[A0045]

Characterizing Cavities in Model Inclusion Fullerenes: A Comparative Study

Francisco Torrens

Departament de Quimica Fisica, Facultat de Quimica, Universitat de Valencia, Dr. Moliner 50, E-46100 Burjassot (Valencia), Spain
Tel.: +34 96 398 3182, Fax: +34 96 398 3156, E-mail: Francisco.Torrens@uv.es

Abstract: The fullerene-82 cavity is selected as a model system 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 molecular surface, 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 the molecular volume. Accurate measures of the molecular volume and surface area have been performed with the pseudorandom Monte Carlo (MCVS) and uniform Monte Carlo (UMCVS) methods. These calculations serve as a reference for the rest of the methods. The SURMO2 method does not recognize the cavity and may not be convenient for intercalation compounds. The programs that detect the cavities never exceed 1-% deviation relative to the reference value for molecular volume and 5% for 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 solvent-accessible surfaces has been calculated. Fullerene-82 is compared with fullerene-60 and -70.

Keywords: molecular cavity, geometric descriptor, topological index, fractal dimension, partition coefficient.

Introduction

Analyses of molecular packing in liquids and crystals [1] 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 [2-5]. Although the existence of these internal cavities has been largely associated with conformational flexibility [6-8], 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 which require mobility in their initial stages, when the crystal structure is not yet distorted by product formation. Internal cavities affect protein stability [9], and play an important role in the close packing of amino acid side chains, particularly in the core of the protein [10,11]. The knowledge of the channels and cavity structures within crystals of globular proteins may be useful [12-20] 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, spheres whose size is determined by the atomic Van der Waals radii represent atoms.

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 [21] and on the topological description of their surfaces [22,23].

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* [24,25] and *solvent-excluding surface* [10] have been used. The corresponding surface areas can be computed from the atomic coordinates by approximate procedures [24,25] or by using analytic algorithms such as those of Gibson and Scheraga [21], Connolly [26,27] or Richmond [28]. The analytic algorithms 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 [29].

Methods to compute atomic and molecular volumes of proteins from the atomic coordinates have also been derived, originally by Richards [4] and Finney [30], and later implemented in automatic algorithms. On the basis of calculations of surface, protein cavities have been detected and analyzed by computer. However, all 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 [31] have described a procedure that is able to detect cavities in a system of interpenetrating spheres, no matter what their size or shape.

In a previous article the (I_h)fullerene-60 (C_{60}, buckminsterfullerene) and (D_{5h})fullerene-70 (C_{70}) molecules were selected as model cavities in order to test various methods for characterizing inclusion molecules [32]. In the present article (C_2)fullerene-82 (C_{82}) has been selected as a model cavity for characterizing inclusion molecules. These inclusion molecules are more suitable than proteins for testing precise algorithms, because of their moderate sizes. The following programs and the methods implemented in them have been used for characterizing the molecular cavities of fullerenes: (1) SURMO2 (surface of molecule), (2) the MS (molecular surface) program by Connolly [26,27], (3) SCAP (solvent-dependent conformational analysis program), (4) the GEPOL (geometry of polyhedron) program, (5) a program for the computation of surface areas of molecules (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 (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 well as 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)

SURMO2-I represents the global shape of a molecule as an envelope of the *van der Waals spheres* of the external atoms [33]. This allows a numerical treatment of integrals defined on the molecular surface, as

$$ I = \int g(\rho, \theta, \phi)\, d\omega \quad (1) $$

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 Equation (1)-type integrals is obtained by calculating the finite sum:

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

To achieve the summation one can take a mesh of points $M_{ij}(\theta, \phi)$ by cutting the surfaces on a single unitary sphere

with a uniform distribution of $2N_1$ parallels and N_2 meridians. If the molecule is cut by an axis passing by the origin, two intersecting points, which correspond to two values of the g function, are obtained: $g+$ and $g-$. Denoting by M_0 the upper pole, M_{1j} the points placed on the first parallel, M_{ij} those of the i-th parallel and M_{N_1j} those placed on the equator, the integral in Equation (1) can be measured as

$$I = \frac{N_2}{3}\left(g_{M_0}^+ + g_{M_0}^-\right) + \frac{7}{6}\sum_{j=0}^{N_2-1}\left(g_{M_{1j}}^+ + g_{M_{1j}}^-\right) + \frac{1}{2}\sum_{j=0}^{N_2-1}\left(g_{M_{N_1j}}^+ + g_{M_{N_1j}}^-\right) + \sum_{i=2}^{N_1-1}\sum_{j=0}^{N_2-1}\left(g_{M_{ij}}^+ + g_{M_{ij}}^-\right) \tag{2}$$

where $\sum_j g_{M_{ij}}$ indicates a sum over all of the mesh points placed on the i-th parallel.

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

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

where r is the radial spherical coordinate, $r = (1/3,\ 7/6,\ 1,\ 1,\dots,1/2)$ and Equation (2) has been used.

The calculation of the *bare* molecular surface S and *accessible surface* areas have been implemented in SURMO-I. In a first approximation, S can be calculated as

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

The deviation from the spherical shape was corrected in SURMO-II by dividing each g^2 value by the cosine of the angle formed by the semiaxis and the normal vector to the surface at this point [32]:

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

Spherical tessellation of the molecular surface (program MS)

The *solvent-accessible surface* area was originally defined and computed by Lee and Richards [24] as the area traced out by the centre 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 [10]. 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. Notice 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. Improving previous algorithms for calculating contact and reentrant surfaces by Greer and Bush [34], and for calculating the solvent-accessible surface area by Shrake and Rupley [25], Connolly [26,27] has developed the MS 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 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 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 = 1.4A) 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, which 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)*.

