Article

Principal Component Analysis of Molecular Geometries of *Cis*- and *Trans*-$C_2H_2X_2$ with X = F or Cl

João Bosco P. da Silva and Mozart N. Ramos*

Departamento de Química Fundamental, Universidade Federal de Pernambuco, Cidade Universitária, 50740-250 Recife - PE, Brazil

Gráficos de escores de PC1 e PC2 mostram como as geometrias calculadas dependem de características da função de onda molecular dos *cis*- e *trans*- difluoro e dicloroetilenos. PC1 e PC2 separam os resultados obtidos com e sem funções de polarização e com e sem a inclusão de correlação eletrônica. A qualidade das geometrias experimentais é analisada projetando-as nos gráficos dos escores. Usando este procedimento, a geometria de Takeo obtida a partir de transições de microondas não se compara com nenhuma dos cálculos *ab initio* para o *cis*-$C_2H_2Cl_2$, ao passo que a geometria de Schäfer obtida por espectroscopia de difração de elétrons está em boa concordância com aquelas de cálculos MP2/cc-pVDZ, MP2/cc-aug-pvDZ e CCD/cc-pVDZ.

PC1 and PC2 score graphs show how calculated molecular geometries depend on characteristics of the molecular wave-functions of *cis*- and *trans*- difluoro- and dichloroethylene. PC1 and PC2 separate the results obtained with or without polarization functions and with or without the inclusion of electronic correlation. The quality of the experimental geometries are analyzed projecting them on the PC score graphs. Using this procedure, Takeo's geometry obtained from microwave transitions does not compares with any of the *ab initio* calculations for *cis*-$C_2H_2Cl_2$, whereas Schäfer's geometry obtaines from gas electron diffraction spectroscopy is in good agreement with the MP2/cc-pVDZ, MP2/cc-aug-pVDZ and CCD/cc-pVDZ calculations.

Keywords: difluoroethylenes, dichloroethylenes, ab initio geometry, experimental geometry, PCA

Introduction

In the last twenty years, we have devoted considerable attention[1-10] to *cis*- and *trans*-dihaloethylenes in order to gain a better understanding of their electronic and vibrational properties. The *cis*- and *trans*-dihaloethylenes ($C_2H_2X_2$) are interesting isomeric species since they contain the same kind and number of chemical bonds. The major difference between them is due to the relative configurations of these bonds within the molecule. In particular, *trans*-dihaloethylenes are intriguing molecules from a spectroscopic point of view because, in spite of their high molecular symmetry, the orientations of their in-plane dipole derivatives are not restricted to the principal symmetry axes. We have shown that these directions are, in general, similar to those expected on the basis of simple chemical valence concepts.[1-3] These studies have also shown, that atomic polar tensors of *cis*- and *trans*-

$C_2H_2X_2$ (X = F or Cl) are very similar since those of the *cis*-compounds are capable of reproducing the experimental vibrational intensities of the *trans*-isomers within the propagated experimental error.[4-6] Furthermore, the electronic structures of *cis*- and *trans*-$C_2H_2Cl_2$ are more similar than those of *cis*- and *trans*-$C_2H_2F_2$ in terms of the intensity parameters of equilibrium charges and charge fluxes.[7,8]

In contrast to chemical intuition, both theoretical and experimental studies[11-14] have revealed that the *cis* isomer is more stable than its corresponding *trans* form, as a consequence of the so called *cis* effect.[15] Theoretical results have shown that a correct interpretation of this effect depends on the precision of the geometric parameters obtained from molecular orbital calculations.[16] However, calculated geometries can be strongly dependent on the calculation level (HF, MP2, CCD or [CCSD(T)]) and basis sets used whereas experimental[17-22] geometries may depend on the experimental technique employed (e.g. gas electron diffraction (GED) or microwave (MW) spectroscopy).

* e-mail: paraiso@ufpe.br

Recently, the molecular geometries of the *trans*-$C_2H_2X_2$ (X = F or Cl) species have been obtained from the microwave transitions observed in high-resolution infrared spectroscopy (IR)[23,24] and are somewhat different from those obtained using GED.[22] For example, the values of the C-Cl, C=C and C-H bond lengths for *trans*-$C_2H_2Cl_2$ obtained from GED[22] are 1.725(2) Å, 1.332(8) Å and 1.092(26) Å respectively whereas their corresponding values using IR[24] are 1.740(3) Å, 1.305(5) Å and 1.078(4) Å.

In order to better understand both the theoretical and experimental changes which occur in the molecular geometries of the *cis*- and *trans*-$C_2H_2X_2$ species, we have performed a multivariate exploratory analysis using Principal Components Analysis (PCA).[25,26] This technique has been successful in analyzing the effects of wave-function modifications on calculated C-H and C-X (X = F or Cl) vibrational frequencies and infrared intensities of the dihaloethylenes.[9,10] For example, all the calculated C-H stretching frequencies can be adequately described by a single principal component whereas bidimensional principal component graphs are sufficiently accurate for a direct comparison of the results of trial wave-functions with the observed results of the vibrational bending frequencies.

