ELSEVIER

Characterizing cavities in model inclusion molecules: A comparative study

Francisco Torrens, José Sánchez-Marín, and Ignacio Nebot-Gil

Departamento de Química Física, Facultad de Química, Universitat de València, Valencia, Spain

We have selected fullerene-60 and -70 cavities as model systems in order to test several methods for characterizing inclusion molecules. The methods are based on different technical foundations such as a square and triangular tessellation of the molecule taken as a unitary sphere, spherical tessellation of the molecular surface, numerical integration of the atomic volumes and surfaces, triangular tessellation of the molecular surface, and a cubic lattice approach to a molecular space. Accurate measures of the molecular volume and surface area have been performed with the pseudo-random Monte Carlo (MCVS) and uniform Monte Carlo (UMCVS) methods. These calculations serve as a reference for the rest of the methods. The SURMO2 and MS methods have not recognized the cavities and may not be convenient for intercalation compounds. The programs that have detected the cavities never exceed 5% deviation relative to the reference values for molecular volume and surface area. The GEPOL algorithm, alone or combined with TOPO, shows results in good agreement with those of the UMCVS reference. The uniform random number generator provides the fastest convergence for UMCVS and a correct estimate of the standard deviations. The effect of the internal cavity on the accessible surfaces has been calculated.

Keywords: molecular cavities, geometric descriptors, topological indices, fractal dimension, solubility, partition coefficient, fullerene

INTRODUCTION

The accurate description of different cavities in molecules is convenient for the study of structured media (i.e., systems that present a time-invariant alternancy of empty and filled space) such as cage molecules (cryptates) and intercalation compounds (graphite, chalcogenides, oxides, clays, fullerenes, polymers, proteins, etc.). Analyses of molecular packing in liquids and crystals[1,2] have shown that cavities, defined as empty spaces, may occur as a result of local packing defects. In proteins, internal cavities large enough to accommodate even xenon atoms have been identified.[3–6] Although the existence of these internal cavities has been largely associated with conformational flexibility,[7–9] their role remains poorly understood.

The characterization of cavities can be applied in studies of enzyme–substrate compatibility; in the chemistry of solvation, cage compounds, and transition metal complexes; in catalysis by inclusion in zeolites; in studies of solid–gas reactions in which channels in the crystal structure allow selective diffusion of the gas in the solid; and in solid-state reactions that require mobility (and hence free space) in their initial stages, when the crystal structure is not yet distorted by product formation. Internal cavities affect protein stability,[10] and play an important role in the close packing of amino acid side chains, particularly in the core of the protein.[5,11,12] The knowledge of the channels and cavity structures within crystals of globular proteins may be useful[13–21] in detecting the binding sites of specific ligands and ions on accessible surfaces of protein molecules within channels; in studying the structure and physical properties of the water near protein surfaces within the channels; for predicting the narrowest sites of the channels for diffusion of molecules of arbitrary sizes, etc.

In physical models of molecules, atoms are represented by spheres whose size is determined by the atomic van der Waals radii. Molecular models are thus commonly represented as systems of fused hard spheres with unequal radii, and there has been an important body of work on the computation of surface area and volume of such systems[22] and on the topological description of their surfaces.[23,24]

For small molecules, the van der Waals surface gives a good description of the molecular shape and the outer surface. In large molecules such as proteins, where most of the van der Waals surface is buried in the interior, other, more suitable definitions such as the solvent-accessible surface[25,26] and

Address reprint requests to: F. Torrens, Departamento de Química Física, Facultad de Química, Universitat de València, Dr. Moliner 50, E-46100 Burjassot (Valencia), Spain.

Received 7 January 1998; revised 2 April 1998; accepted 16 April 1998.

655 Avenue of the Americas, New York, NY 10010

1093-3263/98/$19.00
PII S1093-3263(98)00011-4

solvent-excluding surface[11] have been used. The corresponding surface areas can be computed from the atomic coordinates by approximate procedures[25,26] or by using analytic algorithms such as those of Gibson and Scheraga,[22] Connolly,[27–29] or Richmond.[30] The analytic algorithms have the advantage of being more accurate. They also enable the calculation of first-order derivatives of the surface area, which is an important requirement for the use of surface area functions in optimization procedures.[31]

Methods to compute atomic and molecular volumes of proteins from the atomic coordinates have also been derived, originally by Richards[5] and Finney,[32] and later implemented in automatic algorithms.[27–30] On the basis of calculations of surface, protein cavities have been detected and analyzed by computer, using, e.g., procedures for distinguishing between the surfaces of cavities and outer surfaces.[3,25] However, all of the reported cavity detection algorithms suffer from a number of shortcomings that influence the results. The most common are (1) difficulties in detecting cavities smaller than a certain size, or cavities that are separated by just one layer of spheres from the surface, and (2) the tendency to detect spurious cavities when parts of the outer surface are connected through narrow channels.

Alard and Wodak[33] have described a procedure that is able to detect cavities in a system of interpenetrating spheres, no matter what their size or shape. The procedure requires as input a set of spherical polygons that covers the whole solvent-accessible surface. Such polygons are generated by an analytic algorithm for computing surface areas and volumes that is similar to those of Connolly[27–29] and Richmond.[30] The procedure maps the spherical polygons into sets of disconnected surfaces: the outer surface and the surfaces of all the cavities. It is based on purely topological principles. Moreover, it involves only logical operations and is therefore not subject to numerical imprecision. Once the cavities are detected, their surface areas and volumes are readily computed.

In the present article we have selected the (I_h)fullerene-60 (C_{60}, buckminsterfullerene) and (D_{5h})fullerene-70 (C_{70}) molecules as model cavities in order to test various methods for characterizing inclusion molecules. These inclusion molecules are more suitable than proteins for testing precise algorithms, because of their moderate sizes. We have used the following programs and the methods implemented in them for characterizing the molecular cavities of fullerenes: (1) SURMO2 (surface of molecule), (2) the MS (molecular surface) program by Connolly,[27–29] (3) SCAP (solvent-dependent conformational analysis program), (4) the GEPOL (geometry of polyhedron) program, (5) a program for the computation of the surface areas of molecules (we call it CSAM), (6) the TOPO program, (7) pseudorandom Monte Carlo computation of the molecular volume and surface (MCVS), and (8) uniform Monte Carlo computation of the molecular volume and surface (we call it UMCVS), with a random number generator (RNG) uniformly distributed over a given finite range [a, b].

In the next section, the different methods used in the preceding programs are briefly described, as are the new features that we have included in them. Following that, results are presented and discussed. The last section summarizes our conclusions.

METHODS

Square and triangular tessellation of the molecular surface (program SURMO2)

It is helpful for tessellation methods to represent the global shape of a molecule as an envelope of the van der Waals spheres of the external atoms.[34] This allows for a numerical treatment of integrals defined on the molecular surface, in the form

$$I = \int g(\rho, \theta, \phi) d\omega \tag{1}$$

where g is the position vector of a point in the surface and $d\omega = \sin \theta d\theta\, d\phi$ represents the elementary solid angle. The evaluation of the Eq. (1)-type integrals is obtained by calculating the finite sum:

$$I = \int g\, d\omega \approx \sum_i g_i \delta\omega_i$$

To achieve the summation one can take a mesh of points $M_{ij}(\theta_i, \phi_j)$ as dense as desired by cutting the surfaces on a single unitary sphere with a uniform distribution of $2N_1$ parallels and N_2 meridians. This operation defines a set of $2N_1N_2$ surfaces that are either square or triangular (in the neighborhood of the poles).

If the molecule is cut by an axis passing by the origin, two intersecting points that correspond to two values of the g function are obtained: g^+ and g^-. The axis is forced to cut successively the unitary sphere in the nodes of the mesh previously defined. For each elementary surface, four values of g (only three on the poles) are obtained. The g value selected for each surface element is the mean of the values obtained in its nodes.

Denoting by M_0 the upper pole, M_{1j} the points placed on the first parallel, M_{ij} those of the ith parallel ($1 < i < N_1$), and $M_{N_1 j}$ those placed on the equator ($0 \le j < N_2$), the integral in Eq. (1) can be measured as

$$I = \frac{N_2}{3}\left(g^+_{M_0} + g^-_{M_0}\right) + \frac{7}{6}\sum_{j=0}^{N_2-1}\left(g^+_{M_{1j}} + g^-_{M_{1j}}\right) + \frac{1}{2}\sum_{j=0}^{N_2-1}\left(g^+_{M_{N_1 j}} + g^-_{M_{N_1 j}}\right) + \sum_{i=2}^{N_1-1}\sum_{j=0}^{N_2-1}\left(g^+_{M_{ij}} + g^-_{M_{ij}}\right) \tag{2}$$

where $\sum_j g_{M_{ij}}$ indicates a sum over all of the mesh points placed on the ith parallel. This relationship assumes that the nodes are opposite two by two on the unitary sphere.

The geometric descriptor most directly connected to the shape is the molecular volume:

$$V = \frac{1}{3}\int \rho^3 d\omega$$

where ρ is the radial spherical coordinate. Using Eq. (2) the volume is

$$V = \frac{4\pi}{3 \cdot 2N_1N_2} \sum_{i=0}^{N_1} r_{M_i} \sum_{j=0}^{N_2-1} [(g_{M_{ij}}^{+})^3 + (g_{M_{ij}}^{-})^3]$$

where $r = (1/3, 7/6, 1, 1 \ldots 1/2)$. A g element is represented in Figure 1A.

The SURMO2 program[34] uses the preceding algorithm to calculate the molecular volume and coefficients characterizing the deviation from the spherical shape. Using the same principles, in particular in Eq. (2), we have implemented in it the calculation of the bare molecular surface and accessible surface areas and we have designated this version as SURMO2-I. In a first approximation, S can be calculated as

$$S = \int \rho^2 d\omega = \frac{4\pi}{2N_1N_2} \sum_{i=0}^{N_1} r_{M_i} \sum_{j=0}^{N_2-1} [(g_{M_{ij}}^{+})^2 + (g_{M_{ij}}^{-})^2]$$

We have also corrected the deviation from the spherical shape by dividing each g^2 value by the cosine of the angle formed by the semiaxis and the corresponding normal vector to the surface at this point (Figure 1B):

$$S = \frac{4\pi}{2N_1N_2} \sum_{i=0}^{N_1} r_{M_i} \sum_{j=0}^{N_2-1} \frac{(g_{M_{ij}}^{+})^2 + (g_{M_{ij}}^{-})^2}{\cos \alpha_{M_{ij}}}$$

We have designated this version as SURMO2-II.