Numerical integration of the molecular surface (program SCAP)

The hypothesis of Hopfinger is that one can centre a *solvation sphere* on each group of the molecule [35,36]. The intersecting volume $V°$, between the solvation sphere and the van der Waals spheres of the remaining atoms, is then calculated. This volume allows the calculation of the effective volume of solvation of the group for a given conformation. The model manages up to four parameters for a given solvent: (1) n: maximum number of solvent molecules allowed for filling the solvation sphere; (2) $Dg°$: variation of Gibbs free energy associated with the extraction of one solvent molecule out of the solvation sphere [37,38]; (3) R_v: radius of the solvation sphere; and (4) V_f: free volume available for a solvent molecule in the solvation sphere. 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 centred 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 a group, R, is $$\Delta G_R^o = n\, \Delta g^o \left(1 - \frac{V^o}{V'}\right)$$ and for the extraction of all the solvent molecules out of the solvent spheres of a molecule, the result is $$\Delta G_{ou}^o = \sum_{R=1}^{N} \Delta G_R^o$$ Finally, the solvation free energy of a molecule is minus ΔG_{extr}^o.

However, Pascal [39] found important difficulties in recalculating the volume V'. Hence, the values of V_f [40-45] have been recalculated by means of a computer program [39,46]. Once the ΔG_{solv}^o values have been estimated, one can use them for calculating other properties. The partition coefficient P of a solute between a pair of immiscible solvents [1-octanol (o) and water (w)] is critical for many phenomena [47-54]. With the equation

$$RT \ln P = \Delta G_{solv}^o\ (\text{water}) - \Delta G_{solv}^o\ (1 - \text{octanol})$$

one can calculate the decimal logarithm $\log P$ at a given T, which is taken as 298K.

Accurate triangular tessellation of the molecular surface (program GEPOL)

GEPOL computes the molecular surface as a point distribution [55]. 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 centred on the atoms or groups forming the molecule; (2) the *solvent-accessible surface*, taken as the surface generated by the centre of the solvent, considered as a rigid sphere *(probe sphere)*, when it rolls around the van der Waals surface [24]; and (3) the *solvent-excluding surface* composed of its two parts: the *contact surface* and the *reentrant surface* [10].

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 form the surface has been defined the program calculates the molecular area and volume.

The algorithm creates new spheres in the spaces inaccessible to the solvent. Once the set of spheres has been defined, their spherical surfaces are divided into a set of triangular tessels by using a data-coded generic pentakis dodecahedron. 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, which 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 *i-th* triangle. To calculate the molecular volume, the model first defines a molecular origin. Let $\vec{r}_i$ be the *position vector* of the *i-th* triangle centre and $\vec{n}_i$ be the corresponding normal vector to the surface at this point. The volume is obtained by summing up all solid volumes made by the triangles vector surfaces and the origin of the molecular system: $V = (1/3) \sum_i S_i \vec{n}_i \cdot \vec{r}_i$. It is easy to verify that internal volumes are cancelled out. The coordinates at the centre of each triangle, the elementary surface value and the components of the normal vector to the tessellated surface provide the data concerning the surface in order to facilitate graphical visualization.

The calculation of the molecular *globularity* and *rugosity* indices as well as the *fractal dimension* of the solvent-accessible surface has been implemented in GEPOL [32]. 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 dh; by computing its volume the surface area A can be obtained as $A \approx V/dh$ [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 one has a triangular cut, and on average the areas of triangular cuts tend to be somewhat smaller than the average of $(2/3)a^2$ which is being used. This observation suggests a slightly more elaborate scheme, one in which the polygonal cuts are classified according to their number of sides and using different increments DA_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 *k-th* type of cut for which the increment DA_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 DA_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.

The calculation of the molecular volume V has been implemented in CSAM. This V can be evaluated, in first approximation, by adding the volumes of interior cubes and 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 3.0A, the proper mean contribution in this case is found to be 0.482 (instead of 0.5) of the volume of the cube. This value has been used as the proper contribution from a surface cube: $V = (I + 0.482P)a^3$ [60].

Cubic lattice approach to the molecular space (program TOPO)

TOPO [61] represents the surface of molecules by the external surface of a set of overlapping spheres with appropriate

radii [62], centred on the nuclei of the atoms [63-65]. 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 $V = P \cdot GRID^3$, where P is the number of points within the molecular volume and GRID is the size of the mesh grid.

As a first approximation, the molecular *bare* surface area could be calculated as $S \approx Q \cdot GRID^2$, where Q is the number of points close to the bare surface area (within a distance between R_X and $R_X + GRID$ of any atomic nucleus *X)*. However, the estimate has been improved: 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 $4 \cdot \pi R_X^2 / GRID^2$ 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, $GRID^2 (R_X / R)^2$. The total bare surface area is, therefore, calculated as $S = F \cdot GRID^2$, 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.

Two topological indices of molecular shape can be also calculated: G and G'. Consider S_e as the surface area of a sphere whose volume is equal to the molecular volume V. The ratio $G = S_e / S$ is interpreted as a descriptor of molecular *globularity*. The ratio $G' = S / V$ is interpreted as a descriptor of molecular *rugosity*.