Calculations

A set of *ab initio* molecular orbital calculations was performed with the Gaussian 92[27] and Dalton[28] programs. The Hartree-Fock (HF)[29] and Möller-Plesset of second order (MP2)[30] calculations were carried out using a 2^4 factorial design, where two levels of four factors were investigated: (*i*) the use of basis sets 6-31G or 6-311G; (*ii*) the presence or absence of diffuse functions; (*iii*) the presence or absence of polarization functions; (*iv*) the use, or not, of perturbative Möller-Plesset corrections of second order (MP2) to HF calculations.[16] The MP2 calculations were performed using the frozen-core electron correlation approach. The others were performed using coupled-cluster calculations with double excitations (CCD) and single and double excitations (CCSD) augmented by a perturbational correction for connected triple excitations [CCSD(T)].[31] In order to evaluate the importance of electron correlation for inner-shells, in particular for the dichloroethylene systems, the CCSD(T) calculations were also performed including additional electron correlation for Cl 2s2p core electrons. These calculations result in a data matrix $\mathbf{X_{n,p}}$ composed of 5 variables, which correspond to three bond lengths (C-H, C=C and C-X) and two bond angles (CCH and CCX), and "n" objects, which correspond to the different *ab initio* calculations for each $C_2H_2X_2$ species.

This matrix $\mathbf{X}$ can be taken as a set of "n" calculations represented as a graphic in a 5-dimensional space.

Principal component analysis (PCA) represents a rotation of the original axis system searching a new direction concentrating at maximum the original information and for which one hope to find some kind of patterns present in the original data set. From a practical point of view, this is obtained through the diagonalization of the covariance matrix $\mathbf{X^tX}$ (where $\mathbf{X^t}$ is the transposed of the data matrix $\mathbf{X}$). The eigenvector elements called loadings represent the director cosines, *i.e.*, the contribution of the original axes for the composition of the new axes called principal components. The eigenvalues represent the amount of variance described by the corresponding eigenvectors. The first eigenvector is the first principal component (PC1) and corresponds to the axis for which the objects have the maximum variance. Therefore, PC1 corresponds to the axes for which the objects are at its maximum spread. The second principal component, (PC2), is orthogonal to PC1, and represents the second axis of larger residual variance, *i.e.*, it is the axis of maximum amount of variance not explained by PC1. A projection of the data on these two axes yields a graphical representation of the maximum statistical information that can be compressed into two dimensions, and may help to detect patterns hidden in the original multidimensional data.

In this work the principal component analyses using autoscaled (*i.e.*, each element on a column was subtracted by the average and scaled to unit variance on the column) data were carried out using the chemometrics package Ein*Sight 3.0[32] on a personal microcomputer of the Laboratory of Theoretical and Computational Chemistry of the Departamento de Química Fundamental at the Universidade Federal de Pernambuco (UFPE, Brazil). The M.O. calculations were performed on workstations of UFPE and San Diego Supercomputer Center (SDSC) of the University of California, San Diego (UCSD).

Results and Discussions

In Tables 1 to 4 the optimized geometries for *cis* and *trans*-$C_2H_2X_2$ (X = F and Cl) are shown together with the experimental values.

cis-$C_2H_2F_2$

The score graph in Figure 1 for the *cis*-$C_2H_2F_2$ species shows that the 5-dimensional original space in Table 1 can be accurately represented by two principal components, which describe 95.5% of the total data variance.

Table 1. Optimized geometry of *cis*-C$_2$H$_2$F$_2$. Bond length in Angstrom and bond angles in degrees

Calculations	C-F	C=C	C-H	CCH	CCF
HF/6-31G	1.366	1.308	1.065	124.2	122.1
HF/6-311G	1.364	1.303	1.064	124.3	122.2
HF/6-31++G	1.370	1.310	1.066	124.6	121.9
HF/6-311++G	1.364	1.305	1.064	124.3	122.0
HF/6-31G**	1.324	1.307	1.071	122.9	122.5
HF/6-311G**	1.319	1.306	1.071	122.6	122.8
HF/6-31++G**	1.325	1.303	1.070	123.0	122.5
HF/6-311++G**	1.318	1.307	1.071	122.6	122.8
MP2/6-31G	1.405	1.339	1.083	124.2	122.1
MP2/6-311G	1.404	1.332	1.079	124.3	122.1
MP2/6-31++G	1.416	1.340	1.083	124.9	121.8
MP2/6-311++G	1.408	1.333	1.080	124.7	121.8
MP2/6-31G**	1.349	1.330	1.078	122.9	122.1
MP2/6-311G**	1.338	1.331	1.082	122.2	122.7
MP2/6-31++G**	1.354	1.332	1.078	123.2	122.1
MP2/6-311++G**	1.339	1.332	1.083	122.5	122.5
HF/cc-pVDZ	1.323	1.310	1.078	122.3	123.0
HF/cc-aug-pVDZ	1.326	1.312	1.076	122.7	122.7
MP2/cc-pVDZ	1.343	1.339	1.091	121.7	123.1
MP2/cc-aug-pVDZ	1.356	1.341	1.089	122.9	122.2
CCD/6-31G	1.401	1.339	1.086	124.4	121.8
CCD/6-31G**	1.347	1.327	1.078	123.2	122.0
CCD/cc-pVDZ	1.341	1.336	1.092	122.1	122.9
CCSD(T)/cc-pVTZ	1.337	1.330	1.080	122.3	122.6
CCSD(T)/cc-pVTZ_vib[a]	1.342	1.336	1.083	122.4	122.4
(MW)$_{Laurie}$[b]	1.335	1.324	1.089	124.0	122.1
(MW)$_{Harmony}$[c]	1.337	1.325	1.088	123.9	122.1
(ED)$_{van\ Schaick}$[d]	1.332	1.311	1.100	127	122.5
(ED)$_{Carlos}$[e]	1.335	1.331	1.084	121.6	123.7

