Structure of Molecules and Assemblies

István Hargittai\textsuperscript{a,b,*} and Magdolna Hargittai\textsuperscript{b,#}

\textsuperscript{a}Institute of General and Analytical Chemistry, Budapest Technical University, H-1521 Budapest

\textsuperscript{b}Structural Chemistry Research Group of the Hungarian Academy of Sciences at Eötvös University, H-1431 Budapest, Hungary

Received October 19, 1995; revised December 3, 1995; accepted December 8, 1995

A broad overview is presented about the importance of molecular structure determination and about the structural considerations that come in view when molecular units build assemblies. Structural chemistry and crystallography constitute an integral part of today's supramolecular chemistry as demonstrated by various statements of giants of the field. The challenge of supramolecular chemistry to the structural chemist is in detecting and understanding the structural changes accompanying molecular recognition and assembly formation and in providing assistance for the design of new assemblies and molecular devices with desired properties.

The sphere under the lion's paw in Figure 1 is decorated by a hexagonal pattern but a closer inspection reveals the presence of pentagons among the hexagons.\textsuperscript{1} This example of a »fullerene« illustrates the omnipresence of «chemically» interesting structures. The molecular geometry of buckminsterfullerene was determined in 1991 and the data are collected in Table I.\textsuperscript{2–5} It was not until this direct determination that the structure, and even the existence of buckminsterfullerene could be considered as unequivocal fact.

\textsuperscript{*} For the 1996/97 academic year, Distinguished Visiting Professor, Department of Chemistry, University of North Carolina at Wilmington, Wilmington, NC28403, U.S.A.

\textsuperscript{#} For the 1996/97 academic year, Visiting Scientist, Department of Chemistry, University of North Carolina at Wilmington, Wilmington, NC28403, U.S.A.
Figure 1. Sculpture of guardian lion in the Forbidden City, Beijing, China with a "fullerene" structure under the lion's paw.

TABLE I

Bond lengths in buckminsterfullerene

<table>
<thead>
<tr>
<th>Lengths of shared edges of rings</th>
<th>Gas ED$^a$</th>
<th>Neutron crystallogr.$^b$</th>
<th>X-ray crystallogr.$^c$</th>
<th>Ab initio MO calc.$^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1000 K</td>
<td>5 K</td>
<td>110 K</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$r_g$</td>
<td>$r_a$</td>
<td>$r_a$</td>
<td>$r_a$</td>
</tr>
<tr>
<td>C(5)-C(6),$^e$ Å</td>
<td>1.458(6)</td>
<td>1.455(12)</td>
<td>1.445(5)</td>
<td>1.45</td>
</tr>
<tr>
<td>C(6)-C(6),$^f$ Å</td>
<td>1.401(10)</td>
<td>1.391(18)</td>
<td>1.399(7)</td>
<td>1.39</td>
</tr>
</tbody>
</table>
The importance of the knowledge of molecular geometry is illustrated by a few statements here:

»There is no more basic enterprise in chemistry than the determination of the geometrical structure of a molecule.« {Roald Hoffmann (1983).6}

»The central problem in the identification of a new chemical compound is the determination of its molecular structure.« {Felix Franks (1981).7}

»No one really understands the behaviour of a molecule until he knows its structure that is to say: its size, and shape, and the nature of its bonds.« {C. A. Coulson (1972).8}

»The most important characteristics of a chemical bond is its length.« {L. Pauling, attributed.}

»Form is a diagram of forces.« {D'Arcy W. Thompson (1917).9}

Looking back to the development of our ideas about molecular structure, we listed a few important stages in Table II. This is by far not a complete

