GEPOL: An Improved Description of Molecular Surfaces. I. Building the Spherical Surface Set

Juan Luis Pascual-Ahuir and Estanislao Silla*

Departamento de Química-Física, Universidad de Valencia, 46100 Burjassot, Valencia, Spain

Received 20 September 1989; accepted 15 May 1990

The algorithm used by the program GEPOL to compute the Molecular Surface (MS), as defined by Richards, is presented in detail. GEPOL starts like other algorithms from a set of spheres with van der Waals radii, centered on the atoms or group of atoms of the molecule. GEPOL computes the MS by first searching the spaces inaccessible to the solvent and consequently filling them with a new set of spheres. Here we study the behavior of the method with its parameters, presenting several examples of application.

INTRODUCTION

The concept of superficial area of a solute to determine its thermochemical activity was first introduced by Langmuir in 1925. He thought that the required energy to build a cavity in the solvent was an important factor in relation to solubility, and proportional to the solute area.[1] Afterwards many studies were conducted on the correlation between the area, and the solubility and hydrophobicity of many molecules.[2-3] Furthermore, studies were made on models and methods explaining and calculating the interaction of one molecule with its surroundings, making use of its area as well as the molecular volume.[4-6] On the other hand Linus Pauling postulated that the interactions among molecules of biological interest, such as nucleic acids and proteins, are modulated first and foremost by their size and shape.[7] Therefore biochemistry has been one of the fields of chemistry where the concept of the molecular surface has mostly been utilized.

It is evident that the calculation of molecular surfaces requires suitable methods and a proper definition of the molecular surface, which can be described with simple geometrical models like spheres, ellipsoids, and cylinders,[8-10] or even as being equipotential or equi-electronic density surfaces. However, among the methods the more usual are those which are called the Van der Waals type.[11-15] Under this name are included all molecular surfaces that make use of the Van der Waals radius. Nowadays we can consider there to be three kinds of surfaces of this type. (1) The proper Van der Waals surface (WMS) which is the external surface resulting from a set of spheres centered on the atoms or group of atoms forming the molecule (Fig. 1(a), (2) The surface accessible to the solvent (AMS), defined by Richards and Lee[16] as the surface generated by the center of the solvent, considered as a rigid sphere, when it rolls around the Van der Waals surface. (Fig. 1(b)). (3) the molecular surface defined later by Richards[17] (MS) which is composed of two parts: the contact surface and the reentrant surface. The contact surface is the part of the Van der Waals surface of each atom which is accessible to a probe sphere of a given radius. The reentrant surface is defined as the inward-facing part of the probe sphere when this is simultaneously in contact with more than one atom (Fig. 1(c)).

Recently we have developed a set of efficient and reliable algorithms to calculate the surface and volume of molecules.[18,19] Using these algorithms we have written a program called GEPOL[20] that computes the molecular surface as a distribution of points and calculates the corresponding area and volume. The program starts using a set of spheres centered on the atoms or group of atoms, with the respective radii given in the input. This is enough to compute the WMS and the AMS, but if the MS is the desired surface 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 area and volume. This latter process is controlled by a parameter called NDIV, as defined by us[18,19] and it will be studied more deeply in a future article.

In this article we present the procedure to obtain this set of new spheres and analyze the results obtained for several selected examples. In the first part we explain the method used to build this new set of spheres, as well as how to calculate their radii and coordinates of their centers. Later we discuss the parameters used in the algorithm in relation to the quality of the calculation and

*To whom all correspondence should be addressed.

CCC 0192-8651/90/0901047-14$04.00

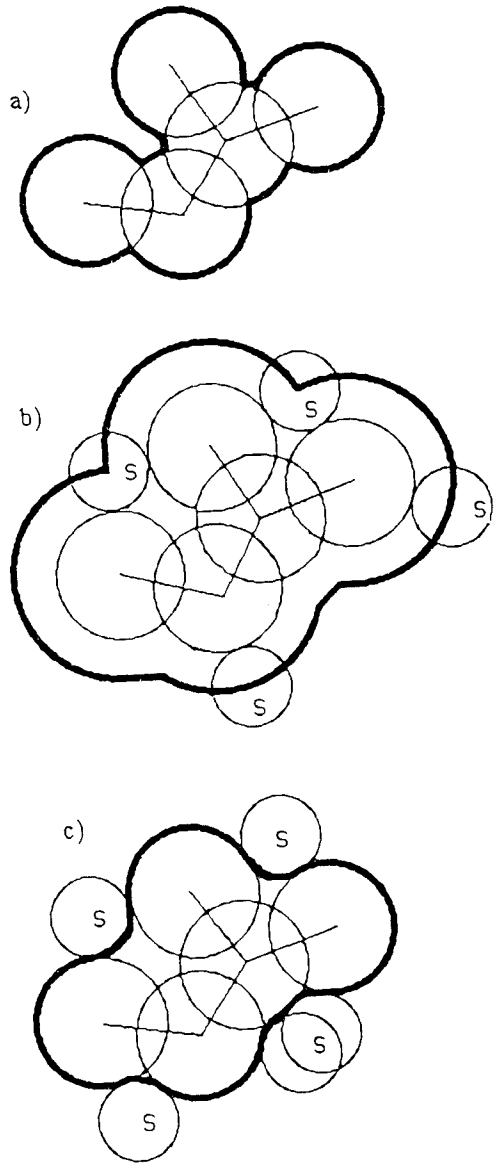


Figure 1. Three definitions of the envelope surface of a molecule. (a) Van der Waals molecular surface (WMS). (b) The accessible molecular surface (AMS). (c) The molecular surface (MS).

the cpu time spent. In the last part we apply the method to several cases where the variation of the area and volume could be significant.

COMPUTATION METHOD

The algorithm for the creation of new spheres in the spaces inaccessible to the solvent works as follows:

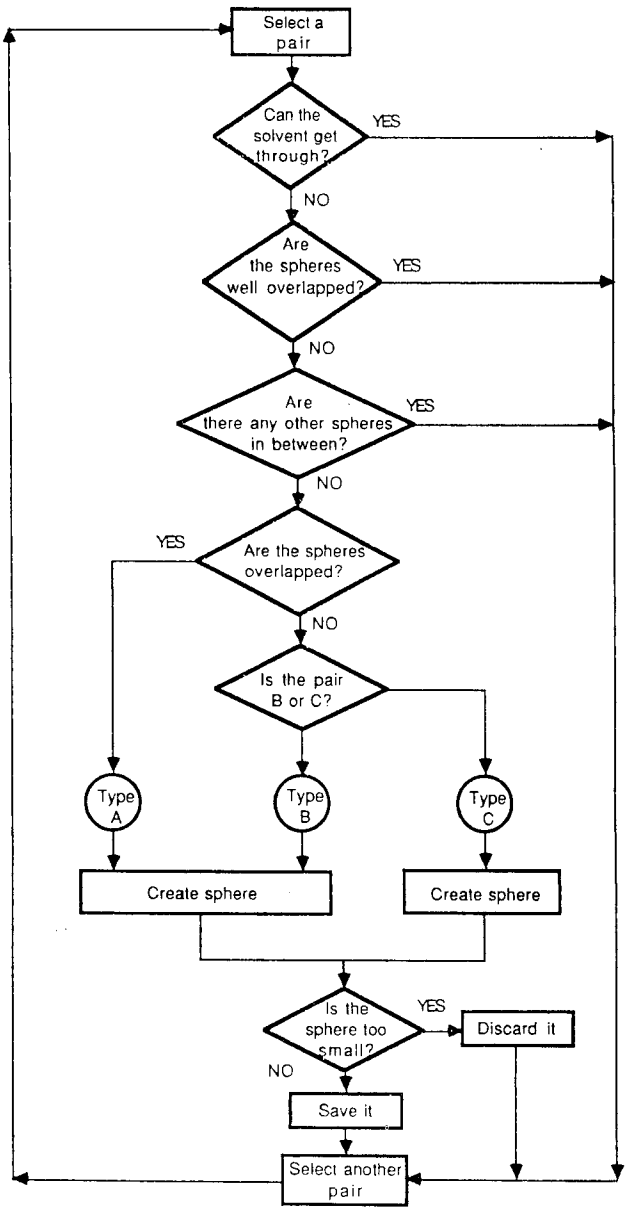
Step 1. The process starts considering all the pairs of spheres that it is possible to get with the original set of spheres. If our molecule had seven atoms and we had centered one sphere over each atom, we would have 21 pairs. From here the whole of computations are made over the pairs.

Step 2. Next, the program carries out the following operations for each pair of spheres, to be explained in detail later. (a) Determine whether between the two spheres of the pair there is some space inaccessible to the solvent that must be filled. (b) If so, the program creates a new sphere between the two spheres of the pair.

Step 3. Once all the pairs given in "step 1" have been analyzed in the "step 2," the program determines again all the pairs that it is possible to get, using now the spheres given in the input and those created in the previous iterations. After eliminating the pairs that were analyzed in previous iterations the process is repeated from step 2.

As can be seen the program works iteratively, creating in each iteration a new set of spheres. The molecular surface defined by the external parts of whole set will tend, in a convergent process, to the molecular surface of Richards.[17] The cycle is stopped when a new iteration does not create new spheres.

The bulk of the program corresponds to step 2 and is described in detail as follows and also is illustrated as a flow diagram in scheme 1.



Scheme 1. Agorithm flow diagram.

How Does the Program Decide if a New Sphere Should be Created?

In this operation the program analyzes whether between the two spheres forming the pair there is a space that is not occupied by another sphere of the original set, and besides it is inaccessible to the solvent. The decision is made applying three tests or conditions to the pair and only when they are fulfilled a new sphere will be created. The three tests are as follows:

1. Can the solvent get through the two spheres? If the answer is no, the pair will pass to the next test, because this answer means that there is a space inaccessible to the solvent. The computation starts calculating the distance d between the centers of the spheres that form the pair and only the pair that fulfill the following relation will pass to the next test:

$$d < (R_G + R_P + 2R_S),$$

where R_S is the radius of the solvent represented by a rigid sphere, and R_G and R_P the radii of the spheres G and P that form the pair.

2. Are the spheres well overlapping? In fact, the volume inaccessible to the solvent between the two spheres decreases continuously when the overlapping angle, as defined in Figure 2, increases. When this angle takes values close to 90°, it becomes so small that the effect of adding more spheres approaches minimum. Actually, the standard version of the program[20] calculates the distance between the two spheres with the overlapping angle fixed at 50°, then selects the pairs with a real distance d larger than the calculated one. The angle of overlapping was fixed to 50° because after several tests it appeared that a larger value did not give a significantly better result, and produced in an increase of cpu time cost.

3. Are there any other spheres in between? Each pair is tested once more before creating a new sphere. The program checks whether a sphere already exists between the pair, obstructing the access of the solvent. If so, a new sphere will not be created. To decide if a sphere I is between the pair GP, the distance between the center of I and the line that joins the centers of G and P, is compared with a value that we call radius of hiding (RHID). This radius is obtained by multiplying the radius of I by the parameter FRADIO given in the input (Fig. 3). FRADIO can take values between 0 and 1 and is used to control the rigidity of this test. In fact, the smaller the value of FRADIO, the closer should the center of I be to the line GP considering that I is between G and P. One should note that when comparing spheres created in the same iteration the condition for hiding is more rigid, i.e., the sphere I hides G from P, when its center is just over the line GP, to avoid the formation

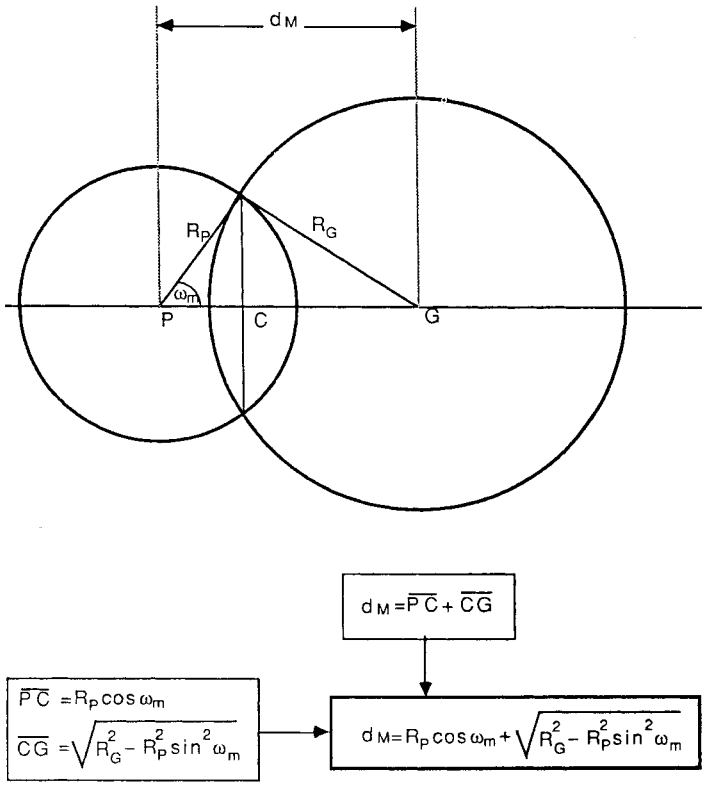


Figure 2. Calculation of the distance between G and P with an overlapping angle equal to ω_m. The angle is always defined from the smallest sphere.