[a]Corrected due to average vibration, see text; [b]Ref. 17; [c]Ref. 18; [d]Ref. 19; [e]Ref. 20.

Table 2. Optimized geometry of *trans*-C$_2$H$_2$F$_2$. Bond length in Angstrom and bond angles in degrees

Calculations	C-F	C=C	C-H	CCH	CCF
HF/6-31G	1.371	1.307	1.065	126.7	119.2
HF/6-311G	1.370	1.303	1.064	126.8	119.3
HF/6-31++G	1.378	1.308	1.066	127.4	118.9
HF/6-311++G	1.372	1.303	1.064	127.1	119.1
HF/6-31G**	1.329	1.306	1.071	125.2	120.2
HF/6-311G**	1.324	1.305	1.071	124.9	120.3
HF/6-31++G**	1.331	1.308	1.071	125.6	120.0
HF/6-311++G**	1.324	1.306	1.071	125.2	120.2
MP2/6-31G	1.410	1.339	1.082	127.0	118.8
MP2/6-311G	1.411	1.331	1.078	127.1	118.9
MP2/6-31++G	1.425	1.339	1.082	128.4	118.1
MP2/6-311++G	1.417	1.332	1.079	128.0	118.2
MP2/6-31G**	1.353	1.330	1.078	125.1	119.8
MP2/6-311G**	1.342	1.331	1.082	124.6	120.2
MP2/6-31++G**	1.361	1.331	1.078	126.1	119.4
MP2/6-311++G**	1.346	1.331	1.082	125.3	119.8
HF/cc-pVDZ	1.327	1.309	1.079	124.8	120.4
HF/cc-aug-pVDZ	1.332	1.311	1.077	125.4	120.2
MP2/cc-pVDZ	1.347	1.339	1.092	124.2	120.5
MP2/cc-aug-pVDZ	1.363	1.340	1.090	125.8	119.5
CCD/6-31G	1.407	1.338	1.085	127.1	118.7
CCD/6-31G**	1.352	1.327	1.078	125.4	119.7
CCD/cc-pVDZ	1.345	1.336	1.093	124.4	120.4
CCSD(T)/cc-pVTZ	1.342	1.330	1.080	124.9	120.1
CCSD(T)/cc-pVTZ_vib[a]	1.346	1.336	1.085	125.1	119.8
(IR)$_{Craig}$[b]	1.352	1.316	1.080	126.3	119.2
(ED)$_{van\ Schaick}$[c]	1.338	1.320	1.088	125	119.8
(ED)$_{Carlos}$[d]	1.334	1.329	1.080	129.3	119.3

[a]Corrected due to average vibration, see text; [b]Ref. 23; [c]Ref 19; [d]Ref. 20.

Table 3. Optimized geometry of *cis*-C$_2$H$_2$Cl$_2$. Bond length in Angstrom and bond angles in degrees

Calculations	C-Cl	C=C	C-H	CCH	CCCl
HF/6-31G	1.783	1.309	1.067	121.7	125.5
HF/6-311G	1.783	1.306	1.065	122.0	125.6
HF/6-31++G	1.784	1.310	1.068	121.7	125.4
HF/6-311++G	1.783	1.307	1.066	121.8	125.6
HF/6-31G**	1.721	1.312	1.072	120.1	125.7
HF/6-311G**	1.724	1.311	1.041	120.5	125.6
HF/6-31++G**	1.721	1.314	1.072	120.2	125.7
HF/6-311++G**	1.723	1.312	1.072	120.3	125.6
MP2/6-31G	1.812	1.340	1.085	121.8	125.2
MP2/6-311G	1.810	1.334	1.081	122.0	125.5
MP2/6-31++G	1.813	1.341	1.085	121.9	125.1
MP2/6-311++G	1.810	1.335	1.082	122.1	125.3
MP2/6-31G**	1.716	1.337	1.079	120.0	125.2
MP2/6-311G**	1.715	1.338	1.083	120.1	125.0
MP2/6-31++G**	1.717	1.339	1.080	120.0	125.0
MP2/6-311++G**	1.715	1.339	1.083	120.1	124.8
HF/cc-pVDZ	1.726	1.316	1.079	120.4	125.4
HF/cc-aug-pVDZ	1.727	1.317	1.077	120.5	125.4
MP2/cc-pVDZ	1.724	1.347	1.093	119.9	124.8
MP2/cc-aug-pVDZ	1.729	1.348	1.092	120.5	124.3
CCD/6-31G	1.816	1.338	1.087	122.0	125.2
CCD/6-31G**	1.723	1.332	1.079	120.2	125.3
CCD/cc-pVDZ	1.731	1.342	1.094	120.1	125.0
CCSD(T)/cc-pVTZ	1.724	1.335	1.080	120.4	124.7
CCSD(T)/cc-pVTZ_vib[a]	1.730	1.339	1.083	120.5	124.5
CCSD(T)/cc-pVTZ_2s2p[b]	1.718	1.336	1.080	120.3	124.7
CCSD(T)/cc-pVTZ_2s2p_vib[c]	1.724	1.340	1.083	120.4	124.5
(ED)$_{Schäfer}$[d]	1.717	1.337	1.096	120.3	124.0
(MW)$_{Takeo}$[e]	1.717	1.319	1.100	123.2	124.2

