

A MOLECULAR AND QUANTUM MECHANICAL STUDY OF INDOLE-3-ACETIC ACID

Rudolf Kiralj (PQ) and Márcia M. C. Ferreira (PQ), Laboratório de Quimiometria Teórica e Aplicada, Instituto de Química, Universidade Estadual de Campinas (UNICAMP), Campinas, SP, 13083-970, Brazil.

E-mails: marcia@iqm.unicamp.br, rudolf@iqm.unicamp.br, URL: <http://lqta.iqm.unicamp.br>



ABSTRACT

Indole-3-acetic acid is the most important auxin, a plant hormone with simple molecular structure. All the attempts to quantify its structure-biological activity relationship were not much successful up to date, mostly due to missing knowledge about some intrinsic molecular properties responsible for intermolecular interactions. Low temperature crystal structure of indole-3-acetic acid is a good reference point for systematic theoretical studies. In this work, the experimental geometry was optimized using molecular mechanics (MMF, SYBIL), semi-empirical methods (MINDO3, MNDO, AM1, PM3) and *ab initio* (about twenty DFT, RHF, and B3LYP basis sets). Bond length matrix $m \times n$ ($m =$ number of bonds, $n =$ number of calculations) and its transpose ($n \times m$) were analyzed utilizing Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA). A carboxylic acid cyclic H-bond dimer from crystal structure was optimized using various methods. In general, *ab initio* bond lengths are closer to experimental ones than those from semi-empirical and molecular mechanics. All the methods do not reproduce well C-OH bond which is the most one affected by dimer's H-bonds. Detailed PCA on both matrices shows that bond lengths are described at most by six Principal Components. The data clustering is similar in PCA and in HCA. Characteristic molecular fragments are recognized among the scores, while similar basis sets of *ab initio* and also methods of the same level of accuracy (e.g. molecular mechanics, semi-empirical, *ab initio*) are distinguished in loadings. Bond lengths of optimized cyclic H-bond dimer are closer to experimental values than those of isolated indole-3-acetic acid, especially considering the carboxylic chain (by more than 0.01 Å). The optimized dimer structures show that the carboxyl group geometry is closely related to H-bond geometry.

1 B. Nigović, S. Antić, B. Kojić-Prodić, R. Kiralj, V. Magnus, B. Salopek-Sondi, *Acta Cryst. B56* (2000) 94–111.

INTRODUCTION

Indole-3-acetic acid (IAA) is an important plant growth hormone (auxin class 1); it consists of indole aromatic ring^{1,2} and acetic acid chain in position 3 (Scheme 1). Geometry of this molecule is an important source of information about IAA's aromaticity, substituent effects, conformational changes due to interaction with its receptor, etc. The question of obtaining as much as possible accurate geometry parameters is then task for both experimentalists and theoreticians. Experimental geometry determined by X-ray crystallography at 102 K is a reliable source of geometry information that satisfies current needs for geometry accuracy. On the other side, arises a question which computational methods to use to reproduce experimental geometries: molecular mechanics (MM), semiempirical methods, or *ab initio* (Hartree-Fock (HF), Hartree-Fock (HF), or some other methods)? Usually experimental geometry is compared with computational results: most of the times the experiment that had been carried out was some diffractometric technique, and the specimen was condensed matter. This way, comparing for example, molecule in crystal (experimental results) with free molecule (theoretical results), an error has been done, and some neighborhood effects (like crystal packing effects) was neglected. IAA molecule binds to other molecule in crystal making hydrogen bond dimer (Scheme 2, Figure 1). Also, IAA molecules are stacked along crystallographic axis *b* (Figure 2). The stacking includes primarily π - π interactions, especially involving atoms N and C4 atoms. These and other interactions in crystal can affect the IAA geometry. Other effect that should be taken account is disadvantage of the experimental technique. It is known that hydrogen bonds as well as bond lengths determined by X-ray diffraction are in general shorter than those from neutron or electron diffraction or some spectroscopic technique. Correction to thermal motion in crystal is only a partial solution of the problem. Various computational techniques act differently from case to case. For example, some method will not always superior to others when some molecular fragment is under question, and when free molecule is studied, and also when some molecular complex is under consideration. IAA geometry was studied in this work: experimental parameters were compared with computational results for free IAA and hydrogen bond IAA dimer. Chemometric methods were utilized to find out which computational methods would be the most appropriate for both isolated and dimer IAA: PCA and HCA. A study on hydrogen bond geometry, using PLS (Partial Least Squares) was also performed. Chemometric methods showed useful information about IAA intrinsic properties as well as on computational techniques.

2 T. L. Gilchrist: *Heterocyclic Chemistry*, 3rd ed., Addison Wesley Longman, Ltd., Harlow, UK, 1997.

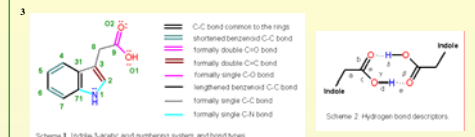


Figure 1. The cyclic H-bond dimer from crystal structure (CSD, INACET03). Hydrogen bonds are drawn as green dashed lines.