Spherical tessellation of the molecular surface (program MS)

The solvent-accessible surface area was originally defined and computed by Lee and Richards[25] as the area traced out by the center of a probe sphere representing a solvent molecule as it is rolled over the surface of the molecule of interest. The solvent-excluding surface has also been defined by Richards.[11] It consists of the part of that van der Waals surface of the atoms that is accessible to the probe sphere (contact surface), connected by a network of concave and saddle-shaped surfaces (reentrant surface) that smoothes over the crevices and pits between the atoms. Note that, unlike the solvent-accessible surface, the solvent-excluding surface is defined as a part of the van der Waals surface. This surface is the boundary of the volume from which a probe sphere is excluded if it is not to experience van der Waals overlap with the atoms.

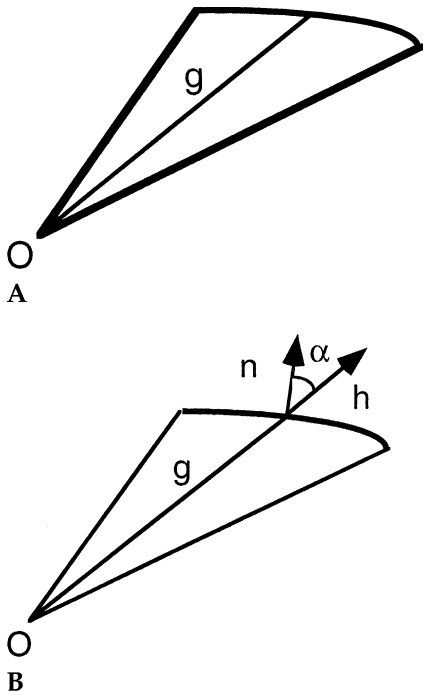

Figure 1. Parameters used to estimate molecular volumes and surface areas with the modified versions of the program SURMO2. (A) Estimate of an element of the molecular volume with SURMO2-I and -II. (B) Estimate of an element of the molecular surface area with SURMO2-II.

Improving previous algorithms for calculating contact and reentrant surfaces by Greer and Bush,[35] and for calculating the solvent-accessible surface area by Shrake and Rupley,[26] Connolly[27–29] has developed the MS (molecular surface) program. This program places dots over the solvent-accessible surface of a molecule in order to calculate the solvent-excluding surface of a molecule.

Connolly[27–29] pointed out that there are two kinds of reentrant surface: (1) the concave reentrant surface, which is generated when the probe sphere simultaneously touches three atoms, and (2) the saddle-shaped reentrant surface, which is generated as the probe sphere rolls along the crevice between two atoms. Connolly[27–29] has also found that the reentrant surface belonging to one probe may be contained in the interior volume of an overlapping probe, and so must be removed.

The analysis of a molecular surface by the MS program is carried out by rolling a probe sphere (e.g., water, radius = 1.4 Å) over the whole molecular surface. For each contact point with the van der Waals surface, there is a dot. The result is a smooth van der Waals surface that represents the accessible zones for the solvent molecule, including the internal cavities. However, those interstices that are too small to accept the probe are eliminated. Thus, the solvent-accessible surface has been calculated. These surfaces can be graphically visualized by means of a series of points (dot surface), provided the density of points is not too small.

Numerical integration of the molecular surface (program SCAP)

The molecule may be fragmented into atoms or groups. The starting hypothesis of Hopfinger[36] and Hopfinger and Battershell[37] is that one can center on each group of the molecule, independently, a solvation sphere.[38] The intersecting volume, V°, among the solvation sphere and the van der Waals spheres of the remaining atoms in the molecule is then calculated (see Figure 2). This volume allows the calculation of the effective volume of solvation of the group for a given conformation of the studied molecule. The model manages up to four parameters for a given solvent: (1) n: maximum number of solvent molecules allowed to fill the solvation sphere; (2) Δg: variation of Gibbs free energy associated with the extraction of one solvent molecule out of the solvation sphere; (3) R_ν: radius of the solvation sphere; and (4) V_f: free volume available for a solvent molecule in the solvation sphere. Another relevant parameter in the model is V_s, the volume of the solvent molecule.

In the present work, the values of n and R_ν have been taken from the minimization of the configuration energy in a force field[36] and the Δg values were taken from Gibson and Scheraga.[38]

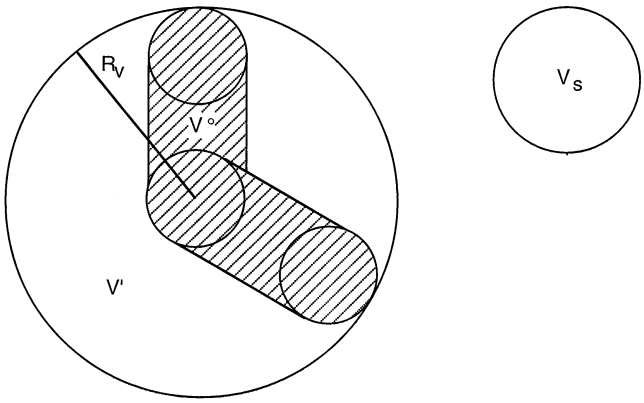

Figure 2. Left: The big circle represents a solvation sphere of radius R_v centered on an atom of the solute molecule. V° is the intersecting volume among the solvation sphere and the van der Waals spheres of the remaining atoms in the molecule. V′ is the volume available for the solvent molecules in the solvation sphere. On the right side, the small circle represents a solvent molecule of volume V_s.

(Other additive fragment contributions have been proposed in the literature.[39])

In the solvation sphere, part of the volume excludes the solvent molecules. This volume consists of the van der Waals volume of the group at which the sphere is centered and of a volume representing the groups bonded to the central group. The latter volume is represented by a number of cylinders. The difference between the total volume of the solvation sphere and the volume excluded to the solvent molecules represents the volume V' that is actually available for the n solvent molecules. Therefore, V_f can be calculated as

$$V_f = \frac{V'}{n} - V_s$$

The variation of Gibbs free energy associated with the extraction of all the solvent molecules out of the solvation sphere of group R is

$$\Delta G_R = n\Delta g\left(1 - \frac{V^\circ}{V'}\right)$$

and for the extraction of all the solvent molecules out of the solvent spheres of a molecule, the result is

$$\Delta G_{extr} = \sum_{R=1}^{N} \Delta G_R$$

Finally, the solvation free energy of a molecule is

$$\Delta G_{solv} = -\Delta G_{extr}$$

Although this is a simple method, Pascal[40] has found important difficulties in recalculating the volume V'. For example, the value of V_f for the (CH) aromatic group solvated in water is 3.3 Å^3 (fitted parameter of Hopfinger[36,37]) and 48.14 Å^3 (calculated using the geometric procedure indicated by Hopfinger). Hence, we have preferred to ensure the coherence of the calculations[41] by calculating the values of V_f by means of SCAP, a computer program developed by Pascal[40]; this program makes use of the subprogram KOROBO,[42] which allows for the calculation of the surfaces and volumes by the numerical integration method of Korobov.

Once the ΔG_{solv} values have been estimated, one can use them for calculating other properties of biological interest. As an example, the hydropathicity[43] (a term encompassing hydrophobicity to hydrophilicity) of a molecule can be quantified in terms of the logarithm of the macroscopic partition coefficient between a pair of immiscible solvents (generally 1-octanol [o] and water [w]), log P. The log P of organic solutes has been found to be useful in correlating and predicting the biological activity of the solute.[44] The lipophilic parameter of Hansch et al.,[45,46] π_X, can be defined as

$$\pi_x = \log \frac{P_X}{P_H}$$

where P_X is the partition coefficient of the molecule substituted by X and P_H is the partition coefficient of the reference molecule.

Hopfinger[36,37] has proposed a method (called π-SCAP) to calculate log P and π_X in terms of structural additive schemes. π-SCAP was initially used to calculate the Gibbs free energy of solvation of molecules (see above). From these data and with the equation

$$-RT \ln P = \Delta\Delta G^\circ_{solv}\ (1\text{-octanol} \leftarrow \text{water})$$

where R is the gas constant, T is the temperature (taken as 298 K), and $\Delta\Delta G^\circ_{solv}$(1-octanol ← water) is the transfer free energy of a given solute to 1-octanol from water, and

$$\Delta\Delta G^\circ_{solv}(1\text{-octanol} \leftarrow \text{water}) = \Delta G^\circ_{solv}(1\text{-octanol}) - \Delta G^\circ_{solv}(\text{water})$$

where ΔG°_{solv}(1-octanol) and ΔG°_{solv}(water) (in kJ · mol^{-1}) are the standard-state free energies of solvation of a given solute considered in 1-octanol and water (respectively), one can calculate

$$\log P = -0.17567\Delta\Delta G^\circ_{solv}(1\text{-octanol} \leftarrow \text{water})$$

The partitioning of a solute between an aqueous phase and an organic phase is critical for many phenomena in biological and medicinal chemistry.[47–50] One of many known examples[51] of a linear free energy relationship relating two partitioning-like processes is

$$\log \frac{1}{C} = 0.75 \log P + 2.30$$

where C is the molar concentration of organic compound necessary to produce a 1:1 complex with bovine serum albumin (BSA) via equilibrium dialysis. This partitioning process is related linearly to log P.

Starting from the SCAP program,[40] which allows for the calculation of the surfaces and volumes by the method of Korobov, we have implemented the following new features[41]: (1) any organic solvent (not only 1-octanol) can be modelled and any organic solvent/water log P can now be calculated; (2) other magnitudes can now be calculated as (a) the molar concentration C of organic compound necessary to produce a 1:1 complex with BSA via equilibrium dialysis; (b) the number

of hydrophilic groups, assuming one lipophilic group, and (c) the hydrophile–lipophile balance (HLB); (3) an atom-to-atom or group-to-group partition of log P and all of the other properties allows the analysis and display of local lipophilicity/hydrophilicity mapped on molecular surfaces. We have normalized the contribution of the atoms or groups, i, to log P, log $(1/C)$, etc., by a factor of V/V_i in order to guarantee that the mean of log P_i be equal to the molecular value.