[a]Corrected due to average vibration, see text; [b]additional electron correlation for Cl 2s2p core electron; [c]additional electron correlation for Cl 2s2p core electron and average vibration corrections; [d]Ref. 22; [e]Ref. 21.

**Table 4.** Optimized geometry of *trans*-C$_2$H$_2$Cl$_2$. Bond length in Angstrom and bond angles in degrees

Calculations	C-Cl	C=C	C-H	CCH	CCCl
HF/6-31G	1.795	1.307	1.066	125.5	121.1
HF/6-311G	1.795	1.304	1.064	126.0	121.2
HF/6-31++G	1.796	1.308	1.067	125.6	121.2
HF/6-311++G	1.794	1.305	1.064	125.9	121.2
HF/6-31G**	1.729	1.311	1.071	123.8	121.7
HF/6-311G**	1.733	1.308	1.071	124.2	121.6
HF/6-31++G**	1.730	1.312	1.071	123.8	121.7
HF/6-311++G**	1.732	1.309	1.071	124.0	121.6
MP2/6-31G	1.824	1.339	1.083	125.6	121.0
MP2/6-311G	1.822	1.333	1.080	126.0	121.1
MP2/6-31++G	1.825	1.340	1.084	125.6	121.0
MP2/6-311++G	1.821	1.334	1.080	125.9	121.0
MP2/6-31G**	1.723	1.335	1.079	123.3	121.7
MP2/6-311G**	1.723	1.336	1.083	123.4	121.5
MP2/6-31++G**	1.724	1.337	1.080	123.3	121.6
MP2/6-311++G**	1.723	1.336	1.083	123.4	121.4
HF/cc-pVDZ	1.735	1.314	1.079	123.9	121.5
HF/cc-aug-pVDZ	1.737	1.314	1.077	124.3	121.3
MP2/cc-pVDZ	1.732	1.345	1.093	123.2	121.2
MP2/cc-aug-pVDZ	1.739	1.346	1.091	123.9	120.9
CCD/6-31G	1.828	1.337	1.086	125.6	121.0
CCD/6-31G**	1.730	1.331	1.079	123.4	121.7
CCD/cc-pVDZ	1.739	1.340	1.093	123.4	121.4
CCSD(T)/cc-pVTZ	1.733	1.333	1.080	123.6	121.2
CCSD(T)/cc-pVTZ_vib[a]	1.738	1.339	1.084	123.7	121.0
CCSD(T)/cc-pVTZ_2s2p[b]	1.727	1.334	1.080	123.4	121.3
CCSD(T)/cc-pVTZ_2s2p_vib[c]	1.732	1.339	1.084	123.6	121.0
(ED)$_{Schäfer}$[d]	1.725	1.332	1.092	124.0	120.8
(IR)$_{Craig}$[e]	1.740	1.305	1.078	125.3	119.9

[a]Corrected due to average vibration, see text; [b]additional electron correlation for Cl 2s2p core electron; [c]additional electron correlation for Cl 2s2p core electron and average vibration corrections; [d]Ref. 22, [e]Ref. 24.
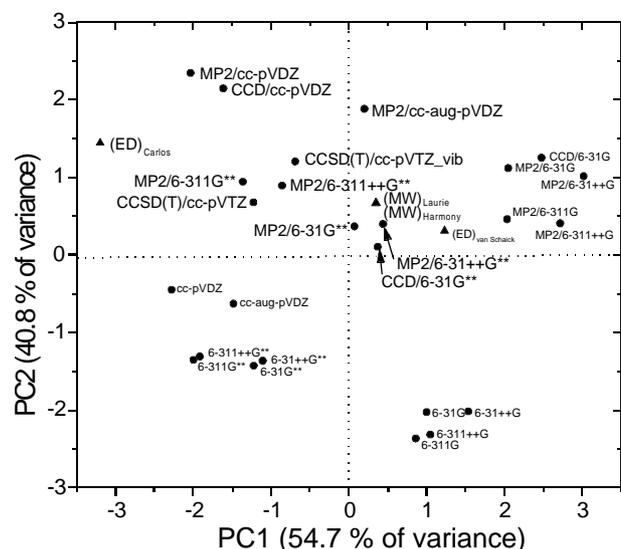
$PC1 = 0.573\ R_{C\text{-}F} + 0.175\ R_{C=C} - 0.037\ R_{C\text{-}H} + 0.561\ A_{CCH} - 0.569\ A_{CCF}$

$PC2 = 0.147\ R_{C\text{-}F} + 0.662\ R_{C=C} + 0.691\ R_{C\text{-}H} - 0.240\ A_{CCH} + 0.070\ A_{CCF}$

Figure 1. Score plot for the optimized geometry of *cis*-$C_2H_2F_2$. The experimental points were projected into the score plot.