TABLE II

Selected Events in the Development of Structural Chemistry Ideas

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democritos (460–370 B.C.E.)</td>
<td>Nothing exists except atoms and empty space; everything else is opinion.</td>
</tr>
<tr>
<td>Kepler 1611</td>
<td>Ubi materia, ibi geometria (Where there is matter, there is geometry)</td>
</tr>
<tr>
<td>Dalton 1808</td>
<td>[Atoms]</td>
</tr>
<tr>
<td>Avogadro 1811</td>
<td>Equal volumes of all gases at the same temperature and pressure contain the same number of molecules</td>
</tr>
<tr>
<td>Gay-Lussac (1778–1850)</td>
<td>We are perhaps not far removed from the time when we shall be able to submit the bulk of chemical phenomena to calculation.</td>
</tr>
<tr>
<td>Pasteur 1848</td>
<td>[Handedness in molecules/crystals]</td>
</tr>
<tr>
<td>Paterno 1869</td>
<td>[Tetrahedral carbon, conformational isomers]</td>
</tr>
<tr>
<td>van't Hoff 1874</td>
<td>[Tetrahedral carbon]</td>
</tr>
<tr>
<td>Butlerov 1861</td>
<td>[Chemical structure (structure = size &amp; shape)]</td>
</tr>
<tr>
<td>Mendeleev 1869</td>
<td>[Periodic Table of the elements]</td>
</tr>
<tr>
<td>Lewis 1916</td>
<td>[Electron pair/covalent bond]</td>
</tr>
<tr>
<td>Pauling 1939</td>
<td>[Geometry &amp; bonding; 0.01% of today's structural information]</td>
</tr>
<tr>
<td>Hassel 1943</td>
<td>[Conformational analysis]</td>
</tr>
<tr>
<td>Bartell 1955</td>
<td>[The meaning of »r«]</td>
</tr>
</tbody>
</table>
list and is certainly biased by our current interests. We would like to single out three entries for special mention here. One is Gay-Lussac’s statement\(^\text{10}\) a hundred and fifty years ago about the growing importance of calculations in chemistry. The other is Linus Pauling’s achievement in the observation of trends and patterns on the basis of data on structure and bonding in the first edition of his *The Nature of the Chemical Bond*\(^\text{11}\) in 1939. It has been estimated\(^\text{12}\) that at that time he possessed merely one hundredth of one percent of the structural information available today, yet his observations have withstood the test of time. When Pauling (Figure 2) was asked\(^\text{13}\) about this and about the discovery of buckminsterfullerene, this is what he had to say: «I am rather surprised that no one had predicted the stability of \(\text{C}_{60}\). I might have done so, especially since I knew about the 60-atom structure with icosahedral symmetry, which occurs in intermetallic compounds. It seems to be difficult for people to formulate new ideas. An example is that from 1873 to 1914 nobody, knowing about the tetrahedral nature of the bonds of the carbon atom, predicted that diamond has the diamond structure.»

The third point we would like to single out here is L. S. Bartell’s pioneering study in 1955\(^\text{14}\) in which he argued forcefully, if also a bit ahead of time, that in the interpretation of interatomic distances originating from various experiments, vibrational effects and other consequences of motion need be considered.

### TABLE III

Techniques for the Determination of Molecular Geometry

<table>
<thead>
<tr>
<th>Medium</th>
<th>Technique</th>
<th>Operational Geometry</th>
<th>Physically Meaningful Geometry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solid</td>
<td>X-ray Diffraction</td>
<td>(r_\alpha)</td>
<td>(r_\alpha)</td>
</tr>
<tr>
<td></td>
<td>Neutron Diffraction</td>
<td>(r_\alpha)</td>
<td>(r_\alpha)</td>
</tr>
<tr>
<td>Gas</td>
<td>Electron Diffraction</td>
<td>(r_\alpha)</td>
<td>(r_\alpha) &amp; (r_\alpha)</td>
</tr>
<tr>
<td></td>
<td>Rotational Spectroscopy</td>
<td>(r_0, r_s)</td>
<td>(r_s)</td>
</tr>
<tr>
<td>Various</td>
<td>NMR Spectroscopy</td>
<td>(r_\alpha)</td>
<td>(r_\alpha)</td>
</tr>
<tr>
<td>Computer</td>
<td>Theoretical Calculations</td>
<td>(r_\alpha)</td>
<td>(r_\alpha)</td>
</tr>
</tbody>
</table>

Today the principal techniques of the determination of molecular geometry are the diffraction methods, high-resolution rotational spectroscopy, NMR spectroscopy, and theoretical calculations. Table III presents a list of these techniques with the indication of the internuclear distance types yielded directly by these techniques, called «operational geometry» and the internuclear distance types that can be obtained from them by various corrections and conversions and that have well-defined physical meaning.
Some explanations of the various distance types follow here (see, e.g., Ref. 15).

