

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/220417262>

# Geometry of Pyrazole: A Test for Ab Initio Calculations.

ARTICLE *in* JOURNAL OF COMPUTATIONAL CHEMISTRY · MARCH 1995

Impact Factor: 3.59 · DOI: 10.1002/jcc.540160302 · Source: DBLP

---

CITATIONS

72

READS

36

## 6 AUTHORS, INCLUDING:



Antonio L. Llamas-Saiz

University of Santiago de Compostela

159 PUBLICATIONS 2,401 CITATIONS

[SEE PROFILE](#)



Otilia Mó

Universidad Autónoma de Madrid

402 PUBLICATIONS 6,371 CITATIONS

[SEE PROFILE](#)



Manuel Yanez

Universidad Autónoma de Madrid

270 PUBLICATIONS 3,741 CITATIONS

[SEE PROFILE](#)



José Elguero

Spanish National Research Council

1,502 PUBLICATIONS 22,151 CITATIONS

[SEE PROFILE](#)

Volume 16 Number 3 March 1995 263–272

# The Geometry of Pyrazole: A Test for *Ab Initio* Calculations

ANTONIO L. LLAMAS-SAIZ\* and CONCEPCIÓN FOCES-FOCES

Departamento de Cristalografía, Instituto Rocasolano, CSIC, Serrano, 119, E-28006 Madrid, Spain

OTILIA MÓ and MANUEL YÁÑEZ

Departamento de Química Física, Universidad Autónoma de Madrid, E-28049 Cantoblanco, Spain

ERIC ELGUERO

Laboratoire d'Hydrologie, ORSTOM, BP 5045, 34032 Montpellier, France

JOSÉ ELGUERO

Instituto de Química Médica, CSIC, Juan de la Cierva, 3, E-28006 Madrid, Spain

Received 11 April 1994; accepted 27 July 1994

## ABSTRACT

*Ab initio* calculations on the structure of pyrazole have been carried out at different levels of accuracy. At the Hartree-Fock (HF) level, the performance of several basis sets, namely 3-21G, 6-31G, 6-31G<sup>\*\*</sup>, and 6-311G<sup>\*\*</sup> was investigated. The influence of electron correlation effects also was studied by carrying out geometry optimizations at the MP2, MP4, and QCISD levels. The performance of a density functional method also was evaluated. We have also investigated the possible influence of the frozen core approximation on the final optimized geometry. Three different statistical analyses were considered in determining which geometry is closest to the experimental microwave geometry—namely Paul Curtin's diagrams, cluster analysis, and multidimensional scaling. From these analyses, we conclude that there is no asymptotic approach to the experimental geometry by increasing the quality of the theoretical model, although, as expected, the more reliable structures are those obtained at the MP2, MP4, and QCISD levels, as well as those obtained by the B3LYP density functional method. We have also found that the values of the rotational constants are a tight criterion to define the quality of a molecular geometry. © 1995 by John Wiley & Sons, Inc.

\*Author to whom all correspondence should be addressed.

## Introduction

Our interest in the geometries of pyrazoles in the solid state (crystallography and neutron diffraction) led us to examine the theoretically calculated geometries.<sup>1</sup> When comparing the geometry determined in the solid state with that calculated for the isolated molecule, everyone is aware of the distortions produced by the crystal packing (for instance, in the case of pyrazole itself by the hydrogen bond network). A less well studied problem concerns the accuracy of the calculated geometry used for comparison.

When a comparison is between a theoretically calculated and a gas phase determined geometry (usually by microwave spectroscopy), the agreement between both structures is usually presented as average differences between bond lengths and bond angles.<sup>2</sup> However, this is not a precise way to define a good geometry, because different structures may lead to similar average deviations. On the other hand, *ab initio* treatments yield in general good estimates of  $r_e$  structures (i.e., structures of vibrationless molecules), whereas spectroscopic geometries are obtained from the observed rotational constants ( $r_0$  structures) or differences in moments of inertia ( $r_s$  structures) for a series of isotopic species, which depend on the vibrational state of the molecule.<sup>3</sup> Because most molecules are in the ground vibrational state ( $v = 0$ ), the most commonly determined rotational constants and the corresponding effective moments of inertia define an  $r_0$  structure. Recently, Smith et al.<sup>4</sup> have proposed a method to calculate theoretical  $r_0$  structures. This method implies, essentially, corrections which are obtained from calculations of a set of systems related to the one under investigation and which permit to take into account both the systematic deficiencies of the theory and the differences between  $r_e$  and  $r_0$  structures with regard to the directly calculated *ab initio* ones.

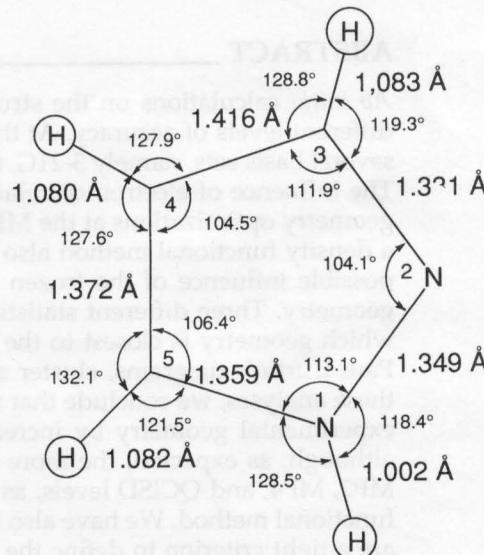
Unfortunately, it is not always easy to find a set of related compounds which fulfill the necessary condition that the relevant environment is as similar as possible to that of the target molecule. A further difficulty is related to the existence of dependent parameters. The coupling between the internal coordinates chosen to define the molecular geometry makes it impossible to define a unique set of parameters to which apply the aforementioned corrections. Pyrazole is a paradigmatic example of the situation because it has five bonds

and five angles but only seven degrees of freedom (considering only nonhydrogen atoms). In such cases, and from the theoretical point of view, only the  $r_e$  structure would be attainable. In this respect, it should be taken into account that the sensitivity of the calculated geometries to the level of theory used is sizeable. In some cases, for instance, the effect of electron correlation on bond lengths and bond angles is as large as 0.03 Å for the former and 2° for the latter.<sup>4</sup> Hence, the question of which calculated geometry is the best still remains to be answered. This automatically implies a second one: Which are the best criteria for preferring one geometry over another?

