J. Phys. Chem. A **2000,** *104,* 11439−11442

On the Distribution of Local Molecular Symmetry in Crystals

Giuseppe Brancato and Francesco Zerbetto*

Dipartimento di Chimica "G. Ciamician", Università degli Studi di Bologna, V. F. Selmi 2, 40126 Bologna, Italy

Received: June 15, 2000; In Final Form: September 19, 2000

Application of the continuous symmetry measure to several hundred crystals that contain benzene, dichloromethane, chloroform, and acetone shows that the probability of finding the structures of these molecules distorted decreases exponentially with the amount of symmetry lowering. The implications of this trend are discussed in terms of B3LYP/6-31G* calculations and Boltzmann population.

1. Introduction

Presently, the interactions that exist in molecular crystals are reasonably well-understood and their description has reached a high level of sophistication; however, a general phenomenological model able to describe the molecular deformations that occur in the condensed phase is still being developed. To this end, database mining of crystal structures has become an increasingly popular and effective tool. To the best of our knowledge, it has been mainly directed toward the analysis of the distribution of bond lengths and bond angles, rather than toward developing rules for what appears to be an effect secondary to the energy lowering. Less crucial as it may be, it is still so general and widespread that it is part of every undergraduate textbook of solid-state chemistry and a firm component of our understanding of this state. The purpose of this work is to focus on the "fall from symmetry" of molecules embedded in crystals. The intent is to extract the global, or better statistical, effect of varying environments on molecules and possibly to establish rules that may be usefully employed to lead to the assembly of crystals. The nearly ubiquitous solvent molecules embedded in crystals are deemed to provide excellent work-horses for extensive exploration since the amount of information gathered for them probably equals or exceeds what is available for any other chemical object in the solid state. In practice, it is here assumed that four of the most common solvents, namely benzene, dichloromethane, chloroform, and acetone, present a statistically significant distribution of data amenable to investigations that can prove the existence of relationships which involve molecular distortion in the form of symmetry lowering. Importantly, the range of energies and structural parameters variations, i.e., changes in bond lengths and bond angles, considered here covers more than the region that is traditionally expected for solvent−solute interactions.

2. Continuous Symmetry Measure and Its Correlation with Energy

To obtain quantitative information about the distortion of a symmetric object requires the definition of a reference, which is here taken as the symmetry − not structure − of the molecule in the isolated phase. A number of approaches have been proposed to quantify a particular aspect of the measure of symmetry,[1] namely chirality. For symmetry in general, the continuous symmetry measure (CSM)[2] is likely to be the most general model proposed to date because it has the additional advantage to generate the reference structure "on the run". The method was successfully employed in a number of cases.[2] For instance, it was found to correlate with (i) the binding activity of biologically active systems, (ii) the π-electron energy in enantiomerization reactions of chiral fullerenes, and (iii) the enantiomeric excesses in organic reactions in an asymmetric synthesis.

In this approach, Avnir and collaborators calculate the CSM or symmetry deviation, S, as the shortest distance that points, i.e., atoms, have to undergo to transform the object they describe into the shape of desired symmetry. In practice, it is calculated as the square of the norm of the vector defined by the difference between the optimal, idealized structure and the actual structure. The optimal structure is not determined a priori but is obtained by numerical optimization techniques[3] for each molecule (herein for each of the several hundred distorted molecules considered). A simple equation connects S to the Cartesian coordinates of the atoms:

$$S = \frac{100}{nD^2}\sum_{i=1}^{n}(p_i - \hat{p}_i)^2 \tag{1}$$

with n being the number of equivalent points, or atoms, D being a normalization factor given by the distance between the center of mass of the molecule and its own farthest atom, p_i being the coordinates of the atoms, and $\hat{p}_i$ being the coordinates of the atoms of the optimal structure that are determined by the numerical optimization techniques. $S = 0$ coincides with perfect symmetry, i.e., no symmetry lowering, and $S = 100$ collapses the system to a point.

In agreement with this approach, in the four cases considered here (benzene, dichloromethane, chloroform, and acetone), only symmetry was imposed on the reference structures. They were D_{6h}, or C_6, for benzene (since the hydrogen atoms are notoriously poorly resolved in diffraction experiments, only carbon atoms were included in the treatment of eq 1 so that for six points the two point groups coincides), C_{2v} for dichloromethane and acetone, and C_{3v} for chloroform. The present results are therefore a measure of symmetry and not of similarity to a predetermined shape.

An important issue that deserves comment is the tendency of CSM to correlate with energy-related quantities such as binding constants, π-electron energies, and reaction yields.[2] As will be seen in the following, something similar is found here.