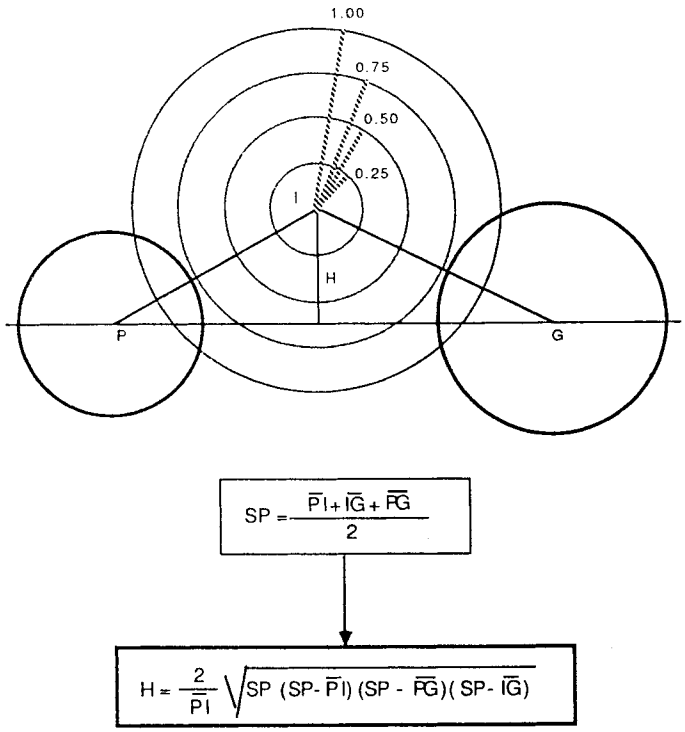


Figure 3. Different radii of hiding for values of FRADIO of 0.25, 0.50, 0.75, and 1.0, and calculation of the distance *H*. See text for further details.

of an asymmetrical surface if the molecule is symmetric.

Creation of New Spheres

The way to create the new spheres depends on the type of the pair. In the Figure 4 the three types of pairs are shown with the new spheres drawn in. The pairs of the type *A* have the spheres overlapping. The difference between *B* and *C* is that in *B* the distance between the surface of the solvent and axis *GP* is larger than the radius of the smallest sphere that we allow to be created. The radius of this smallest sphere is called RMIN and is calculated by multiplying FRADIO by the average of the radii of all initial spheres.

If the pair is of the type *A* or *B* the new sphere will have its center at a point on the axis *GP* and equidistant from the two spherical surfaces of the pair, and with such a radius that it will be tangent to the solvent. The methods for calculating the coordinates and radius of the new sphere for these types of pair are shown in the Figures 5 and 6. As can be seen in the Figure 7, for the pairs of the *C* type the center will be in the intersection of the surface of the biggest sphere of the pair with the axis *GP* and the radius will be calculated being tangent to the solvent.

The addition of new spheres having too small a radius encumbers the computation without producing significant improvement in the results. For this reason, a new test is used in our algorithm: if the radius of the new sphere is smaller than RMIN the sphere will be discarded. Thus, the use of RMIN is a decisive factor to stop the iterative process that the program follows.

STUDY OF THE VALUES OF FRADIO

As we saw in the previous section FRADIO is the parameter that determines the values of RMIN and RHID and so controls the number of spheres created in the process of smoothing the molecular surface. With the aim of studying the behavior of the method with RMIN and RHID, we have made a modified version of GEPOL, that allows independent values of RMIN and RHID to be given using two different FRADIO.

For this analysis we have used a structure that we call NAD-LOOP, where NAD is the coenzyme nicotinamideadenine dinucleotide and LOOP is a polypeptide loop of the LADH (Liver Alcohol Dehydrogenase) involved in the binding of the coenzyme. The association of the NAD and this loop of the LADH allows the LADH to catalyze the reaction of alcohols to aldehydes or ketones.[21] This structure was not selected for its special biochemical interest but as a real structure with which to test the method. NAD-LOOP is composed of 165 nonhydrogen atoms, the initial spheres of radius 1.8 Å being centered on them. The carte-

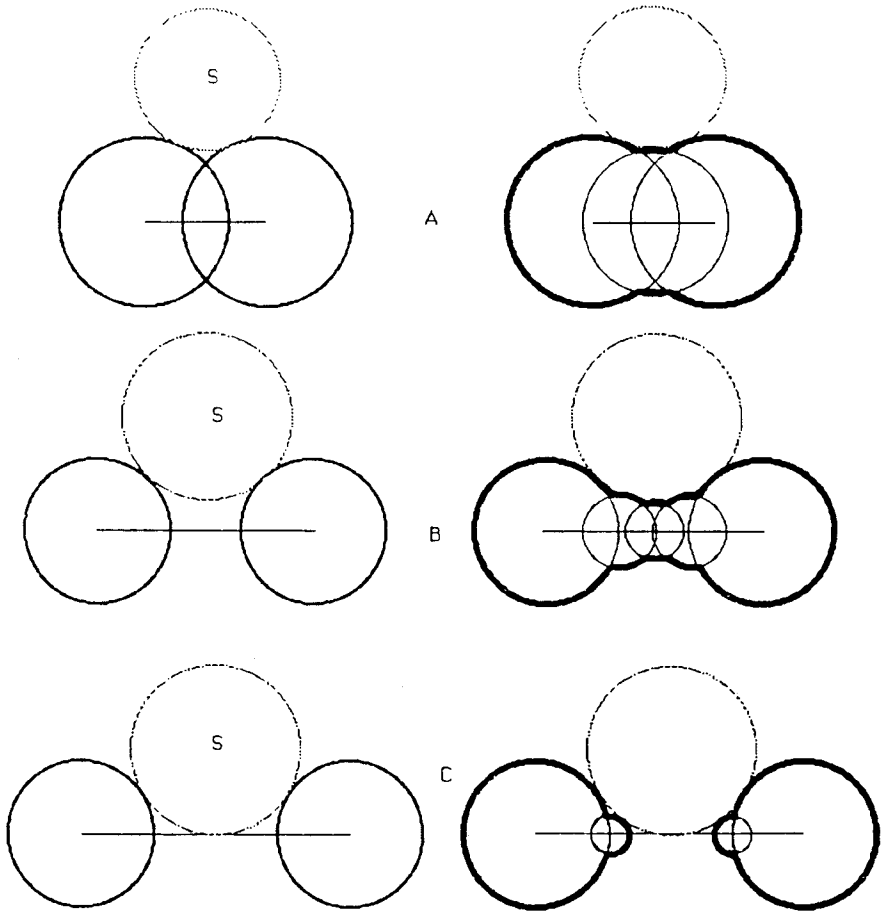


Figure 4. The three different pairs of spheres considered by the algorithm. The new spheres created in each are shown, as well as the resulting MS (bold line). S is the solvent or probe sphere.

sian coordinates were obtained from a molecular dynamic calculation.[22] The radius of the solvent was taken equal to 1.5 Å.