The aim of this article is to discuss these problems using as a model the pyrazole geometry ( $C_3H_4N_2$ ) for which an accurate microwave spectroscopy (MW) geometry is available (Fig. 1).<sup>5</sup>

## Computational Details

The geometry of pyrazole was optimized at different levels of accuracy by using the Gaussian 90 series of programs.<sup>6</sup> Because one of the aims of this article is to analyze the reliability of the different approaches usually employed in *ab initio* calculations, we have carried out these geometry optimizations first at the HF level by using five different basis sets—namely, 3-21G, 6-31G, 6-31G\*, 6-31G\*\*, and 6-311G\*\*. Electron correlation effects



**FIGURE 1.** The microwave structure of pyrazole with atom numbering.

were taken into account by using either the Møller-Plesset perturbation theory at second and fourth orders (MP2 and MP4) or the QCISD formalism. At the second order, we have discussed the results obtained with a 6-31G\*\* basis set, which is the most commonly used for these kinds of geometry optimizations, and those obtained with a 6-311G\*\* basis, which was specially designed for use with correlated wave functions and to which an additional set of polarization functions on the heavy atoms was added to obtain a 6-311G(2d, p) basis set. We have also studied the influence of diffuse functions in the basis set by performing geometry optimizations at the MP2/6-311 + G\*\* level. To investigate also the effect of the frozen core approximation on the optimized geometries, the MP2 optimizations have been carried out with and without keeping frozen the core electrons. The different energies (in Hartrees) obtained with the different procedures are collected in Table I.

Because, as we shall discuss later, the differences between MP2/6-31G\*\* and MP2/6-311G(2d, p) geometries are not significant, the MP4 and QCISD geometry optimizations were carried out only on the smaller (6-31G\*\*) basis set. The MP4 optimization was restricted to the internal coordinates which define the azole ring (i.e., the C—H and N—H bond lengths), and the angles involving these bonds were kept constant and equal to the values obtained in the MP2/6-31G\*\* optimization. In all the other cases, the only constraint imposed during the optimization was the planarity of the ring (consistent with the MW structure).<sup>5</sup>

**TABLE I.**  
**Calculated Energies (Hartrees) for Pyrazole.**

Theoretical level	Energy
HF / 3-21G	-223.52552 <sup>a</sup>
HF / 6-31G	-224.70975 <sup>b</sup>
HF / 6-31G**	-224.80328
HF / 6-311G**	-224.84880
MP2(FC) / 6-31G**	-225.53642
MP2(FULL) / 6-31G**	-225.55920
MP2 / 6-311 + G(d, p)	-225.62289
MP2(FULL) / 6-311G(2d, p)	-225.75942
MP4(SDTQ) / 6-31G**	-225.59268 <sup>c</sup>
QCISD / 6-31G**	-225.55861

<sup>a</sup> From M. Alcamí, O. Mó, and M. Yáñez, *J. Phys. Chem.*, **93**, 3929 (1989).

<sup>b</sup> From O. Mó, J. L. G. de Paz, and M. Yáñez, *J. Phys. Chem.*, **90**, 5597 (1986).

<sup>c</sup> Value obtained by a geometry optimization restricted to the pentagon of heavy atoms.

When this work was almost finished, a geometry optimization of pyrazole using density functional theory was reported.<sup>7</sup> For the sake of completeness, we have considered it of interest to include these results, obtained using the 6-31G\* basis set, in our statistical analyses. Because, as we shall discuss in forthcoming sections, this reported geometry is not significantly different from those obtained at the MP2, MP4, and QCISD levels, it would be of interest to investigate also the performance of this theoretical scheme. Hence, we have also included in the statistical analysis which follows the optimized geometries (/6-31G\* and /6-31G\*\*) obtained with the Becke3LYP functional implemented in the Gaussian-92-DFT (density functional theory) package. These DFT optimized geometries were graciously provided to us by Professor P. v. R. Schleyer.

Another aspect of the geometry optimization problem, which is particularly relevant in this case, is the possible coupling between the internal coordinates chosen to define the molecular geometry.<sup>8</sup> To investigate this point, we have used for the MP2/6-31G\*\* optimizations two different sets of internal coordinates. In the first case, the azole ring is described as a chain of atoms which correspond to a set of strongly coupled internal coordinates. In the second case, the ring is defined as a set of atoms radially bonded to a dummy atom located in the middle of the ring, which is a less coupled coordinate system.

All optimizations at the HF level were followed by evaluation of the harmonic vibrational frequencies by the analytical second derivatives technique to assure that the optimized geometries correspond to true minima of the potential energy surface.