10.1021/jp002156r CCC: $19.00

TABLE 1: Summary of the Data for Benzene, Chloroform, Dichloromethane, and Acetone

	benzene	chloroform	dichloromethane	acetone
no. of structures	269	261	248	165
length range (Å)	$1.35 \le r_{CC} \le 1.44$	$1.73 \le r_{CCl} \le 1.85$	$1.73 \le r_{CCl} \le 1.85$	$1.47 \le r_{CC} \le 1.57$
angle range (deg)	any	$95 \le$ ClCCl ≤ 121	$95 \le$ ClCCl ≤ 121	$106 \le$ CCC ≤ 126
lowest *S* value	0.001	0.	<0.001	0.001
largest *S* value	1.5212	0.1169	0.0433	1.2387
A	68.8	205.9	150.4	84.5
slope, *b* (kcal mol^{-1})	34.88	115.14	not obtained	19.65
intercept, *c* (kcal mol^{-1})	2.19	1.72	not obtained	1.94
correlation coefficient, *r*	0.88	0.69	0.25	0.93
maximum ΔE (kcal mol^{-1})	57.51	15.83	13.01	25.00
average ΔE (kcal mol^{-1})	3.21	2.40	2.18	2.47
T (K)	255	281	not obtained	117

While the previously reported cases strengthen and are strengthened by the present one, it may be useful to investigate under which conditions, in the case treated here, a correlation must positively exist. The algebraic treatment described below is not used in the actual calculations but is proposed with the aim of understanding how one can express an energy increase due to distortion in terms of a symmetry-related quantity.

In general, to a molecular distortion must correspond an energy of distortion, ΔE_{dist}. The existence of a linear relationship between the *S* and energy is, however, not to be expected and when found is nontrivial because of the different forces that act on individual molecules. To exemplify with the case of benzene, its π-electron system can form both stacks with other π-electron-containing species or T-shaped complexes with electron-poor hydrogen atoms. The two adducts will tend to distort the molecule in different, apparently unrelated, directions.

To express the energy of a generic distortion one can start with the harmonic oscillator approximation and use the normal modes of vibrations. Other approaches are possible, but the discussion of their nature is outside the scope of the present paper. The treatment presented here is rather similar in that it provides the center of the Franck−Condon envelope in electronic transitions.[4] For a quadratic potential, the energy of distortion increases as one-half of the square of the difference, *Z*, between the initial and final points, multiplied by the force constant, k_i, of each mode:

$$\Delta E_{\text{dist}} = \frac{1}{2}\sum k_j Z_j^2 = \frac{1}{2}4\pi^2 c^2 \sum m_j \tilde{\nu}_j^2 Z_j^2 \qquad (2)$$

where *c* is the speed of light, m_j is the mass of the *j*th vibration, $\tilde{\nu}_j$ is the *j*th vibrational wavenumbers, and the sum goes over all the normal modes. Z_j can be written as:

$$Z_j = \sum_i (p_i - \hat{p}_i) L_i^{(j)} \qquad (3)$$

with the sum extending over all the atom positions and $L^{(j)}$ being the *j*th normal mode.

Both eqs 1 and 2 contain the square of a vector $(p_i - \hat{p}_i)$; eq 2 contains it through eq 3. Energy and CSM appear therefore to be related. Apart from several proportionality constants, the difference between the two equations is the presence of the normal modes, *L*, which, however, are orthonormal so that in a dot product such as that of eq 1 they effectively disappear.

In simple terms, a proportionality relation exists between energy and *S* when the symmetry decrease expressed by CSM can be decomposed along spatial directions, each of which gives a proportional energy contribution. It is concluded that the correlation between CSMs and the energy-related quantities used in this work originates from geometrical displacements along common preferential orthogonal directions (which actually may or may not be the vibrational normal modes).

