Combined Computational and Chemometric Study of 1*H*-Indole-3-Acetic Acid

RUDOLF KIRALJ, MÁRCIA M. C. FERREIRA

Laboratório de Quimiometria Teórica e Aplicada, Instituto de Química, Universidade Estadual de Campinas, Campinas 13083-862, Sao Paulo, Brazil

Received 17 November 2001; revised 14 February 2003

DOI 10.1002/qua.10693

ABSTRACT: Geometries of free 1*H*-indole-3-acetic acid (IAA) and IAA hydrogen bond dimer were optimized at several computational levels: molecular mechanics, semiempirical methods, ab initio density functional theory with SVWN (Slater-Vosko-Wilk-Nusair) and Becke's three-parameter exchange functional and the gradient-corrected functional of Lee, Yang, and Paar (B3LYP) functionals, and Hartree–Fock (HF). Bond length matrices $\mathbf{X}(mxn)$ (m = number of bonds, n = number of experimental determinations and theoretical calculations) and their transposes $\mathbf{X}(nxm)$ for IAA monomer, IAA dimer, and hydrogen bond ring (angles included) were analyzed using principal component analysis (PCA) and hierarchical cluster analysis (HCA). Ab initio methods prove to be superior to molecular mechanics and semiempirical methods: SVWN methods are the best for monomer, and B3LYP are best for dimer geometry optimization. The B3LYP and HF methods can be used equally well for optimization of the dimer ring geometry. Other hydrogen bond and aromaticity structural parameters exhibit preference either for B3LYP or SVWN methods. © 2003 Wiley Periodicals, Inc. Int J Quantum Chem 95: 237–251, 2003

Key words: carboxylic acid; chemometrics; computational methods; cyclic dimer; (hetero)aromaticity; indole-3-acetic acid

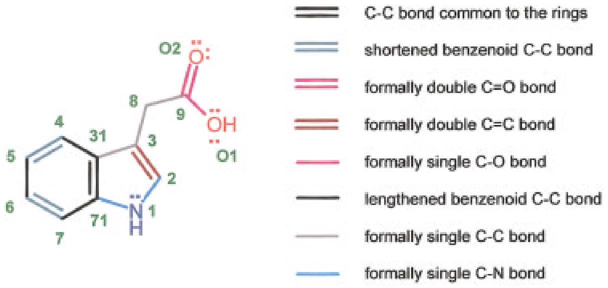
Introduction

1*H*-Indole-3-acetic acid (IAA) is the first isolated and the most important plant hormone from auxin class [1]. Its quantitative structure–activity relationship/quantitative structure–property relationships (QSAR/QSPR) [2] are not simple, and the structure of an IAA receptor have just appeared recently [3]. Consisting of an indole ring [4, 5] and acetic acid chain at position 3, IAA has a simple molecular structure (shown in Scheme 1). This schematic representation of formal single and double bonds that are not in the benzene ring is in accord with common structural formula for in-

---

Correspondence to: M. M. C. Ferreira; e-mail: marcia@iqm.unicamp.br.

Contract grant sponsor: FAPESP.



SCHEME 1. Indole-3-acetic acid numbering system and bond types. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

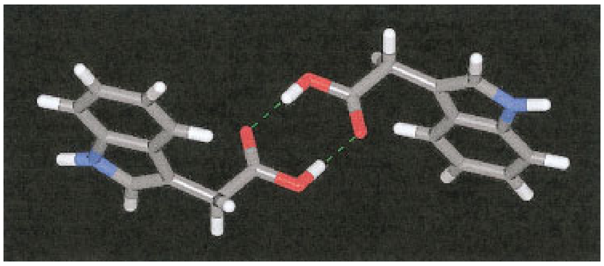
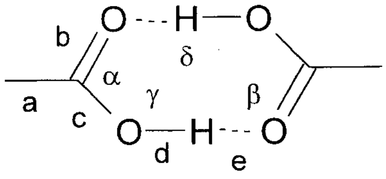


FIGURE 1. The cyclic bond dimmer from crystal structure (CSD: INACET03). Hydrogen bonds are drawn as green dashed lines. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

dole, indene, and indole-3-acetic acid [6], chemical behavior of IAA [5] and bond lengths from the crystal structure of IAA [2] compared with standard single and double CC, CN, and CO bond lengths from gas–phase determination of small molecules [7] or crystal structures of organic compounds [8]. The benzene ring in indole, although traditionally drawn as delocalized, in Scheme 1 appears as the Kekulé structure closer to bond lengths from crystal structure of IAA at 103 K [2]. In fact, crystal structure determination of many IAA derivatives and some ab initio calculations on IAA [2] revealed systematically shortening of the bonds C4—C5 and C6—C7. Several papers [4, 5, 9–13] on aromaticity of indole and its isomers clearly showed that indole is a highly aromatic system, with the nitrogen free electron pair participating in electron delocalization as a requirement for indole's aromaticity [4] (the lone pair is formally equivalent to a double bond [11]). The position of nitrogen in the pyrrole ring makes the indole more aromatic than its isomers [9, 13]. The presence of a substituent at position 3 does not change significantly the aromaticity properties of the indole system.

Precise geometry of IAA contains information about its aromaticity, substituent effects, and possible conformational changes due to interaction with its receptor, for example. Experimental IAA structure [2] like any determined at low temperature [14], is a reliable source of structural information. How does one decide which computational methods to use to reproduce experimental geometries, and to apply for molecules from the same class—molecular mechanics (MM), semiempirical (SEM) methods, or ab initio, for example, density functional (DFT), or Hartree–Fock (HF)? Ab initio results are not always better than those from SEM or MM, and they can provide significantly different optimized geometries [15]. As a minimum requirement for ab initio methods treating isolated molecules are d-type functions (included in 3-21G* or 6-31G* split-valence Gaussian basis sets), whereas use of large basis sets do not result in geometry closer to the experimental [16]. It is desirable to use diffuse functions in geometry optimization of a molecular aggregation [17]. Basis set superposition error (BSSE) should be minimized or even eliminated for comparison of energetical and structural parameters from experiment and calculations [15, 17]. Use of polarization function as a necessary condition for ab initio (restricted Hartree–Fock [RHF]) treatment of IAA geometry optimization was confirmed by Ramek et al. [18].

A heteroaromatic indole ring, and a carboxylic side chain [2] are the two systems that combine to form IAA. The ring is the component that participates in $\pi \ldots \pi$ stacking interactions, and so IAA crystallizes with picric acid [19], various nucleobases [20–22] and other (hetero)aromatic systems [23–25]. 1*H*-Indole-3-acetic acid molecules in crystal of IAA [2] are stacked (primarily N and C4 atoms) along shortest crystallographic axis *b* (Fig. 1), with approximately 4 Å vertical separation of indole planes. As the monofunctional carboxylic acids regularly form cyclic dimers (Scheme 2) with the center of symmetry in the middle of the eight-membered planar ring with the two hydrogen bonds [26, 27], IAA can be characterized in



SCHEME 2. Hydrogen bond descriptors.

