratchell, Chemom. Intell. Lab. Systems, 6(1989) 105.
21. J. M. Barnard, G. M. Downs, J. Chem. Inf. Comput. Sci. 32 (6) (1992), 644.
22. G. M. Down, P. Willett, in Advanced Computer-Assisted Techniques in Drug Discovery, H. van de Waterbeemd, Ed., VCH Publishers, Weinheim, 1994, pp. 111-130.
23. Clustan Ltd., 16 Kingsburgh Road, Edinburgh, UK. http://www.clustan.com.
24. SAS Institute Inc., SAS Campus Drive, Cary, NC 27513, USA. http://www.sas.com.
25. Barnard Chemical Information Ltd., 46 Uppergate Road, Stannington, Sheffield S6 6BX, UK. http://www.bci.gb.com.
26. Chemical Computing Group Inc., 1010 Sherbrooke Street West, Suite 910, Montreal, Quebec H3A 2R7, Canada. http://www.chemcomp.com.
27. Daylight Chemical Information Systems Inc., 441 Greg Avenue, Santa Fe, NM 87501, USA. http://www.daylight.com.
28. MDL Information Systems, Inc., 14600 Catalina Street, San Leandro, CA 94577, USA. http:// www.mdl.com.
29. Accelrys (formerly Molecular Simulations Inc.), 9685 Scranton Road, San Diego, CA 92121- 3752, USA. http://www.accelrys.com.
30. Tripos Inc., 1699 South Hanley Road, St. Louis, MO 63144, USA. http://www.tripos.com.
31.W. A. Warr, in Computational Methods for the Analysis of Molecular Diversity, P. Willett, Ed., Perspectives in Drug Discovery and Design, Vol. 7/8, Kluwer/ESCOM, Dordrecht, The Netherlands, 1997, pp. 115-130. Commercial Software Systems for Diversity Analysis.
32.M.J Frisch; G.W Trucks;H.B Schlegel; et al., 2009. Gaussian 09, Rev. A. 1 Gaussian, Inc., Wallingford CT.
33.N. Özdemir, M. Dinçer, A .Çukurovalı, O. Büyükgüngör, J.Mol.Model,15 (2009) 1435.
34.T.Teslova, C.Corredor, R. Livingstone, T. Spataru, R.L. Birke, J.R. Lombardi, M.V
Canamares, M. Leona, J.Raman, Spectros.38(2007) 802.
35.C.Corredor, T. Teslova, M.V. Canamares, Z.Chen, J.Zhang, J.R. Lombardi, M. Leona, Vibr. Spectrosc. 49(2009)190.
36.N.Surendra babu and Didugu Jayaprakash .International journal of Scientific research.4(6) (2015) 615-620.