3. Results and Discussion

The presence of symmetry in the molecules is a requirement for the CSM treatment. Benzene, dichloromethane, chloroform, and acetone were selected because of the presence of at least a 2-fold axis in their heavy atom frames. The Cambridge crystallographic database[5] was searched for their structures. To avoid systematic errors, a minimum of care must be exerted in the selection of the structures. Ideal structures, such as, for instance, the perfect 6-fold symmetry of benzene molecules, were not considered because of the reasonable assumption that their coordinates had been refined using a rigid body model. Also overly distorted structures were discarded. The ranges of bond lengths and angles considered in the search are given in Table 1. They are centered at about the experimental or the B3LYP/631G* optimized values. For benzene, structures were accepted when their nominally double CC bonds, r_{CC}, were $1.35 \le r_{CC} \le 1.44$ Å. For chloroform and dichloromethane, the same values were $1.73 \le r_{CCl} \le 1.85$ Å and $95° \le \alpha_{ClCCl} \le 121°$, with α the angle. The acetone structures were considered when their two nominally single CC bonds were $1.47 \le r_{CC} \le 1.57$ Å and the angle was $106° \le \alpha_{CCC} \le 126°$. For each molecule the number of structures is large enough to provide a statistically significant variety of cases from which to extract general trends (vide infra). Within these constraints, the largest *S* value is obtained for benzene, while the range of values obtained for dichloromethane is the narrowest. Attempts to correlate the individual *S* values to geometrical quantities of the crystals, such as interatomic distances or bond angles with neighboring systems, gave inconclusive results and will not be commented upon.

Figure 1 shows the distributions of *S* for the four solvents. To assist the eye, they are presented as histograms although histograms were not actually used in determining the nature of the distribution (see below). Several distributions were tested, but in the end, it was found that all the four molecules have a probability distribution of the distortion, *P*(*S*), that does not differ in a statistically significant way (acceptance at 95% confidence or more) from the exponential one (see refs 3 and 6 for a discussion of the statistical test used in the work):

$$P(S) = \exp(-aS) \qquad (4)$$

Table 1 gives *a* for each molecule (they range from 68.8 to 205.9).

The exponential form of the distributions can be chemically and physically nontrivial. Two interpretations come to mind. The less significant explanation could be summarized as “unusual events are rare”. Indeed, the symmetry fall could be