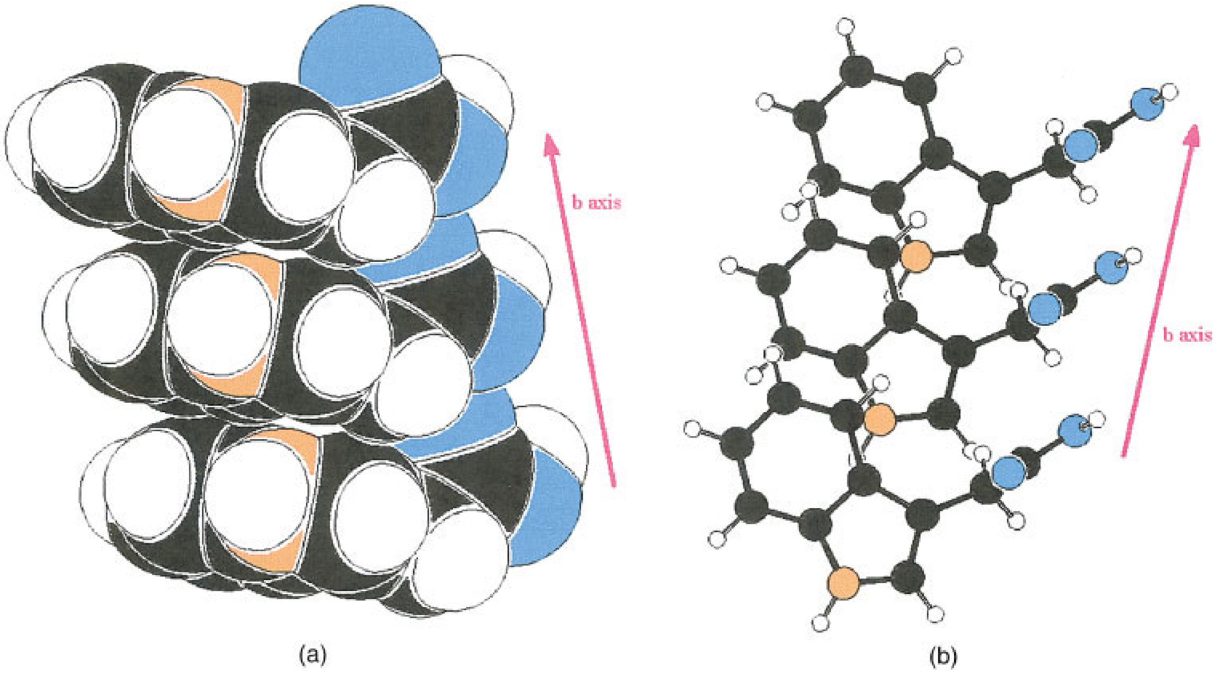


FIGURE 2. (a) Stacking interactions between neighboring IAA molecules in crystal structure. The stacking occurs along the *b* crystallographic axis. (b) Carboxylic groups are also stacked along the same axis. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

the very same way (Fig. 2). These hydrogen bond rings, arranged in stacks also along axis *b*, are stabilized through electrostatic interactions, and the residues (indoles) are irrelevant to the stabilization of dimers [27]. These and other intermolecular interactions do not necessarily represent unimportant crystal packing effects [28], ranging from 0.01 to 0.02 Å for bond lengths. Other difficulties in using the experimental structure is the bond length shortening due to: vibrations in crystal (correction to thermal motion should be applied [14]); and X-ray diffraction (location of the electron density centroids instead of atomic centers), which determines hydrogen atoms with low precision [14]. Estimated standard deviation of experimental bond lengths should be taken into account when comparing experimental and calculated bond length values.

This article focuses on the use of chemometric methods PCA (Principal Component Analysis) and HCA (Hierarchical Cluster Analysis) to determine: the best calculation methods for geometry of IAA monomer, dimer, and its hydrogen bond ring (Scheme 2), and some aromaticity parameters, by an HCA–PCA procedure [7]; and the effects of hydrogen bonding on IAA geometry. The geometry of IAA monomer and dimer was optimized using several MM, SEM, and ab initio computational methods. As the first step of study, correction to BSSE was not applied.

Methods

EXPERIMENTAL GEOMETRY

Experimental fractional atomic coordinates [1] (Cambridge Structural Database or CSD code: INACET03) were used to calculate bond lengths uncorrected and corrected to thermal motion in crystal using PLATON [29].

GEOMETRY OPTIMIZATION OF THE FREE IAA

Semiempirical MINDO/3 [30], MNDO [31], AM1 [32], and PM3 [33] methods incorporated into MOPAC 6.0 [34] were used. Computational methods from Titan [35] were: MMFF94 [36] and SYBIL [37] MM; HF-DFT methods [38] as local density (SVWN functional [39]) and as HF–nonlocal density methods (B3LYP functional) [40, 41] with several basis sets. Additionally, Mulliken and Löwdin bond orders were calculated for experimental and

HF 6-31G** optimized IAA geometry, as well as for HF 6-31G** optimized benzene structure.

GEOMETRY OPTIMIZATION OF THE IAA CYCLIC DIMER

The experimental geometry of the IAA dimer was optimized with most of MM, SEM, and ab initio methods applied to the monomer. Other levels of computation for monomer, due to limited computer time, were not applied for the dimer.

CHEMOMETRICS

Description of Methods

Some of the most suitable methods of multivariate data analysis are PCA [42–44] and HCA [43], based on the correlation among variables. 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 autoscaling (the scaled variables have zero mean and unity variance). Preprocessed **X** is decomposed in PCA as:

$$\mathbf{X} = \mathbf{T}\mathbf{V}^{\mathrm{T}}. \tag{1}$$

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 the subsequent components are orthogonal to one another and describe the maximum of the remaining variance. Once the redundancy is removed, only the first few principal components are required to describe most of the information contained in the original data. Another important multivariate method of data analysis is HCA [43]. 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

$$S_{kl} = 1.000 - \frac{d_{kl}}{d_{\max}} \tag{2}$$

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.

Chemometric Application to Bond Length Data

The bond lengths for monomer and dimer IAA, and hydrogen bond parameters (distances and angles as defined in Scheme 2), experimental (both uncorrected and corrected to thermal motion) and calculated, were studied using HCA (incremental linkage method) and PCA. These data matrices **X** were used: **X**(14, 20) for monomer and for dimer (14 = number of bonds, 20 = number of experimental and theoretical methods); **X**(14, 31) for monomer (31 = the number of experimental and theoretical methods including those not in common for dimer); **X**(10, 20) for the hydrogen bond ring (10 = the number of hydrogen bond ring geometrical parameters, 20 = the number of experimental and theoretical methods). Performed also on the transposed matrices were PCA and HCA. All chemometric methods were performed using software Pirouette [45].

Results and Discussion

IAA BOND TYPES

The CC, CN, and CO bond lengths in organic compounds can be classified as single, double, triple, and partical double or triple (delocalized, conjugated, etc.) according to standard bond lengths data [7, 8]. The following standards are used in this work: 1.535 and 1.339 Å for CC bonds, 1.471 and 1.276 Å for CN bonds, and 1.425 and 1.208 Å for CO bonds. Although Scheme 1 is not an exact molecular representation, its bond length classification reflects rather well the IAA's behavior (see the Introduction). Using the bond length standards (including benzene bond length 1.397 Å [46]) compared with the IAA experimental bond lengths data (Table I), and calculated Mulliken/Lowdin (M/L) bond orders, it is possible to refine more the IAA

TABLE I

Bond length data (Å) for IAA monomer.