The so-called operational parameters are the direct output of experimental studies. They do not have well-defined physical meaning. The most important and common ones are listed here.

$r_a$  Effective internuclear distance, obtained directly from the analysis of electron diffraction intensities. Its conversion into $r_g$ distance (see below) is simple with a very good approximation, $r_g \approx r_a + l^2/r_a$, where $l$ is the mean vibrational amplitude. In other words, there is no need to use $r_a$ in any comparison; it is preferable to use $r_g$.  

---

Figure 2. Linus Pauling on the cover of the first issue of *The Chemical Intelligencer*. 
Effective internuclear distance, obtained from the rotational constants; usually refers to the ground vibrational state. Since it depends strongly on the isotopic composition, it may differ from the equilibrium distance by a couple of hundredths of an Ångström.

Effective internuclear distance determined from the isotopic substitution coordinates of the respective atoms. Since it depends slightly on the isotopic compositions, it may differ from the equilibrium distance by a few thousandths of an Ångström.

Internuclear distances with well-defined physical meaning are the following:

- \( r_o \): Equilibrium internuclear distance between equilibrium nuclear positions in the minimum position of the potential energy function. No experiment yields directly this parameter. All computed geometries, in principle, correspond to this distance, but only in principle, of course. Basis-set choice, approximations, and all computational conditions may influence the results.

- \( r_g \): Distance-average incorporating the effect of all vibrations at temperature \( T \). This is the parameter attainable in a straightforward way from electron diffraction.

- \( r_{a0}/r_z \): Distance between average nuclear positions in the ground vibrational state; \( r_{a0} \) and \( r_z \) have the same meaning; \( r_z \) originates from rotational spectra applying vibrational corrections. \( r_{a0} \) is the distance between average nuclear positions averaged over all vibrational states at temperature \( T \). \( r_a \) and \( r_{a0} \) are obtained from electron diffraction applying vibrational corrections.

Since of these techniques, gas-phase electron diffraction is little known, and since it is the principal experimental technique of our research group, we would like to devote a few words to it. In fact, the cover illustrations (Figures 3a and 3b) of a two-volume treatise about the stereochemical applications of this technique\(^{16}\) summarize concisely what we have to say. Let's consider first the technique (Figure 3a, descending from the upper right corner). The essence of the experiment is the production of a diffraction pattern by letting a beam of fast monochromatic electrons onto a beam of molecules. By separating the molecular contribution to the total electron scattering pattern from the rest, and Fourier transforming it, we obtain something like a probability density distribution of intramolecular internuclear distances. The internuclear distances can be read off directly from this distribution for simple systems. The analysis is usually done on the basis of the molecular contribution to the electron scattering pattern and model building is an important part of this analysis. A few important applications are illustrated in Figure 3b (ascending from the lower right corner), as the determination of the structure of monomers and dimers, the determination of structural changes in series of related molecules, conformational analysis, and comparison of gaseous and crystalline molecular structures (this latter from crystallographic studies).
Computational techniques have become an equal partner to the experimental physical techniques in the determination of molecular structure. Just as it is important to distinguish and scrutinize the meaning of parameters originating from different physical techniques, it is mandatory, for any demanding comparison, to consider the difference in the physical meaning of the computed and experimentally determined structures. «For truly accurate comparison experimental bond lengths should be compared with computed ones only following necessary corrections, bringing all information involved in the comparison to a common denominator.»17. The differences in the distances due to the difference in physical meaning may easily exceed the precision of modern structure determination, and may even be greater, by orders of magnitude, for fluxional molecules. An in-depth and critical survey of all techniques of molecular structure determination, along with a discussion of the application of accurate structural information is available in a monograph Accurate Molecular Structures, prepared for the International Union of Crystallography.18

