Journal of Mathematical Chemistry 6(1991)351–358

NUMERICAL COMPUTATION OF SURFACE AREAS OF MOLECULES

Peter SENN
Physical Chemistry Laboratory, ETH-Zentrum, CHN H32, CH-8092 Zürich, Switzerland

Received 13 August 1990; revised 28 November 1990

Abstract

For a cubical tesselation of a finite region of space which contains an irregularly shaped surface, a fairly accurate estimate of the surface area is $\frac{2}{3}Na^2$, where N is the total number of cubes cut by the surface and a is the length of the edges of the cubes. An estimate of slightly improved accuracy can be obtained by using different increments to the surface area, depending on the number of edges of the corresponding cube cut by the surface and the number of vertices on either side of the surface.

1. Introduction

The volume and the surface area of a molecule play an important role in semi-empirical theories relating physical properties and chemical and biological activity to molecular structure [1]. There exists a variety of definitions of the "surface" of a molecule, and these definitions may be based on different physical concepts [2]. In many cases, the solvent accessible surface of a molecule or the portion of its surface with steric complementarity with respect to the reactive site of a given reagent is of interest [3].

The volume of a molecule can readily be computed by decomposing the space it occupies using a cubical tesselation. One starts with a fairly large size of cubes which would provide merely a rough estimate of its volume. An improved estimate can then be obtained by decomposing the cubes which are cut by the surface into eight smaller cubes called octants, and those octants which are cut by the surface are again decomposed into eight smaller cubes and so on, until the volume can be determined with the desired accuracy. This rather efficient technique of progressively dividing cubes into smaller ones can be used for the calculation of volumes of spatial structures of arbitrary shape [4]. The problem of computing surfaces can be reduced to the computation of volumes by converting the surface into a sheet of finite uniform thickness h and computing its volume from which the surface area A can be obtained as $A \sim V/h$. However, for surfaces of arbitrary shape, the construction of a sheet of finite uniform thickness may be far from trivial. In this case, the method outlined below might be appropriate.

2. Method

It can be shown that if a cube is randomly oriented in space, then its height averaged over all orientations in space is $\frac{3}{2}a$, where a denotes the length of the edges of the cube [5,6]. From this result we conclude that if a plane cuts a randomly oriented cube, then on the average the area of the cut will be as follows:

$$\Delta A \sim \tfrac{2}{3}a^2. \tag{1}$$

This means that an estimate of the surface area of an object can readily be obtained in a manner analogous to the computation of volumes. In the calculation of volumes, the cubes in the interior contribute a^3 to the volume and the cubes cut by the surface each contribute one half of a^3. For the computation of surface areas, only the cubes cut by the surface contribute to A and the increment is $\frac{2}{3}a^2$. Of course, this method is not exact, regardless of the smallest size of cubes considered. Take, for example, a flat surface parallel to one pair of faces of the cubes used in the cubical tesselation of space. In this case, the correct increment to the surface area would be a^2 per cube cut by the surface. The proper increment in this case is considerably different from the average value of $\frac{2}{3}a^2$ being used. An improved estimate of the surface area can be expected from a scheme where we differentiate among different types of cuts. Consider again the example with the flat surface discussed above. In this case, all cubes are cut such that there are four vertices on each side of the surface, and we could argue that for this type of cut the proper increment should be a^2 instead of $\frac{2}{3}a^2$.

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 an average the areas of triangular cuts tend to be somewhat smaller than the average of $\frac{2}{3}a^2$ which is being used. This observation suggests a slightly more elaborate scheme, where the polygonal cuts are classified according to their number of sides and using different increments ΔA_k for the different types of cuts. Let n_e be the number of sides of the cut and n_v the smaller one of the number of vertices located on the same side of the plane. Then, in general, $n_v = n_e - 2$. There exists, however, a fifth case where $n_v = n_e = 4$. For the estimate of the surface area according to (1), the symbol A_1 will be used, i.e.