The first principal component, PC1, describes 54.7% of the variance. It is dominated by the C-F (+0.57) bond length and the CCH (+0.56) and CCF (-0.57) bond angles (see equation of PC1 in Figure 1). In this Figure, we can observe that PC1 separates the calculations containing polarization functions (at left), which have near zero or negative scores, and the calculations without polarization functions (at right, positive scores). This arrangement means that the *ab initio* calculations with polarization functions have the smallest numerical values for the C-F bond length and the CCH bond angle, which have positive coefficients in the PC1 equation, and the highest numerical value for the CCF angle (negative coefficients in PC1). For example, the C-F, CCH and CCF values for the MP2/6-311G calculation are 1.404Å, 124.3° and 122.1° respectively, whereas their corresponding values are 1.338 Å, 122.2° and 122.7° for the MP2/6-311G** calculation. On the other hand, PC2 describes 40.8% of the total data variance. It is dominated by the C-H (+0.69) and C=C (+0.66) bond lengths. This second principal component separates the calculations including electron correlation (MP2, CCD and CCSD(T)), which have positive scores, from those at the HF level (negative scores). In this case, *ab initio* calculations without electronic correlations produce the smallest numerical values for the C-H and C=C bond lengths. For example, these values are 1.071Å (C-H) and 1.307Å (C=C) for the HF/6-311++G** calculation, whereas their corresponding values are 1.082Å and 1.332Å for the MP2/6-311++G** calculation, respectively. These values for the more sophisticated

CCSD(T)/cc-pVTZ vib average (*i.e.*, for geometrical corrections due to average vibrations) calculation are 1.083Å (C-H) and 1.336Å (C=C), respectively, thus very similar to the MP2/6-311++G** calculation.

In Figure 1, the experimental geometries were inserted substituting autoscaled experimental values in equations of PC1 and PC2. This procedure will also be adopted for the other dihaloethylenes. In Table 1 we can note that the microwave (MW) geometries from Laurie and Pence[17] and from Harmony *et al.*[18] are very similar and appear superimposed in Figure 1. They are very close to those using the MP2/6-311++G**, MP2/6-31G** and CCD/6-31G** calculations. van Schaick's geometry[19] using gas electron diffraction spectroscopy (GED) is situated at the right and near to MP2/6-31++G** calculation. On the other hand, the geometry of Carlos *et al.*,[20] also using GED, is very far from this group and practically isolated. It appears at the left and near the top as consequence of both a large CCF bond angle (123.7°) and small CCH bond angle (121.6°), corresponding to a negative score of PC1, and also a large C=C bond length (1.331Å) with a positive score in PC2.

trans-$C_2H_2F_2$

The score graph in Figure 2 for the *trans*-$C_2H_2F_2$ species reveals that the original 5-dimensional space in Table 2 can be adequately represented by two principal components, which describe 98.1% of the total data variance.
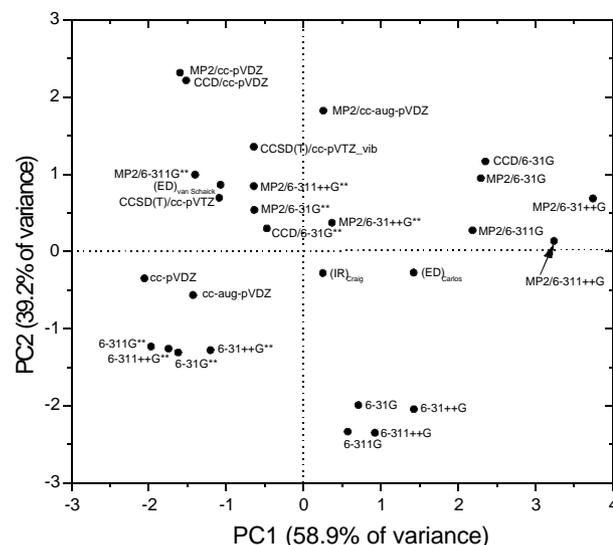
$PC1 = 0.571\ R_{C\text{-}F} + 0.197\ R_{C=C} - 0.017\ R_{C\text{-}H} + 0.550\ A_{CCH} - 0.576\ A_{CCF}$

$PC2 = 0.087\ R_{C\text{-}F} + 0.664\ R_{C=C} + 0.705\ R_{C\text{-}H} - 0.221\ A_{CCH} + 0.080\ A_{CCF}$

Figure 2. Score plot for the optimized geometry of *trans*-$C_2H_2F_2$. The experimental points were projected into the score plot.