The energy requirements of geometrical changes merit special attention. It has been estimated19 that, for a carbon-carbon chain, a typical bond stretch of 0.1 Å requires about 15 kJ/mole, a bond angle deformation of 10 degrees about
5 kJ/mol, and a torsional distortion of 15 degrees about 1 kJ/mol. These changes do not occur separately but should be considered as parts of the overall structure relaxation, and should be treated in a concerted way, especially beyond certain accuracy requirements.

There are well-documented cases of gas/solid structure differences, involving the consequences of intermolecular interactions in the crystal.²⁰–²² Kitaigorodskii (Figure 4) did pioneering studies of these interactions (see, e.g., Ref. 23), and his relatively simple geometrical model successfully served for predicting the occurrence frequencies of three-dimensional space groups among crystal structures. Both symmetry considerations and considerations of optimal space utilization have been considered (for further discussion, see, e.g., Ref. 24).

The packing of molecules in the crystal is best accomplished by complementarity of shapes, the protrusions fitting in the dents and vice versa. Already Lucretius noticed in his De Rerum Natura, 2000 years ago, that «Things whose textures have such a mutual correspondence, that cavities fit solids, the cavities of the first the solids of the second, the cavities of the second the solids of the first, form the closest union.»²⁵ The same principle is stressed today, as illustrated, for example, by a recent observation,²⁷ concerning the association of helical peptides and ion channels: «... The various apolar side chains are not particularly selective with respect to packing except for spatial considerations. Bulges try to fit into grooves. ...»
In the process of molecules getting together via interactions, however weak they may be as compared with the usual covalent bond, these interactions may lead to intramolecular structural changes. It is then a question of the accuracy of our detection techniques whether these changes become appreciable or not. Hilgenfeld and Saenger called attention to the importance of the subtle structural changes in inclusion phenomena already in 1982: «Subtle changes of molecular structure may result in severe changes of inclusion behaviour of a potential host molecule due to the complicated interplay of weak intermolecular forces that govern host-guest complex formation.» We would like to mention our own contribution to stressing the »Importance of Small Structural Differences« back in 1987.

These studies have gained added importance (see Refs. 30 and 31) with the emergence of Supramolecular Chemistry, signified by the 1987 Nobel prize in Chemistry to Donald J. Cram (Figure 5a), Jean-Marie Lehn (Figure 5b), and the late Charles J. Pedersen (1904–1989) »for their development and use of molecules with structure-specific interactions of high selectivity.« The pioneering move was the preparation of the crown ethers which then accommodated various guests in a mode of interaction weaker than the covalent...
bond. A nice example of such interactions and their structures was presented by Dr. I. Matijasíc and colleagues at this Meeting.\textsuperscript{32}

Hydrogen bond formation is a conspicuously important intramolecular and intermolecular interaction, and it is weaker than the covalent bond. We have seen interesting examples of its crystallographic investigation at this Meeting in the reports by Gordana Pavlović,\textsuperscript{33} Marijana Gavranic,\textsuperscript{34} and Tomislav Friščić,\textsuperscript{35} and others.

In our research group we have been interested in probing into the geometrical consequences of intramolecular hydrogen bond formation in the rest of the molecule. We consider these to be model systems in which our objects are free of any other interaction since they are carried out on the isolated molecule in the gaseous state. We hope that these studies facilitate the understanding of the structural consequences of the formation of supramolecular assemblies.