Figure 2. Stacking interactions between neighboring IAA molecules in crystal structure. The stacking occurs along the *b* crystallographic axis.

METHODOLOGY

1) Experimental geometry:

Experimental bond lengths¹ were corrected to thermal motion in crystal using PLATON³.

2) PM3 geometry optimization of free IAA:

Semiempirical MINDO3, MNDO, AM1, and PM3 methods incorporated into MOPAC 6.0⁴ were used with experimental starting geometry at RHF, UHF and CI level and PRECISE geometry optimization. Other computational methods were used in Titan⁵, MMFF94 and SYBIL molecular mechanics force fields, HF with several basis sets starting with STO-3G up to cc-p-VTZ, DFT (VWN functional) with a few basis sets, and B3LYP.

3) PM3 geometry optimization of IAA dimer:

Starting experimental geometry of IAA dimer was optimized with semiempirical MNDO, AM1 and PM3 in Titan. Also, DFT and B3LYP (6-31G** basis set) and SYBIL calculations were carried out using the same software. MMFF94 calculations were performed by optimizing the experimental geometry and also by Monte Carlo conformational search. In both cases the hydrogen bonded complex rearranged into van der Waals association.

5) Chemometrics:

The results from the calculations were bond lengths for isolated and dimer IAA, and hydrogen bond parameters (distances and angles as defined in Scheme 2) which are not presented here. Experimental (both uncorrected and corrected to thermal motion) and computational results were studied using HCA (incremental linkage method) and PCA on regular ($m \times n$) and transpose ($n \times m$) autocorrelated data sets for isolated IAA, dimer IAA, and hydrogen bond ring. Finally, PLS (leave-one out crossvalidation) was used to related H...O distance as a function of the COO geometry.

3 PLATON, A Multipurpose Crystallographic Tool, v. 31000, 2000, Spek, A. L. Utrecht University, Utrecht.
4 Stewart, J. J. P. MOPAC, version 6.0, 1990, Frank, Selter Res. Lab, Air Force Academy, Colorado Springs, Colorado
5 Titan, version 1.2000, Wavfunction, Inc., Irvine, CA.

RESULTS

1) HCA and PCA of the regular data set for isolated IAA:

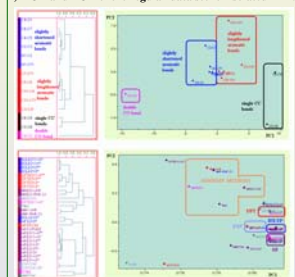


Figure 3. Results of the HCA (left) and PCA (right) analysis for free IAA. Dendrogram for samples (IAA bonds) shows clusters and sub-clusters based on bond length. Double C=C bond is isolated, while the rest makes three sub-clusters: long C-C bonds including methylene group, then aromatic bonds slightly lengthened and slightly shortened with respect to benzene bond length (approximately 1.4 Å). The dendrogram clearly reveals that C-N bonds are relatively short and almost of equal length due to participation of N in electron delocalization. Bonds C4-C5 and C6-C7 are almost identical, what happens also to C31-C4 and C71-C7, exhibiting slight bond length alteration with respect to benzene. Bond C7-C71 is longer than benzene bond due to its position (belong to the both rings). Such a clustering is also observed in PCA. The variables (computational methods) exhibit more complex clustering. Most of them are grouped in HCA as well as in PCA in several groups: semi-empirical, experimental, DFT, B3LYP, HF, HF methods seem to be the closest ones to the experiment. PCA shows that the first PC describes 97.96% of the total variance, while the next PCA contain less than 1% of original information.

2) HCA and PCA of the transpose data set for isolated IAA:

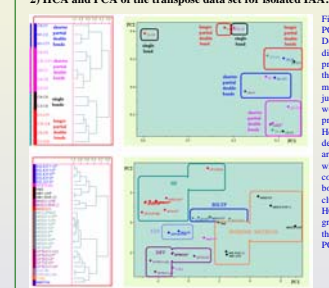


Figure 4. Results of the HCA (left) and PCA (right) analysis for free IAA. Dendrogram for samples is not much different than the analogous one in the previous analysis. The main difference is that the double C=C bond is not isolated now. There are three clusters like before, just this time the property in question would not be bond length but electronic property, electron delocalization effects. Here bonds are variables, meaning delocalization mechanisms, while samples are computational methods, the quality of which can be characterized by computational results on some types of bonds (in this case bonds of IAA). Bond clustering in PCA is in general like in HCA. Computational methods are grouped in both HCA and PCA clearer than in the previous analysis. The first six PCs describe 95.08% of the total variance.

3) HCA and PCA of the regular data set for IAA dimer:

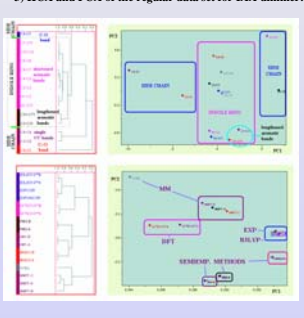


Figure 6. Results of HCA and PCA for IAA dimer. Clustering of bonds in both HCA and PCA is similar to the previous cases. In the present analysis clustering of the bonds into indole ring bonds and side chain bonds is more pronounced. This is reasonable as two strong hydrogen bonds of the dimer change the side chain geometry, especially of the carboxylic group. Methodologies are grouped as MM, DFT, B3LYP, EXP, SEMEMP. The first PC describes 96.76% of the total variance.