Because, as we shall show later, the rotational constants obtained at the MP2/6-31G\*\* and MP2/6-311G(2d, p) levels of accuracy, respectively, are similar, the rotational constants of the different isotopic pyrazoles were evaluated only at the former level. To do so, and for the sake of consistency with the experimental microwave description of pyrazole, all C—D and N—D bonds have been shortened by 0.003 Å.<sup>5</sup>

## Results and Discussion

### GEOMETRIES AND ROTATIONAL CONSTANTS

The geometries of pyrazole are reported in Table II. Only the data corresponding to the microwave

**TABLE II.**  
**Experimental (MW) and Theoretical Geometries (Bond Lengths in Å, Bond Angles in Degrees) of Pyrazole.**

Method	N <sub>1</sub> N <sub>2</sub>	N <sub>2</sub> C <sub>3</sub>	C <sub>3</sub> C <sub>4</sub>	C <sub>4</sub> C <sub>5</sub>	C <sub>5</sub> N <sub>1</sub>	N <sub>1</sub> H <sub>1</sub>	C <sub>3</sub> H <sub>3</sub>	C <sub>4</sub> H <sub>4</sub>	C <sub>5</sub> H <sub>5</sub>
MW <sup>a</sup>	1.349(10)	1.331(5)	1.416(5)	1.372(5)	1.359(5)	1.002(10)	1.083(10)	1.080(10)	1.082(10)
3-21G	1.379	1.313	1.419	1.363	1.353	—	—	—	—
6-31G	1.3519	1.3146	1.4174	1.3671	1.3528	0.9877	1.0649	1.0653	1.0656
6-31G*	1.330	1.302	1.413	1.363	1.341	—	—	—	—
6-31G**	1.3299	1.3021	1.4130	1.3626	1.3411	0.9914	1.0714	1.0692	1.0702
6-311G**	1.3277	1.3000	1.4144	1.3624	1.3413	0.9912	1.0713	1.0690	1.0702
MP2/6-31G**	1.3478	1.3474	1.4055	1.3846	1.3599	1.0067	1.0773	1.0756	1.0764
MP2/6-311G(2d, p)	1.3419	1.3443	1.4040	1.3832	1.3562	1.0081	1.0791	1.0780	1.0785
MP2/6-311 + G(d, p)	1.3430	1.3476	1.4103	1.3891	1.3615	1.0096	1.0811	1.0797	1.0803
MP4/6-31G**	1.3583	1.3449	1.4146	1.3868	1.3634				
QCISD / 6-31G**	1.3531	1.3324	1.4160	1.3790	1.3587	1.0038	1.0778	1.0757	1.0767
DZVP2	1.340	1.337	1.410	1.385	1.357	1.022	1.093	1.091	1.092
B3LYP / 6-31G*	1.350	1.333	1.414	1.381	1.359	1.009	1.082	1.080	1.081
B3LYP / 6-31G**	1.350	1.333	1.414	1.381	1.359	1.008	1.082	1.079	1.080
Method	N <sub>1</sub> N <sub>2</sub> C <sub>3</sub>	N <sub>2</sub> C <sub>3</sub> C <sub>4</sub>	C <sub>3</sub> C <sub>4</sub> C <sub>5</sub>	C <sub>4</sub> C <sub>5</sub> N <sub>1</sub>	C <sub>5</sub> N <sub>1</sub> N <sub>2</sub>	N <sub>2</sub> C <sub>3</sub> H <sub>3</sub>	C <sub>3</sub> C <sub>4</sub> H <sub>4</sub>	C <sub>4</sub> C <sub>5</sub> H <sub>5</sub>	C <sub>5</sub> N <sub>1</sub> H <sub>1</sub>
MW <sup>a</sup>	104.1(3)	111.9(3)	104.5(3)	106.4(3)	113.1(3)	119.3(5)	127.9(5)	132.1(5)	128.5(5)
3-21G	104.5	118.8	105.0	107.1	111.6	—	—	—	—
6-31G	105.23	111.12	104.88	106.60	112.16	120.42	127.47	131.14	128.56
6-31G*	104.9	111.7	103.9	106.6	112.8	—	—	—	—
6-31G**	104.99	111.68	103.91	106.63	112.79	120.05	128.33	131.58	127.84
6-311G**	105.13	111.63	103.85	106.63	112.76	120.06	128.37	131.54	127.83
MP2/6-31G**	103.38	112.09	104.89	105.73	113.91	119.14	128.16	132.39	127.53
MP2/6-311G(2d, p)	103.73	111.85	104.82	105.83	113.77	119.19	128.26	132.27	127.41
MP2/6-311 + G(d, p)	103.83	111.83	104.73	105.74	113.87	119.17	128.26	132.50	127.53
MP4/6-31G**	103.43	112.31	104.61	106.07	113.58				
QCISD / 6-31G**	103.80	112.19	104.44	106.25	113.32	119.48	128.30	131.99	127.75
DZVP2	104.3	111.7	104.6	105.9	113.5	—	—	—	—
B3LYP / 6-31G*	103.88	112.15	104.48	106.11	113.38	119.40	128.26	132.06	127.89
B3LYP / 6-31G**	103.91	112.12	104.49	106.13	113.34	119.47	128.26	131.99	127.80

<sup>a</sup>For the errors, see Landolt-Börnstein, New Series, Vol. II / 7, p. 228 (compound no. 371); K. H. and H. A. Hellwege, Eds., Structure Data of Free Polyatomic Molecules, Springer-Verlag, Berlin, 1976.

geometry<sup>5</sup> and those of the 3-21G and 6-31G\* calculations<sup>9</sup> are from the literature. The experimental (MW)<sup>5</sup> and calculated rotational constants are reported in Table III.

Table II apparently shows that the most significant changes regarding optimized bond lengths and, to a lesser degree, optimized bond angles take place when electron correlation effects are taken into account. However, as we shall show in the next section, increasing the accuracy of the theoretical model, either by enlarging the basis set or by including electron correlation effects, does not result in a parallel improvement of the agreement between theoretical and experimental geometrical parameters. On the other hand, there is not a unique criterion with which to decide on the goodness of an optimized geometry.