Accurate triangular tessellation of the molecular surface (program GEPOL)

The GEPOL (geometry of polyhedron) program, developed in this laboratory by Silla et al.[52–54] computes the molecular surface as a point distribution. Each point represents a piece of the surface, called a *tessel*. Three kinds of envelope surfaces can be computed: (1) the van der Waals molecular surface, taken as the external surface resulting from a set of intersecting spheres centered on the atoms or groups forming the molecule; (2) the solvent-accessible surface, taken as the surface generated by the center of the solvent, considered as a rigid sphere (probe sphere), when it rolls around the van der Waals surface[25]; and (3) the solvent-excluding surface, composed of its two parts: the contact surface and the reentrant surface as described above.[11]

The program computes the chosen molecular surface as a point distribution and calculates the corresponding area and volume. The model starts by setting a sphere on each atom. If the solvent-excluding surface is desired, the program will fill the spaces inaccessible to the probe sphere by creating a new set of spheres among the original set. Once the set of spheres that forms the surface has been defined the program calculates the molecular area and volume.

The algorithm for the creation of new spheres in the spaces inaccessible to the solvent has been reported in Ref. 53. Once the set of spheres has been defined, the spherical surfaces are divided into a set of triangular tessels by using a data-coded generic pentakis dodecahedron.[54] A simple geometric procedure is used to eliminate all triangles found at the intersection volume of the spheres. The division can be increased by means of the NDIV tessellation parameter that specifies the division level for the triangles on the surface. If NDIV is 1, 2, 3, 4, or 5, the number of spherical triangles per sphere will be 60, 240, 960, 3 840, and 15 360, respectively.

The molecular surface area is easily obtained by summing up the areas of all triangles: $S = \sum_i S_i$, where S_i is the area of the ith triangle. To calculate the molecular volume, the model first defines a molecular origin. Let $\overline{r_i}$ be the position vector of the ith triangle center and $\overline{n_i}$ be the corresponding normal vector to the surface at this point. The volume is obtained by summing up all solid volumes made by the triangle vector surfaces and the origin of the molecular system: $V = (1/3) \sum_i S_i \overline{n_i} \cdot \overline{r_i}$. It is easy to verify that internal volumes are canceled out.

The data concerning the surface are provided by the coordinates at the center of each triangle, the elementary surface value, and the components of the normal vector to the tessellated surface in order to facilitate graphical visualization.[55]

In this work, we have implemented into the GEPOL program[52–54] the calculation of the molecular globularity and rugosity indices as well as the fractal dimension of the solvent-accessible surface (see below for details). The algorithm is fast and efficient.

Cubic lattice approach to the molecular surface (program CSAM)

The volume of a molecule can readily be computed by decomposing the space it occupies, using a cubic lattice.[56–59] The problem of computing surfaces can be reduced to the computation of volumes by converting the surface into a sheet of finite uniform thickness δh; by computing its volume the surface area A can be obtained as $A \approx V/\delta h$.[60] It can be shown that if a plane cuts a randomly oriented cube, then on average the area of the cut will be

$$\Delta A \approx \frac{2}{3} a^2$$

This means that an estimate of the surface area of an object can readily be obtained: only the cubes cut by the surface contribute to A and the increment is $(2/3)a^2$. Of course, this method is not exact. The cut between a plane and a cube is a polygon with from three to six sides. If, for example, one corner is located on one side of the plane and the remaining seven corners on the opposite side, then we have a triangular cut, and on average the areas of triangular cuts tend to be somewhat smaller than the average of $(2/3)a^2$ that is being used. This observation suggests a slightly more elaborate scheme, one in which the polygonal cuts are classified according to number of sides and using different increments ΔA_k for the five different types of cuts. The area is now

$$A = \sum_{k=1}^{5} N_k \Delta A_k$$

where N_k represents the number of cubes cut by the surface with the kth type of cut for which the increment ΔA_k is added to the surface area A. Senn[60] has reported the classification scheme for the preceding five cases, together with the computed values for ΔA_k, and has written a computer program for the computation of surface areas of molecules (CSAM) with spherical atoms. The limiting parameter a is the length of the edges of the cubes used in the cubical tessellation of the molecular space.

We have implemented in the CSAM program[60] the calculation of the molecular volume V. It can be evaluated, in first approximation, by adding the volumes of interior cubes and about half of the volumes of surface cubes, i.e., $V = (I + 0.5P)a^3$, where I is the number of cubes in the inside of the molecular surface and P is the number of cubes partially included in this surface. This approximation becomes asymptotically correct as the unit length a becomes smaller. Of course, this method is not exact. When the cube is a surface one, it is expected to contribute to the molecular volume by about half of the volume of the cube. However, exact mean contribution is slightly less than half of the volume of the cube because of the convex curvature of spheres. For a representative sphere of radius 3.0 Å, the proper mean contribution in this case is found to be 0.482 (instead of 0.5) of the volume of the cube. We have used this value taken from Ref. 59 as the proper contribution from a surface cube: $V = (I + 0.482P)a^3$.

Cubic lattice approach to the molecular space (program TOPO)

The TOPO program for the theoretical simulation of molecular shape has been described elsewhere.[61] The surface of molecules can be represented by the external surface of a set of overlapping spheres with appropriate radii, centered on the nuclei of the atoms.[62–64] If the radii are those of van der Waals,[65] the molecular bare (solvent-free) surface is obtained[62,63]; if the radii are those of van der Waals plus the effective radius of a solvent molecule, the solvent-accessible molecular surface is now obtained.[25]

The characterization of the molecular shape is helped by the calculation of geometric descriptors and topological indices of the molecules.[62,63] The molecule is treated as a solid in space, defined by tracing spheres about the atomic nuclei. It is computationally enclosed in a graduated rectangular box and the geometric descriptors evaluated by counting points within the solid or close to chosen surfaces. The molecular volume V is concurrently approximated as $V = P \cdot \mathrm{GRID}^3$, where P is the number of points within the molecular volume (within a distance R_X of any atomic nucleus X) and GRID is the size of the mesh grid.

As a first approximation, the molecular bare surface area could be calculated as $S \approx Q \cdot \mathrm{GRID}^2$, where Q is the number of points close to the bare surface area (within a distance between R_X and $R_X + \mathrm{GRID}$ of any atomic nucleus X). However, the estimate has been improved[63]: if the point falls exactly on the surface of one of the atomic spheres, it accounts indeed for GRID^2 units of area on the molecular bare surface. This is because the total surface of atom X can accommodate $4\pi R_X^2/\mathrm{GRID}^2$ points. When a point falls beyond the surface, it represents GRID^2 units of area on the surface of a sphere of radius $R > R_X$, not on the surface of atom X. On the surface of X it accounts only for a fraction of this quantity, namely, $\mathrm{GRID}^2(R_X/R)^2$. The total bare surface area is, therefore, calculated as $S = F \cdot \mathrm{GRID}^2$, where F is the sum of elements AF defined as $AF = R_X^2/R^2(I)$ for those points close enough to the surface of any atom X. R_X^2 is the squared radius of atom X and $R^2(I)$ is the squared distance of point I from the atomic nucleus X. Meyer[63] has written a program including the OEPP (one element per point) subprogram that allows the calculation of the molecular volume and bare surface area. We have written a version of this subprogram (ATOEPP) that allows the atom-to-atom partition analysis of these geometric descriptors into their atomic component parts.[61]

Two topological indices of molecular shape can now be calculated: G and G'.[63] The ratio $G = S_e/S$ has been interpreted as a descriptor of molecular globularity; S_e is the surface area of a sphere whose volume is equal to the molecular volume V. The ratio $G' = S/V$ has been interpreted as a descriptor of molecular rugosity.

The importance of the solvent validates the assumption that the properties of the systems solvated in water are strongly related to the contact surface between solute and water molecules.[25,31] We can propose a new molecular geometric descriptor, that is, the solvent-accessible surface AS.[25] This surface is denoted when a probe sphere is allowed to roll on the outside while maintaining contact with the molecular bare surface. The continuous sheet defined by the locus of the center of the probe is the accessible surface. Alternatively, the accessible surface is calculated as the molecular bare surface by using pseudoatoms, whose van der Waals radii were increased by the radius R of the probe. The accessibility is a dimensionless quantity varying between 0 and 1 and represents the ratio of the accessible surface area in a particular structure to the accessible surface area of the same atom when isolated from the molecule.

Fractal surfaces[66] provide a means for characterizing the irregularity of molecular surfaces. The area of the solvent-accessible surface, AS, depends on the value of the probe radius, R. The fractal dimension D of the molecules may be obtained according to Lewis and Rees[67] as

$$D = 2 - \frac{d(\log \mathrm{As})}{d(\log R)}$$

The fractal dimension D provides a quantitative indication of the degree of surface accessibility toward different solvents. The larger the D value the faster the accessible surface area drops with an increase in solvent molecular size.

A version of the TOPO program has been implemented into the AMYR program for the theoretical simulation of molecular associations and chemical reactions.[68–70] The algorithm allows for the characterization of molecular units and aggregates. All of the geometric descriptors and topological indices but fractal dimension can now be calculated. In this work, a version of the algorithm for the calculation of the molecular globularity and rugosity indices as well as the fractal dimension of the solvent-accessible surface has been implemented into the GEPOL program[52–54] for the calculation of the molecular volume and surface.

The Monte Carlo measure of molecular volume and surface (program MCVS)

The Monte Carlo (MC) measure of the volume of some region in d-dimensional space (i.e., the union or intersection of some spheres) may be performed as follows[71]:

1. Build a routine able to detect whether a point is inside the region.
2. Find a rectangular box including the region.
3. Generate N points uniformly distributed in the box, and count the number N_i of points falling inside the region.

Let $p = N_i/N$ and $q = 1 - p$. Thus, the MC estimate of the volume is $V = Wp$, W being the volume of the box, and the corresponding standard deviation is $\sigma = W(pq/N)^{1/2}$.