Another molecular geometric descriptor was proposed by Lee and Richards [24]: the *solvent-accessible surface*, or, simply, accessible surface, AS. This AS is defined by means of a *probe sphere*, which is allowed to roll on the outside while maintaining contact with the molecular bare surface. The AS can be calculated in the same way as the molecular bare surface by mean of pseudoatoms, whose van der Waals radii have been increased by the radius R of the probe. Fractal surfaces [32,66] provide a means for characterizing the irregularity of molecular surfaces. The area of the 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

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

A version of the TOPO program has been implemented in the AMYR program for the theoretical simulation of molecular associations and chemical reactions [68-71] and in the GEPOL program [55] for the calculation of the molecular volume and surface.

The Monte Carlo measure of the 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 [72]:

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 [73]:

1. Let S be the surface of the *i-th* sphere and T be the sum of all these surfaces.

i

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 *i-th* sphere has a probability equal to S_i/T of being selected.

b. Generate a *d*-dimensional point $\bar{x}$ distributed uniformly in the *i-th* sphere and norm it to obtain $\langle \bar{x} - \bar{c}_i | \bar{x} - \bar{c}_i \rangle = R_i^2$.

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.

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 $[a,b]$ [74,75]. We have used the LOIUAB URNG [74]. 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 $S_i = (A \, S_{i-1} + B) \, MOD \, 2^M$. 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. Transformation 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 $S_0 = 123456789 \left(1 + 2^{32}\right)$, 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 [32]. Notice 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 $M_1 = M_2 = 72$ cut the unitary sphere in $2 M_1 M_2 = 10368$ surface elements. For the MS algorithm the density of surface points, which has been taken as 100 points·A^{-2}, gives 26 251 total surface points for C_{82}. No limiting parameter is given as 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 2 074 points for C_{82}. The CSAM and TOPO methods have as limiting parameter the size of the mesh grid, which has been taken as 0.1A. The molecular volume and surface were measured accurately with the MCVS and UMCVS methods, using $N = 10^7$ observations for

each measure. This represents 14 730 points·A^{-3} for C_{82}. The van der Waals radius for the carbon atom has been taken as 1.76A for all the methods [32].

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

Descriptor	Based on:	Limiting Parameter	Usual	This work
SURMO2	Square/triangular tessellation	N_1 and N_2	12	72
MS	Spherical tessellation	Density of surface points[a]	10	100
SCAP	Numerical integration	b	b	b
GEPOL	Triangular tessellation	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	10^7
UMCVS	Uniform Monte Carlo	Number of observations	10^4	10^7

a Density of surface points (points·A^{-2}).

b No limiting parameter is given as an input for this model.

c Size of the mesh grid (A).