The first principal component, PC1, describes 58.9% of the variance, while PC2 contains 39.2% of the remaining variance. Analogously to what was found for cis-$C_2H_2F_2$, PC1 is dominated by the C-F bond length (+0.57) and the CCH (+0.55) and CCF (-0.58) bond angles, whereas PC2 is dominated by the C-H (+0.71) and C=C (+0.66) bond lengths. It is also interesting to note that the coefficients of the PC1 and PC2 equations in $trans$-$C_2H_2F_2$ are also very similar to those found in cis-$C_2H_2F_2$. As consequence, here also PC1 separates the calculations with polarization functions with respect to those without them, which have positive scores and are at the right in Figure 2. PC2 separates the calculations with electronic correlation (MP2, CCD and CCSD(T)) from those at HF level.

Since $trans$-$C_2H_2F_2$ is non-polar, the geometries are not directly accessed from the MW spectrum. Here three experimental geometries are available: those obtained by van Schaick et $al.$[19] and Carlos et $al.$[20] using the GED technique and that from Craig et $al.$[23] using infrared spectroscopy. The GED geometry from van Schaick et $al.$ is close those obtained using the CCSD(T)/cc-pVTZ, MP2/6-311G**, MP2/6-31G** and MP2/6-311++G** calculations. Craig's geometry is reasonably close to those using the CCD/6-31G** and MP2/6-31++G** calculations. Again Carlos' geometry is far from those obtained using more elaborate calculations. This is mainly due to the large CCH bond angle of 129.3°. The CCH values obtained from Craig et $al.$ and from van Schaick et $al.$ are 126.3° and 125°, respectively.

cis-$C_2H_2Cl_2$

The score graph in Figure 3 for the cis-$C_2H_2Cl_2$ species shows that the 5-dimensional original space in Table 3 can be adequately represented by two principal components.

The first principal component, PC1, describes 58.7% of the total data variance, while PC2 contains 34.3% of the variance. Therefore, PC1 and PC2 describe 93.0% of the total variance. Their coefficients are different from those found for cis- and $trans$-$C_2H_2F_2$. PC1 is dominated by the C=C (-0.51) and C-H (-0.50) bond lengths and the CCCl (+0.51) bond angle. It is important to point out that the C-Cl (+0.31) bond length and the CCH (+0.37) bond angle are small but can not be ignored. In contrast to cis- and $trans$-$C_2H_2F_2$, here PC1 separates the calculations with or without electronic correlation; the HF calculations have positive scores and are located at the right of PC1. PC2 is mainly dominated by the C-Cl (+0.65) bond length and the CCH (+0.58) bond angle.

Four groups can be roughly identified in Figure 3. On

$$PC1 = 0.306\ R_{C-Cl} - 0.508\ R_{C-C} - 0.501\ R_{C-H} + 0.369\ A_{CCH} + 0.512\ A_{CCCl}$$
$$PC2 = 0.649\ R_{C-Cl} + 0.343\ R_{C-C} + 0.308\ R_{C-H} + 0.583\ A_{CCH} - 0.166\ A_{CCCl}$$

Figure 3. Score plot for the optimized geometry of cis-$C_2H_2Cl_2$. The experimental points were projected into the score plot.

opposite sides of the PC1 axis, (I) HF calculations without polarization functions at right and (II) calculations including electronic correlation without polarization functions appearing near the top, (III) HF calculations including polarization functions, located at the bottom part of the plot and (IV) calculations with electronic correlation including polarization functions possessing negative scores in PC1 and PC2.

Two experimental geometries were inserted in Figure 3. The geometry of Takeo et $al.$[21] using microwave (MW) transitions and Schäfer et $al.$[22] ones using gas electron diffraction (GED) spectroscopy. The first is isolated from all the theoretical calculations. This is mainly due to the large CCH angle (123.2°) and the low value of the C=C bond length (1.319Å). Schäfer's geometry is close the MP2/cc-pVDZ, MP2/cc-aug-pVDZ and CCD/cc-pVDZ calculations. This suggests the necessity of using correlated basis sets in electronic correlation calculations to adequately reproduce the experimental geometry. Here the MP2/6-nG** and CCD/6-nG** (n = 31 or 311) calculations are not close to the experimental ones, in contrast to what was observed for cis- and $trans$-$C_2H_2F_2$.

$trans$-$C_2H_2Cl_2$

Figure 4 shows the score graph for the $trans$-$C_2H_2Cl_2$ species. The first two principal components explain 92.9% of the total data variance of the calculated angles and bond lengths. PC1 describes 50.7% of this variance and PC2 contains 42.2% of the variance.

$$PC1 = 0.557\ R_{C\text{-}Cl} - 0.302\ R_{C=C} - 0.347\ R_{C\text{-}H} + 0.617\ A_{CCH} - 0.312\ A_{CCCl}$$
$$PC2 = 0.265\ R_{C\text{-}Cl} + 0.583\ R_{C=C} + 0.553\ R_{C\text{-}H} + 0.092\ A_{CCH} - 0.525\ A_{CCCl}$$

Figure 4. Score plot for the optimized geometry of *trans*-$C_2H_2Cl_2$. The experimental points were projected into the score plot.