The MC measure of the surface area of the union of n spheres may be performed as follows[72]:

1. Let S_i be the surface of the ith sphere and T be the sum of all these surfaces.
2. Generate N points distributed uniformly on the total surface of all the spheres. This is performed with the following substeps:
   a. Select one of the n spheres, such that each ith sphere has a probability equal to S_i/T of being selected.
   b. Generate a d-dimensional point $\bar{x}$ distributed uniformly in the ith sphere and norm it to obtain $\langle \bar{x} - \overline{c_i} | \bar{x} - \overline{c_i} \rangle = R_i^2$.
   c. If $\bar{x}$ falls inside one of the $n - 1$ other spheres, $\bar{x}$ does not pertain to the surface of the union. If $\bar{x}$ falls outside, $\bar{x}$ pertains to the surface of this union.
3. Count the number N_u of points lying on the surface of the union.

Table 1. Foundations of different programs used to characterize molecular cavities

Descriptor	Based on:	Limiting Parameter	Usual	This work
SURMO2	Square/triangular tesselation	N_1 and N_2	12	72
MS	Spherical tesselation	Density of surface points[a]	10	100
SCAP	Numerical integration	*b*	*b*	*b*
GEPOL	Triangular tesselation	NDIV	5	5
CSAM	Cubic lattice	Mesh grid[c]	0.5	0.1
TOPO	Cubic lattice	Mesh grid[c]	0.5	0.1
MCVS	Pseudorandom Monte Carlo	Number of observations	10^4	10^6
UMCVS	Uniform Monte Carlo	Number of observations	10^4	10^6

[a] Density of surface points (points · Å^{-2}).
[b] No limiting parameter is given as an input for this model.
[c] Size of the mesh grid (Å).

Table 2. Geometric descriptors and topological indices for fullerene-60, calculated by different methods[a]

	Molecule + cavity			Cavity-free molecule						
Property	SURMO2-I[b]	SURMO2-II[c]	MS	SCAP	GEPOL	CSAM	TOPO	Combined[d]	MCVS	UMCVS
V[e]	548.4	548.4	*f*	497.9	526.8	525.7	521.7	521.7	522.9 (≥0.4)	521.8 (0.4)
S[g]	323.6	354.0	352.6	399.4	401.5	398.1	381.9	401.5	403.2 (≥0.6)	402.0 (0.6)
AS[h]	513.0	527.4	525.3	520.8	531.4	530.2	520.1	531.4	537.0 (≥1.2)	530.6 (1.2)
AS′[i]	946	954	951	942	954	953	945	954	979 (≥3)	953 (3)
G[j]	1.0012	0.9153	*f*	0.7606	0.7856	0.7913	0.8206	0.7806	0.7785 (≥0.0016)	0.7797 (0.0016)
G'[k]	0.5901	0.6455	*f*	0.8022	0.7621	0.7573	0.7320	0.7696	0.7711 (≥0.0017)	0.7704 (0.0017)
D[l]	1.407	1.425	1.425	1.426	1.431	1.430	1.420	1.431	1.417	1.431

[a] Statistical errors for some methods are reported in parentheses.
[b] Program SURMO2 modified for the calculation of molecular spherical surface areas.
[c] Program SURMO2 after correcting the deviation from the spherical shape.
[d] The combined method calculates the molecular volume with program TOPO and the molecular surface areas with program GEPOL.
[e] Molecular volume (Å^3).
[f] This method does not allow the calculation of this property.
[g] Molecular surface area (Å^2).
[h] Water-accessible surface area (Å^2).
[i] Side-chain accessible-surface area (Å^2).
[j] Molecular globularity.
[k] Molecular rugosity (Å^{-1}).
[l] Fractal dimension of the solvent-accessible surface.

Let $p = N_u/N$ and $q = 1 - p$. Thus, the MC estimate of the surface area is $S = Tp$ and the corresponding standard deviation is $\sigma = T(pq/N)^{1/2}$.

The uniform Monte Carlo method (program UMCVS)

To ensure an unbiased sampling of the whole space of parameters, the random number generator (RNG) of the MCVS numeric algorithm has been substituted in the uniform Monte Carlo (MC) program UMCVS by a uniform random number generator (URNG) that yields uniformly distributed real numbers over a given finite range [*a*, *b*].[73,74] We have used the LOIUAB[72] URNG. This subprogram returns a real random number following the uniform law $U(a, b)$. The LOIUAB algorithm has three steps:

1. Update S_i as: $S_i = (AS_{i-1} + B)\text{MOD } 2^M$. Here MOD has the usual meaning in programming languages.
2. Perform floating-point normalization of S_i between 0 and 1 (real division of S_i by 2^M).
3. Transform from [0, 1] to [*a*, *b*].

The values adopted in the database of the program are $A = 13^{13}$, $B = 0$, and $M = 59$. The initiation step is $S_0 = 123456789(1 + 2^{32})$, which shows a period of 2^{57}.

The LOIUAB method guarantees the portability of both source code and results. The consequence in terms of the representation of big integers and of their conversion into double precision is that they are not machine dependent.

CALCULATION RESULTS AND DISCUSSION

Geometric descriptors

A summary of the basis of various programs and algorithms used in the present work to characterize the molecular cavity of

fullerenes is presented in Table 1, along with usual values for the limiting parameters and the value used in this work. Note that some methods were designed for proteins and so the usual limiting parameters have been correspondingly increased whenever possible for the smaller molecules discussed in this work.

For SURMO2 the limiting parameters $N_1 = N_2 = 72$ cut the unitary sphere in $2N_1N_2 = 10\ 368$ surface elements. For the MS algorithm the density of surface points, which has been taken as 100 points · Å^{-2}, gives 19 460 total surface points for C_{60} and 22 929 total surface points for C_{70}. No limiting parameter is given as an input for the SCAP model. For the GEPOL program, variable NDIV controls the tessellation level and has been taken as its maximum value of 5; this gives 1 472 points for C_{60} and 1 670 points for C_{70}. The CSAM and TOPO methods have as limiting parameter the size of the mesh grid (edge length), which has been taken as 0.1 Å. The molecular volume and surface were measured accurately with the MCVS and UMCVS methods, using $N = 10^6$ observations for each measure. This represents 19 164 points · Å^{-3} for C_{60} and 16 499 points · Å^{-3} for C_{70}. The van der Waals radius for the carbon atom has been taken as 1.76 Å for all of the methods.[65]

The geometries of C_{60} and C_{70} optimized with the MM2-87 program[75,76] have been taken from the literature.[77] Each fullerene is brought into its principal inertial coordinate system. The length x of the molecule is defined as the maximum length, the height z as its minimum thickness, and its width y is measured at right angles to the axes indicated by the length and height. The origin is the center of mass for each fullerene. The C_{60} structure is given in Figure 3A. The molecular volumes of C_{60} are shown in Figure 4A. The total volume V_t is the sum of the molecular (V_m) and cavity (V_c) volumes: $V_t = V_m + V_c$. The molecular surfaces of C_{60} are shown in Figure 4B. The molecular surface S_m is the sum of the external (S_e) and cavity (S_c) surfaces: $S_m = S_e + S_c$.

The geometric descriptors and topological indices calculated for C_{60} are given in Table 2. The SURMO2 and MS methods are unable to recognize the fullerene cavity. Hence, the calculated volume V is a measure of the total volume V_t, and is 548.4 Å^3. The rest of the procedures recognize the cavity, and the value of the molecular volume V_m is available and lies in the range 498 − 527 Å^3. These results compare well with the accurate UMCVS reference calculation, which gives a result of 521.8 Å^3. The external surface area, $S_e = 323.6$ Å^2, is estimated by SURMO2-I, which was modified for the calculation of molecular spherical surface areas. This result is improved when the correction of the deviation from the spherical shape is taken into account with SURMO2-II ($S_e = 354.0$ Å^2), which is in good agreement with MS ($S_e = 352.6$ Å^2).

However, the actual (external plus internal) molecular surface area, S_m, lies in the range 382–403 Å^2 (programs from SCAP to UMCVS), which compares well with the UMCVS reference (402.0 Å^2). Not surprisingly, the globularity index G is rather greater as calculated by SURMO2 (close to unity, as expected for a sphere) compared with the rest of the methods (in the range 0.76–0.82), and the rugosity index G' is rather smaller. Note that the internal cavity effect is difficult to appreciate in the context of the water-accessible surface area AS and of the side-chain accessible-surface area AS′, because of the small cavity contribution ($AS_{cav} \approx 3$ Å^2 and $AS'_{cav} \approx 1$ Å^2). Similarly, the cavity effect in the fractal dimension of the solvent-accessible surface is calculated as negligible in C_{60} and C_{70}.

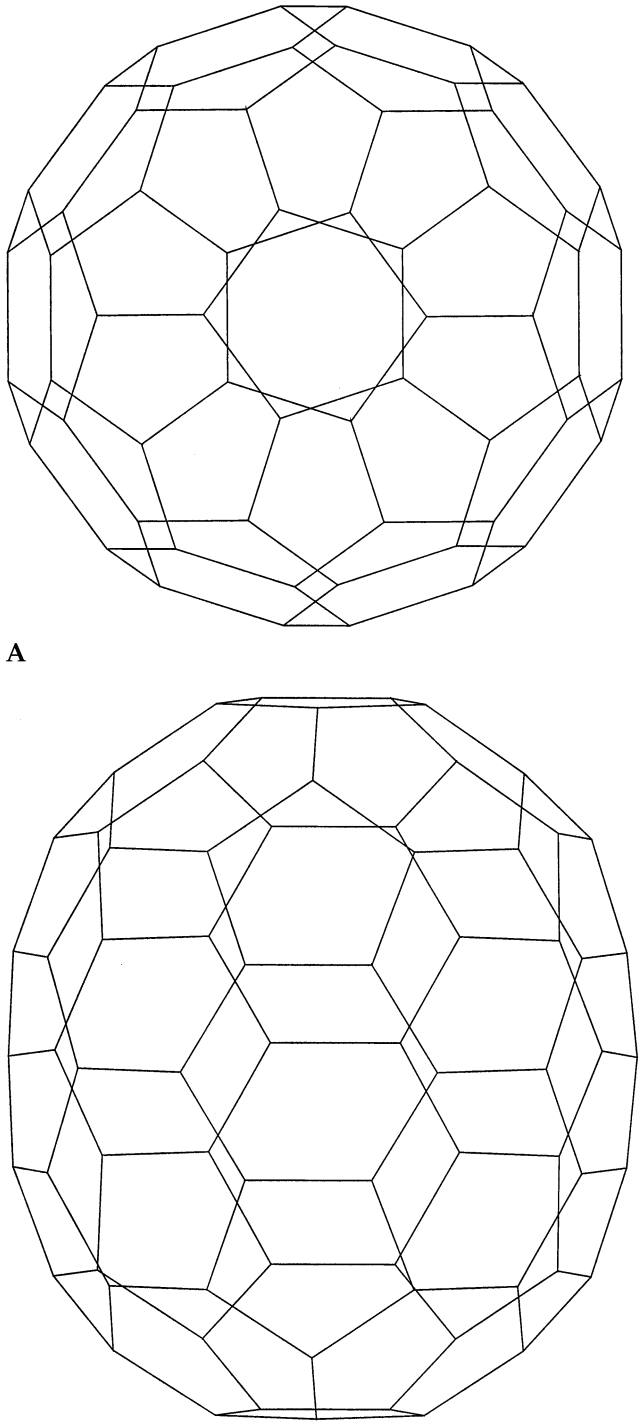

Figure 3. Molecular images of fullerenes. (A) (I_h)fullerene-60 (C_{60}, buckminsterfullerene); (B) (D_{5h})fullerene-70 (C_{70}).