The geometry of C_{82} has been optimized with the MM2-87 program [76]. The molecule 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 centre of mass for the molecule. The C_{82} structure is given in Figure 1. The molecular volumes of C_{82} are shown in Figure 2a. 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_{82} are shown in Figure 2b. The molecular surface S_m, is the sum of the external (S_e) and cavity (S_c) surfaces: $S_m = S_e + S_c$.
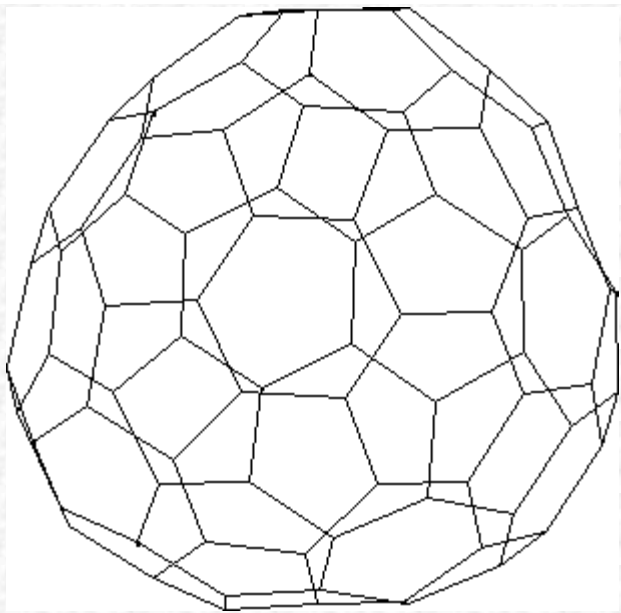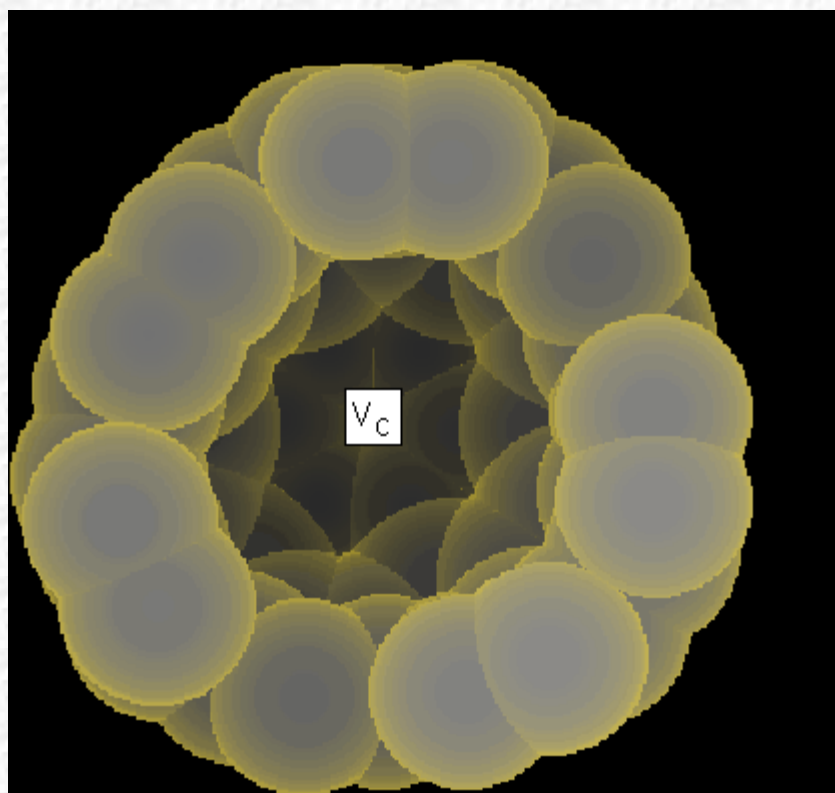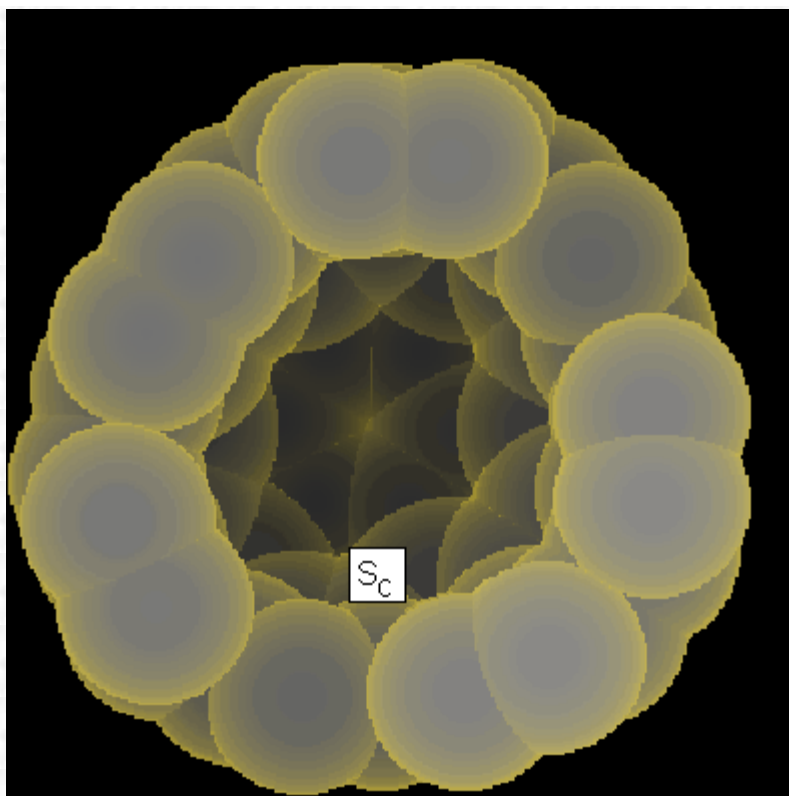
Figure 1. Molecular image of (C_2)fullerene-82 (C_{82}).

V_C

(a)

(b)

Figure 2. Molecular volumes and surfaces of (C_2)fullerene-82: a) Molecular volume V_m, cavity volume V_c, and total volume V_t, b) external surface S_e, cavity surface S_c, and molecular surface S_m.

The geometric descriptors and topological indices calculated for C_{82} are given in Table 2. The SURMO2 method is unable to recognize the fullerene cavity. Hence, the calculated volume V is a measure of the total volume V_t, and is $730.6A^3$. The rest of procedures recognize the cavity, and the value of the molecular volume V_m is available and lies in the range $679-686A^3$. These results compare well with the accurate UMCVS reference calculation, which gives a result of $(678.9\pm0.5)A^3$. The external surface area, $S_e = 391.3\ A^2$, is estimated by SURMO2-I, which has been modified for the calculation of molecular spherical surface areas [32]. This result is improved when the correction of the deviation from the spherical shape is taken into account with SURMO2-II ($S_e = 429.4 A^2$).

Table 2. Geometric descriptors and topological indices for fullerene-82 calculated with different methods.[a]

| Property | Molecule+cavity | | | Cavity free molecule | | | | |
	SURMO-I[b]	SURMO-II[c]	SCAP	MS	GEPOL	CSAM	TOPO	Combined[d]
V^e	730.6	730.6	678.5	f	686.4	683.8	678.8	678.8
S^g	391.3	429.4	503.6	504.9	505.9	500.3	477.6	505.9
AS^h	598.0	617.0	633.8	627.8	628.8	627.6	616.5	628.8
AS'^i	1061	1073	1078	1072.7	1072	1072	1063	1072
G^j	1.0025	0.9135	0.7415	f	0.7438	0.7502	0.7820	0.7383

G'k	0.5356	0.5878	0.7423	f	0.7371	0.7317	0.7036	0.7453
D^l	1.44	1.46	1.49	1.48	1.48	1.48	1.47	1.48