4) HCA and PCA of the transpose data set for IAA dimer:

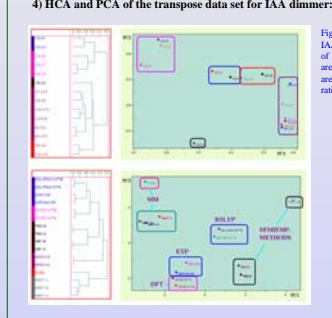


Figure 7. Results of HCA and PCA analysis for IAA dimer. The first five PCs contain 96.46% of the original information. The methodologies are grouped similar to other analyses. The bonds are grouped in a way quite complicate to rationalize.

5) HCA and PCA of the regular data set for hydrogen bond ring:

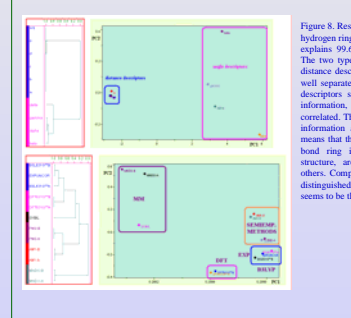


Figure 8. Results of HCA and PCA analysis for hydrogen ring of the IAA dimer. The first PC explains 99.67% of the original information. The two types of hydrogen bond descriptors, distance descriptors and angle descriptors, are well separated both in HCA and PCA. Bond descriptors seem to contain the very same information, by other words, they are highly correlated. The first four PCs describe the full information ($\approx 100\%$ of the variance). This means that the descriptors, since the hydrogen bond ring is a nearly planar and closed structure, are linear combinations of each other. Computational methodologies are well distinguished in both HCA and PCA. B3LYP seems to be the closest one to the experiment.

6) HCA and PCA of the transpose data set for hydrogen bond ring:

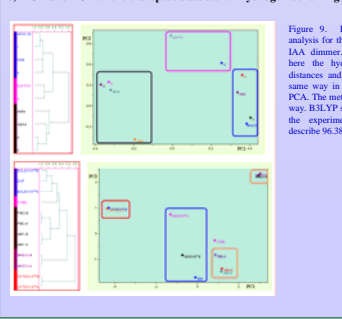


Figure 9. Results of HCA and PCA analysis for the hydrogen bond ring of the IAA dimer. Unlike the previous cases, here the hydrogen bond descriptors - distances and angles - are mixed in the same way in a similar way in HCA and PCA. The methods are grouped in a usual way. B3LYP seems to be the closest one to the experiment. The first four PCs describe 96.38% of the total variance.

7) PLS models for hydrogen bond geometry:

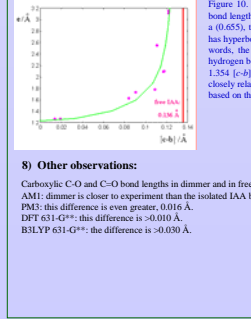


Figure 10. H...O distance as a function of differences between C-O and C=O bond lengths. Variable ϵ is moderately to highly correlated to $[\epsilon - \delta]$ ($r=0.737$), to a (0.655), to $[\epsilon - \delta]^2$ (0.878) and to ϵ^2 (0.776). The non-linearity here presented has hyperbolic shape, with vertical asymptote at the value of free IAA. By other words, the degree of carboxylic group delocalization is closely connected to hydrogen bond strength. A quite satisfactory PLS model was built: $\epsilon = 0.312a + 1.354[\epsilon - \delta] - 0.648\epsilon^2$, 3 PCs, SEC=0.245 Å, Q=0.909, R=0.944. Angle δ is closely related to $[\epsilon - \delta]$ (-0.551) and to $[\epsilon - \delta]^2$ (-0.815). PLS model for δ , based on these two variables, gave for 1 PC, R=0.849, Q=0.766, SEP=5.59°.

8) Other observations:

Carboxylic C-O and C=O bond lengths in dimer and in free IAA. AM1 dimer is closer to experiment than the isolated IAA by 0.008 Å. PM3: this difference is even greater, 0.016 Å. DFT 631-G** : this difference is -0.010 Å. B3LYP 631-G** : this difference is -0.030 Å.

CONCLUSION

Chemometrics-computational study on bond lengths of free IAA reveals that: -optimized IAA dimer is significantly closer to crystal structure dimer than free IAA in terms of geometry parameters; it can be observed especially in side group geometry -the best method for IAA dimer geometry optimization is B3LYP (HF was not performed). Hydrogen bond ring studies pointed out that: -hydrogen bond descriptors are closely related to each other, even at PLS level -electron delocalization of the carboxylic group and hydrogen bond geometry are quantitatively connected, exhibiting that the charge transfer inside the ring has a cyclic character -the best method for geometry optimization of the hydrogen bond ring is B3LYP (HF was not performed).

ACKNOWLEDGEMENT

The authors acknowledge FAPESP for the financial support