The comparison between the GEPOL and TOPO programs is of special interest because the former is an efficient algorithm but does not perform an atom-to-atom partition analysis of the geometric descriptors and topological indices of molecules.[61] It is interesting, for the reasons discussed below, to

Table 3. Geometric descriptors and topological indices for fullerene-70, calculated by different methods[a]

	Molecule + cavity			Cavity-free molecule						
Property	SURMO2-I[b]	SURMO2-II[c]	MS	SCAP	GEPOL	CSAM	TOPO	Combined[d]	MCVS	UMCVS
V[e]	645.6	645.6	*f*	578.6	612.2	610.6	606.1	606.1	610.5 (≥0.4)	606.1 (0.4)
S[g]	360.3	396.4	408.2	458.7	459.8	456.0	435.8	459.8	456.4 (≥0.6)	460.2 (0.6)
AS[h]	559.8	577.4	582.4	579.1	585.6	585.0	572.1	585.6	587.8 (≥1.3)	585.8 (1.3)
AS′[i]	1 010	1 021	1 015	1 009	1 021	1 022	1 012	1 021	1 016 (≥3)	1 023 (3)
G[j]	1.0025	0.9113	*f*	0.7321	0.7583	0.7633	0.7947	0.7533	0.7625 (≥0.0013)	0.7526 (0.0013)
G[k]	0.5581	0.6140	*f*	0.7928	0.7511	0.7468	0.7190	0.7586	0.7476 (≥0.0015)	0.7593 (0.0015)
D[l]	1.428	1.448	1.452	1.455	1.460	1.457	1.446	1.460	1.468	1.458

[a] Statistical errors for some methods are reported in parentheses.
[b] Program SURMO2 modified for the calculation of molecular spherical surface areas.
[c] Program SURMO2 after correcting the deviation from the spherical shape.
[d] The combined method calculates the molecular volume with program TOPO and the molecular surface areas with program GEPOL.
[e] Molecular volume (Å^3).
[f] This method does not allow the calculation of this property.
[g] Molecular surface area (Å^2).
[h] Water-accessible surface area (Å^2).
[i] Side-chain accessible-surface area (Å^2).
[j] Molecular globularity.
[k] Molecular rugosity (Å^{-1}).
[l] Fractal dimension of the solvent-accessible surface.

Table 4. Mean relative deviations (percent) of the geometric descriptors and topological indices for fullerene-60 and -70, calculated with different cavity-free molecule methods relative to the UMCVS reference calculations

Property	SCAP	GEPOL	CSAM	TOPO	Average[a]	Combined[b]	MCVS
V[c]	−5	1	0.7	−0.01	2	−0.01	0.5
S[d]	−0.5	−0.1	−1	−5	2	−0.1	−0.3
AS[e]	−2	0.03	−0.09	−2	1	0.03	0.7
AS′[f]	−1	−0.05	−0.03	−0.9	0.5	−0.05	1
G[g]	−3	0.8	1	6	3	0.1	0.4
G[h]	4	−1	−2	−5	3	−0.1	−1
D[i]	−0.3	0.07	−0.07	−0.8	−0.2	0.07	−0.1

[a] Mean of the relative deviations (percent) in absolute value for the first four methods.
[b] The combined method calculates the molecular volume with program TOPO and the molecular surface areas with program GEPOL.
[c] Molecular volume (Å^3).
[d] Molecular surface area (Å^2).
[e] Water-accessible surface area (Å^2).
[f] Side-chain accessible-surface area (Å^2).
[g] Molecular globularity.
[h] Molecular rugosity (Å^{-1}).
[i] Fractal dimension of the solvent-accessible surface.

propose a combined method that calculates the molecular volume with program TOPO and the molecular surface areas with the program GEPOL. The resulting G and G' values show the best agreement with the UMCVS reference.

Both Monte Carlo (MC) methods provide statistical estimations of the standard deviation of the C_{60} molecular volume [$\sigma(V) = 0.18$ Å^3]. This quantity basically depends on the number of observations and is exact for UMCVS and approximate for MCVS. From the standard deviation σ we have estimated the statistical absolute error E as $E = t_{95\%,N}s$, where $t_{95\%,N}$ is an integral of the Student distribution, which is easy to calculate by looking up the corresponding value of t in Fisher tables for a 95% confidence band and N degrees of freedom (number of observations).[78] This statistical parameter is a function of N. It approaches an asymptotic limit of 1.96 when N approaches infinity. It is not very different from 1.96 for $N \geq 50$. Thus, with the number of observations given in Table 1 ($N = 10^6$) we estimate the statistical absolute error as $E(V) = 0.4$ Å^3 and the relative error as $E_r(V) = 0.08\%$. We have repeated the same procedure for the molecular surface area [$\sigma(S) = 0.28$ Å^2] and we have obtained a statistical absolute error of $E(S) = 0.6$ Å^2. The relative error of the molecular surface area (0.15%) is about double that of the molecular volume. This reflects the fact that the molecular surface area is a more difficult magnitude to calculate than the molecular volume.[62,63] It should be remarked that this statistical error is

Table 5. Geometric descriptors and topological indices for the cavities of fullerene-60 and -70

Cavity	V^a	S^b	AS^c	AS′^{d}	G^e	G'^f	D^g
Fullerene-60	26.6	49.4	3.2	1.4	0.87	1.86	2.75
Fullerene-70	39.5	52.0	8.4	2.5	1.08	1.32	2.89

[a] Molecular volume (Å^3).
[b] Molecular surface area (Å^2).
[c] Water-accessible surface area (Å^2).
[d] Side-chain accessible-surface area (Å^2).
[e] Molecular globularity.
[f] Molecular rugosity (Å^{-1}).
[g] Fractal dimension of the solvent-accessible surface.

Table 6. Solvation descriptors for fullerene-60 and -70

Fullerene	$\Delta G_{solv,w}$[a]	$\Delta G_{solv,o}$[b]	log P (SCAP)[c]	log P (CDHI)[d]	log 1/C (SCAP)[e]	log 1/C (CDHI)[d]	$V_{cav,w}$[f]	$S_{cav,w}$[g]	$V_{cav,o}$[h]	$S_{cav,o}$[i]
-60	−15.60	−128.72	19.9	13.8	17.2	12.7	1 322.5	595.3	2 526.5	906.2
-70	−18.09	−148.40	22.9	15.8	19.5	14.2	1 486.3	643.2	2 793.6	969.3

[a] Gibbs free energy of solvation in water (kJ · mol^{-1}).
[b] Gibbs free energy of solvation in 1-octanol (kJ · mol^{-1}).
[c] P is the 1-octanol/water partition coefficient.
[d] CDHI: calculations carried out with a method developed by Kantola et al.[79]
[e] C is the molar concentration necessary to produce a 1:1 complex with bovine serum albumin via equilibrium dialysis.
[f] Solvent cavity volume in water (Å^3).
[g] Solvent cavity surface area in water (Å^2).
[h] Solvent cavity volume in 1-octanol (Å^3).
[i] Solvent cavity surface area in 1-octanol (Å^2).

not the only source of error. When comparing the MCVS results with those of the UMCVS reference we can see a bias of 1.1 Å^3 for the molecular volume [rather greater than the statistical absolute error of $E(V) = 0.4$ Å^3] and 1.2 Å^2 for the molecular surface area [rather greater than $E(S) = 0.6$ Å^2]. This feature is a consequence of the nonuniform random number generator (RNG) used in MCVS, which results in an underestimation of standard deviations and errors.

The structure of C_{70} is shown in Figure 3B. The partitioning of molecular volumes and surfaces of this fullerene is shown in Figures 4C and 4D. The geometric descriptors and topological indices calculated for this molecule are shown in Table 3. These results show general trends that are similar to those found for C_{60}. The total volume calculated by SURMO2-II is greater than the actual molecular volume determined by the rest of the methods. The external surface area S, as determined by SURMO2-II and MS, is smaller than the actual molecular surface area. When comparing with C_{60}, the greatest spherical character of this molecule is accompanied with a greater molecular globularity G, as calculated by UMCVS. Surprisingly, the molecular rugosity, $G' = S/V$, is slightly greater for the spherical C_{60} as calculated by UMCVS. This is due to the relatively high mean atomic contribution to the surface area (6.70 Å^2 per carbon atom) when compared with C_{70} (6.57 Å^2 per atom). The fractal dimension is slightly greater for C_{70}, indicating that the solvent-accessible surface area is more irregular than for C_{60}.

In comparing the MCVS results for σ and error with the UMCVS reference, we obtain a bias of 4.4 Å^3 for the molecular volume and −3.8 Å^2 for the molecular surface area. This bias is one order of magnitude greater than the statistical absolute error and corroborates the bias observed for C_{60}.