It may be also observed that at the second-order correlation treatment, the N<sub>1</sub>—N<sub>2</sub> and N<sub>2</sub>—C<sub>3</sub> bond lengths are predicted to be of equal length. This is the most significant difference with respect to both the experimental MW values and the results obtained at the HF level.

This shortcoming of MP2 calculations disappears when higher order terms are included in the theoretical treatment. As shown in Table II, at the MP4 and at the QCISD levels, both bond lengths are clearly different and close to the experimental values. It can be also observed that at the MP4 and QCISD levels, bond angles are similar to the MP2 ones. Significantly, these effects are also nicely reproduced by DFT. The B3LYP optimized geometry is actually closer to the MP4 and QCISD optimized geometries than that at MP2. The inclusion

**TABLE III.**  
**Rotational Constants (MHz) of Isotopic Pyrazoles.**

Rotational constants	Parent	1D	3D	4D	5D	$1^{-15}\text{N}$	$2^{-15}\text{N}$	$3^{-13}\text{C}$	$4^{-13}\text{C}$	$5^{-13}\text{C}$
A (Exp.)	9618.770	9455.230	9435.783	9566.176	9537.304	9488.641	9618.304	9457.566	9582.358	9571.671
A (Calc.)	SCF <sup>a</sup> 9877.78									
	MP2 <sup>b</sup> 9645.33	9469.62	9424.47	9596.77	9539.03	9518.13	9642.81	9454.61	9612.71	9584.35
	MP2 <sup>c</sup> 9642.67									
	MP2 <sup>d</sup> 9626.18	9448.28	9404.42	9576.48	9520.17	9497.90	9623.81	9435.99	9592.92	9565.74
	MP2 <sup>e</sup> 9590.77									
	QCISD 9649.26									
	B3LYP 9638									
B (Exp.)	9412.535	8859.733	8774.190	8617.847	8677.868	9339.851	9180.010	9340.094	9193.394	9223.004
B (Calc.)	SCF <sup>a</sup> 9645.74									
	MP2 <sup>b</sup> 9406.44	8864.75	8805.79	8614.17	8694.35	9331.58	9173.15	9362.57	9184.15	9227.63
	MP2 <sup>c</sup> 9439.04									
	MP2 <sup>d</sup> 9385.57	8843.80	8783.54	8592.34	8671.11	9312.21	9152.62	9340.66	9164.46	9206.73
	MP2 <sup>e</sup> 9375.76									
	QCISD 9388.48									
	B3LYP 9400									
C (Exp.)	4755.853	4572.847	4545.189	4532.324	4542.381	4705.424	4695.610	4697.835	4690.511	4695.634
C (Calc.)	SCF <sup>a</sup> 4880.19									
	MP2 <sup>b</sup> 4762.19	4578.60	4552.32	4539.48	4548.56	4711.97	4701.06	4703.93	4696.77	4701.30
	MP2 <sup>c</sup> 4769.88									
	MP2 <sup>d</sup> 4752.17	4568.03	4541.69	4528.88	4537.91	4702.07	4691.15	4694.04	4686.90	4691.40
	MP2 <sup>e</sup> 4741.02									
	QCISD 4758.54									
	B3LYP 4759									

<sup>a</sup> Values obtained using a 6-31G\* basis set.

<sup>b</sup> Values obtained in a MP2/6-31G\*\* geometry optimization, including all electrons.

<sup>c</sup> Values obtained in a MP2/6-311G(2d, p) geometry optimization, including all electrons.

<sup>d</sup> Values obtained in a MP2/6-31G\*\* geometry optimization by using the frozen core approximation.

<sup>e</sup> Values obtained in a MP2/6-311 + G(d, p) geometry optimization by using the frozen core approximation.

of diffuse functions in the basis set results in a slight enlargement of the ring size because all bonds lengths become longer. The obvious consequence is that all rotational constants become smaller and are in poorer agreement with the experimental ones.

The agreement between the rotational constants of the different isotopic pyrazoles considered and the experimental values is good as far as the MP2 calculations are concerned. Actually, the average deviation is always smaller than 0.3%, whereas at the HF level it is at least 10 times larger.

The frozen core approximation in MP2 calculations has a negligible effect on the optimized geometries as far as the individual bond lengths and bond angles are concerned. The former differ by 0.001 Å or less and the latter by 0.01° or less. However, the effect on the rotational constants is sizeable, with differences of the order of 21 MHz (see Table III). But the most striking result is that the rotational constants obtained by using this

approximation are closer to the experimental ones (average deviation 0.08%) than those obtained when all electrons are included in the correlated calculation (average deviation 0.3%).

When the two different sets of internal coordinates defined in the previous section are used, one obtains apparently the same geometry (bond lengths differ by less than 0.0005 Å and bond angles by less than 0.001°), with total energies which differ less than  $10^{-8}$  Hartrees but which present rotational constants which differ by 1–2 MHz. However, these differences disappear by imposing tight conditions for the convergency criteria when the strongly coupled set of internal coordinates is used.