	Cavity free molecule	
Property	MCVS	UMCVS
V^e	681.4 (±0.5)	678.9 (0.5)
S^g	508.4 (±0.7)	505.5 (0.7)
AS^h	632.5 (±1.5)	629.7 (1.5)
AS'^{i}	1071 (±3)	1074 (3)
G^j	0.7365 (±0.0014)	0.7390 (0.0014)
G'^{k}	0.7461 (±0.0016)	0.7445 (0.0016)
D^l	1.49 (±0.04)	1.48 (0.04)

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 (A^3).

f This method does not allow the calculation of this property.

g Molecular surface area (A^2).

h Water-accessible surface area (A^2).

i Side-chain accessible-surface area (A^2).

j Molecular globularity.

k Molecular rugosity (A^{-1}).

l Fractal dimension of the solvent-accessible surface.

However, the actual (external plus internal) molecular surface area, S_m, lies in the range 478-508A^2 (programs from SCAP to UMCVS), which compares well with the UMCVS reference (505.5±0.7A^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 methods, and the rugosity index G' is rather smaller. Notice 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 ($AS_{cav} \approx 13$ A^2 and $AS'_{cav} \approx 1$ A^2). Similarly, the cavity effect in the fractal dimension of the solvent-accessible surface is calculated as negligible in C_{82}.

The comparison between the GEPOL and TOPO programs has a 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 to propose a combined method that calculates the molecular volume with program TOPO and the molecular surface areas with program GEPOL [32]. In particular, 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_{82} molecular volume

3

[$\sigma(V) = 0.26$ A]. This quantity basically depends on the number of observations and is exact for UMCVS and approximate for MCVS. From the standard deviation s we have estimated the statistical absolute error E as $E = t_{95\%,N}\sigma$, 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) [77]. This statistical parameter is a function of N. It approaches an asymptotic limit of 1.96 when N approximates infinity. It is not very different from 1.96 for $N \geq 50$. Thus, with the number of observations given in Table 1 the statistical absolute error has been estimated as $E(V) = 0.5$ A^3 and the relative error as $E_r(V) = 0.07\%$. The same procedure has been repeated for the molecular surface area [$\sigma(S) = 0.36$ A^2] and a statistical absolute error of $E(S) = 0.7$ A^2 has been obtained. The relative error of the molecular surface area (0.14%) is doubled when compared with the molecular volume. This reflects the fact that molecular surface area is a more difficult magnitude to calculate than molecular volume [63,64]. It should be remarked that this statistical error is not the only source of error in MCVS. When comparing the MCVS results with the UMCVS references a bias of 2.5A^3 can be seen for the molecular volume [5 times greater than E(V)] and 2.9A^2 for the molecular surface area [4 times greater than E(S)]. 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. This bias corroborates those observed for C$_{60}$ and C$_{70}$ [32].

The mean relative deviations (percent) of the geometric descriptors and topological indices calculated for C$_{60}$, C$_{70}$ and C$_{82}$ with different cavity-sensitive methods (from SCAP to TOPO) relative to the UMCVS reference are reported in Table 3. Notice that only three sets of data have been averaged, so that when a calculated relative deviation is rather low [e.g., $E_r(V) = -0.01\%$ for TOPO] 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 1% 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 [$|E_r(AS)| \leq 2\%$], side-chain accessible-surface area [$|E_r(AS')| \leq 1\%$], 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 the algorithms, even starting from different points of view, are producing consistent results. It should be noted that TOPO gives the best molecular volume but the greatest errors for the surface-related properties. However, GEPOL shows, in general, the best results and the outcome is remarkably good for all the surface-related properties. In comparing CSAM and TOPO, the former, designed for dealing with the surface area, reveals a lower error for this property. However, TOPO, conceived for handling molecular volume, shows the least error of all methods for this property. It is thus appealing to combine GEPOL and TOPO. The combined (S from GEPOL/V from TOPO) method shows relative errors of less than 0.1% for all the descriptors in the three fullerenes and the smallest errors for indices G and G'.

Table 3. Mean relative deviations (percent) of the geometric descriptors and topological indices for fullerene-60, -70 and -82, calculated with different cavity-free molecule methods relative to the UMCVS reference calculations.

Property	SCAP	MS	GEPOL	CSAM	TOPO	average[a]	Combined[b]	MCVS
V[c]	-0.06	-	1	0.7	-0.01	0.4	-0.01	0.4
S[d]	-0.2	0.007	-0.04	-1	-5	1	-0.04	0.01
AS[e]	-0.8	-0.6	-0.008	-0.2	-2	0.7	-0.008	0.7
AS'[f]	-0.7	-0.4	-0.1	-0.1	-1	0.5	-0.1	0.6
G[g]	0.2	-	0.7	1	6	2	0.04	0.3
G'[h]	-0.1	-	-1	-2	-5	2	-0.03	-0.4
D[i]	-0.1	-0.3	0.07	-0.09	-0.8	0.3	0.07	0.06

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 (A^3).

d Molecular surface area (A^2).

e Water-accessible surface area (A^2).

f Side-chain accessible-surface area (A^2).

g Molecular globularity.

h Molecular rugosity (A^{-1}).

i Fractal dimension of the solvent-accessible surface.