PC1 is dominated by the C-Cl (+0.56) bond length, and the CCH (+0.62) and CCCl (-0.31) bond angles. Essentially, PC1 separates the calculations with or without polarization functions. This means that the *ab initio* calculations with polarization functions have the smallest numerical values for the C-Cl bond length and the CCH bond angle, which have positive coefficients, and the highest ones for the CCCl angle which has a negative coefficient in the equation for PC1. For example, the C-Cl, CCH and CCCl parameters for the MP2/6-311G calculation are 1.822Å, 126.0° and 121.1° respectively, whereas the corresponding values are 1.723 Å, 123.4° and 121.5° for the MP2/6-311G** calculation. PC2 is dominated by the C=C (+0.58), C-H (+0.55) bond lengths and the CCCl (-0.53) bond angle and separetes the calculations with and without electron correlation. For example, in PC2, HF calculations with or without polarization functions have negative scores.

Two experimental geometries were inserted in Figure 4. The geometry of Craig *et al.*[24] obtained from infrared (IR) spectroscopy and Schäfer *et al.*'s geometry[22] obtained from gas electron diffraction (GED) spectroscopy. Analogous to what was found for *cis*-$C_2H_2Cl_2$, the latter is relatively close to the MP2/cc-aug-pVDZ, CCSD(T)/cc-pVTZ-vib, CCSD(T)/cc-pVTZ-2s2p-vib, MP2/cc-pVDZ and CCD/ccpVDZ, *i.e.*, to the more sophisticated calculations. Craig's geometry in turn, is close the MP2 calculations with basis sets without polarization functions. In particular, these experimental geometries mainly differ on the values of the bond lengths.

Conclusions

The results of the principal component analysis (PCA) reported here reveal in a convincing way how calculated molecular geometries depend on the characteristics of the molecular wave-functions of *cis*- and *trans*- difluoro- and dichloroethylene. This can be better visualized through bidimensional graphs. In other words, these graphs indicate that the 5-dimensional original space (three bond lengths and two angle bonds) is adequately represented by only two principal components (PC1 and PC2) in describing the total data variance. The coefficients of the PC1 and PC2 equations are very similar for *cis*- and *trans*-$C_2H_2F_2$ and *trans*-$C_2H_2Cl_2$, in contrast to what occurs in *cis*-$C_2H_2Cl_2$.

Our results reveal that the presence or not of polarization functions and the inclusion or not of electronic correlation in the *ab initio* calculations are the two main effects explaining the total data variance for the geometry. The inclusion of polarization functions in the basis set decreases both the C-X (X = F or Cl) bond length and the CCH bond angle, whereas the inclusion of electronic correlation (MP2, CCD or CCSD(T)) increases both the C=C and C-H bond lengths. The simultaneous inclusion of these effects is essential to obtain calculated geometries in good agreement with the experimental ones. The use of 6-31G or 6-311G basis sets with or without difuse functions seem to have smaller effects. From the bidimensional PCA graphs, it was possible to analyze how (di)similar are these experimental geometries (obtained from different techniques) compared to the calculated ones. For example, the microwave geometries compare very well with the CCD/6-31G**, MP2/6-31G** and MP2/6-31++G** calculations, whereas the experimental values obtained from gas electron diffraction (GED) are not close to these calculations when considering the bidimensional graph of *cis*-$C_2H_2F_2$. On the other hand, the GED geometries from van Schaick *et al.*[19] for *trans*-$C_2H_2F_2$ are in good agreement with theoretical calculations when polarization functions and electronic correlations are simultaneously used, in contrast to the GED geometry from Carlos *et al.*.[20] For the dichloroethylene species the GED geometries from Schafer *et al.*[22] seem to be the best since they appear close to the higher level calculations.

Acknowledgements

References

1. Suto, E.; Ramos, M. N.; Bruns, R. E.; *J. Phys. Chem.* **1993**, *97*, 6161.

2. da Silva, J. B. P.; Ramos, M. N.; Suto, E.; Bruns, R. E.; *J. Phys. Chem.* **1997**, *101*, 6293.

3. Ramos, M. N.; Neto, B. B.; Bruns, R. E.; *J. Phys. Chem.* **1985**, *89*, 4979.

4. Kagel, R. O.; Powell, D. L.; Overend, J.; Ramos, M. N.; Bassi, A. B. M. S.; Bruns, R. E.; *J. Chem. Phys.* **1983**, *78*, 7029.

5. Hopper, M. J.; Overend, J.; Ramos, M. N.; Bassi, A. B. M. S.; Bruns, R. E.; *J. Chem. Phys.* **1983**, *79*, 17.

6. Kagel, R.O.; Powell, D. L.; Overend, J.; Hopper, M. J.; Ramos, M. N.; Bassi, A. B. M. S.; Bruns, R. E.; *J. Phys. Chem.* **1984**, *88*, 521.

7. da Silva, J. B. P.; Ramos, M. N.; *J. Mol. Struct. (TheoChem)* **2001**, *539*, 83.