In the Table I we indicate with the symbol + the calculations that have been made. The calculations corresponding to smaller values of FRADIO-(RHID) and FRADIO(RMIN) have not been made because they do not give more information and they require a large amount of cpu time. For the value of FRADIO(RMIN) equal to 0.9 there is no creation of spheres.

In the Figure 8 are shown the values of area and volume as a function of the cpu time for all of the calculations given in Table I. Due to the filling of the space inaccessible to the solvent, the process of smoothing implies an increase to the volume and a decrease of the area. In these figures we can see that there is a limit of efficiency shown with a continuous line. The values corresponding to the diagonal of Table I are drawn in bold type. We can see that they lie close to the continuous line, which is not a surprise because in this way both FRADIO (RMIN) and FRADIO(RHID) give a similar degree of freedom in the process of selection of the pairs. This effect which we have observed in other cases is the reason for combining them in one, FRADIO, that controls both factors at the same time.

In the Figure 8 we can also see that when FRADIO decreases the area and volume approach a limit. This limit (FRADIO equal to 0.0) corresponds to the continuous molecular surface defined by Richards.[17] Consequently, there is a direct relation between the computational effort and the quality of the result. After these calculations and some others that we have made, we recommend a value for FRADIO of 0.5 as giving a good ratio between the cpu time used and the result obtained.

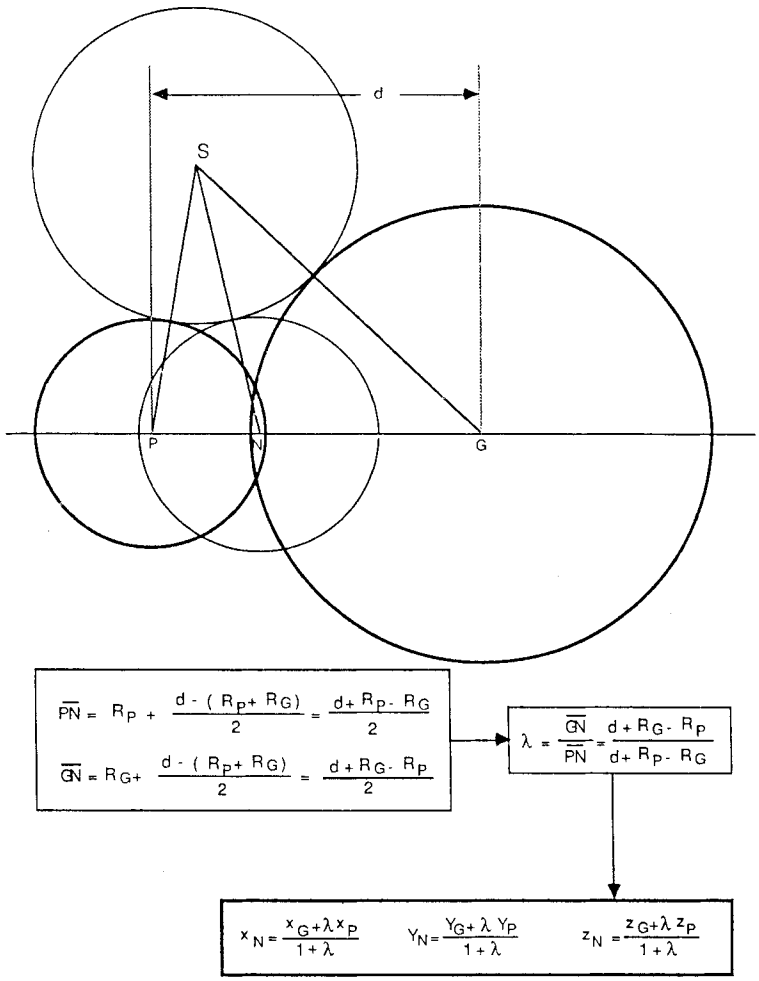


Figure 5. Calculation of the coordinates of the new sphere N in a pair of type A. The same formulae are used for a pair of type B. R_S solvent radius; R_P radius of sphere P; R_G radius of sphere G.

SOME EXAMPLES

We have selected a set of examples where the molecular surface could be a decisive influence on the properties under study. Nevertheless we do not try to reach any conclusions about these properties, but only to show the capability of GEPOL to describe the molecular surface. We start analyzing the variation of the molecular surface as a function of the distance between two spheres, for example as if dissociating in solution. The second case corresponds to an internal rotation of octanol, showing the importance of the kind of surface used. And the last one is the use of this method in the study of the fractality, which is one of the new ways to use a molecular surface.

Separation of Two Spheres

We have selected this simple model because it is one of the few cases compatible with analytical calculation and thus we can determine the capability of the method to compute the molecular area and volume. Also we would like to show the different behavior of the area and the volume, depending on the molecular surface definition, in this process that can be seen like a simple model of dissociation. This difference can be important in some methods that use these kinds of surfaces to study the reactivity.[23]

The radius of each sphere is 2.4 Å and the initial distance between the spheres is 1.5 Å. The solvent is represented by a sphere with a radius of 1.5 Å. In the Figure 9(a) can be seen the values of the area of the van der Waals surface (WMS) as well as the molecular surface (MS) calculated using GEPOL, along with the exact values of the MS calculated analytically. To calculate the MS, given the small size of the system, we have used a value of FRADIO equal to 0.2. Figure 9(b) represents the same calculations showing the volume enclosed by the surfaces.