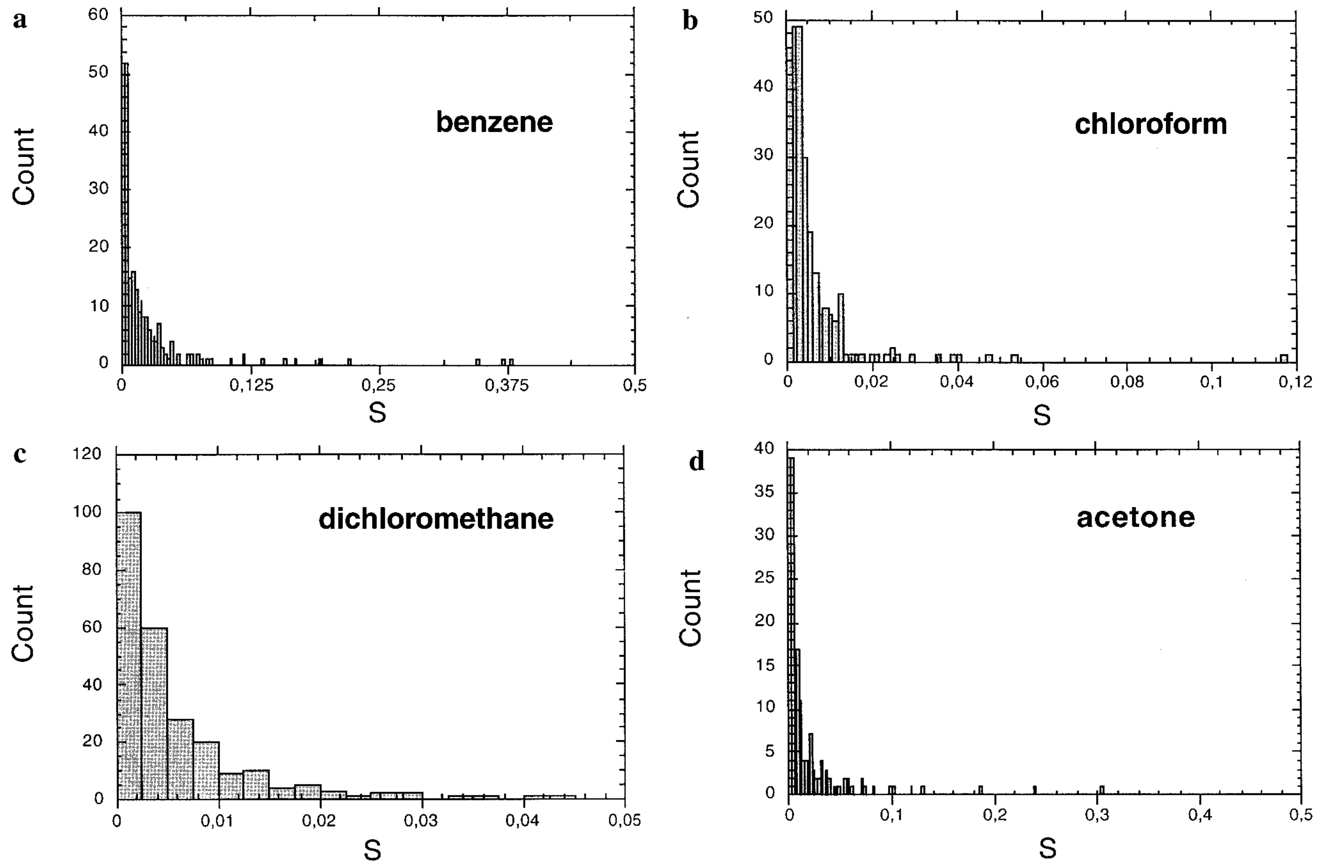

Figure 1. Histograms of the *S* values for (a) benzene, (b) chloroform, (c) dichloromethane, and (d) acetone. The histograms are provided for sake of simplicity of representation but are not used in determining the probability distribution of the symmetry lowering. The even longer tails of benzene and acetone are not shown.

envisaged as a decay process that is exponential in nature. An implication of the present result is that the data gathered from the Cambridge database provide significant statistics.

Alternatively, it is possible to take a stronger stance and attempt to provide a more physical picture. If the energy of the distortions can be separated from that of other processes, the probability of having a molecular deformation must follow the Boltzmann distribution which is, of course, exponential. In practice, the probability of populating, or more practically finding in the Cambridge database, a deformation of value S is given by $\exp(-\Delta E/k_BT)$, where ΔE is the energy of the distortion and T is the temperature at which the crystal has been grown. Most crystals are grown at room temperature. For a statistical set of data of a given molecule, T needs not be the room temperature but might be an effective temperature. For instance, the distortions due to the hydrogen atoms, which are not considered in the symmetry measure since their positions are usually not determined, may play a role in the final effective temperature.

The energies of distortions, ΔE, were obtained as the energy difference with the optimized B3LYP/6-31G*[7-9] structure. The hydrogen atoms were added both rigidly and optimizing their positions using molecular mechanics (the results did not change significantly and the latter are given). The energy ranges of the structures exceed 10 kcal mol^{-1}. A linear fit of S versus ΔE to give:

$$\Delta E = bS + c \tag{5}$$

was performed. The values of b and c are shown in Table 1. The plots of the fits are given as Supporting Information. As expected the results differ for the various molecules. The larger b the higher the energy cost to achieve the same S. The nonzero value of c includes the component of distortions that does not lead to symmetry lowering. For instance, in the case of benzene, an atom located on the 6-fold axis interacts with the molecule modifying its structural parameters but cannot decrease the symmetry. It is important to notice the large b-to-c ratio. In the three of the four systems investigated, the CSM scales linearly with the energy. This result is in keeping with eqs 2 and 3. The poorer fits (however, the correlation coefficients, r, were as high as 0.88 for benzene and 0.93 for acetone) can be due to the lack of a common set of orthonormal coordinates in eqs 2 and 3 or to the presence of anharmonic effects. Interestingly, a search of the Cambridge database coupled with molecular mechanics and quantum chemical calculations found that benzene is far from rigid and therefore quite possibly far from the harmonic behavior.[10] For dichloromethane, the fit gave a scattered plot, $r = 0.25$, and the corresponding values are not given. This dramatic failure could also be due to the small number, 3, of atoms that are used or to the fact that the range of bond lengths and bond angles that we used gave quite a small range of S values.

One can now set the right member of eq 4 equal to the Boltzmann distribution where the energy is given by eq 5. In the limit of c sufficiently small, it is possible to find the effective temperature. Surprisingly, benzene and chloroform gave an effective temperature (255 and 281 K) very close to the room temperature − the temperature most crystals are grown at! This is an unanticipated result. Applicability of the Boltzmann law means that the symmetry distortion is a degree of freedom that can be separated from all the others, just in the same way the various vibrational degrees of freedom are separable from each other. It also implies that for benzene (or chloroform) the distortions calculated from the structures found in the Cambridge database can be described by a set of common orthonormal coordinates. Even in the case of acetone, which gave a much lower effective temperature, the Boltzmann law may still hold. In acetone, in fact, torsion of the methyl groups is accomplished easily and they can contain a sizable part of the distortion and/or energy. In the crystal structure their positions are not determined and their contribution could increase the effective temperature substantially.