8. Ramos, M. N.; Fausto, R., Teixeira-Dias, J. J. C.; Castiglioni, C.; Gussoni, M.; Zerbi, G.; *J. Mol. Struct.* **1991**, *248*, 281.

9. da Silva, J. B. P.; Ramos, M. N.; Bruns, R. E.; *Spectrochim. Acta A*, **1997**, *53*, 733.

10. Ramos, M. N.; da Silva, J. B. P.; Bruns, R. E.; *Spectrochim. Acta A* **1997**, *53*, 1563.

11. Graig, N. C.; Entemann, E. A.; *J. Am. Chem. Soc.* **1961**, *83*, 3047.

12. Craig, N. C.; Overrend, J.; *J. Chem. Phys.* **1967**, *51*, 1127.

13. Ambartzmuian, R. V.; Chekalin, N. V.; Doljikov, V. S.; Letokhov, V. S.; Lokman, V. N.; *Opt. Commun.* **1976**, *18*, 400;

14. Craig, N. C.; Piper, L. G.; Wheeler, V. L.; *J. Phys. Chem.* **1971**, *75*, 1453.

15. Yamamoto, T.; Tomoda, S.; *Chem. Lett.* **1997**, *10*, 1069.

16. da Silva, J. B. P.; *J. Braz. Chem. Soc.* **2000**, *11*, 219.

17. Laurie, V. W.; Pence, D. T.; *J. Chem. Phys.* **1963**, *38*, 2693.

18. Harmony, M. D.; Laurie, V. W.; Kuczkowski, R. L.; Schwendeman, R. H.; Ramsay, D. A.; Lovas, F. J.; Lafferty, W. J.; Maki, A. G.; *J. Phys. Chem. Ref. Data* **1979**, *8*, 619.

19. van Schaick, E. J. M.; Mijlhoff, F. C.; Renes, G.; Geise, H. J.; *J. Mol. Struct.* **1974**, *21*, 17.

20. Carlos, J. L.; Karl, R. R.; Bauer, S. H.; *J. Chem. Soc., Faraday Trans.2* **1974**, *2*, 177.

21. Takeo, H.; Sugie, M.; Matsumura, C.; *J. Mol. Struct.* **1988**, *190*, 205.

22. Schäfer, L.; Ewbank, J. D.; Siam, K.; Paul, D. W.; Monts, D. L.; *J. Mol. Struct.* **1986**, *145*, 135.

23. Craig, N. C.; Abiog, O. P.; Hu, B.; Stone, S. C.; Lafferty, W. J.; Xu, L.; *J. Phys. Chem.* **1996**, *100*, 5310.

24. Craig, N. C.; Appleman, R. A.; Barnes, H. E.; Morales, E.; Smith, J. A.; Klee, S.; Lock, M.; Mellau, G. C.; *J. Phys. Chem.* **1998**, *102*, 6745.

25. Mardia, K. V.; Kent, J. T.; Bibby, J. M.; *Multivariate Analysis*, Academic Press: London, 1979, chapter 8.

26. Suto, E.; Ferreira; M. M. C.; Bruns, R. E.; *J. Comput. Chem.* **1991**, *12*, 885.

27. Frisch, M. J.; Binkley, J. S.; Schlegel, H. B.; Raghavachari, K.; Melius, C. F.; Martin, R. L.; Stewart, J. J. P.; Bobrowicz, F. W.; Rohlfing, C. M.; Kahn, L. R.; Defrees, D. J.; Seeger, R.; Whiteside, R. A.; Fox, D. J.; Fleuder, E. M.; Pople, J. A.; *Gaussian 92, Revision C*; Gaussian, Inc.: Pittsburg PA, 1992.

28. Helgaker, T.; Aa. Jensen, H. J.; Jørgensen, P.; Olsen, J.; Ruud, K.; Ågren, H.; Auer, A. A.; Bak, K. L.; Bakken, V.; Christiansen, O.; Coriani, S.; Dahle, P.; Dalskov, E. K.; Enevoldsen, T.; Fernandez, B.; Hättig, C.; Hald, K.; Halkier, A.; Heiberg, H.; Hettema, H.; Jonsson, D.; Kirpekar, S.; Kobayashi, R.; Koch, H.; Mikkelsen, K. V.; Norman, P.; Packer, M. J.; Pedersen, T. B.; Ruden, T. A.; Sanchez, A.; Saue, T.; Sauer, S. P. A.; Schimmelpfennig, B.; Sylvester-Hvid, K. O.; Taylor, P. R.; Vahtras, O.; *Dalton, a Molecular Electronic Structure Program, Release 1.2*; 2001.

29. Dunning Jr., T. H., Hay, P. J.; Schaefer III, H. F. eds.; *Electronic Structure Theory*, Plenun Press: New York, 1977, p.12.

30. Moller, C.; Plesset, M.S.; *Phys. Rev.* **1934**, *46*, 618.

31. Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M.; *Chem. Phys. Lett.* **1989**, *157*, 479.

32. *Ein*Sight 3.0*; Infometrix, 2200 Sixth Avenue, Suite 833, Seattle: WA 98121, 1991.