Comparison of selected parameters of 2-nitrophenol\textsuperscript{36} and nitroresorcinol\textsuperscript{37,38} on the one hand with the parent molecules, phenol\textsuperscript{39} and nitrobenzene,\textsuperscript{40} on the other, reveals the geometrical consequences of intramolecular hydrogen bond formation in these systems. The consequences are displayed as changes in the parameters. There are marked bond lengthenings and shortenings, angular openings, and tilts (Table IV). We have paralleled our ex-

\begin{table}
\centering
\caption{Differences of Selected Parameters\textsuperscript{a} Demonstrating Geometrical Consequences of Intramolecular Hydrogen-Bond Formation}
\begin{tabular}{lccccc}
\hline
Parameter & \multicolumn{2}{c}{2-Nitrophenol/Phenol} & \multicolumn{2}{c}{2-Nitroresorcinol/Phenol} \\
 & Electron Diffraction & Calculation MP2(FC)/6-31G* & Electron Diffraction & Calculation MP2(FC)/6-31G* \\
\hline
O-C, Å & -0.022 & -0.024 & -0.027 & -0.024 \\
(O)C-C-N, Å & +0.012 & +0.014 & +0.027 & +0.028 \\
O-C-C(N), deg & +2.7 & +2.8 & +1.6 & +1.9 \\
CO tilt, deg & +1.6 & +1.3 & +1 & +1.2 \\
\hline
Parameter & \multicolumn{2}{c}{2-Nitrophenol/nitrobenzene} & \multicolumn{2}{c}{2-Nitroresorcinol/Nitrobenzene} \\
 & Electron Diffraction & Calculation MP2(FC)/6-31G* & Electron Diffraction & Calculation MP2(FC)/6-31G* \\
\hline
N-C, Å & -0.022 & -0.015 & -0.037 & -0.032 \\
(O)C-C(N), Å & +0.011 & +0.018 & +0.026 & +0.033 \\
N-C-C(O), deg & +2.5 & +2.4 & +3.1 & +2.6 \\
O-N-C(O), deg & +0.9 & +0.7 & +2.0 & +2.0 \\
CN tilt, deg & +1.5 & +1.8 & – & – \\
\hline
\end{tabular}
\textsuperscript{a} For references, see text
\end{table}
perimental gas-phase electron diffraction analyses by *ab initio* molecular orbital calculations at the MP2/6-31G* level to establish the consistency of our findings and also to prepare ground for further studies in which we are combining electron diffraction experimental data and computed information in the analysis. This may be necessary for the study of systems of lower symmetry where additional assumptions serve as constraints in the electron diffraction analysis. Mindful of the differences in the physical meaning of the parameters, we incorporate *parameter differences* only rather than the parameter themselves in the concerted analyses. The studies of 4,6-dinitroresorcinol and salicylaldehyde are further examples. The computational investigation of o-trifluoromethylphenol has demonstrated the hydrogen bond formation in a less symmetrical configuration by electron density distribution (Figure 6) and by the changes of bond lengths in the rest of the molecule.

Coming back to the fundamental intermolecular interactions and the ideas of Lucretius and Kitaigorodskii, it is worth quoting a very resonant description of host-guest complexation. According to Cram and Cram, the hosts are organic molecules containing convergent binding sites and the guests are molecules or ions containing divergent binding sites.

![Figure 6](image-url)

*Figure 6. The cross-section of electron density distribution of two models of trifluoromethylphenol. (a) The minimum and (b) a saddle point. The plane of the cross-section is defined by the nuclei of the hydroxy hydrogen and the fluorine closest to it and by the geometrical center of the molecule. The electron density contour lines denote electron densities from 0.013 to 0.200 e/au³. The length of the H...F hydrogen bond is 1.98 Å and geometrical changes are also indicated (in ångström units) as compared with phenol and trifluorobenzene.*
In all this, molecular recognition is the central phenomenon, and it is here that structural considerations appear to be decisive in intermolecular interactions making supramolecular chemistry. According to Lehn, "Molecular recognition involves the specific interaction of one part of a molecule with a particular part of another molecule and it implies the (molecular) storage and (supramolecular) retrieval of molecular structural information." Snježana Antolič made molecular recognition the starting point in her discussions of structure-activity correlations presented at this Meeting.