From the calculation results in Table 2 and Reference 33, referring to the total (SURMO2) and cavity-sensitive (SCAP to UMCVS) molecular shapes, the geometric descriptors and topological indices for the cavities of the three fullerenes have been calculated. The data are given in Table 4 and show that for C_{60}, the molecular volume V, surface area S and water-accessible surface area AS are smaller than for C_{70} and for C_{82} while the side-chain accessible-surface area AS' remains almost constant (1A) for all the fullerenes. It should be noted that, for the three fullerenes, $S > AS > AS'$. This is because a water molecule with an effective radius of 1.41 A and a volume of about 18 A^3 has little space to move inside the cavity while a *probe sphere* representing a protein side chain with a radius of 3.5 A and a volume of about 180 A^3 cannot be contained inside the C_{60}, C_{70} or C_{82} cavity. In particular, for C_{82}, the fractal dimension of the cavity solvent-accessible surface is strongly greater than for C_{70} and for C_{60}, indicating that, for the former, this cavity surface is rather more irregular.

Table 4. Geometric descriptors and topological indices for the cavities of fullerene-60, -70 and –82.

Cavity	V^a	S^b	AS^c	AS'^d	G^e	G'^f	D^g
Fullerene-60	26.6	48.0	3.2	0	0.8978	1.8045	2.75
Fullerene-70	39.5	63.8	8.4	2	0.8791	1.6152	2.89
Fullerene-82	51.7	76.1	12.7	1	0.8819	1.4720	5.21

a Molecular volume (A^3).

b Molecular surface area (A^2).

c Water accessible surface area (A^2).

d Side-chain accessible surface area (A^2).

e Molecular globularity.

f Molecular rugosity (A^{-1}).

g Fractal dimension of the solvent-accessible surface.

Solvation descriptors

The solvation descriptors calculated with program SCAP for the three fullerenes are listed in Table 5. The negative Gibbs free energy of solvation is slightly increased on going from C_{60} (15.60kJ·mol^{-1}) to C_{82} (20.86kJ·mol^{-1}). However, the negative Gibbs free energy of solvation in 1-octanol is rather increased from C_{60} (128.7kJ·mol^{-1}) to C_{82} (172.4kJ·mol^{-1}), so that the 1-octanol/water partition coefficient P_o is smaller for C_{60} ($\log P_o = 19.9$) than for C_{70} and

for C_{82} (26.6).

Table 5. Solvation descriptors for fullerene-60, -70 and -82.

Fullerene	$DG_{s,w}$[a]	$DG_{s,o}$[b]	$DG_{s,ch}$[c]	$DG_{s,cf}$[d]	log P_o (GSCAP)[e]	log P_o (CDHI)[f]	log P_{ch} (GSCAP)[g]	log P_{ch}[h]	log P_{cf} (GSCAP)[i]	log P_{cf}[h]
fullerene-60	-15.60	-128.7	-78.48	-103.2	19.9	13.8	11.1	11.6	15.4	21.0
fullerene-70	-18.09	-148.4	-91.20	-119.9	22.9	15.8	12.8	13.6	17.9	24.4
fullerene-82	-20.86	-172.4	-106.2	-139.2	26.6	17.6	15.0	16.1	20.8	28.6

a Gibbs free energy of solvation in water (kJ·mol^{-1}).

b Gibbs free energy of solvation in 1-octanol (kJ·mol^{-1}).

c Gibbs free energy of solvation in cyclohexane (kJ·mol^{-1}).

d Gibbs free energy of solvation in chloroform (kJ·mol^{-1}).

e P_O is the 1-octanol/water partition coefficient.

f CDHI: calculations carried out with a method developed by Kantola *et al*.

g P_{ch} is the cyclohexane/water partition coefficient.

h Calculations carried out with a method developed by Leo *et al*.

i P_{cf} is the chloroform/water partition coefficient.

The $\log P_o$ results can be compared with data obtained with a method developed by Kantola *et al*. [78]. 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; a, b and g are fitting parameters dependent only on the atomic number of atom *i;* and Dq represents the atomic net charges [79], which we have computed with the POLAR algorithm [80]. The comparison between SCAP 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, SCAP also takes into account the functional group to which each atom belongs in the molecule. A computer program, called CDHI, which uses the method of Kantola *et al*., has been written [81]. The contribution of each atom to the molecular surface area is calculated with the TOPO algorithm [61]. The calculated $\log P_o$ values do not agree exactly with the CDHI values. However, both the trend ($C_{82} > C_{70} > C_{60}$) and the basic information that one obtains from the results is the same. In fact, all the three fullerenes will remain in the organic solvent, as one would expect.

SCAP has allowed calculating other organic solvent/water partition coefficients [40-45] for cyclohehexane (ch) and chloroform (cf). $\log P_{ch}$ and $\log P_{cf}$ show the same trend and $\log P_o$. In particular, the results, especially for cyclohexane/water, compare better than $\log P_o$ with calculations carried out with a method proposed by Leo *et al*. [82].

Conclusions

In this article we have selected C_{82} as a model cavity in order to test eight different methods for characterizing inclusion molecules. From these results and previous work on C_{60} and C_{70}, the following conclusions can be drawn.

1. The use of SURMO2, which do not recognize the cavities, may not be convenient for intercalation compounds, since this method gives an estimate of the global fraction of occupied space within the volume, not allowing distinction among different niches [32,83].