4. Conclusions

The target of this work was to discover some rules that describe the effect of crystal assembly on constituent molecules. This knowledge could, ultimately, be used to predict the probability of formation of a crystal. Probably the simplest of the effects one may consider is the distortion from perfect symmetry. Ideally, the route to easily understood and applicable rules for crystal assembly could evolve from the analysis of a large number of structures differentiated by a smooth variation of one molecular parameter at a time. The fabrication of crystals is, however, not at all a matter of routine. The rational sampling of the various crystal parameters has thus to be replaced by the statistical examination of the crystals available. Investigation of the CSM distribution of the structures of four solvents in crystals coupled with B3LYP/6-31G* calculation has shown that: (i) the probability of finding a symmetry-lowered molecular structure decreases exponentially with the size of the deformation and (ii) the exponential probability can be linked to the Boltzmann Law.

These two statements are easily put to practical use. Molecular simulations of parts of crystals can provide an independent estimation of the energy and the size of the symmetry lowering due to the interaction between neighboring molecules which can then be related to the probability of crystal assembly.

Acknowledgment. We thank Dario Braga and Giorgio Orlandi for interesting discussions and helpful comments on the topic of this work, Prof. D. Avnir for his useful comments on the manuscript, and all the referees for helping us clarify our work.

Supporting Information Available: Plots of S vs ΔE for benzene, chloroform, dichloromethane, and acetone; reference codes of the structures used in this work; and Fortran program to determine numerically the CSM for a set of six points. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Murray-Rust, P.; Rüsti, H. B.; Dunitz, J. D. *Acta Crystallogr. B* **1978**, *34*, 1787. Rassat, A. *Compt. Rend. Acad. Sci. Paris II* **1984**, *299*, 53. Gilat, G. *Chem. Phys. Lett.* **198**5, *121*, 9. Maruani, J.; Mezey, P. G. *Mol. Phys.* **1990**, *69*, 97. Buda, A. B.; Mislow, K. *J. Am. Chem. Soc.* **1992**, *114*, 6006. Ruch, E. *Acc. Chem. Res.* **1972**, *5*, 49.

(2) Zabrodsky Hel-Or, H.; Peleg, S.; Avnir, D. *J. Am. Chem. Soc.* **1992**, *114*, 7843.; Zabrodsky Hel-Or, H.; Peleg, S.; Avnir, D. *J. Am. Chem. Soc.* **1993**, *115*, 8278. Zabrodsky Hel-Or, H.; Peleg, S.; Avnir, D. *J. Am. Chem. Soc.* **1995**, *117*, 462. Keinan, S.; Avnir, D. *J. Am. Chem. Soc.* **1998**, *120*, 6152. Gao, D.; Schefzick, S.; Lipkowitz, K. B. *J. Am. Chem. Soc.* **1999**, *121*, 9481. Keinan, S.; Avnir, D. *J. Am. Chem. Soc.* **2000**, *122*, 4378−4384. Pinto, Y.; Fowler, P. W.; Mitchell, D.; Avnir, D. *J. Phys. Chem. B* **1998**, *102*, 5776.

(3) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes, The Art of Scientific Computing*, 2nd ed.; Cambridge University Press: Cambridge, 1992.

(4) Zerbetto, F. *J. Phys. Chem.* **1994**, *98*, 13157−13161. Zerbetto, F. *J. Am. Chem. Soc.* **1995**, *117*, 1621−1624.

(5) Cambridge Structural Database, Cambridge Crystallographic Data Centre, Cambridge, U.K.

(6) By inspection of the histograms, the non-Gaussian shape of the probability distribution appears evident and the question arises of the type(s) of distribution. While histograms can be instructive, they can hardly be used for quantitative work. The reason lies in the sensitivity of a fitting to the binning of the data.[3] Different bin sizes can result in different fitting parameters. Distribution of data can, however, be investigated with the Kolmogorov−Smirnov statistical test.[3] The test works on the normalized cumulative frequencies and finds the greatest discrepancy between the observed and expected values which is called the *D*-statistic. Comparison of the calculated *D*-statistic against the critical *D*-statistic for a given sample size allows to accept or reject the null hypothesis. In practice, if the calculated value is lower than the tabulated critical *D*-statistic, the form of the distribution is statistically acceptable. In agreement with the visual inspection of the data, the Gaussian distribution was found to be not statistically acceptable and was rejected at the 1% level or less.

(7) Becke, A. D. *J. Chem. Phys.* **1988**, *88*, 1053.

(8) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.

(9) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94*, revision B.2; Gaussian, Inc.: Pittsburgh, PA, 1995.

(10) Lipkowitz, K. B.; Peterson, M. A. *J. Comput. Chem.* **1993**, *14*, 121.