$$A \sim A_1 = \tfrac{2}{3}Na^2, \tag{2}$$

and A_2 denotes the estimate obtained from the modified scheme outlined above:

$$A \sim A_2 = \sum_{k=1}^{5} N_k \Delta A_k, \tag{3}$$

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 A. Obviously,

Table 1

Classification scheme for cuts of a cube by a plane. The cuts are polygons with from three to six sides. The quantitities n_e and n_v identify the types of cuts which are numbered from 1 to 5, where the index k indicates the type. Here, n_e indicates the number of sides of the boundaries of the polygonal cuts and n_v represents the lesser of the number of vertices located on the same side of the plane. The quantities ΔA_k represent the average area for the different types of cuts and p_k represents the probability with which the corresponding cuts arise, where $\sum_k p_k = 1$ and, obviously, $\sum_k p_k \Delta A_k = \frac{2}{3}a^2$

k	n_v	n_e	ΔA_k[a]	p_k
1	1	3	$0.137474172a^2$	0.27982572
2	2	4	$0.622185068a^2$	0.33209571
3	3	5	$1.05368062a^2$	0.18693034
4	4	6	$1.2512237a^2$	0.04644769
5	4	4	$1.07621810a^2$	0.15470054

[a]The last digits indicated for ΔA_k may be in error by ± 2.

$N_1 + N_2 + N_3 + N_4 + N_5 = N$. In table 1, the classification scheme for the above five cases is shown together with the computed values for ΔA_k. The simplest method for the computation of the averages ΔA_k consists of a Monte Carlo algorithm where a cube is cut by a large number of randomly oriented planes. The cuts are subsequently classified according to the scheme presented in table 1, the areas of the cuts are computed analytically for each random cut, and the quantities ΔA_k are obtained as separate averages for the five different cases. The number of cuts to be computed in a Monte Carlo calculation for the accuracy of ΔA_k indicated in table 1 would be excessive. The actual numbers shown in table 1 were obtained by numerical integration. The values obtained were in agreement with less accurate estimates obtained from Monte Carlo calculations.

3. Results and discussion

For a large number of spheres with randomly chosen radii r and random locations of their centers, estimates A_1 and A_2 of their surface areas have been computed. In each case, the quantities $\bar{\eta} = A_1/4\pi r^2$ and $\eta = A_2/4\pi r^2$ were determined. Figure 1 depicts the ratio $\bar{\eta}$ as a function of the variable ρ, where $\rho = r/a$. According to fig. 1, the accuracy of A_1 as an estimate of the area of a surface of a sphere with radius r can be expected to be 1% or better if $\rho = r/a > 10$.

Notice that for curved surfaces, more cases than indicated in table 1 would have to be taken into account. Examples of types of cuts not considered in table 1 would be the ones where the same edge is cut twice by the surface. For surfaces with small curvature and/or, equivalently, for small values of a, such cases are