No.	Method	N1—C71	N1—C2	C2—C3	C3—C31	C31—C71	C31—C4	C4—C5	C5—C6	C6—C7	C7—C71	C3—C8	C8—C9	C9—O2	C9—O1
1	Experimental[a]	1.380	1.366	1.367	1.434	1.413	1.404	1.381	1.404	1.378	1.388	1.492	1.506	1.224	1.314
2	Experimental[b]	1.382	1.368	1.370	1.436	1.415	1.406	1.383	1.406	1.381	1.390	1.494	1.508	1.226	1.316
3	MMFF94	1.375	1.369	1.385	1.442	1.397	1.414	1.406	1.392	1.397	1.404	1.505	1.516	1.219	1.357
4	SYBIL	1.360	1.364	1.399	1.398	1.389	1.394	1.400	1.402	1.400	1.394	1.526	1.503	1.221	1.333
5	MNDO	1.400	1.399	1.395	1.458	1.441	1.417	1.395	1.423	1.394	1.416	1.495	1.535	1.231	1.357
6	AM1	1.399	1.393	1.397	1.448	1.447	1.396	1.387	1.407	1.386	1.403	1.466	1.498	1.235	1.364
7	PM3	1.415	1.410	1.381	1.443	1.420	1.400	1.380	1.408	1.380	1.402	1.478	1.513	1.218	1.353
8	HF STO-3G	1.396	1.399	1.346	1.457	1.400	1.406	1.370	1.410	1.370	1.403	1.520	1.548	1.216	1.392
9	HF 3-21G*	1.377	1.384	1.350	1.450	1.402	1.395	1.375	1.403	1.375	1.391	1.507	1.504	1.202	1.361
10	HF 6-31G*	1.371	1.371	1.351	1.447	1.400	1.400	1.375	1.405	1.375	1.395	1.507	1.511	1.188	1.331
11	HF 6-31G**	1.371	1.371	1.350	1.447	1.400	1.400	1.375	1.405	1.375	1.395	1.506	1.511	1.188	1.330
12	SVWN 3-21G*	1.375	1.383	1.376	1.434	1.428	1.397	1.388	1.406	1.388	1.391	1.479	1.493	1.231	1.367
13	SVWN 6-31G*	1.369	1.372	1.373	1.428	1.418	1.398	1.385	1.403	1.387	1.391	1.479	1.493	1.215	1.344
14	SVWN 6-31G**	1.369	1.372	1.374	1.428	1.418	1.398	1.385	1.403	1.386	1.391	1.478	1.492	1.215	1.344
15	B3LYP 3-21G*	1.387	1.398	1.374	1.450	1.427	1.406	1.390	1.412	1.392	1.399	1.510	1.519	1.229	1.380
16	B3LYP 6-31G*	1.382	1.382	1.372	1.444	1.422	1.407	1.390	1.411	1.391	1.399	1.506	1.520	1.211	1.356
17	B3LYP 6-31G**	1.382	1.382	1.372	1.444	1.422	1.407	1.389	1.411	1.390	1.399	1.505	1.520	1.212	1.355
18	B3LYP 6-311G*	1.379	1.379	1.369	1.441	1.419	1.404	1.385	1.407	1.386	1.396	1.503	1.517	1.204	1.354
19	B3LYP 6-311G**	1.378	1.379	1.369	1.441	1.419	1.404	1.385	1.407	1.386	1.396	1.504	1.518	1.204	1.354
20	B3LYP 6-311+G**	1.380	1.381	1.369	1.441	1.419	1.405	1.386	1.407	1.388	1.396	1.503	1.518	1.207	1.355
21	MINDO/3[c]	1.374	1.365	1.397	1.480	1.462	1.430	1.392	1.423	1.391	1.427	1.500	1.513	1.216	1.333
22	HF 6-311G*[c]	1.370	1.370	1.347	1.446	1.397	1.398	1.373	1.402	1.373	1.392	1.501	1.510	1.183	1.327
23	HF 6-31+G*[c]	1.372	1.374	1.352	1.450	1.401	1.403	1.378	1.407	1.378	1.397	1.508	1.515	1.191	1.327
24	HF 6-311+G*[c]	1.371	1.370	1.348	1.447	1.398	1.399	1.374	1.403	1.374	1.393	1.503	1.512	1.185	1.328
25	HF 6-311G**[c]	1.370	1.370	1.347	1.446	1.398	1.398	1.373	1.402	1.373	1.392	1.503	1.511	1.183	1.328
26	HF 6-31+G**[c]	1.372	1.374	1.352	1.450	1.401	1.403	1.378	1.407	1.378	1.397	1.508	1.515	1.191	1.327
27	HF 6-311+G**[c]	1.371	1.370	1.348	1.447	1.398	1.399	1.374	1.403	1.374	1.393	1.503	1.512	1.185	1.328
28	HF cc-p-VDZ[c]	1.366	1.367	1.349	1.442	1.398	1.397	1.374	1.403	1.373	1.393	1.498	1.499	1.185	1.329
29	SVWN 6-311+G*[c]	1.367	1.370	1.369	1.425	1.415	1.395	1.381	1.399	1.382	1.387	1.417	1.488	1.208	1.344
30	SVWN 6-311+G**[c]	1.365	1.363	1.366	1.426	1.417	1.394	1.383	1.399	1.382	1.386	1.468	1.484	1.209	1.338
31	B3LYP 6-311+G*[c]	1.381	1.380	1.370	1.443	1.420	1.405	1.387	1.408	1.388	1.397	1.505	1.517	1.205	1.354

[a] Experimental values, uncorrected to thermal motion. Estimated standard deviation for all bonds is 0.003 Å.

[b] Experimental values, corrected to thermal motion.

[c] Calculation not performed for IAA hydrogen bond dimer.

TABLE II

Bond length data (Å) for IAA dimer.

No.	Method	N1—C71	N1—C2	C2—C3	C3—C31	C31—C71	C31—C4	C4—C5	C5—C6	C6—C7	C7—C71	C3—C8	C8—C9	C9—O2	C9—O1
1	Experimental[a]	1.380	1.366	1.367	1.434	1.413	1.404	1.381	1.404	1.378	1.388	1.492	1.506	1.224	1.314
2	Experimental[b]	1.382	1.368	1.370	1.436	1.415	1.406	1.383	1.406	1.381	1.390	1.494	1.508	1.226	1.316
3	MMFF94	1.369	1.366	1.380	1.431	1.392	1.403	1.398	1.389	1.391	1.394	1.499	1.524	1.227	1.354
4	SYBIL	1.360	1.364	1.399	1.397	1.389	1.394	1.400	1.402	1.400	1.394	1.525	1.503	1.221	1.332
5	MNDO	1.400	1.400	1.394	1.459	1.441	1.417	1.396	1.423	1.395	1.416	1.495	1.536	1.231	1.355
6	AM1	1.399	1.393	1.397	1.448	1.447	1.396	1.387	1.407	1.386	1.403	1.466	1.500	1.239	1.357
7	PM3	1.415	1.411	1.381	1.443	1.420	1.400	1.381	1.408	1.380	1.401	1.478	1.515	1.228	1.337
8	HF STO-3G	1.397	1.400	1.346	1.457	1.399	1.406	1.371	1.411	1.371	1.404	1.520	1.552	1.235	1.351
9	HF 3-21G*	1.378	1.384	1.351	1.449	1.402	1.395	1.375	1.404	1.376	1.391	1.506	1.502	1.223	1.323
10	HF 6-31G*	1.371	1.370	1.350	1.448	1.400	1.399	1.375	1.404	1.374	1.394	1.507	1.511	1.206	1.313
11	HF 6-31G**	1.372	1.371	1.351	1.448	1.400	1.399	1.375	1.405	1.375	1.395	1.507	1.511	1.202	1.305
12	SVWN 3-21G*	1.376	1.385	1.376	1.437	1.427	1.398	1.386	1.407	1.388	1.392	1.489	1.497	1.271	1.304
13	SVWN 6-31G*	1.370	1.372	1.371	1.429	1.416	1.398	1.383	1.405	1.385	1.391	1.488	1.499	1.263	1.280
14	SVWN 6-31G**	1.371	1.371	1.373	1.430	1.418	1.400	1.384	1.404	1.386	1.392	1.486	1.499	1.255	1.277
15	B3LYP 3-21G*	1.387	1.395	1.375	1.452	1.429	1.406	1.391	1.412	1.391	1.399	1.511	1.517	1.251	1.337
16	B3LYP 6-31G*	1.382	1.382	1.373	1.444	1.423	1.407	1.390	1.411	1.391	1.399	1.507	1.520	1.230	1.323
17	B3LYP 6-31G**	1.383	1.382	1.372	1.446	1.423	1.408	1.391	1.411	1.391	1.400	1.507	1.519	1.233	1.317
18	B3LYP 6-311G*	1.378	1.379	1.369	1.441	1.418	1.404	1.385	1.407	1.386	1.396	1.504	1.518	1.223	1.321
19	B3LYP 6-311G**	1.382	1.383	1.372	1.445	1.422	1.408	1.390	1.411	1.391	1.401	1.506	1.521	1.228	1.320
20	B3LYP 6-311+G**	1.379	1.379	1.369	1.441	1.418	1.404	1.385	1.407	1.385	1.396	1.505	1.520	1.220	1.321

[a] Experimental values, uncorrected to thermal motion. Estimated standard deviation for all bonds is 0.003 Å.

[b] Experimental values, corrected to thermal motion.

TABLE III

Distance and angle descriptors for the IAA dimer hydrogen bond ring.