The mean relative deviations (percent) of the geometric descriptors and topological indices, calculated for both fullerenes with different cavity-sensitive methods (from SCAP to TOPO) relative to the UMCVS reference, are reported in Table 4. Note that we have averaged only two sets of data, so that when a calculated relative deviation is rather low [e.g., $E_r(V) = -0.01\%$ for the TOPO program] it could be underestimated. It can be seen from the data that differences in molecular volume V, relative to those obtained with the UMCVS reference, never exceed 5% in spite of the sometimes relevant differences in method of calculation. The same observation can be made for the molecular surface area [$|E_r(S)| \leq 5\%$], water-accessible surface area [$|E_r(\mathrm{AS})| \leq 2\%$], side-chain accessible-surface area [$|E_r(\mathrm{AS}')| \leq 1\%$], molecular globularity [$E_r(G) \leq 6\%$], molecular rugosity [$|E_r(G')| \leq 5\%$], and fractal dimension [$|E_r(D)| \leq 0.8\%$]. This is encouraging, since it means that all of the algorithms, even starting from different points of view, are producing consistent results. It should be noted that the SCAP method gives, in general, the greater errors, especially for the molecular volume. The GEPOL algorithm shows, in general, the best results and the outcome is remarkably good for all of the surface areas. In comparing the CSAM and TOPO programs, the former, designed for dealing with surface area, reveals a lower error in calculating this property. Note also that the TOPO program, conceived for handling molecular volume, shows the least error of all methods for this property. It is thus

Table 7. Geometric and solvation descriptors and topological indices for fullerene-60 and -70: Atom–atom partition for each carbon atom

Fullerene	Atom type	V^a	S^b	G^c	G'^c	AS^e	Accessibility^f	D^g	log P (SCAP)^h	lop P (CDHI)^i	log 1/C (SCAP)^j	log 1/C (CDHI)^i
-60	*a*	8.7	6.4	3.21	0.73	8.6	7.5	1.41	19.9	13.8	17.2	12.7
-70	*a*	8.7	6.4	3.20	0.73	8.8	7.8	1.41	23.7	15.8	20.1	14.2
	b	8.7	6.3	3.24	0.73	8.9	7.8	1.42	23.5	15.9	19.9	14.2
	c	8.7	6.4	3.20	0.74	8.7	7.6	1.41	23.2	15.8	19.7	14.1
	d	8.6	6.1	3.33	0.71	7.7	6.7	1.48	22.6	15.9	19.2	14.2
	e	8.7	6.0	3.42	0.69	7.0	6.2	1.55	22.1	15.9	18.9	14.2

[a] Molecular volume (Å^3).
[b] Molecular surface area (Å^2).
[c] Molecular globularity.
[d] Molecular rugosity (Å^{-1}).
[e] Water-accessible surface area (Å^2).
[f] Accessibility (percent) of the accessible surface.
[g] Fractal dimension of the solvent-accessible surface.
[h] P is the normalized 1-octanol/water partition coefficient.
[i] CDHI: calculations carried out with a method developed by Kantola et al.[79]
[j] C is the normalized molar concentration necessary produce a 1:1 complex with bovine serum albumin via equilibrium dialysis.

appealing to combine both methods. The combined (S from GEPOL/V from TOPO) method shows relative errors of less than 0.1% for all of the descriptors in C_{60} and C_{70}.

From the calculation results in Tables 2 and 3, referring to the total (SURMO2 and MS methods) and cavity-sensitive (SCAP to UMCVS algorithms) molecular shape, we have calculated the geometric descriptors and topological indices for the cavities of both fullerenes. The data are given in Table 5 and show that for C_{60}, the molecular volume V, surface area S, water-accessible surface area AS, and side-chain accessible-surface area AS′ are smaller than for C_{70}. It should be noted that, for both molecules, $S > \text{AS} > \text{AS}'$. This is because a water molecule with an effective radius of 1.41 Å and a volume of about 18 Å^3 has little space to move inside the cavity while a probe sphere representing a protein side chain with a radius of 3.5 Å and a volume of about 180 Å^3 cannot be contained inside the C_{60} or C_{70} cavity. For C_{70}, the fractal dimension of the cavity solvent-accessible surface is slightly greater than for C_{60}, indicating that, for the former, this cavity surface is somewhat more irregular. Because the cavities of C_{60} and C_{70} are in fact closed holes, these values are of interest solely as numerical tests for the methods; they are of little relevance for ordinary chemistry.

Solvation Descriptors

The solvation descriptors calculated with the program SCAP for both fullerenes as described previously are listed in Table 6. The negative Gibbs free energy of solvation is slightly increased on going from C_{60} (-15.60 kJ · mol^{-1}) to C_{70} (-18.09 kJ · mol^{-1}). However, the negative Gibbs free energy of solvation in 1-octanol is rather increased from C_{60} (-128.72 kJ · mol^{-1}) to C_{70} (-148.40 kJ · mol^{-1}), so that the 1-octanol/water partition coefficient P is smaller for C_{60} (log P = 19.9) than for C_{70} (22.9). The transfer free energy to 1-octanol from water, $\Delta\Delta G^\circ_{\text{solv}}$(1-octanol ← water), can be easily obtained by subtracting $\Delta G^\circ_{\text{solv}}$(water) from $\Delta G^\circ_{\text{solv}}$(1-octanol).

The log P and log 1/C results can be compared with data obtained with a method developed by Kantola et al.[79] This method allows one to obtain conformationally dependent hydrophobic indices based on atomic contributions. Kantola et al. fitted the following expression for the 1-octanol/water partition coefficient:

$$\log P = \sum_i \alpha_i(N)S_i + \beta_i(N)S_i(\Delta q_i)^2 + \gamma_i(N)\Delta q_i$$

where S_i is the contribution of atom i to the molecular surface area; α, β, and γ are fitting parameters dependent only on the atomic number of atom i; and Δq represents the atomic net charges,[80] which we have computed with the POLAR program (described elsewhere).[61] The comparison between the SCAP program and the method of Kantola et al. has special interest. The latter assigns a set of parameters for each atom, depending only on its atomic number and not on the surrounding atoms in the molecule. Instead, the SCAP program also takes into account the functional group to which each atom belongs in the molecule. We have written a computer program, called CDHI, that uses the method of Kantola et al. and we have implemented in it an atom–atom partition analysis of log P and log 1/C near selected atoms. The contribution of each atom to the molecular surface area is calculated with the TOPO algorithm.[61] The CDHI program also calculates other magnitudes such as the molar concentration C of organic compound necessary to produce a 1:1 complex with bovine serum albumin (BSA) via equilibrium dialysis, and the hydrophile–lipophile balance (HLB).[81]

The calculated log P and log 1/C values do not agree exactly with the CDHI values. However, both the trend (C_{70} greater than C_{60}) and the basic information that one obtains from the results is the same. In fact, all of the C_{60} and C_{70} molecules will remain in the organic solvent, as one would expect, and they do not form complexes with BSA in appreciable quantities. On going from C_{60} to C_{70} the solvent cavity volume and surface area in water are increased 12.4% ($V_{\text{cav,w}}$) and 8.0% ($S_{\text{cav,w}}$). However, this augmentation is smaller for the solvent cavity volume and surface area in 1-octanol ($V_{\text{cav,o}}$, 10.6%; $S_{\text{cav,o}}$, 7.0%).

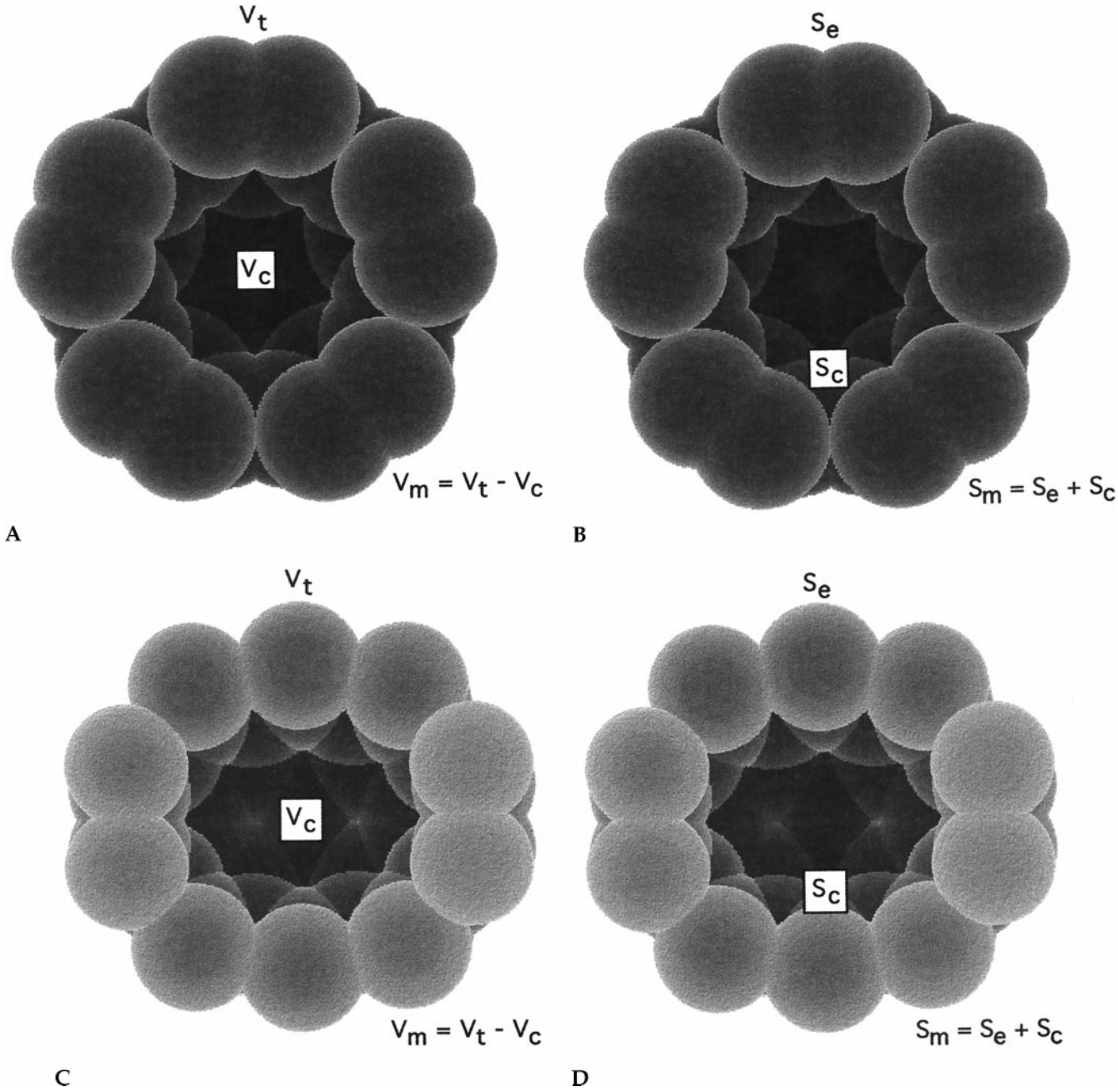

Figure 4. Molecular volumes and surfaces of fullerenes. (A) Molecular volume V_m, cavity volume V_c, and total volume V_t of (I_h)fullerene-60; (B) external surface S_e, cavity surface S_c, and molecular surface S_m of (I_h)fullerene-60; (C) molecular volume V_m, cavity volume V_c, and total volume V_t of (D_{5h})fullerene-70; and (D) external surface S_e, cavity surface S_c, and molecular surface S_m of (D_{5h})fullerene-70. Only the shells of atoms placed near to the equator are shown in order to make visible the cavity.