### STATISTICAL ANALYSIS

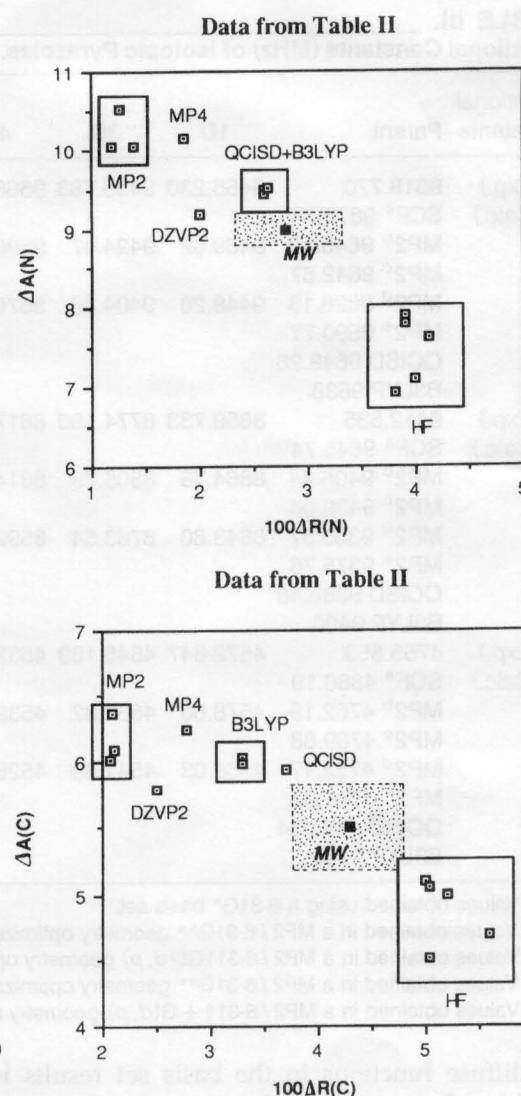
The problem of determining which theoretical geometry is closest to the experimental one is not a trivial matter even in the case of a planar pen-

tagon. As mentioned earlier, a planar pentagon has seven degrees of freedom and there are several combinations (but not any one) of the five bond lengths and the five bond angles which could be used (for instance, four lengths and three adjacent angles. We have used three different approaches to this problem.

### Paul Curtin's Diagrams

We have advocated the use of these diagrams to classify pyrazole geometries.<sup>1, 10</sup> The two kinds of diagrams both use differences in bond angles versus differences in bond lengths:  $\Delta A(N)$  versus  $\Delta R(CN)$  and  $\Delta A(C)$  versus  $\Delta R(CC)$  (to avoid small numbers,  $\Delta R$  has to be multiplied by 100, and thus  $\Delta A$  is in degrees whereas  $\Delta R$  is in hundredths of Å; i.e., pm). This approach used four angles ( $C_5N_1N_2/N_1N_2C_3$ ;  $N_2C_3C_4/C_4C_5N_1$ ) and four distances ( $N_1C_5/N_2C_3$ ;  $C_3C_4/C_4C_5$ )—that is, eight coordinates, one more than necessary. It neglects one angle ( $C_3C_4C_5$ ) and one bond ( $N_1N_2$ ). The result is shown in Figure 2. The shaded area around the MW geometry corresponds to an error of  $\pm 3^\circ$  and  $\pm 0.005 \text{ \AA}$  (Table II).

The HF calculations form a set. Although the subsets corresponding to the inclusion (SVP) or not (SV) of polarization can be distinguished, the former is closer to the experimental results. It is worth noticing that HF geometries appear concentrated in the lower right corner of the diagram shown in Figure 2, indicating that the deviations in bond angles are relatively smaller than the deviations in bond lengths. The introduction of electron correlation effects (MP2 calculations) does not result in displacement of the points closer to the MW experimental point but in a diagonal shift which moves the calculated geometries into the upper left corner of the diagram. In other words, the inclusion of correlation contributions results in a significant decrease of the deviations affecting the bond lengths, whereas deviations in bond angles are slightly greater than at the HF level. Figure 2 also shows that a further improvement of the model by taking into account energy correlation contributions up to the fourth order moves slightly the calculated point closer to the experimental one. The displacement shows that MP4 bond lengths are slightly better than MP2 ones, whereas improvement in bond angles is less significant. A similar but more pronounced affect is observed for QCISD optimized structures, which are the closest ones to the experimental point. Significantly, the



**FIGURE 2.** Paul Curtin's diagrams. (a)  $\Delta A(N)$  versus  $100\Delta R(N)$ . (b)  $\Delta A(C)$  versus  $100\Delta R(CC)$ .

DFT optimized geometries are also close to the experimental one and only slightly worse than the QCISD geometry, as far as the  $\Delta A(C)$  versus  $\Delta R(CC)$  diagram (Fig. 2b) is concerned.

The most important conclusion of this analysis is that if one takes into account the uncertainties affecting the experimental geometry (which define the shaded area in Fig. 2), both the QCISD and B3LYP optimized structures are in excellent agreement with the MW structure. Furthermore, in both cases the corresponding rotational constants are also close to the experimental ones (see Table III), the differences being 0.2–0.3% for A, 0.1–0.2% for B, and 0.06% for C. These differences are not larger

than those found by Smith et al.<sup>4</sup> for the theoretical  $r_0$  structure of vinyl fluoride and vinyl alcohol.