No.	Method	*a*/Å	*b*/Å	*c*/Å	*d*/Å	*e*/Å	$\lvert b - c \rvert$/Å	α/°	β/°	γ/°	δ/°
1	Experimental[a]	1.506	1.224	1.315	0.916	1.728	0.091	122.95	125.62	109.66	178.22
2	Experimental[b]	1.506 (3)	1.224 (3)	1.315 (3)	0.92 (3)	1.73 (3)	0.091 (4)	122.95 (19)	125.6 (9)	109.7 (16)	178 (2)
3	MMFF94	1.524	1.224	1.354	0.992	1.811	0.127	119.97	140.18	106.57	155.38
4	SYBIL	1.503	1.221	1.332	0.950	2.541	0.111	122.39	115.85	111.14	148.80
5	MNDO	1.536	1.233	1.355	0.950	3.130	0.122	118.87	142.58	115.90	160.65
6	AM1	1.500	1.239	1.357	0.976	2.094	0.118	116.93	136.36	110.39	173.80
7	PM3	1.515	1.228	1.337	0.967	1.774	0.109	117.32	130.24	112.19	179.35
8	HF STO-3G	1.552	1.235	1.351	1.012	1.510	0.117	123.97	127.04	108.46	179.45
9	HF 3-21G*	1.502	1.223	1.323	0.994	1.651	0.101	123.66	135.41	113.47	167.35
10	HF 6-31G*	1.511	1.206	1.313	0.966	1.818	0.107	123.74	130.06	110.38	175.65
11	HF 6-31G**	1.511	1.202	1.305	0.962	1.815	0.103	123.59	129.87	110.93	175.57
12	SVWN 3-21G*	1.497	1.271	1.304	1.155	1.324	0.033	125.44	121.21	114.57	178.10
13	SVWN 6-31G*	1.499	1.262	1.280	1.168	1.252	0.018	125.14	116.92	114.11	175.99
14	SVWN 6-31G**	1.499	1.277	1.260	1.154	1.259	0.017	124.88	117.74	114.19	176.77
15	B3LYP 3-21G*	1.517	1.251	1.337	1.045	1.562	0.086	124.43	128.27	111.82	175.43
16	B3LYP 6-31G*	1.520	1.230	1.323	1.005	1.677	0.093	124.34	125.55	109.81	179.67
17	B3LYP6-31G**	1.519	1.233	1.317	1.007	1.623	0.084	124.35	125.36	109.99	179.66
18	B3LYP 6-311G*	1.518	1.223	1.321	0.995	1.698	0.099	124.25	126.96	107.44	174.74
19	B3LYP 6-311G**	1.521	1.228	1.320	1.002	1.647	0.092	124.51	128.01	109.70	177.33
20	B3LYP 6-311+G**	1.520	1.220	1.321	0.995	1.741	0.101	123.41	126.22	110.89	178.30

[a] Experimental values, uncorrected to thermal motion, as were used in PLS models.

[b] Experimental values rounded together with estimated standard deviations (in parentheses).

bond classification given in Scheme 1 (with differences smaller than 0.06 for X-ray and HF 6-31G** optimized geometries). There are the following bond groups in IAA: (1) slightly shortened single C3—C8 and C8—C9 bonds (M/L orders ≈1) due to the presence of neighboring π-systems (indol and carbonyl, respectively); (2) lengthened double C9—O2 (M/L orders 1.8/2.1) and shortened single C9—O1 bond (M/L orders 1.0/1.3) as constitutive parts of resonance-assisted hydrogen bond ring [47]; (3) delocalized pyrrole CC bonds: C2—C3 is significantly different from pure single bond (M/L orders 1.7/1.6), whereas C3—C31 is between single and benzene bond (M/L orders 1.2); (4) CN bonds are obviously delocalized (1.37–1.38 Å, M/L orders 1.0–1.3) due to the participation of the nitrogen lone pair in electron delocalization; (5) benzenelike bonds: C31—C71, C4—C31, C7—C71, and C5—C6 are slightly lengthened with respect to benzene bond (M/L orders 1.3–1.4 compared with 1.5 for benzene), and bonds C4—C5 and C6—C6 are shortening (M/L orders above 1.5); this arrangement of bonds is a consequence of benzene ring deformation due to electron delocalization and of fusion of the two rings. The reader can find more detailed description of experimental IAA geometry in work by Nigović et al. [2]. The difference between data matrices for monomer (Table I) and dimer (Table II) confirms that the most sensitive part of IAA is the carboxyl group (differences up to 0.067 Å) when ab initio methods are employed. Other hydrogen bond parameters from different calculation methods and the experimental structure determination (Table III) support this finding. All these observations on IAA bond lengths are the platform for further PCA–HCA studies in this work.

HCA–PCA COMPARISON OF IAA MONOMER WITH DIMER

Coupled HCA and PCA for finding the best calculation method for some property [7] consists of simultaneous inspection of HCA dendograms and 2D or 3D PCA plots and finding out what these plots have in common with respect to the closest neighborhood of some reference (usually experimental) method. These plots frequently exhibit clustering of methods or objects (molecules, bonds, molecular fragments, etc.) according to some common property. This information might also be use-

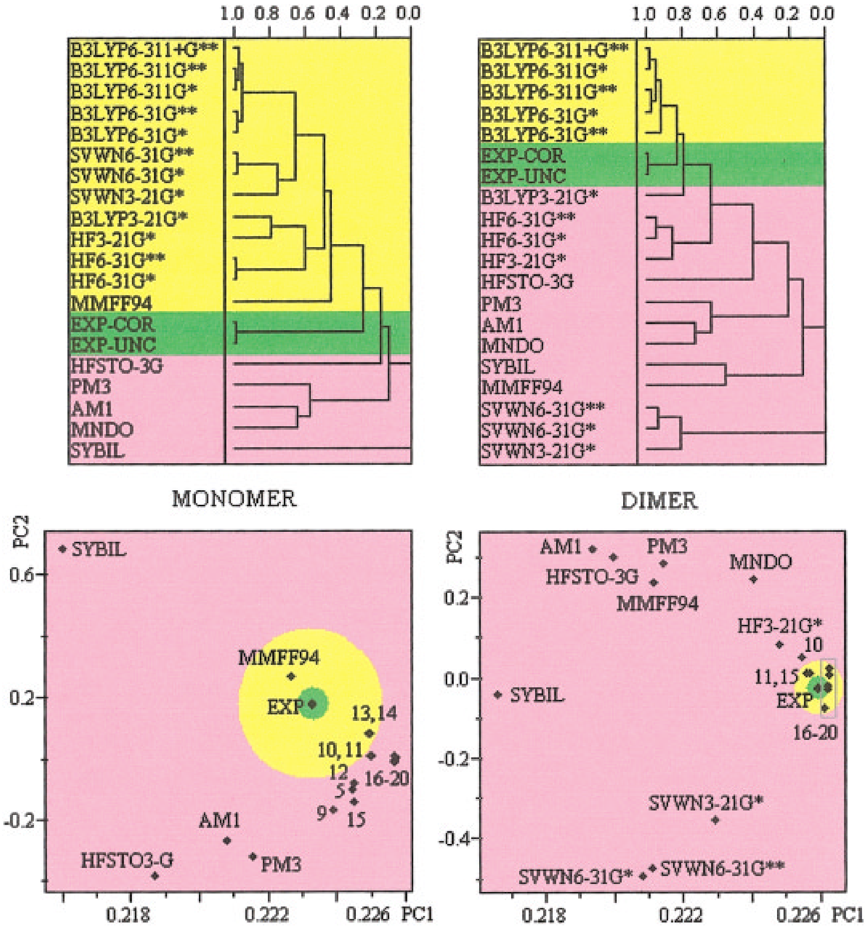


FIGURE 3. The HCA dendograms on variables (top) and PCA loadings plots (bottom) for IAA monomer and dimer. Green: experimental bond lengths (EXP) uncorrected (EXP-UNC) and corrected (EXP-COR) to thermal motion in crystal. Yellow field: calculation methods considered best for calculation of IAA bond lengths. Rose field: calculation methods not recommended to use for IAA bond lengths. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

ful, although not necessarily in every case. In this work, it is interesting to investigate what effect is provoked by hydrogen bonds in IAA dimer, with respect to the monomer. Dendograms on variables (Fig. 3, top), from HCA analysis using **X**(14, 20) data matrices, show that a large variety of ab initio methods (yellow field) are close to the experiment (green field) in case of monomer, whereas only B3LYP methods (except 3-21G*) are the closest to the experiment. Both SEM and MM methods are usually far from the experiment (rose field) and therefore not recommendable. The methods closest to the experiment make a cluster, with the highest similarity indices with respect the experiment. Principle component analysis for monomer and dimer shows that only the first PC (PC1) is adequate to describe over 97% of original data (Table IV). This is why 2D loading plots can be utilized (Fig. 3, bottom) to draw a circle around the experiment and include in it only the closest methods. For monomer, this circle contains MMFF94 and SVWN (6-31G* and 6-31G**) methods. The circle for dimer is much smaller, with ab initio methods more concentrated around the experiment; the closest methods are SVWN 6-31G** and all B3LYP methods. Both PCA and HCA plots with clustering of bonds can also help in this case. In Figure 4 it is evident that there are five types of IAA bonds, in agreement with the above bond type discussion: CO bonds (rose), single side-chain CC bonds (gray), NC bonds (green), delocalized shortened (blue), and lengthened (yellow) bonds. The dendogram for dimer is more ordered, with both CO bonds forming an isolated cluster, than that for the monomer. The PCA scores plot for dimer (Fig. 4, bottom) is also more ordered than for the

TABLE IV
PCA results for IAA monomer and dimer.