The energy costs of molecular recognition interactions is a crucial consideration. There is a gradual transition from supermolecules to supramolecular assemblies as supramolecular chemistry comprises two broad areas (Lehn): 1. oligomolecular supermolecules and 2. supramolecular assemblies, such as layers, films, membranes, vesicles, micelles, microemulsions, gels, mesomorphic phases, solid state species, etc.

Investigating the transition from the isolated molecule to the molecular aggregation provides unique means of learning about the interactions of molecules in the assembly, and, in particular, about how the assembling process alters the structure of the individual molecule.

Possibly the most important outcome of such studies is the information about how biological systems operate. According to Cram, "The exquisite chemical activities of biological processes depend largely on complexation involving large numbers of weak but additive interactions. ... Enzymic catalysis, immune responses, genetic information storage, retrieval and replication can all be modelled."

Crystallography has a special place in supramolecular research because the molecular crystal is the supramolecule par excellence. Quoting Jack Dunitz (Figure 7), "... molecules in a periodic arrangement by just the same kind of non-covalent bonding interactions as are responsible for molecular recognition and complexation at all levels... the crystallization process itself is an impressive display of supramolecular self-assembly, involving specific molecular recognition at an amazing level of precision. Long-range periodicity is a product of directionally specific short-range interactions... Crystals are ordered supramolecular systems.« Furthermore, "If a crystal is a supramolecule, then polymorphic modifications are supra-isomers and polymorphism is a kind of supra-isomerism; thus, diamond, graphite, and the fullerenes are an extreme example of a family of supra-isomers.« Very similar concepts are expressed with very similar words by Jean-Marie Lehn in his discussion of molecular recognition-directed self-assembly of ordered solid-state structures." The control of the arrangement of molecules in the solid state depends on the intermolecular interactions and on the packing factors. It is a problem of supramolecular nature... The solid is thus... a very large supermolecule, whose formation is based on molecular recognition and...
self-organization processes. Accordingly, the crystal represents the ultimate of the extended but periodic supramolecular entity. Solid state polymorphs may then be considered as supramolecular isomers and the conversion of one polymorphic form into another one as a supramolecular isomerization.

Molecular recognition effects provide an entry into crystal engineering... Hydrogen bonding patterns may direct structure formation in the solid as well as in solution...

If the mode of molecular packing in molecular crystals is so important then the test of our understanding these materials is indeed whether crystal structures can be predicted simply on the basis of chemical composition or not. There are many subquestions of this question that can be handled successfully but the fundamental question «Are crystal structures predictable» can yet be answered in the negative only. Peter Zorkii has formulated eloquently the difficulties by saying that «... the mechanism of formation of primary molecular agglomerates is often memorized in the final arrangement of molecules and the resulting crystal hardly corresponds to a global minimum of the multidimensional surface describing the dependence of the potential energy on the structure parameters.»
Rather than by grand design, it is by painstaking steps, by the determination of the structure of isolated molecules, molecular assemblies, and crystals, by the observation of trends and patterns, the enrichment of data banks and by ever growing computational studies and model building is what seems to pave the way of a fuller understanding of the structure of molecules and assemblies.

REFERENCES

25. Lucretius, *De Rerum Natura* quoted by Dunitz.  
SAŽETAK

Struktura molekula i njihovih skupova

István Hargittai i Magdolna Hargittai

Dan je opsežan prikaz važnosti određivanja molekularne strukture i strukturnih promišljanja koja nadolaze kad molekularne jedinice grade skupove. Strukturna kemija i kristalografija sastavni su dio današnje supramolekularne kemije što pokažuju brojna očitovanja velikana tog područja. Izazov supramolekularne kemije strukturnom kemičaru je u otkrivanju i razumijevanju strukturnih promjena koje prate molekularno prepoznavanje i stvaranje skupova i u potpomaganju oblikovanja novih skupova i molekularnih pronalazaka sa željenim svojstvima.