Atomic Analysis of Descriptors

An atom-to-atom partition analysis of the geometric and solvation descriptors and topological indices for the fullerene molecules is reported in Table 7. C_{60} has an icosahedral symmetry, with all 60 carbon atoms occupying equivalent sites.[82] However, for C_{70}, a projected substructure is shown in Figure 5, in which the five nonequivalent carbon atoms are labeled *a*–*e*, following previous notation.[83,84] Note that the *a*, *b*, *c*, *d*, and *e* carbon atoms are bound in this order (*a* next to *b* next to *c* . . .) in the C_{70} structure. It should be remarked that atoms labeled *a*–*d* join one pentagon with two hexagons while type *e* atoms are clearly different because they join three hexagons.

The contribution to the molecular volume *V* is similar for all carbon atoms (in the range 8.6–8.7 Å^3) in both fullerenes. However, the atomic contributions to the molecular surface area *S*, indices *G* and *G′*, water-accessible surface area AS, the accessibility of this surface *D*, and normalized log *P* and log 1/*C* appear clearly grouped in two sets. The first group is made

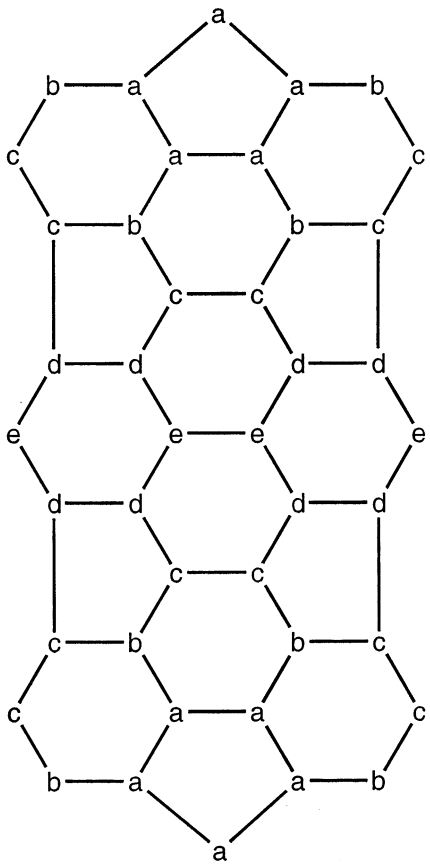

Figure 5. Flat projection of approximately three-tenths of the structure of (D_{5h})fullerene-70, showing the five nonequivalent carbon atom types. The fivefold symmetry axis of rotation is through the center of the top and bottom pentagons.

up of the *a*–*c* atoms of C_{70}, which present similar results. On the other hand, the calculated properties for atoms with type *e* in C_{70} are clearly different, corresponding to the fact that these atoms join three hexagons. The C_{70} *d* atoms present an intermediate behavior between both extremes, as expected from their intermediate position between *e* and *c*–*b*–*a* atoms.

The contribution to the fractal dimension is slightly greater for the C_{70} *e* atoms, indicating that, for this molecule, the irregularity of the solvent-accessible surface is maximal near the type *e* atoms. Note that the normalized contributions to log *P* and log 1/*C* as calculated with program SCAP can be used to distinguish between both groups of atoms, while those of program CDHI cannot.

CONCLUSIONS

In this article we have selected C_{60} and C_{70} as model cavities in order to test eight different methods for characterizing inclusion molecules. From the results of our work, the following conclusions can be drawn. The reader should keep in mind, however, that these conclusions are established from results obtained with the limiting parameter values shown in Table 1.

1. The use of the SURMO2 and MS methods, which do not recognize the cavities, may not be convenient for intercalation compounds, since these methods give an estimate of the global fraction of occupied space within the volume, not allowing distinction among different niches.

2. The programs that do recognize the cavities (SCAP, GEPOL, CSAM, and TOPO) never exceed 5% relative deviation in molecular volume and surface area, in spite of the sometimes relevant differences in method of calculation. This means that the choice should rely mainly on other possibilities offered by the different methods, such as computational performance, possibility of fragment analysis, etc. The GEPOL algorithm shows, in general, the best results and the outcome is remarkably good for all of the surface areas. The combined (GEPOL/TOPO) method shows relative errors below 0.1% for all of the descriptors in C_{60} and C_{70}.

3. The molecular volume and surface area were measured accurately with the MCVS and UMCVS methods. UMCVS measures the molecular volume and surface areas with high precision, so that the standard deviation is divided by 10 each time the number of points (and the CPU time) is multiplied by 100. This may be compared with the CSAM and TOPO methods, using a three-dimensional grid, for which a multiplication of the CPU time by about 1 000 is required to divide the error by 10.

4. The URNG (uniformly distributed over a given finite range [*a*, *b*]) in UMCVS provides the fastest convergence for the algorithm and a better estimate of the standard deviations. The use of a pseudorandom generator in MCVS produces a bias in the calculated properties and underestimates their standard deviations and errors, so that it cannot be recommended for high-precision predictions or as a benchmark maker.

5. In C_{60} and C_{70} the effect of including the internal cavity surface in the calculation of the solvent-accessible surfaces is small but, evidently, if the cavity were much bigger (e.g., fullerene-240, -540, and -960) this effect would be much greater, because the solvent molecule would then have room inside; at present, this effect can be corrected with an additional calculation carried out by use of a method that does not recognize the cavity.

REFERENCES

1 Gavezzotti, A. The calculation of molecular volumes and the use of volume analysis in the investigation of structured media and of solid-state organic reactivity. *J. Am. Chem. Soc.* 1983, **105,** 5220–5225
2 Lipkowski, J. In: *Inclusion Compounds* (J.L. Atwood, J.E.D. Davies, and D.D. MacNichol, eds.). Academic Press, London, 1984, Vol. 1
3 Rashin, A.A., Iofin, M., and Honig, B. Internal cavities and buried waters in globular proteins. *Biochemistry* 1986, **25,** 3619–3625
4 Schoenborn, B.P. *J. Mol. Biol.* 1969, **45,** 297
5 Richards, F.M. *J. Mol. Biol.* 1974, **82,** 1
6 Tilton, R.F., Jr., Kuntz, I.D., Jr., and Petsko, G.A. Cavities in proteins: Structure of a metmyoglobin–xenon complex solved to 1.9 angstroms. *Biochemistry* 1984, **23,** 2849–2857
7 Lumry, R. and Rosenberg, A. *Coll. Int. C.N.R.S.* 1975, **246,** 53
8 Richards, F.M. Packing defects, cavities, volume fluctuations and access to the interior of proteins, including some general comments on surface area and protein structure. *Carlsberg Res. Commun.* 1979, **44,** 47–64