	Monomer				Dimer			
	X(14, 20)		**X**(20, 14)		**X**(14, 20)		**X**(20, 14)	
PCs	% Variance	% Cumul. var.	% Variance	% Cumul. var.	% Variance	% Cumul. var.	% Variance	% Cumul. var.
PC1	98.11	98.11	36.38	36.38	97.64	97.64	37.43	37.43
PC2	0.75	98.86	26.07	62.45	0.93	98.58	23.96	61.39
PC3	0.54	99.40	16.51	78.95	0.67	99.25	16.77	78.16
PC4	0.27	99.67	6.83	85.88	0.43	99.68	8.78	86.95
PC5	0.15	99.81	4.72	90.50	0.11	99.79	4.75	91.70

monomer: the clusters are better defined, and the bond lengths increase along PC1 (due to atom types and bond types) and PC2 (bond length variations inside the same bond type). Thus, calculation methods treating the dimer result in molecular geometry closer to experimental than those optimizing the monomer geometry. Both PCA and HCA on data matrices **X**(20, 14) for monomer and dimer (Fig. 5) can give useful information also. Dendograms on samples exhibit rather small cluster of SVWN (monomer) and B3LYP (dimer) methods as the closest to the experiment. Only the first five PCs contain over 90% of the total variance (Table IV), and so at least 3D space (PC1—PC2—PC3) is needed to investigate the neighborhood around the experiment. The best methods for monomer are SVWN (6-31G* and 6-31G**) and

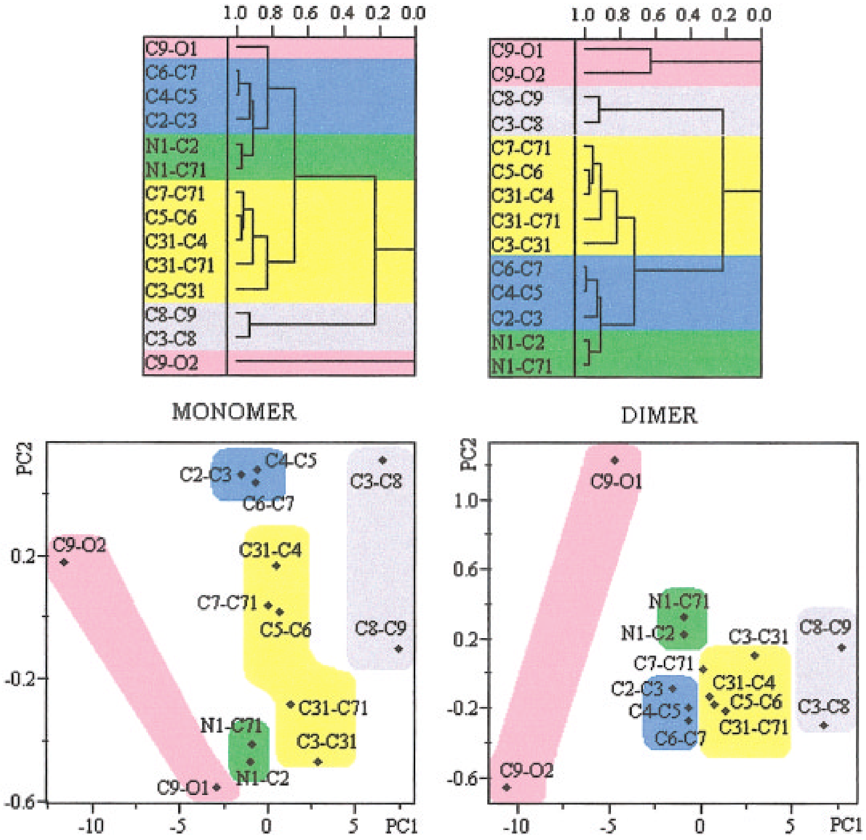


FIGURE 4. The HCA dendograms on samples (top) and PCA score plots (bottom) for IAA monomer and dimer. Coloring is the same as in Figure 3. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

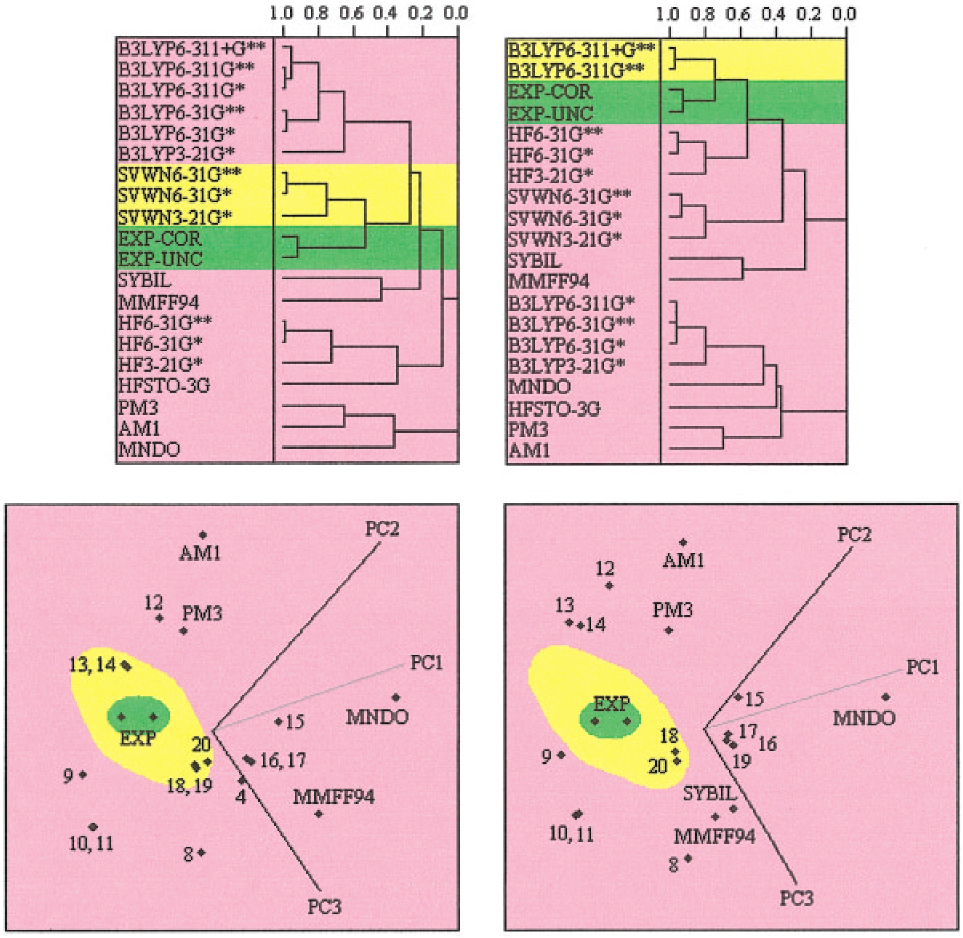


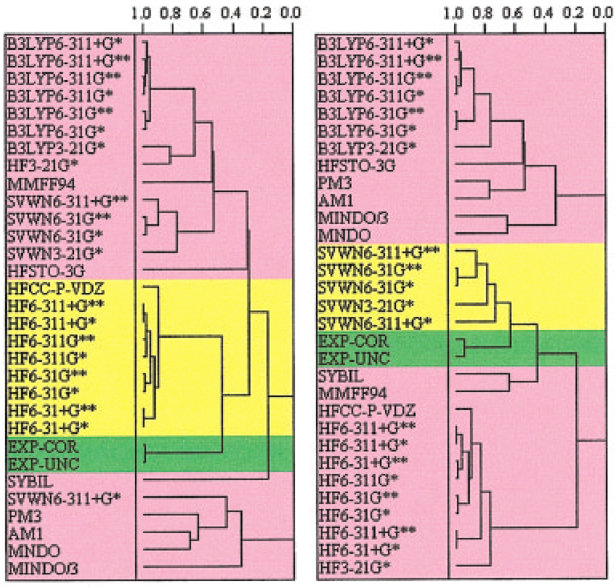
FIGURE 5. The HCA dendograms on samples (top) and PCA scores plots (bottom) for IAA monomer and dimer. Coloring is the same as in Figure 3. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

some B3LYP methods. For dimer, only these B3LYP methods remain. The HCA and PCA plots with bonds do not provide useful information besides showing two distinct clusters of formally double (C9—O2, C2—C3, C4—C5, and C6—C7) and single bonds (dendogram not shown). The final result of the

TABLE V

PCA results for IAA monomer and dimer hydrogen bond ring.