2. The programs that do recognize the cavities (SCAP, MS, GEPOL, CSAM and TOPO) never exceed 1% relative deviation in molecular volume and 5% in 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*. GEPOL shows, in general, the best results and the outcome is remarkably good for the entire surface areas and surface-related properties. The combined (GEPOL/TOPO) method shows relative errors below 0.1% for all of the descriptors in C_{60}, C_{70} and C_{82}.

3. The molecular volume and surface area were measured accurately with MCVS and UMCVS. The 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 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}, C_{70} and C_{82} 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 and Notes

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. Rashin, A. A.; Iofin, M.; Honig, B. Internal cavities and buried waters in globular proteins. *Biochemistry* **1986**, *25*, 3619-3625.

3. Schoenborn, B. P. *J. Mol. Biol*. **1969**, *45*, 297.

4. Richards, F. M. *J. Mol. Biol*. **1974**, *82*, 1-14.

5. Tilton, R. F. Jr.; Kuntz, I. D. Jr.; Petsko, G. A. Cavities in proteins: Structure of a metmyoglobin-xenon complex solved to 1.9 angstrom. *Biochemistry* **1984**, *23*, 2849-2857.

6. Lumry, R.; Rosenberg, A. *Coll. Int. C.N.R.S*. **1975**, *246*, 55-63.

7. 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*, 47-63.

8. Smith, J. L.; Hendrickson, W. A.; Honzatko, R. B.; Sheriff, S. Structural heterogeneity in protein crystals.

Biochemistry **1986**, *25*, 5018-5027.

9. Dill, K. A. Dominant forces in protein folding. *Biochemistry* **1990**, *29*, 7133-7155.

10. Richards, F. M. Areas, volumes, packing, and protein structure. *Annu. Rev. Biophys. Bioeng*. **1977**, *6*, 151-176.

11. Hubbard, S. J.; Gross, K. H.; Argos, P. Intramolecular cavities in globular proteins. *Protein Eng*. **1994**, *7*, 613-626.

12. Pesek, J. J.; 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.

13. Eisenman, G.; Oberhauser, A. *Biophys. J*. **1988**, *53*, A631.

14. Hannon, R. A.; Ford, G. C. Analysis of lattice contacts in protein crystals. *Biochem. Soc. Trans*. **1988**, *16*, 961-962.

15. Olah, G. A.; Huang, H. W.; Liu, W.; Wu, Y. Location of ion-binding sites in the gramicidin channel by X-ray diffraction. *J. Mol. Biol*. **1991**, *218*, 847-858.

16 Leach, A. R.; Kuntz, I. D. Conformational analysis of flexible ligands in macromolecular receptor sites. *J. Comput. Chem*. **1992**, *13*, 730-748.

17. Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated docking with grid-based energy evaluation. *J. Comput. Chem*. **1992**, *13*, 505-524.

18. Shoichet, B. K.; Bodian, D. L.; Kuntz, I. D. Molecular docking using shape descriptors. *J. Comput. Chem*. **1992**, *13*, 380-397.

19. Lewis, R. A.; Roe, D. C.; Huang, C.; Ferrin, T. E.; Langridge, R.; Kuntz, I. D. Automated site-directed drug design using molecular lattices. *J. Mol. Graphics* **1992**, *10*, 66-78.

20. Kuntz, I. D. Structure-based strategies for drug design and discovery. *Science* **1992**, *257*, 1078-1082.

21. Gibson, K. D.; 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.

22. Arteca, G. A.; Mezey, P. G. Shape characterization of some molecular model surfaces. *J. Comput. Chem*. **1988**, *9*, 554-563.

23. Arteca, G. A.; Mezey, P. G. Shape group theory of van der Waals surfaces. *J. Math. Chem*. **1989**, *3*, 43-71.

24. Lee, B.; Richards, F. M. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol*. **1971**, *55*, 379-400.

25. Shrake, A.; Rupley, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol*. **1973**, *79*, 351-371.

26. Connolly, M. L. Solvent-accessible surfaces of proteins and nucleic acids. *Science* **1983**, *221*, 709-713.

27. Connolly, M. L. Analytical molecular surface calculation. *J. Appl. Crystallogr*. **1983**, *16*, 548-558.

28. 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.

29. Wodak, S. J.; Janin, J. Analytical approximation to the accessible surface area of proteins. *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 1736-1740.

30. Finney, J. L. Volume occupation, environment and accessibility in proteins. The problem of the protein surface. *J. Mol. Biol.* **1975**, *96*, 721-732.

31. Alard, P.; Wodak, S. J. Detection of cavities in a set of interpenetrating spheres. *J. Comput. Chem.* **1991**, *12*, 918-922.

32. Torrens, F.; Sanchez-Marin, J.; Nebot-Gil, I. Characterizing cavities in model inclusion molecules: A comparative study. *J. Mol. Graphics Mod.* **1998**, *16*, 57-71.

33. Terryn, B.; 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.

34. Greer, J.; Bush, B. L. Macromolecular shape and surface maps by solvent exclusion. *Proc. Natl. Acad. Sci. USA* **1978**, *75*, 303-307.

35. Hopfinger, A. J. Polymer-solvent interactions for homopolypeptides in aqueous solution. *Macromolecules* **1971**, *4*, 731-737.

36. Hopfinger, A. J.; 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.

37. Gibson, K. D.; Scheraga, H. A. Minimization of polypeptide energy, I. Preliminary structures of bovine pancreatic ribonuclease S-peptide. *Proc. Natl. Acad. Sci. USA* **1967**, *58*, 420-427.