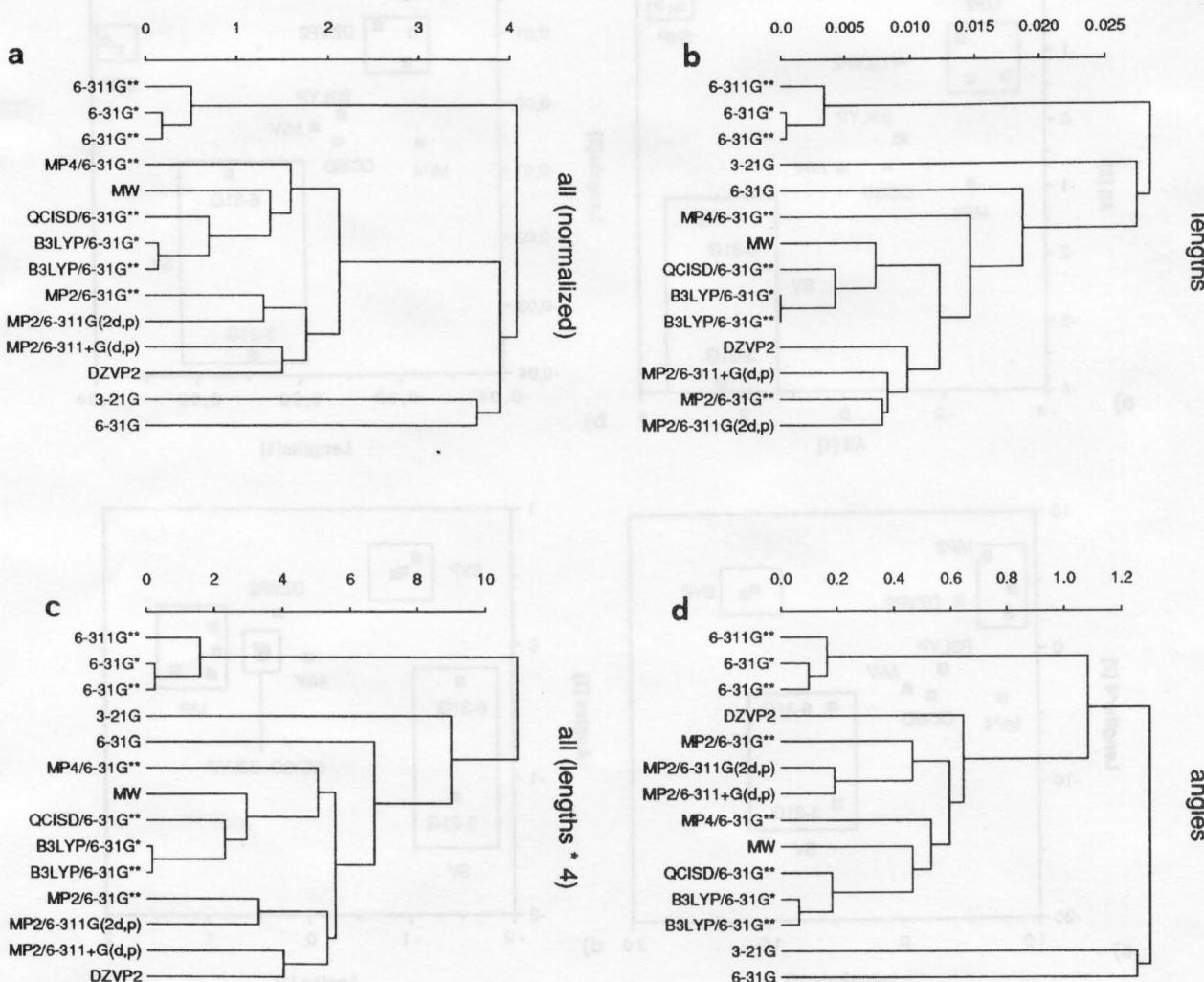
### Cluster Analysis

This is the standard procedure to determine similarities (Euclidean distances and single linkage procedures were selected).<sup>11</sup> We have carried out three such analyses: (1) using the five bond lengths (Fig. 3b), (2) using the five bond angles (Fig. 3d), and (3) using the 10 data (Fig. 3a). In the first two cases, the natural variables were used, but in the last case, the mixing of variables of different units ( $\text{\AA}$ , degrees) obliges us to carry out a scaling of these variables. Considering the differences in energy involved, we have repeated the last analysis

by multiplying the weight of the lengths by 4—that is, in the cluster the importance of similarity in bond lengths weight four times that of similarity in bond angles (Fig. 3c).

Within small differences, the four analyses are similar (the scaling of bond lengths, Fig. 3c, produces a tree nearly identical to that obtained using bond lengths alone, Fig. 3b). The MW geometry appears to be close to the QCISD and B3LYP geometries and different from the SV (3-21G and 6-31G) and SVP (6-31G\*, 6-31G\*\*, and 6-311G\*\*) geometries. The remaining calculations (DZVP2, MP2, and MP4) occupy intermediate positions.

These results quantify somehow those obtained using Paul Curtin's diagrams. As mentioned earlier, the diagonal shifting of the MP2, MP4, QCISD,



**FIGURE 3.** Cluster analysis. (a) All the 10 values. (b) Distances. (c) All the 10 values with a 4 : 1 distances / angles weighting scheme. (d) Angles.

and B3LYP calculations with regard to the HF ones is a consequence of the improvement in the description of the bond lengths.