	Monomer				Dimer hydrogen bond ring			
	X(14, 31)		**X**(31, 14)		**X**(10, 20)		**X**(20, 10)	
PCs	% Variance	% Cumul. var.	% Variance	% Cumul. var.	% Variance	% Cumul. var.	% Variance	% Cumul. var.
PC1	98.17	98.17	41.83	41.83	97.64	97.64	52.25	52.25
PC2	0.63	98.80	23.41	65.23	0.93	98.58	17.01	69.27
PC3	0.51	99.31	13.78	79.02	0.67	99.25	10.50	79.76
PC4	0.28	99.59	9.30	88.95	0.43	99.68	7.45	87.22
PC5	0.16	99.75	3.47	92.41	0.11	99.79	5.77	92.98



(a) (b)

FIGURE 6. The HCA dendograms on variables (a) and on samples (b) for IAA monomer using **X**(14, 31) and **X**(31, 14) bond lengths data matrices, respectively. Coloring is the same as in Figure 3. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

complete HCA–PCA analysis reveals a quite visible effect of hydrogen bonds on the choice of calculation method that would be the best for IAA bond lengths: DFT methods with SVWN functional (6-31G* and 6-31G**) for monomer, and DFT-B3LYP methods (6-311G* and 6-311+G**) for dimer.

HCA–PCA ANALYSIS OF EXTENDED DATA FOR MONOMER

What is there are more bond lengths data, from SEM, MM, or ab initio, that can be added to **X**(14, 20) or **X**(20, 14) data matrices? Would they change the results of the previous HCA–PCA analysis? For this purpose, HCA and PCA analysis (Table V) on the complete data set for monomer (Table I) was performed, by using matrices **X**(14, 31) and **X**(31, 14). Visual inspection of dendograms shows that all HF (Hartree-Fock) methods except STO-3G are closest to the experiment for the former matrix, and all SVWN methods for the latter matrix (Fig. 6). Inspection of the 2D and 3D PCA plots (not shown) reveals that some B3LYP (6-311G*, 6-311G*, and 6-31+G**) and SVWN (6-31G* and 6-31G**) methods are the best when **X**(14, 31) matrix is used, and SVWN 3-21G* and B3LYP 3-21G* methods when the analysis is performed on **X**(31, 14) matrix. The methods that appear with the highest frequency (twice) are SVWN 6-31G* and SVWN 6-31G**, the same methods as when calculation methods for IAA monomer and dimer were compared.

HCA–PCA OF THE HYDROGEN BOND RING

In Figure 7 are shown the dendograms for the dimer hydrogen bond ring data (Table III). It is clear that when performing HCA on **X**(10, 20) data matrix, some B3LYP (6-311+G**, 6-31G**, 6-31G*) methods and surprisingly HF STO-3G appear as the closest ones to the experiment. When **X**(20, 10) data matrix is used in the same analysis, some HF (6-31G**, 6-31G*, 3-21G*) methods seem to be the best ones. When PCA is carried out on the same data matrices (Table V), 2D gives B3LYP (6-311+G**, 6-31G**, 6-31G*) and HF STO-3G as the best for **X**(10, 20) matrix. The best methods when PCA is applied to **X**(20, 10) are HF6-31G* and HF6-31G**, as shown in Figure 7, bottom. Eliminating HF STO-3G from the methods with maximal frequency (two), one obtains as two choices: B3LYP (6-311+G**, 6-31G**, 6-31G*) and HF (6-31G**, 6-31G*).

REPRODUCIBILITY OF PARTICULAR HYDROGEN BOND AND AROMATICITY PARAMETERS

After finding the best methods to reproduce experimental bond lengths, one can ask the following question about some structure correlation for the hydrogen bond ring or some hydrogen bond or aromaticity parameters calculated from bond lengths: What are the best methods in such cases? Some examples of new relationships are presented in Figure 8. The hydrogen-acceptor distance e (Scheme 2) is a hyperbolic function of the difference between C═O and C—O bond lengths $|b - c|$ [Fig. 8(a)]. Theoretically, the free carboxylate ion in gas–phase, that is, without the presence of hydrogen bonds, could determine $e = 1.015$ Å (reaches the O—H bond length from neutron diffraction [8]), meaning that $|b - c| = 0$ and the hydrogen ion left the parent O atom (during the dissociation of the carboxylate group) and bound to the O atom of the other IAA molecule of the dimer. On the other hand, when $|b - c| = 0.150$ Å (the value for free acetic acid [6]), e goes to infinity. This theoretical e, called p, is defined as $p = 0.15225/(0.15 - |b - c|)$ Å, and its correlation with e is presented in Figure 8(b).

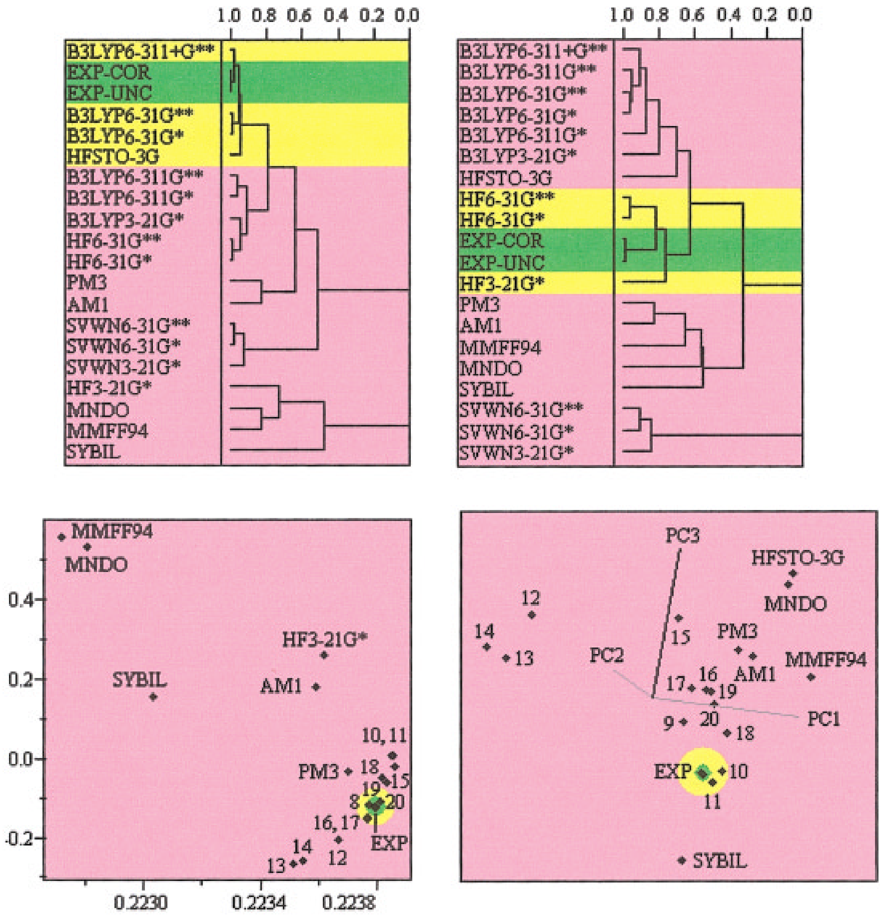


FIGURE 7. Top: HCA dendograms on variables (left) and on samples (right) for the dimer hydrogen bond ring using **X**(14, 31) and **X**(31, 14) bond lengths data matrices, respectively. Bottom: 2D loadings and 3D scores plot generated by PCA on the same data matrices, respectively. Coloring is the same as in Figure 3. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