38. Rekker, R. F. *The Hydrophobic Fragmental Constant*, Elsevier: Amsterdam, 1976.

39. Pascal, P. Program SCAP, Universite Henry Poincare-Nancy I, 1991.

40. Torrens, F.; Sanchez-Marin, J.; Nebot-Gil, I. A universal model for the calculation of all organic solvent/water partition coefficients. *J. Chromatogr. A* **1998**, *827*, 345-358.

41. Torrens, F. Universal organic solvent-water partition coefficient model. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 236-240.

42. Torrens, F. Calculation of partition coefficient and hydrophobic moment of the secondary structure of lysozyme. *J. Chromatogr. A* in press.

43. Torrens, F. A new tool for rational chiral-drug design. *J. Pharm. Biomed. Anal.* submitted for publication.

44. Torrens, F.; Soria, V. Calculated stationary-mobile phase distribution coefficient for polystyrene standards. *J. Chromatogr. A* submitted for publication.

45. Torrens, F. Application of SCAP to rational chiral-drug design. *Chirality* submitted for publication.

46. Huron, M. J.; 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.

47. Kyte, J.; Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105-132.

48. Franke, R.; Dove, S.; Kuhne, B. *Eur. J. Med. Chem.* **1977**, *14*, 363-374.

49. Giesen, D. J.; Gu, M. Z.; Cramer, C. J.; Truhlar, D. G. A universal organic solvation model. *J. Org. Chem*. **1996**, *61*, 8720-8721.

50 Giesen, D. J.; Chambers, C. C.; Cramer, C. J.; Truhlar, D. G. Solvation model for chloroform based on class IV atomic charges. *J. Phys. Chem. B* **1997**, *101*, 2061-2069.

51. Leo, A.; Hansch, C.; Elkins, D. Partition coefficients and their uses. *Chem. Rev.* **1971**, *71*, 525-616.

52. Klopman, G.; Iroff, L. D. Calculation of partition coefficients by the charge density method. *J. Comput. Chem*., **1981**, *2*, 157-160.

53. Leo, A.; Jow, P. Y. C.; Silipo, C.; Hansch, C. *J. Med. Chem*. **1975**, *18*, 865-868.

54. Hansch, C.; Leo, A. J. *Substituent Constants for Correlation Analysis in Chemistry and Biology*, John Wiley and Sons: New York, 1979.

55. Pascual-Ahuir, J. L.; Silla, E.; Tomasi, J.; 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.

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.; 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.; Go, 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.; Orti, E.; Sanchez-Marin, J. Vectorized TOPO program for the theoretical simulation of molecular shape. *J. Chim. Phys. Phys.-Chim. Biol*. **1991**, *88*, 2435-2441.

62. Bondi, A. Van der Waals volumes and radii. *J. Phys. Chem.* **1964**, *68*, 441-451.

63. Meyer, A. Y. Molecular mechanics and molecular shape. Part 1. Van der Waals descriptors of simple molecules. *J. Chem. Soc., Perkin Trans. 2* **1985**, 1161-1169.

64. 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.

65. 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.

66. Mandelbrot, B. B. *The Fractal Geometry of Nature*. Freeman, San Francisco, 1983.

67. Lewis, M.; 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.; Orti, E.; Sanchez-Marin, J. Pair-potential calculation of molecular associations: A vectorized version. *Comput. Phys. Commun.* **1991**, *66*, 341-362.

71. Torrens, F.; Sanchez-Marin, J.; Nebot-Gil, I. AMYR 2: A new version of a computer program for pair potential calculation of molecular associations. *Comput. Phys. Commun*. **1998**, *115*, 87-89.

72. Hammersley, J. M.; Handscomb, D. C. In: *Monte Carlo Methods*, Chapman and Hall: London, 1983, Chap. 5.

73. Petitjean, M. On the analytical calculation of van der Waals surfaces and volumes: Some numerical aspects. *J. Comput. Chem*. **1994**, *15*, 507-523.

74 Luscher, M. A portable high-quality random number generator for lattice field theory simulations. *Comput. Phys. Comun*. **1994**, *79*, 100-110.

75. James, F. RANLUX: A Fortran implementation of high-quality pseudorandom number generator of Luscher. *Comput. Phys. Comun*. **1994**, *79*, 111-114.

76. 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.

77. Deming, W. E. *Statistical Adjustment of Data*. Dover: New York, 1964.

78. Kantola, A.: Villar, H. O.; Loew, G. H. Atom based parametrization for a conformationally dependent hydrophobic index. *J. Comput. Chem*. **1991**, *12*, 681-689.

79. Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity: A rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219-3228.

80. Torrens, F.; Sanchez-Marin J.; Nebot-Gil, I. Interacting induced dipoles polarization model for molecular polarizabilities. Reference molecules, amino acids and model peptides, *J. Mol. Struct. (Theochem)*, **1999**, *463*, 27-39.

81. The CDHI program is available from the authors on request.

82. Leo, A.; Hansch, C. Linear Free-Energy Relationships between Partitioning Solvent Systems. *J. Org. Chem*. **1971**, *36*, 1539-1544.

83. Torrens, F. Fractal dimension of different structural-type zeolites and of the active sites. *Top. Catal*. submitted for publication.