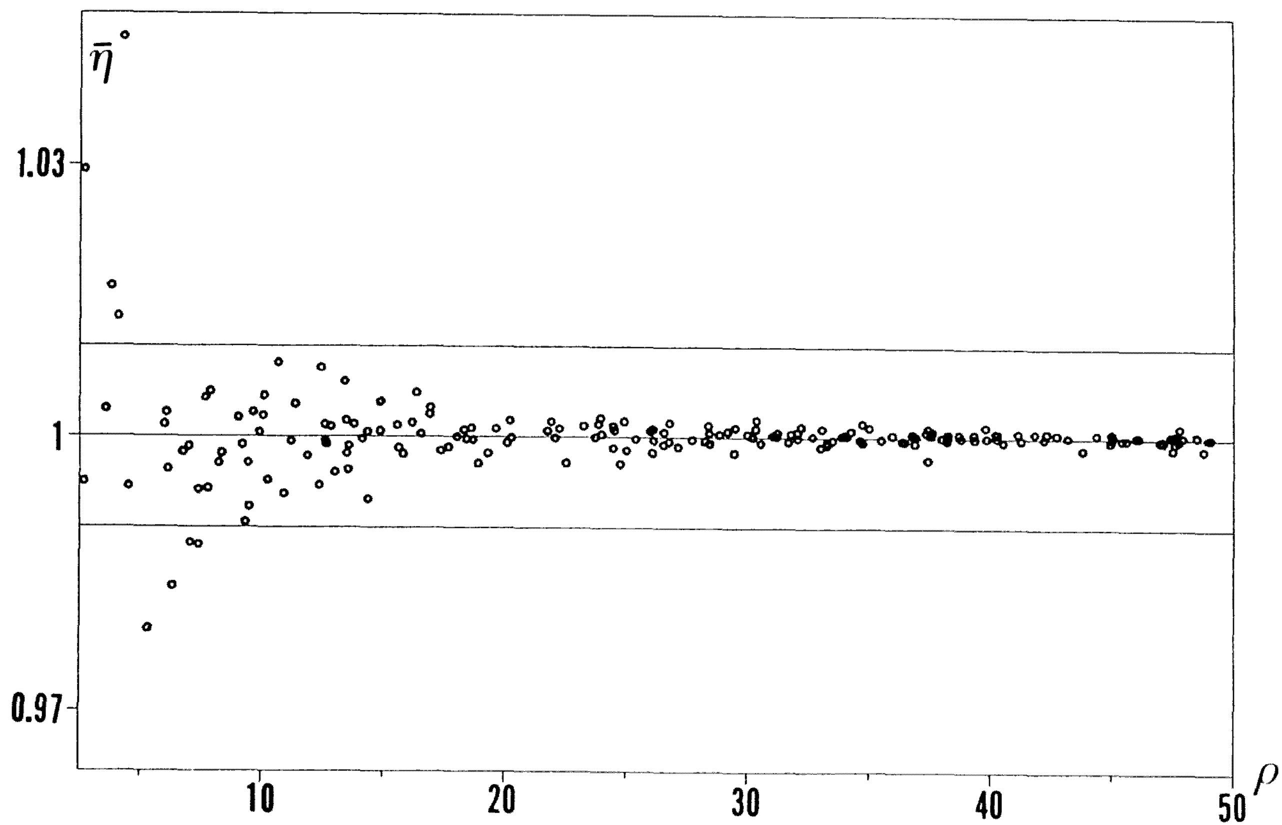

Fig. 1. The ratio $\bar{\eta} = A_1/4\pi r^2$ computed for a large number of "random" spheres as a function of the variables $\rho = r/a$, where r represents the radius of the corresponding sphere and a is the length of the edges of the cubes used in the cubical tesselation of space. The three equidistant horizontal lines are at $\bar{\eta} = 1$ and $\bar{\eta} = 1 \pm 0.01$.

rarely encountered. If the above algorithms are being implemented in computer programs, such additional cases may nevertheless have to be considered.

When a cube with random orientation in space is cut by a horizontal plane, then the probability of having n_b corners below the plane is exactly the same as the probability of having $8 - n_b$ corners below the plane. For closed surfaces such as spheres or surfaces of molecules, the locations of the vertices are such that they are either interior or exterior or, on rare occasions, they may be on the surface. The probability that one corner is located inside the surface with the remaining seven being located outside is found to be somewhat greater than the probability of having one corner outside the closed surface, in contrast to what is observed for cuts by a plane. Since the present statistical model is based on cuts by a plane, it might be possible to introduce a correction based on the skewness of the distribution of the number of corners located inside the surface.

Previously, it was shown that considerable errors could be made using A_1 as an estimate for the surface area when the surface is completely flat. Using A_2 as an estimate for the surface area, for a flat surface oriented parallel to a pair of faces of the cubes we would obtain $N_1 = N_2 = N_3 = N_4 = 0$ and $N_5 \neq 0$, such that $A_2 = N_5 \Delta A_5 = 1.076 N_5$ and the error would be roughly 7.5%, which is less than a quarter of the error resulting from using A_1 as an estimate. Figure 2 shows the quantity η as a function of ρ, where $\eta = A_2/4\pi r^2$ for a large number of "random" spheres. A comparison of figs. 1 and 2 reveals that the actual improvement resulting from using A_2 instead of A_1 as an estimate of the surface area is rather modest. It should be pointed out that the classification of the cuts is relatively time consuming such that the additional work in computing A_2 may not be worthwhile.