The parameter $D = [d^2 + e^2 - 2\, d\, e \cos \delta]^{1/2}$ is the O . . . O (donor–acceptor) distance, and its correlation with e is shown in Figure 8(c). These three plots exhibit regularity even when various calculation methods were applied; SEM and MM are methods with maximum deviation around a curve or line. The B3LYP methods seem to be the closest to the experimental result and are recommended for the IAA hydrogen bond studies. Based on the dimer experimental structure [2], a delocalization model of the IAA hydrogen bond dimer is proposed [48] in the light of the resonance theory [49, 50] (Fig. 9). Resonance structures II, V, and VI have small but important nonzero contributions (in terms of the contribution errors derived from experimental errors) to the resonance hybrid. They might be responsible for significant difference between $|b - c|$ for free and dimer acid, which is in the case of formic acid equal to 0.103 Å [6]. The reader may also note that all methods reproduce longer C9═O2 bonds in dimer than in monomer and shorter C9—O1 in dimer than in monomer. The structure B3LYP 6-31G** dimer reveals interesting electronic and steric relationships between donor, hydrogen, and acceptor, which is a characteristic of medium-strength hydrogen bonds in organic compounds (Fig. 10). Aromaticity parameters for indole ring can also reveal which methods are the best for calculation of these parameters in monomer and dimer. Julg's structural aromaticity index [51, 52] is an important parameter describing bond length equalization, a structural descriptor of the (hetero)aromaticity phenomenon. It is defined as $A = 1 - 255(\sigma/Q)^2$ where σ, Q are standard deviation and the mean of the (hetero)aromatic bond lengths under study, respectively. For ideally delocalized

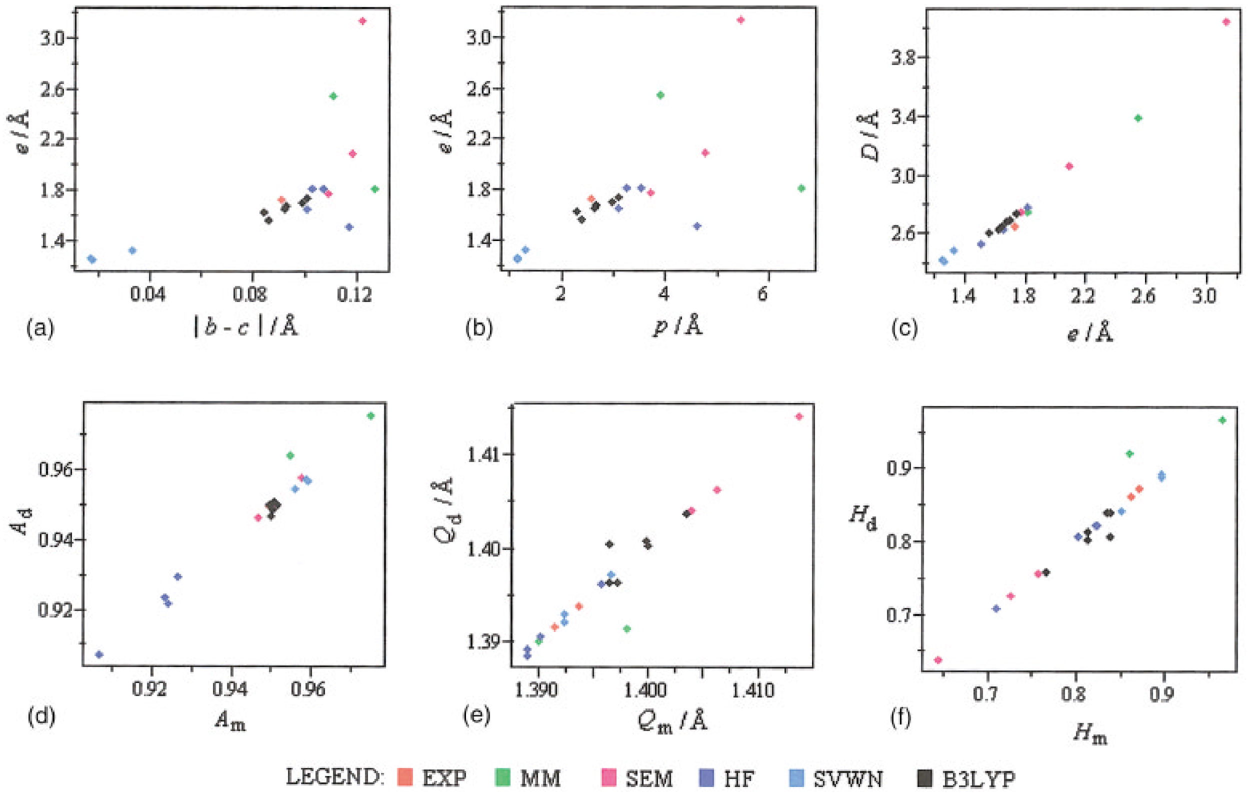


FIGURE 8. (a–c) Relationships between some hydrogen bond structural parameters. (d–f) Relationships between some structural aromaticity indices for monomer and dimer. See text for definition of parameters. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

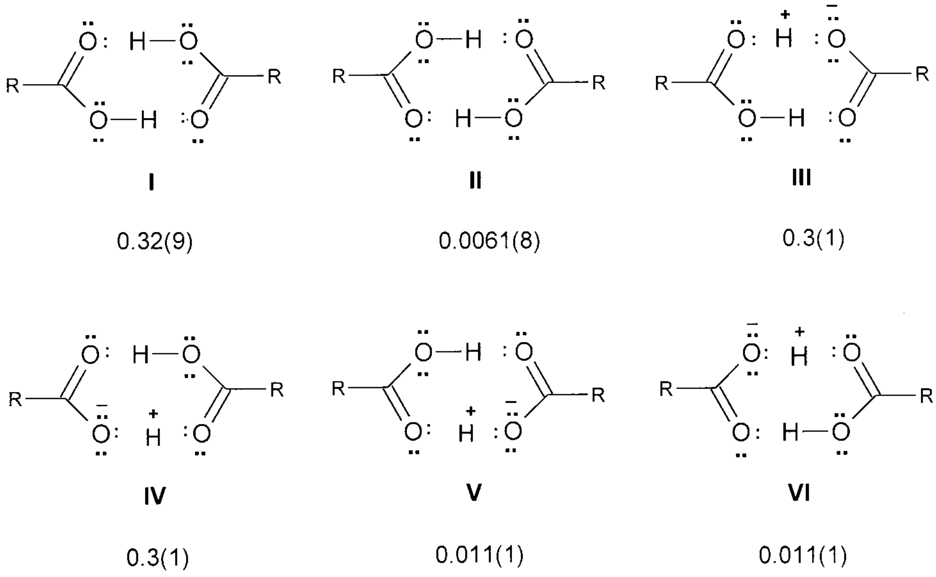


FIGURE 9. Proposed resonance structures of the IAA hydrogen bond dimer and their contributions to the hybrid.

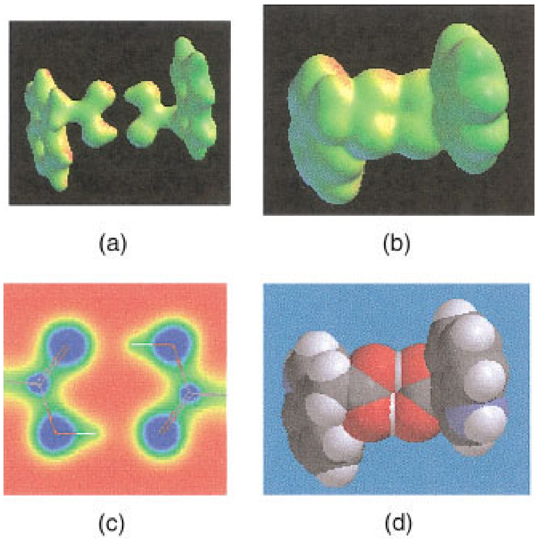


FIGURE 10. The B3LYP 6-31G**2 IAA dimer. (a) Bond surface model (0.08 electrons/au^3 electron isodensity surface). (b) Size surface model (0.002 electrons/au^3 electron isodensity surface). (c) Slice model (electron isodensity cross-section) in the plane of the hydrogen bond ring. (d) Space-filling model. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

systems, such as benzene in vacuum, $A = 1$. Indole bond lengths show that indole, benzene, and pyrrole rings are highly aromatic systems, because A is 0.94, 0.98, and 0.90, respectively [48]. The HOMA (Harmonic Oscillator Model of Aromaticity) aromaticity index [53] H, is another important structural aromaticity index. It is defined for indole ring as $H = 1 - (257.7\ H_{CC} + 93.52\ H_{CN})/10$, where H_{CC} and H_{CN} are equal to sums $\Sigma_i(B_i - B_0)$ over CC and CN bond lengths, respectively. Parameter B_i is the actual CN or CC bond length, and B_0 is the standard bond parameter (1.334 for CN and 1.388 Å for CC bond). Plots of monomer against dimer aromaticity parameter [m and d indices in Fig. 8(d–f)] show the best calculation methods for each parameter: B3LYP methods are best for the Julg's index A, and SVWN methods for Q and H, both for monomer and dimer.