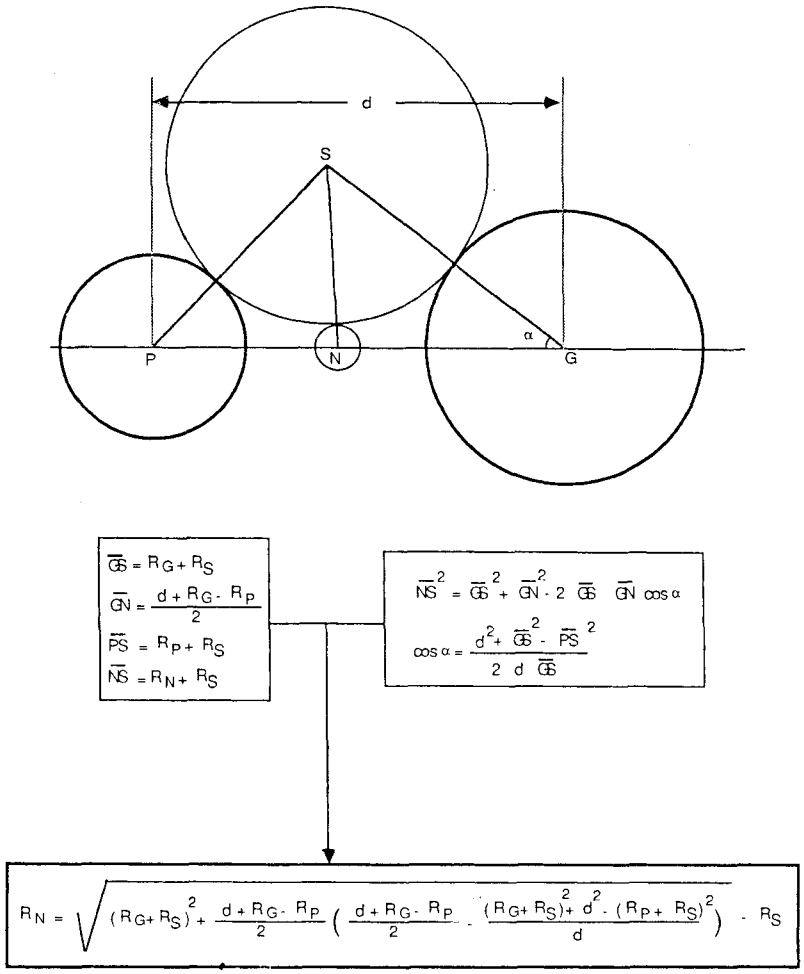


Figure 6. Calculation of the radius of the new sphere N in a pair of type B. The same formulae are used for the pairs A. α angle defined always from the bigger sphere. R_S solvent radius; R_P radius of sphere P; R_G radius of sphere G.

From the graphs it can be seen that the results obtained for the MS are good, the difference with the exact values being no larger than 1.5% in any case. GEPOL always overestimates the value of area and underestimates the value of volume. This is hardly surprising given the strategy used by GEPOL to build the MS. In fact, the smoothing process implies the filling up of space inaccessible to the solvent in successive steps, increasing in each one the molecular volume and decreasing the molecular area until reaching the limit given by the surface of Richards.

As we know the exact value of the area we can set up a comparison with the Connolly's method,[13] which is the most widely used nowadays. By using this method you get a basically similar area though the strategy used is very different. The differences between the exact value and those obtained with the MSDOT and GEPOL methods are shown in Figure 10. The MSDOT results show bigger mistakes than our results for almost all the cases. The MSDOT results show large downward fluctuations which confirm its oscillating nature.[29] In the WMS (Fig. 10(A)) the relative error for our program is never larger than 0.2% while Connolly's program goes up to 2.6%. In the MS (Fig. 10(B)) the relative error for GEPOL is never larger than 1.5% while MSDOT goes up to 2.6%. The CPU time consumed by our program is always smaller. The overall calculations shows that MSDOT uses a

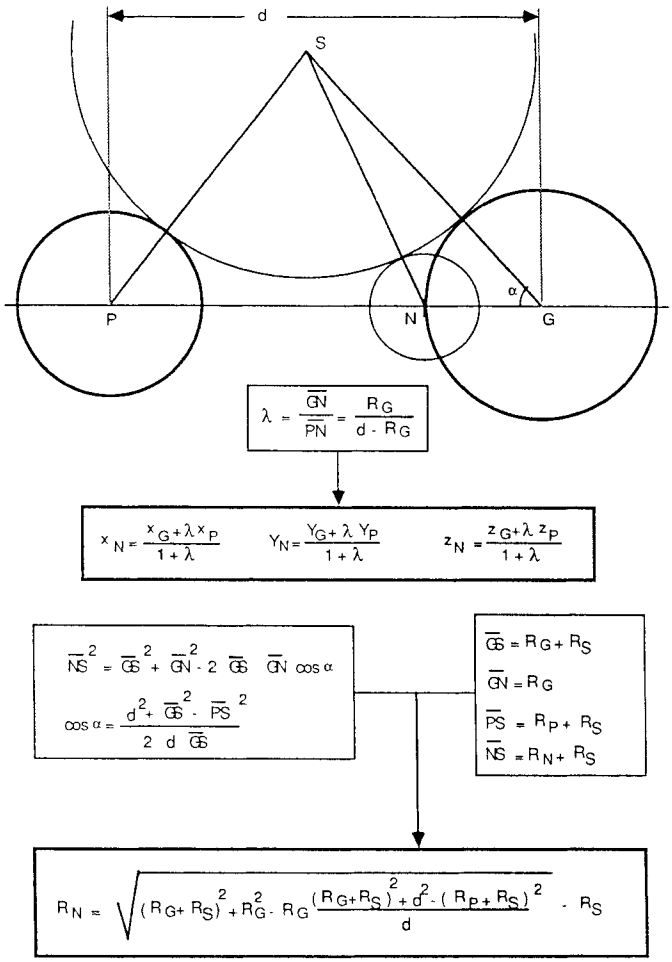


Figure 7. Calculation of the radius and the coordinates of the new sphere N in a pair of type C. α angle defined always from the bigger sphere. R_S solvent radius; R_P radius of sphere P; R_G radius of sphere G.

42.5% more time-consuming than GEPOL (41% in the WMS and 44% in the MS).

In Figure 9 it is also possible to see the differences of the behavior of the WMS in respect to the MS. The area of the WMS is bigger than the MS up to a distance of 5.5 Å between the centers of the spheres; after that it takes smaller values. That is logical because the WMS is formed for two isolated spheres at a distance of 4.8 Å (twice the radius of the spheres) and for the MS that occurs at a distance equal to 7.8 Å (4.8 plus the diameter of the solvent). This difference in behavior is important because the interaction of the solvent with the molecule is through the surface[6,18] and any variation of that is reflected in the interaction. Also the variation of the volume enclosed by the MS has a maximum but its values are always larger than in the WMS.

Internal Rotation

The study of the energetic barrier in an internal rotation is a common objective for a theoretical chemist. In solution the changes of the area and volume during an internal rotation produces variations in the values of the electrostatic interaction, energy of cavitation, etc. that can alter the form of the barrier. GEPOL has demonstrated its capacity to describe these changes, using both MS and WMS,[19] versus other well known method.[12,13] Here we will show the differences in the area and volume of octanol molecule when either type of