9 Smith, J.L., Hendrickson, W.A., Honzatko, R.B., and Sheriff, S. Structural heterogeneity in protein crystals. *Biochemistry* 1986, **25,** 5018–5027
10 Dill, K.A. Dominant forces in protein folding. *Biochemistry* 1990, **29,** 7133–7155
11 Richards, F.M. Areas, volumes, packing, and protein structure. *Annu. Rev. Biophys. Bioeng.* 1977, **6,** 151–176
12 Hubbard, S.J., Gross, K.H., and Argos, P. Intramolecular cavities in globular proteins. *Protein Eng.* 1994, **7,** 613–626
13 Pesek, J.J. and Schneider, J.F. The detection of mercury, lead, and methylmercury binding sites on lysozyme by carbon-13 NMR chemical shifts of the carboxylate groups. *J. Inorg. Biochem.* 1988, **32,** 233–238
14 Eisenman, G. and Oberhauser, A. *Biophys. J.* 1988, **53,** A631
15 Hannon, R.A. and Ford, G.C. Analysis of lattice contacts in protein crystals. *Biochem. Soc. Trans.* 1988, **16,** 961–962
16 Olah, G.A., Huang, H.W., Liu, W., and Wu, Y. Location of ion-binding sites in the gramicidin channel by X-ray diffraction. *J. Mol. Biol.* 1991, **218,** 847–858
17 Leach, A.R. and Kuntz, I.D. Conformational analysis of flexible ligands in macromolecular receptor sites. *J. Comput. Chem.* 1992, **13,** 730–748
18 Meng, E.C., Shoichet, B.K., and Kuntz, I.D. Automated docking with grid-based energy evaluation. *J. Comput. Chem.* 1992, **13,** 505–524
19 Shoichet, B.K., Bodian, D.L., and Kuntz, I.D. Molecular docking using shape descriptors. *J. Comput. Chem.* 1992, **13,** 380–397
20 Lewis, R.A., Roe, D.C., Huang, C., Ferrin, T.E., Langridge, R., and Kuntz, I. D. Automated site-directed drug design using molecular lattices. *J. Mol. Graphics* 1992, **10,** 66–78
21 Kuntz, I.D. Structure-based strategies for drug design and discovery. *Science* 1992, **257,** 1078–1082
22 Gibson, K.D. and Scheraga, H.A. Exact calculation of the volume and surface area of fused hard-sphere molecules with unequal radii. *Mol. Phys.* 1987, **62,** 1247–1265
23 Arteca, G.A. and Mezey, P.G. Shape characterization of some molecular model surfaces. *J. Comput. Chem.* 1988, **9,** 554–563
24 Arteca, G.A. and Mezey, P.G. Shape group theory of van der Waals surfaces. *J. Math. Chem.* 1989, **3,** 43–71
25 Lee, B. and Richards, F.M. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* 1971, **55,** 379–400
26 Shrake, A. and Rupley, J.A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* 1973, **79,** 351–371
27 Connolly, M.L. Solvent-accessible surfaces of proteins and nucleic acids. *Science* 1983, **221,** 709–713
28 Connolly, M.L. Analytical molecular surface calculation. *J. Appl. Crystallogr.* 1983, **16,** 548–558
29 Connolly, M.L. *QCPE Bull.* 1981, **1,** 75
30 Richmond, T.J. Solvent accessible surface area and excluded volume in proteins: Analytical equations for overlapping spheres and implications for hydrophobic effect. *J. Mol. Biol.* 1984, **178,** 63–90
31 Wodak, S.J. and Janin, J. Analytical approximation to the accessible surface area of proteins. *Proc. Natl. Acad. Sci. U.S.A.* 1980, **77,** 1736–1740
32 Finney, J.L. Volume occupation, environment and accessibility in proteins. The problem of the protein surface. *J. Mol. Biol.* 1975, **96,** 721–732
33 Alard, P. and Wodak, S.J. Detection of cavities in a set of interpenetrating spheres. *J. Comput. Chem.* 1991, **12,** 918–922
34 Terryn, B. and Barriol, J. On the evaluation of the usual quantities or coefficients related to the shape of a molecule approximated on the basis of the van der Waals radii. *J. Chim. Phys. Phys.-Chim. Biol.* 1981, **78,** 207–212
35 Greer, J. and Bush, B.L. Macromolecular shape and surface maps by solvent exclusion. *Proc. Natl. Acad. Sci. U.S.A.* 1978, **75,** 303–307
36 Hopfinger, A.J. Polymer–solvent interactions for homopolypeptides in aqueous solution. *Macromolecules* 1971, **4,** 731–737
37 Hopfinger, A.J. and Battershell, R. D. Application of SCAP to drug design. 1. Prediction of octanol–water partition coefficients using solvent-dependent conformational analyses. *J. Med. Chem.* 1976, **19,** 569–573
38 Gibson, K.D. and Scheraga, H.A. Minimization of polypeptide energy. I. Preliminary structures of bovine pancreatic ribonuclease S-peptide. *Proc. Natl. Acad. Sci. U.S.A.* 1967, **58,** 420–427
39 Rekker, R.F. *The Hydrophobic Fragmental Constant.* Elsevier, Amsterdam, 1976
40 Pascal, P. *Program SCAP*. Université Henry-Poincaré-Nancy I, Nancy, France, 1991
41 Torrens, F., Sánchez-Marín, J., and Nebot-Gil, I. SCAP: A universal program for the calculation of solubility in organic solvents in water and partition coefficient. In: *Solvent Selection for Pesticide Residue Analysis* (J. Fournier, ed.). FECS, Angers, France (in press)
42 Huron, M.J. and Claverie, P. Calculation of the interaction energy of one molecule with its whole surrounding. I. Method and application to pure nonpolar compounds. *J. Phys. Chem.* 1972, **76,** 2123–2133
43 Kyte, J. and Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 1982, **157,** 105–132
44 Klopman, G. and Iroff, L.D. Calculation of partition coefficients by the charge density method. *J. Comput. Chem.* 1981, **2,** 157–160
45 Leo, A., Jow, P.Y.C., Silipo, C., and Hansch, C. *J. Med. Chem.* 1975, **18,** 865–868
46 Hansch, C. and Leo, A.J. *Substituent Constants for Correlation Analysis in Chemistry and Biology*. John Wiley & Sons, New York, 1979
47 Franke, R., Dove, S., and Kuhne, B. *Eur. J. Med. Chem.* 1977, **14,** 363–374
48 Giesen, D.J., Gu, M.Z., Cramer, C.J., and Truhlar, D.G. A universal organic solvation model. *J. Org. Chem.* 1996, **61,** 8720–8721
49 Giesen, D.J., Chambers, C.C., Cramer, C.J., and Truhlar, D.G. Solvation model for chloroform based on class IV atomic charges. *J. Phys. Chem. B* 1997, **101,** 2061–2069
50 Leo, A., Hansch, C., and Elkins, D. Partition coefficients and their uses. *Chem. Rev.* 1971, **71,** 525–616
51 Helmer, F., Kiehs, K., and Hansch, C. The linear free-energy relation between partition coefficients and the binding, and conformational perturbation of macromolecules by small organic compounds. *Biochemistry* 1968, **7,** 2858–2863

52 Pascual-Ahuir, J.L., Silla, E., Tomasi, J., and Bonaccorsi, R. Electrostatic interaction of a solute with a continuum. Improved description of the cavity and of the surface cavity bound charge distribution. *J. Comput. Chem.* 1987, **8,** 778–787
53 Pascual-Ahuir, J.L. and Silla, E. GEPOL: An improved description of molecular surfaces. I. Building the spherical surface set. *J. Comput. Chem.* 1990, **11,** 1047–1060
54 Silla, E., Tuñón, I., and Pascual-Ahuir, J.L. GEPOL: An improved description of molecular surfaces. II. Computing the molecular area and volume. *J. Comput. Chem.* 1991, **12,** 1077–1088
55 Silla, E., Villar, F., Nilsson, O., Pascual-Ahuir, J.L., and Tapia, O. Molecular volumes and surfaces of biomacromolecules via GEPOL: A fast and efficient algorithm. *J. Mol. Graphics* 1990, **8,** 168–172
56 Muller, J.J. Calculation of scattering curves for macromolecules in solution and comparison with results of methods using effective atomic scattering factors. *J. Appl. Crystallogr.* 1983, **16,** 74–82
57 Pavlov, M.Yu., and Fedorov, B.A. Improved technique for calculating X-ray scattering intensity of biopolymers in solution: evaluation of the form, volume, and surface of a particle. *Biopolymers* 1983, **22,** 1507–1522
58 Connolly, M.L. Computation of molecular volume. *J. Am. Chem. Soc.* 1985, **107,** 1118–1124
59 Higo, J. and Gõ, N. Algorithm for rapid calculation of excluded volume of large molecules. *J. Comput. Chem.* 1989, **10,** 376–379
60 Senn, P. Numerical computation of surface areas of molecules. *J. Math. Chem.* 1991, **6,** 351–358
61 Torrens, F., Ortí, E., and Sánchez-Marín, J. Vectorized TOPO program for the theoretical simulation of molecular shape. *J. Chim. Phys. Phys.-Chim. Biol.* 1991, **88,** 2435–2441
62 Meyer, A. Y. Molecular mechanics and molecular shape. 1. van der Waals descriptors of simple molecules. *J. Chem. Soc., Perkin Trans. 2* 1985, 1161–1169
63 Meyer, A.Y. Molecular mechanics and molecular shape. V. On the computation of the bare surface area of molecules. *J. Comput. Chem.* 1988, **9,** 18–24
64 Hermann, R.B. Theory of hydrophobic bonding. II. The correlation of hydrocarbon solubility in water with solvent cavity surface area. *J. Phys. Chem.* 1972, **76,** 2754–2759
65 Bondi, A. van der Waals volumes and radii. *J. Phys. Chem.* 1964, **68,** 441–451
66 Mandelbrot, B.B. *The Fractal Geometry of Nature.* Freeman, San Francisco, 1983
67 Lewis, M. and Rees, D.C. Fractal surfaces of proteins. *Science* 1985, **230,** 1163–1165
68 Fraga, S. A semiempirical formulation for the study of molecular interactions. *J. Comput. Chem.* 1982, **3,** 329–334
69 Fraga, S. Molecular associations. *Comput. Phys. Commun.* 1983, **29,** 351–359
70 Torrens, F., Ortí, E., and Sánchez-Marín, J. Pair-potential calculation of molecular associations: A vectorized version. *Comput. Phys. Commun.* 1991, **66,** 341–362
71 Hammersley, J.M. and Handscomb, D.C. In: *Monte Carlo Methods*, Chapman & Hall, London, 1983, Chap. 5
72 Petitjean, M. On the analytical calculation of van der Waals surfaces and volumes: Some numerical aspects. *J. Comput. Chem.* 1994, **15,** 507–523
73 Luscher, M. A portable high-quality random number generator for lattice field theory simulations. *Comput. Phys. Commun.* 1994, **79,** 100–110
74 James, F. RANLUX: A Fortran impolementation of high-quality pseudorandom number generator of Luscher. *Comput. Phys. Commun.* 1994, **79,** 111–114
75 Allinger, N.L. Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V_1 and V_2 torsional terms. *J. Am. Chem. Soc.* 1977, **99,** 8127–8134
76 Allinger, N.L. and Yuh, Y.H. MM2, QCPE Program No. 395
77 Froimowitz, M. Molecular geometries and heats of formation of C_{60} and C_{70} as computed by MM2-87. *J. Comput. Chem.* 1991, **12,** 1129–1133
78 Deming, W.E. *Statistical Adjustment of Data.* Dover, New York, 1964
79 Kantola, A., Villar, H.O., and Loew, G.H. Atom based parametrization for a conformationally dependent hydrophobic index. *J. Comput. Chem.* 1991, **12,** 681–689
80 Gasteiger, J. and Marsili, M. Iterative partial equalization of orbital electronegativity: A rapid access to atomic charges. *Tetrahedron* 1980, **36,** 3219–3228
81 The CDHI program is available from the authors on request.
82 Kroto, H.W., Heath, J.R., O'Brien, S.C., Curl, R.F., and Smalley, R.E. C_{60}: Buckminsterfullerene. *Nature (London)* 1985, **318,** 162–163
83 Heath, J.R., O'Brien, S.C., Zhang, Q., Liu, Y., Curl, R.F., Kroto, H.W., Tittel, F.K., and Smalley, R.E. Lanthanum complexes of spheroidal carbon shells. *J. Am. Chem. Soc.* 1985, **107,** 7779–7780
84 Taylor, R., Hare, J.P., Abdul-Sada, A.K., and Kroto, H.W. Isolation, separation and characterisation of the fullerenes C_{60} and C_{70}: The third form of carbon. *J. Chem. Soc., Chem. Commun.* 1990, 1423–1425