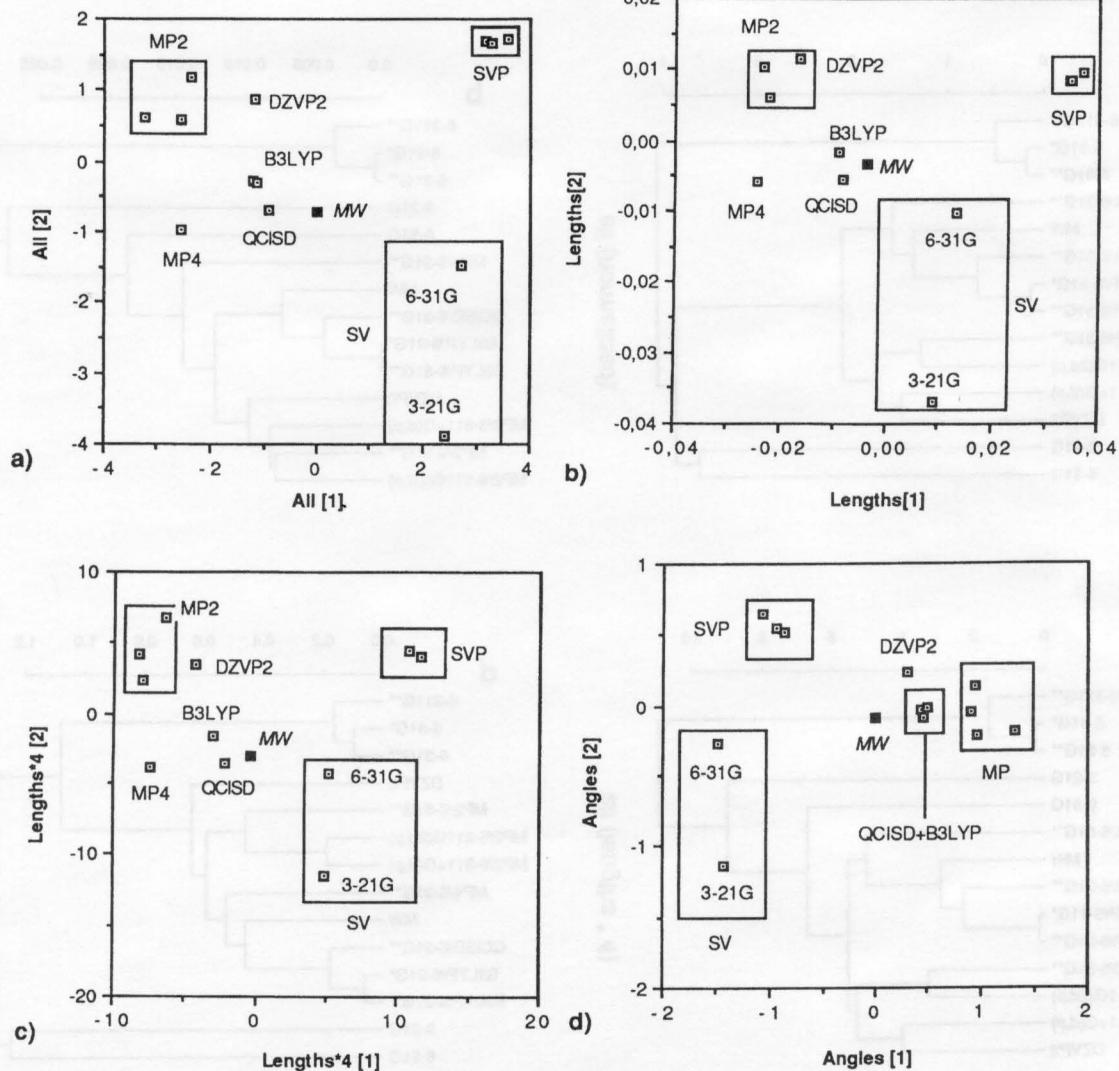
### Multidimensional Scaling

This procedure is described in ref. 12. In all cases, standardized variables have been used. As in Paul Curtin's diagrams (Fig. 2), the results of the multidimensional scaling (MDS) (Fig. 4) show the experimental point in the middle of the plot whatever the coordinates used are. The main difference between MDS and Paul Curtin's plots is the clear separation between SV and SVP sets in the former. Otherwise, they are similar (compare

Figs. 2a and 2b with Fig. 4a), proving that Paul Curtin's plots are an acceptable projection of the multidimensional space necessary to represent the pyrazole geometry. Figures 4a, 4b, and 4c are more alike between themselves than Figure 4d (angles). The superiority of QCISD and B3LYP calculations, as far as reproducing the MW geometry, is clearly apparent in these figures.

### THE ENERGIES INVOLVED IN GEOMETRY DEFORMATIONS

Estimating these energies is difficult because they depend on the nature of the atoms involved; nevertheless, a rough estimation for the C—C bond



**FIGURE 4.** Multidimensional scaling. (a) All the 10 values. (b) Distances. (c) All the 10 values with a 4 : 1 distances/angles weighting scheme. (d) Angles.

and the C—C—C angle is that a 0.1-Å distortion costs 15 kJ/mol<sup>-1</sup> whereas a 10° distortion costs 4 kJ/mol<sup>-1</sup> (ref. 13; an independent estimation is that to shorten or lengthen a C—C bond 0.1 Å from the equilibrium geometry costs 21 and 9 kJ/mol<sup>-1</sup>).<sup>14</sup> Then 0.01 Å corresponds to 1.5 kJ/mol<sup>-1</sup> and 1° to 0.4 kJ mol<sup>-1</sup>. In a description of the results of 3-21G calculations of pyrazole and imidazole,<sup>8</sup> the authors conclude that for pyrazole the largest absolute deviations from experiment (MW)<sup>5</sup> of the 3-21G-optimized geometrical parameters are 0.03 Å (4.5 kJ/mol<sup>-1</sup>) in the N<sub>1</sub>N<sub>2</sub> bond length and 1.5° (0.6 kJ/mol<sup>-1</sup>) in the C<sub>5</sub>N<sub>1</sub>N<sub>2</sub> bond angle. In light of the previous discussion, we can conclude that the optimized geometries which lead to smaller energetic effects arising from geometry deformations are the QCISD and B3LYP geometries followed by the MP4 ones, because they minimize the deviations affecting the bond lengths (which are those of higher energy cost). If the estimations presented in this article are correct, we may conclude that the structural deformations found at the MP2, MP4, QCISD, and B3LYP levels account for energies ranging from 2.4 kJ/mol<sup>-1</sup> for bond lengths and 1.3 kJ/mol<sup>-1</sup> for bond angles.