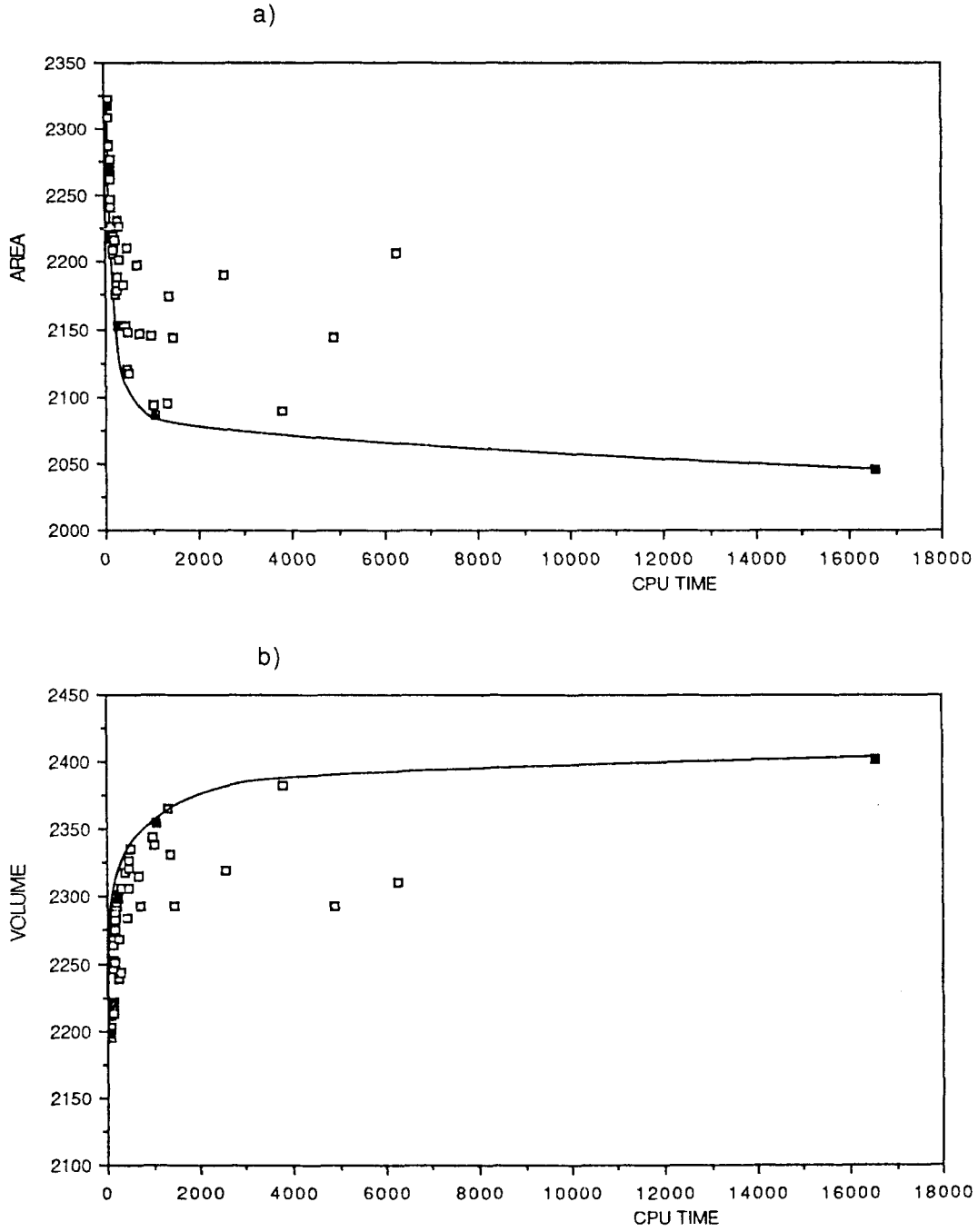


Figure 8. Values of the area (a) and the volume (b) in the system NAD-LOOP in function of the cpu time in seconds (Micro VAX II). See text for further details.

surface is used. Figure 11 shows the differences between the areas obtained with the WMS and MS, as well as the volumes enclosed by them, as a function of the angle of rotation. The angle of rotation corresponds to the rotation around the bond C4—C5 (values used are between 0° and 180°, in increments of 10°). Standard values were used for the geometry,[24] a solvent sphere of 1.5 Å and a value of FRADIO of 0.5.

From the area and volume curves, it can be seen that the differences between the two kinds of surfaces are significant. Besides that it is interesting to note that these differences depend on the angle and, thus, on the molecular geometry. In fact the difference between the areas increases almost 50% when the angle changes from 10 to 70. If the envelope surface of a molecule modulates the interaction of the molecule with the sur-

Table I. The symbol + marks the calculations performed for the system NAD-LOOP.

FRADIO (RHID)	FRADIO(RMIN) .1	.2	.3	.4	.5	.6	.7	.8	.9
.1	–	–	–	–	–	–	+	+	+
.2	–	–	–	–	–	–	+	+	–
.3	–	–	–	–	–	–	+	+	–
.4	–	–	–	+	–	+	+	+	–
.5	–	–	–	–	+	+	+	+	+
.6	–	–	+	+	+	+	+	+	–
.7	–	–	+	+	+	+	+	+	–
.8	–	+	+	+	+	+	+	+	–
.9	+	+	+	+	+	+	+	+	–
1.0	+	+	+	+	+	+	+	+	–