For benzene, the exact Van der Waals surface area can be determined analytically. A rather lengthy calculation shows that, using the Van der Waals radii $r_C = 1.7$ Å and $r_H = 1.2$ Å with bond lengths of 1.4 Å and 1.1 Å for the C–H and the C–H bonds, respectively, the area of the Van der Waals surface is

$$A = \left\{ \pi \left\{ \frac{522.66}{11} - (749.088)^{1/2} \right\} + 40.8 \int_0^{u+v} \arctan\left(\tfrac{1}{2} x[(u-v-x)(x-u-v)]^{-1/2} \right) \mathrm{d}x \right\} \text{Å}^2 \tag{4}$$

$$= 110.3764209 \text{ Å}^2,$$

where $u = 1.05$ and $v = (9/5)^{1/2}$. In table 2, the relative errors in percent for the Van der Waals surface area estimated as A_1 and as A_2 are shown for four random orientations of the molecule in space. For $a = 0.128$ Å, the Van der Walls radii are about ten times larger than a and, according to the above, the accuracy of A_1 as an estimate

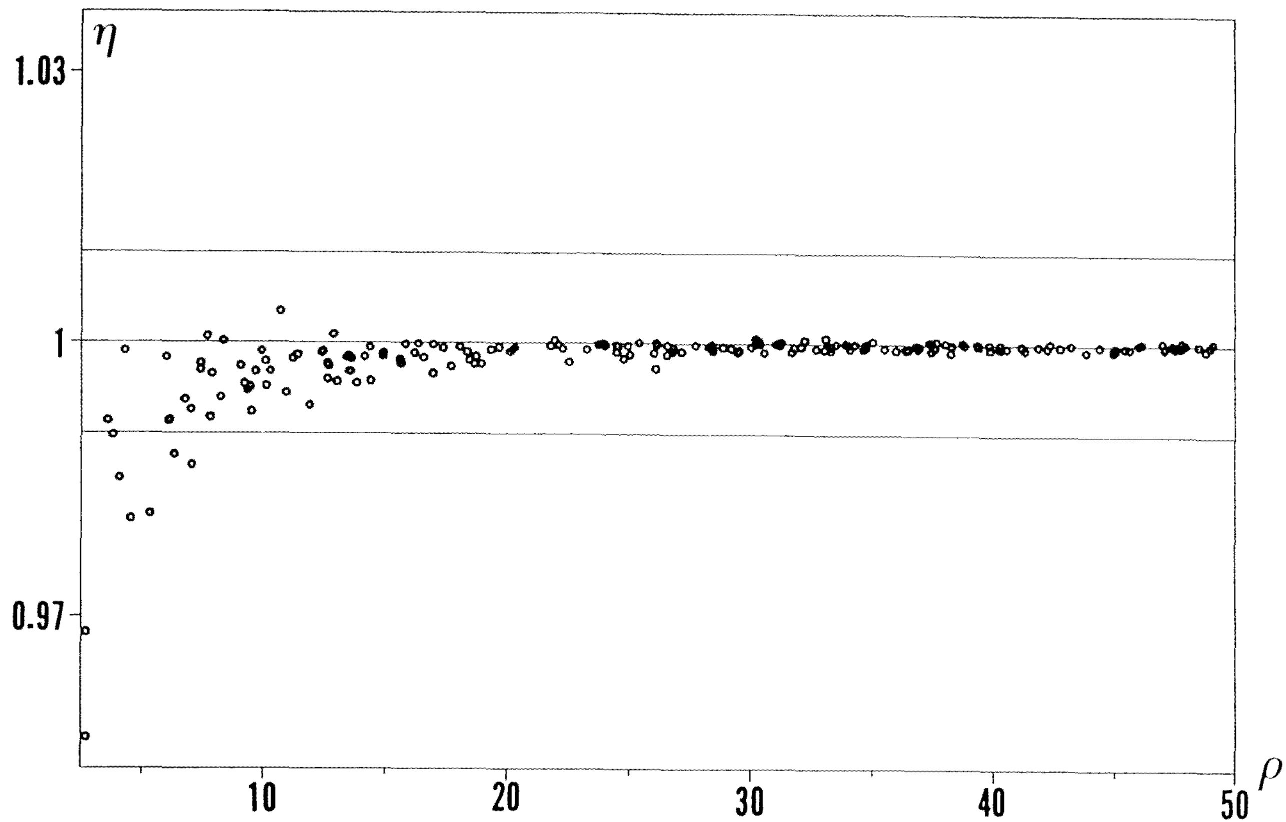

Fig. 2. The ratio $\eta = A_2/4\pi r^2$ computed for a large number of "random" spheres as a function of the variable $\rho = r/a$, where r represents the radius of the corresponding sphere and a is the length of the edges of the cubes used in the cubical tesselation of space. The three equidistant horizontal lines are at $\eta = 1$ and $\eta = 1 \pm 0.01$. For small values of ρ, the estimates A_1 and A_2 tend to be somewhat smaller than the actual surface areas, i.e. in the majority of cases, $\bar{\eta} < 1$ and $\eta < 1$ if ρ is small.

Table 2

Relative errors in percent of the computed surface areas A_1 and A_2 for benzene in four random orientations O_1, O_2, O_3, and O_4 of the molecule in space. The relative errors are computed as follows:

$$\frac{A_{\text{exact}} - A_{\text{calc}}}{A_{\text{exact}}} \times 100\%,$$

where A_{exact} denotes the exact surface area and $A_{\text{calc}} = A_1$ (upper entries) or $A_{\text{calc}} = A_2$ (lower entries)

a [Å]	O_1	O_2	O_3	O_4
1.024	14.591	11.989	13.670	15.530
	8.167	5.000	5.000	10.067
0.512	3.760	3.889	5.692	4.193
	0.250	1.517	3.100	1.042
0.256	1.715	1.610	1.741	1.704
	0.883	− 0.067	0.408	0.408
0.128	1.002	1.029	1.237	0.923
	0.507	− 0.067	0.349	0.201
0.064	0.519	0.611	0.643	0.471
	0.037	− 0.213	− 0.195	− 0.119
0.032	0.327	0.422	0.455	0.285
	− 0.032	− 0.300	− 0.273	− 0.247