## Concluding Remarks

It is well established that although a general improvement in the description of molecular geometries can be expected at higher levels of theory, there is no asymptotic approach to the experimental geometry with the increase of the quality of the theoretical treatment. The situation is more complicated, however. The HF/3-2-21G method is clearly unsuitable, but the 6-31G basis set is a good compromise. The inclusion of polarization modifies the calculated geometry, although the differences among the three basis sets used (6-31G, 6-31G, and 6-311G\*\*) are minor.

The agreement between calculated and experimental bond lengths increases when electron correlation effects are included in the theoretical treatment, although for the bond angles the opposite is found. Importantly, these effects do not counterbalance each other, as revealed by the values of the corresponding rotational constants, which are a tight criterion with which to define the quality of a molecular geometry. Actually, even if the HF/6-31G and MP2/6-31G\*\* structures can be considered in good agreement with the experimental one

as far as individual bond lengths and individual bond angles are concerned, the errors affecting the corresponding rotational constants are one order of magnitude smaller (0.3%) when electron correlation effects are taken into account. One may then conclude that the MP2 structures are much more reliable than the HF ones, even though the description of the bond angles is not as good as that found at the HF level (if one assumes the experimental values to be correct). It should also be noticed that when the set of internal coordinates chosen to describe the molecular geometry is strongly coupled, it is necessary to use tight convergency criteria to obtain reliable structures.

According to the three different statistical analyses presented in this article, MP2/6-31G\*\*, MP2/6-311G(2d, p), and MP2/6-311 + G(d, p) structures appear always in the same group, showing their similarity. However, the latter two exhibit rotational constants which are similar but not closer to the experimental ones than those of the former. Hence, enlarging the basis set from 6-31G to 6-311G(2d, p) or 6-311 + G(d, p), which implies a considerable increase of the computational effort, does not lead to any significant improvement of the optimized geometry. The improvement in the description of the molecular geometry by including correlation contributions up to fourth order (MP4) is almost negligible as far as bond angles are concerned and small, but sizeable, as far as bond lengths are considered. However, the improvements found do not justify at all the enormous increase in the computational cost. The most significant improvement is attained at the QCISD level, which is not dramatically expensive. If one considers reliability and computational effort, the best choice will be the B3LYP density functional procedure, which yields optimized geometries close to the QCISD geometry and to the experimental MW one. Finally, those structures which, independently of the statistical analysis used, are predicted to be close to the experimental one present rotational constants which differ from the experimental values by less than 0.3%.

## References

1. A. L. Llamas-Saiz, C. Foces-Foces, and J. Elguero, *J. Mol. Struct.*, **319**, 231 (1994).
2. W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. A. Pople, *Ab Initio Molecular Orbital Theory*, John Wiley & Sons, New York, 1986, p. 137.
3. (a) M. L. Harmony, V. W. Laurie, R. L. Kuczkowski, R. H. Schwendeman, D. A. Ramsay, F. J. Lovas, W. J. Lafferty,

- and A. G. Maki, *J. Phys. Chem. Ref. Data*, **8**, 619 (1979); (b) H. D. Rudolph, *Struct. Chem.*, **2**, 581 (1991); (c) M. Hargittai and I. Hargittai, *Int. J. Quantum Chem.*, **44**, 1057 (1992).
4. B. J. Smith, D. Coffey, Jr., and L. Radom, *J. Chem. Phys.*, **97**, 6113 (1992); (b) B. J. Smith and L. Radom, *J. Am. Chem. Soc.*, **112**, 7525 (1990).
  5. L. Nygaard, D. Christen, J. T. Nielsen, E. J. Pedersen, O. Snerling, E. Vestergaard, and G. O. Sørensen, *J. Mol. Struct.*, **22**, 401 (1974).
  6. M. J. Frisch, M. Head-Gordon, G. W. Trucks, J. B. Foresman, H. B. Schlegel, K. Raghavachari, M. Robb, J. S. Binkley, C. Gonzalez, D. J. Defrees, D. J. Fox, R. A. Whiteside, R. Seeger, C. F. Melius, J. Baker, R. L. Martin, L. R. Kahn, J. J. P. Stewart, S. Topiol, and J. A. Pople, Gaussian 90, Revision I, Gaussian Inc., Pittsburgh, PA, 1990.
  7. P. Mátyus, K. Fuji, and K. Tanaka, *Tetrahedron*, **50**, 2405 (1994).
  8. T. Clark, *A Handbook of Computational Chemistry*, John Wiley & Sons, New York, 1985.
  9. J. M. Bofill, S. Olivella, and A. Solé, *J. Am. Chem. Soc.*, **111**, 7740 (1989).
  10. C. Foces-Foces, A. L. Llamas-Saiz, R. M. Claramunt, C. López, and J. Elguero, *J. Chem. Soc., Chem. Commun.*, 1143 (1994).
  11. B. Everitt, *Cluster Analysis* (2nd ed.), Halsted, New York, 1980.
  12. (a) K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*, Academic Press, London, 1979; (b) M. L. Davison, *Multidimensional Scaling*, John Wiley & Sons, New York, 1983.
  13. I. Hargittai, personal communication.
  14. U. Burkert and N. L. Allinger, *Molecular Mechanics*, ACS Monograph 177, American Chemical Society, Washington, DC 1982, p. 39.

**REFERENCES AND NOTES**

Table 1 contains a list of references to the literature on the use of QM/MM calculations for the study of biological systems. A few additional references are included in the text. We have tried to include all relevant literature, but it is likely that some papers have been omitted. The references are grouped into three main categories: (1) general methods, (2) applications to protein-ligand complexes, and (3) applications to protein-protein complexes.