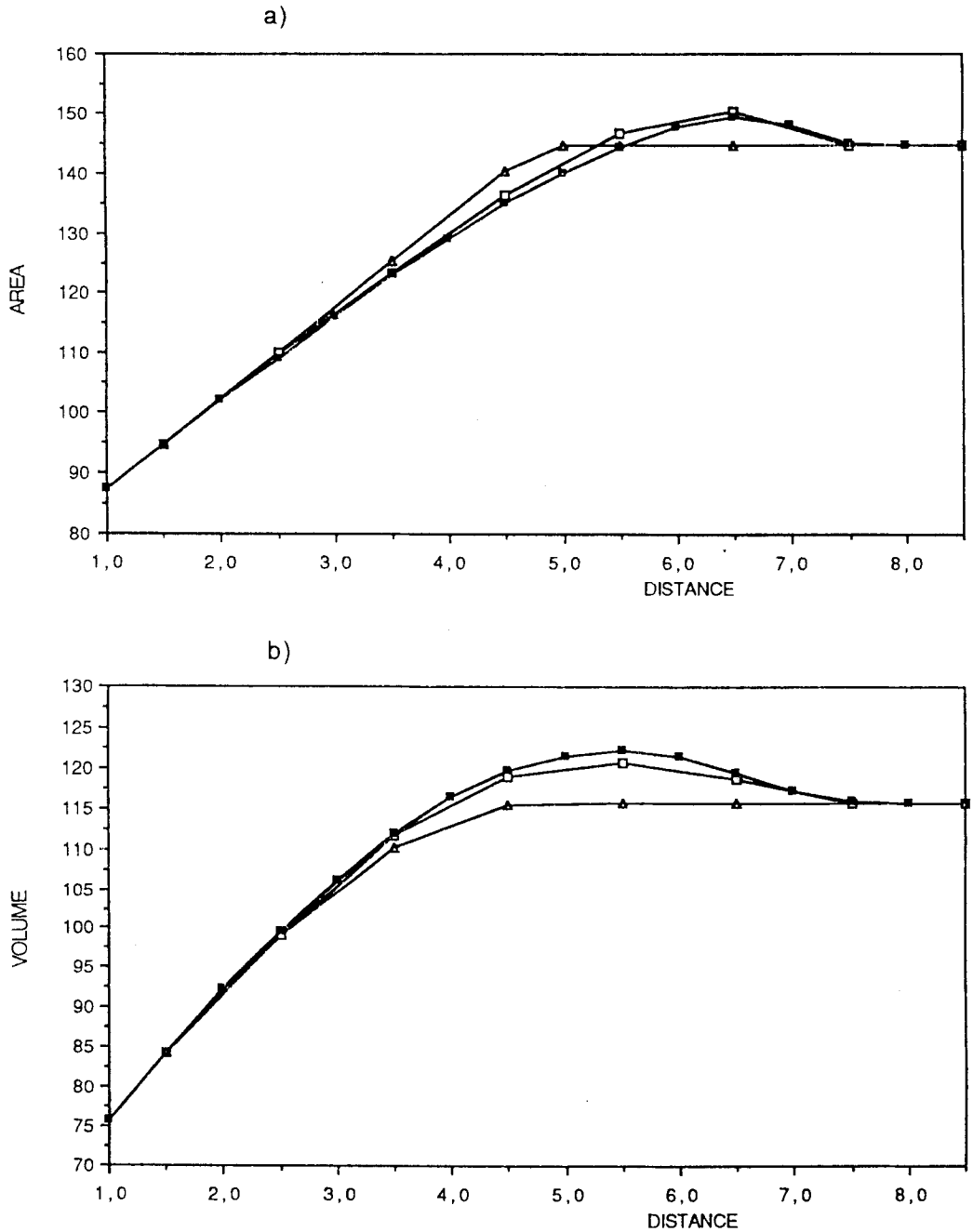


Figure 9. Values of the area (Å^2) and the volume (Å^3) enclosed by the surface formed initially for two spheres, versus the distance (Å) between their centers. △ WMS GEPOL; □ MS GEPOL; ■ MS Analytical.

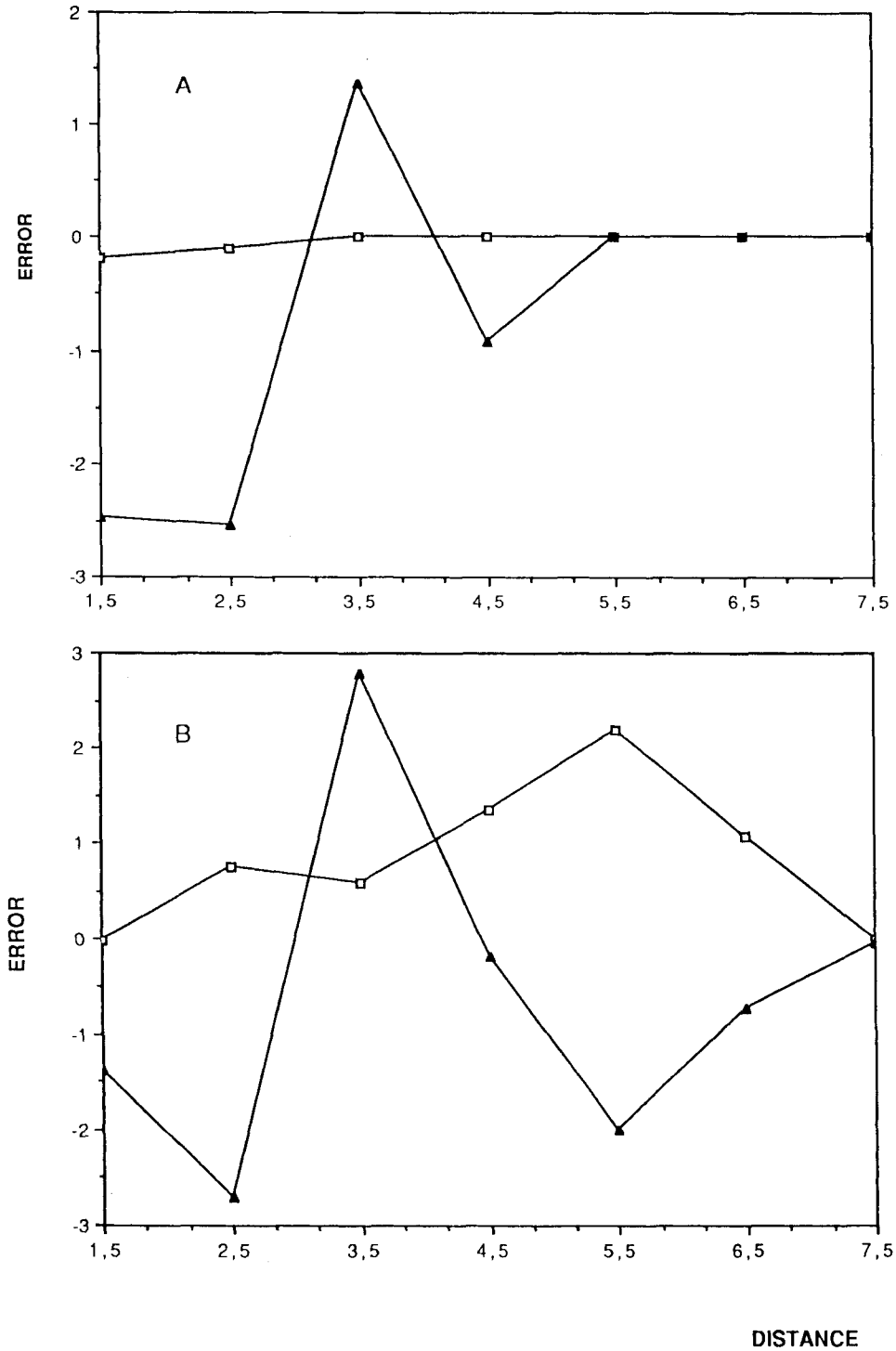


Figure 10. Absolute error in the area calculation for a two spheres model versus the distance (Å) between their centers. ▲ MSDOT; □ GEPOL. (A) van der Waals Molecular Surface; (B) Richards Molecular Surface.

roundings,[6,18] the type of molecular surface used will be vital in calculating the energy of interaction. Consequently, the choice of the surface can favor one conformation over others.

The difference between the WMS and the MS areas as a function of the rotation angle have been calculated according to the Connolly's method and our own. Figure 12 shows the comparison between both methods. The MSDOT results show big fluctuations which ratifies its oscillating nature[29] whereas the GEPOL results show in a better way the continuous variation that the molecular area has to go through when changing the rotation angle.