0.016	0.241	0.349	0.352	0.190
	− 0.078	− 0.304	− 0.310	− 0.286

should be roughly 1%. According to table 2, the same applies for A_1 as an estimate of the area of the Van der Waals surface of benzene. According to ref. [4], using cubes with edges of 0.125 Å length, the surface area of benzene could be computed from the volume of a sheet of uniform thickness of 0.1 Å with an average relative error of roughly 0.05%, which indicates that the present method is less accurate by a factor of approximately 20 than the method used in ref. [4]. If A_2 were used as an estimate, a slightly more favourable outcome would be obtained from a comparison between the present method and the method used in ref. [4]. In this case, we would obtain a factor of six or seven instead of twenty.

Several other molecules have been computed. For example, for the planar molecule xanthan hydride ($C_2H_2N_2S_3$), using the structural parameters reported by Stanford [7], an average of 132.972 Å^2 was obtained from 55 random orientations in space for A_1, while for A_2 the average was 133.296 Å, where the smallest size of cubes considered had edges of 0.04 Å length. The standard deviations from the mean were 0.806 Å^2 and 0.513 Å^2 for A_1 and A_2, respectively.

In conclusion, it can be said that with respect to the attainable accuracy the present method cannot compete with other methods presently in use which are of similar simplicity. However, in cases where it is difficult to construct a sheet of uniform thickness, the method outlined above may provide estimates of the surface areas of molecules with an adequate accuracy.

A source listing on paper of a simple computer program in FORTRAN, with which the above method of computing surface areas of molecules has been tested, can be obtained from the author upon request.

References

[1] F.M. Richards, Ann. Rev. Biophys. Bioeng. 6(1977)151.
[2] M.L. Connolly, Science 221(1983)709.
[3] H.R. Karfunkel and V. Eyraud, J. Comp. Chem. 10(1989)628.
[4] J. Higo and N. Go, J. Comp. Chem. 10(1989)376.
[5] P. Senn, SIAM Rev. 31(1989)319.
[6] J.G. Watson, SIAM Rev. 32(1990)304.
[7] R.H. Stanford, Jr., Acta Cryst. 16(1963)1157. The Van der Waals radii used for C, H, N, and S were 1.7 Å, 1.2 Å, 1.55 Å, and 1.8 Å, respectively.