Conclusions

Chemometric analysis of the computational results on bond lengths and other structural parameters of IAA monomer and dimer, and IAA dimer hydrogen bond ring clearly shows that ab initio methods are superior to MM and SEM methods: SVWN are the best for monomer, and B3LYP for dimer geometry optimization. Both B3LYP and HF can be used equally well for optimization of the dimer ring geometry. Other hydrogen bond and aromaticity structural parameters exhibit preference for either B3LYP or SVWN methods. Besides, grouping of bonds and methods into clusters is in accordance with structural properties of IAA and characteristics of the methods (methods of the same kind tend to make clusters). The results in this article confirm that the hydrogen bonds in the IAA dimer are resonance-assisted medium-strength hydrogen bonds, where carboxylic groups play role of resonance system.

ACKNOWLEDGMENT

The authors acknowledge FAPESP for financial support.

References

1. Arteca, R. Plant Growth Substances: Principles and Applications; Chapman & Hall: New York, 1996.
2. Nigović, B.; Antolić, S.; Kojić-Prodić, B.; Kiralj, R.; Magnus, V.; Salopek-Sondi, B. Acta Cryst 2000, B56, 94–111.
3. Woo, E. J.; Marshall, J.; Bauly, J.; Chen, J. G.; Venis, M.; Napier, R. M.; Pickersgill, R. W. EMBO J 2002, 21, 2877–2885.
4. Davies, D. T. Aromatic Heterocyclic Chemistry; Oxford: Oxford, 1995; Chapter 7.
5. Gilchrist, T. L. Heterocyclic Chemistry, 3rd edition; Addison Wesley Longman; Harlow, UK, 1997.
6. Lide, D. R., Ed. CRC Handbook of Chemistry and Physics 1999–1998, 78th edition; CRC Press: Boca Raton, 1997.
7. Kiralj, R.; Ferreira, M. M. C. J Chem Inf Comput Sci (to be published).
8. Allen, F. H.; Kennard, O.; Watson, D. G.; Brammer, L.; Orpen, A. G.; Taylor, R. J Chem Soc Perkin Trans II 1987, S1–S19.
9. Hess, B. A., Jr.; Schaad, L. J Tetrahedron Lett 1977, 535–538.
10. Chacko, E.; Bornstein, J.; Sardella, D. J Tetrahedron Lett 1977, 1095–1098.
11. George, P. Chem Rev 1975, 75, 85–111.
12. Dewar, M. J. S.; Gleicher, G. J. J Phys Chem 1966, 44, 759–773.
13. Trinajstić, N. Molekularne orbitale u kemiji; Skolska knjiga; Zagreb; 1974; p. 90.
14. Glusker, J. P.; Lewis, M.; Rossi, M. Crystal Structure Analysis for Chemists and Biologists; VCH: New York, 1994.
15. Dannenberg, J. J. J Mol Struct (Theochem) 1997, 401, 279–286.
16. Titan Tutorial and User Guide; Schrodinger: Portland, OR; and Wavefunction; Irvine, CA, 1999.
17. Müller-Dethlefs, K.; Hobza, P. Chem Rev 2000, 100, 143–167.

18. Ramek, M.; Tomić, S.; Kojić-Prodić, B. Int J Quant Chem Quant Biol Symp 1995, 22, 75–81.
19. Nagata, H.; In, Y.; Doi, M.; Ishida, T.; Wakahara, A. Acta Cryst 1995, B51, 1051–1058.
20. Ishida, T.; Usami, H.; Inoue, M.; Yamagata, Y.; Tomita, K. Biochem Biophys Res Comm 1982, 107, 746–750.
21. Yamagata, Y.; Kato, M.; Fujii, S. Chem Pharm Bull 1994, 42, 2385–2387.
22. Ishida, T.; Inoue, M. Biochem Biophys Res Comm 1981, 99, 149–154.
23. Sakaki, T.; Sogo, A.; Wakahara, A.; Kanai, T.; Fujiwara, T.; Tomita, K.-I. Acta Cryst 1976, 32, 3235–3242.
24. Lynch, D. E.; Latif, T.; Smith, G.; Byriel, K. A.; Kennard, C. H. L.; Parsons, S. Aust J Chem 1998, 51, 403–408.
25. Lynch, D. E.; Mistry, N.; Smith, G.; Byriel, K. A.; Kennard, C. H. L. Aust J Chem 1998, 51, 813–818.
26. Gavezzotti, A. Acc Chem Res 1994, 27, 309–314.
27. Bernstein, J.; Etter, M. C.; Leiserowitz, L. In Structure Correlation, Vol. 2; Bürgi, H.-B.; Dunitz, J. D., Eds.; VCH: New York, 1994; Chapter 11.
28. Bürgi, H.-B. Acta Cryst 1998, A54, 873–885.
29. Spek, A. L. PLATON, A Multipurpose Crystallographic Tool, v. 31000; Utrecht University: Utrecht, 2000.
30. Bingham, R. C.; Dewar, M. J. S.; Lo, D. H. J Am Chem Soc 1975, 97, 1285–1293.
31. Dewar, M. J. S.; Thiel, W. J Am Chem Soc 1977, 99, 4899–4907.
32. Dewar, M. J. S.; Zoebisch, E. G.; Healy, G. F.; Stewart, J. J. P. J Am Chem Soc 1985, 107, 3902–3909.
33. Stewart, J. J. P. J Comp Chem 1989, 10, 209–220.
34. Stewart, J. J. P. MOPAC 6.0; Frank Seiler Research Laboratory, Air Force Academy; Colorado Springs, CO, 1990.
35. Titan v. 1; Wavefunction; Irvine, CA, 2000.
36. Halgren, T. A. J Comp Chem 1996, 17, 450–519.
37. Clark, M.; Cramer, R. D., III; van Opdensch, N. J Comp Chem 1989, 10, 982–1012.
38. Jones, R. O.; Gunnarsson, O. Rev Mod Phys 1989, 61, 689–746.
39. Vosko, S. H.; Wilk, L.; Nusair, M. Can J Phys 1980, 58, 1200–1211.
40. Becke, A. D. Phys Rev 1988, A38, 3098–3100.
41. Lee, C.; Yang, W.; Parr, R. G. Phys Rev 1988, B37, 785–789.
42. Malinowski, E. R. Factor Analysis in Chemistry, 2nd edition; Wiley: New York, 1991.
43. Beebe, K. R.; Pell, R. J.; Seasholtz, M. B. Chemometrics: A Practical Guide; Wiley: New York, 1998.
44. Ferreira, M. M. C.; Antunes, A. M.; Melgo, M. S.; Volpe, P. L. O. Quím Nova 1999, 22, 724–731.
45. Pirouette, Multivariate Data Analysis for IBM PS Systems, v. 2.7; Infometrix: Woodinville, WA, 2000.
46. Jeffrey, G. A.; Ruble, J. R.; McMullan, R. K.; Pople, J. A. Proc R Soc London, 1987, A414, 47–57.
47. Bertolasi, V.; Gilli, P.; Ferreti, V.; Gilli, G. Acta Cryst 1995, B51, 1004–1015.
48. Ferreira, M. M. C.; Kiralj, R. (unpublished results).
49. Pauling, L. The Nature of the Chemical Bond, 3rd edition; Cornell University Press: Ithaca, 1972; p. 452.
50. Jeffrey, G. A. An Introduction to Hydrogen Bonding; Oxford University Press: New York, 1997; p. 17.
51. Lewis, D.; Peters, D. Facts and Theories of Aromaticity, 1st edition; MacMillan; London, 1975; p. 10–14.
52. Schleyer, P. R.; Freeman, P. K.; Jiao, H.; Goldfuss, B. Angew Chem Int Ed Engl 1995, 34, 337–340.
53. Kruszewski, J.; Krygowski, T. M. Can J Chem 1975, 53, 945–951.