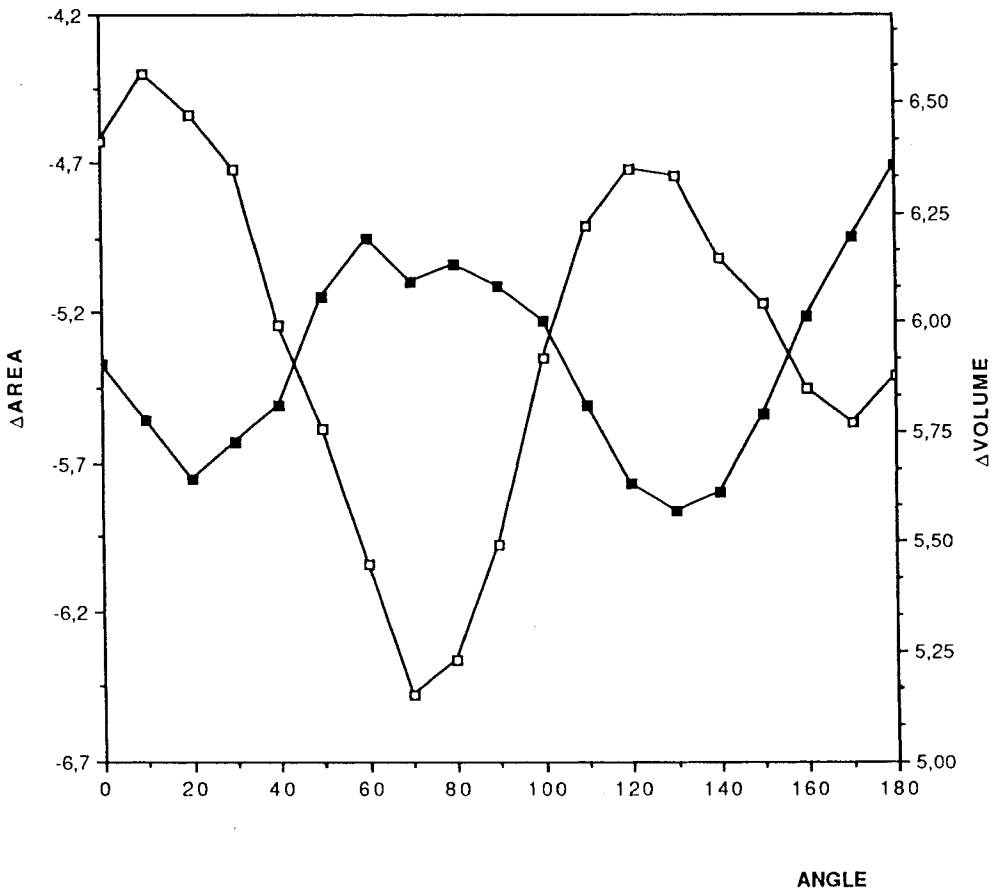


Figure 11. Difference between the MS and the WMS (MS minus WMS) for an internal rotation of the octanol molecule. □ area (Å^2); ■ volume (Å^3).

Fractality of the Surface

The fractality is a good index to characterize the irregularities of the molecular surface, which have been shown to be related to the reactivity of the biomolecules. In fact, a correlation between regions of high fractal index and those involved in protein-protein interactions has been found.[25-27] This index can be calculated using the formula of Lewis and Reed.[25]

$$2 - D = \text{dlog}\,(A)/\text{dlog}\,(R)$$

where D is the index of fractality and A is the area of the MS calculated for a given radius of probe R.

We have calculated the global fractality of the Retinol Binding Protein (RBP), only in order to show the capability of GEPOL working with structures of this size. This molecule is formed by 1438 nonhydrogen atoms, where we have centered the spheres with the following radii: 1.4 Å for the oxygens, 1.5 Å for the nitrogens, and 1.9 Å for the carbons. The geometry was taken from a molecular dynamic simulation[28] and we used a value of FRADIO equal to 0.6. Values for the area have been obtained for radii of the probe from 0.5 to 3 Å. In the Figure 13 are shown the values of fractality obtained for the different radii. This graph exhibits similar behavior to that obtained by Lewis and Reed for other proteins.

CONCLUSION

In conclusion, we have presented the procedure used by the program GEPOL to calculate the Molecular Surface starting from a set of spheres centered on the atoms or group of atoms of the molecule. GEPOL computes the MS by first searching the spaces inaccessible to the solvent and then filling them with a new set of spheres, this strategy being one of the fundamental differences between GEPOL and other methods.[12,13] The parameters used by this method are princi-

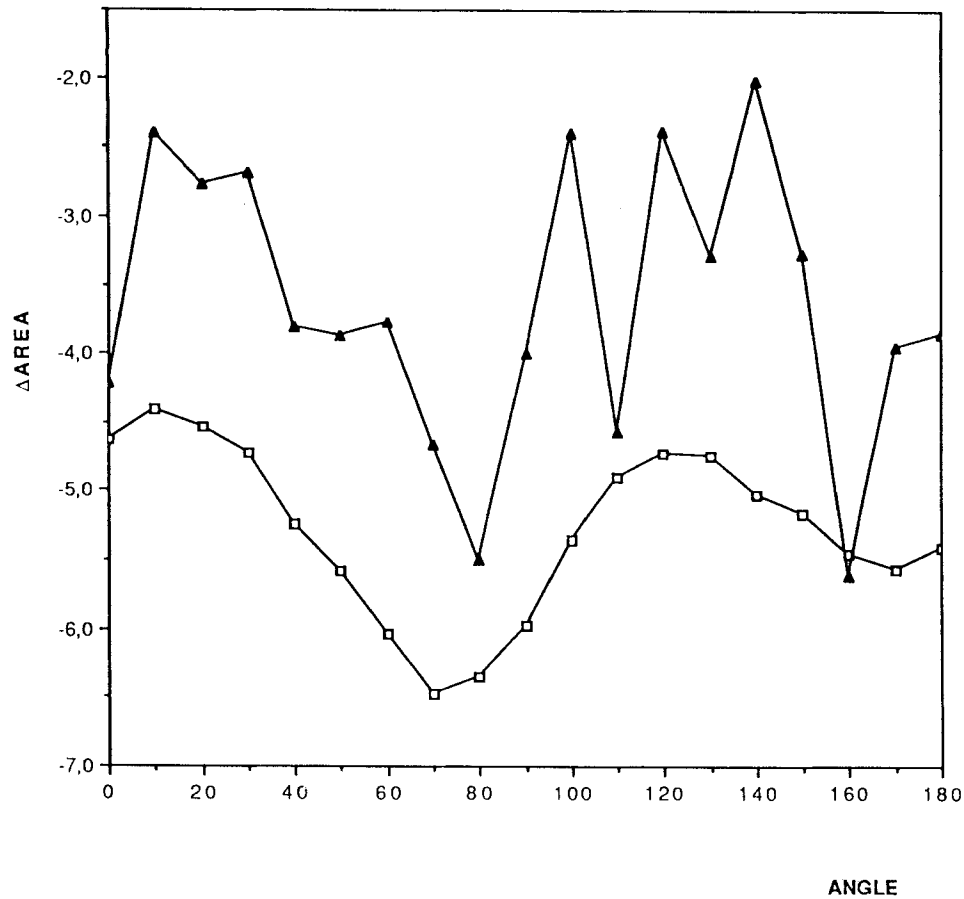


Figure 12. Difference between the MS and the WMS (MS minus WMS) for an internal rotation of the octanol molecule. ▲ MSDOT; □ GEPOL.

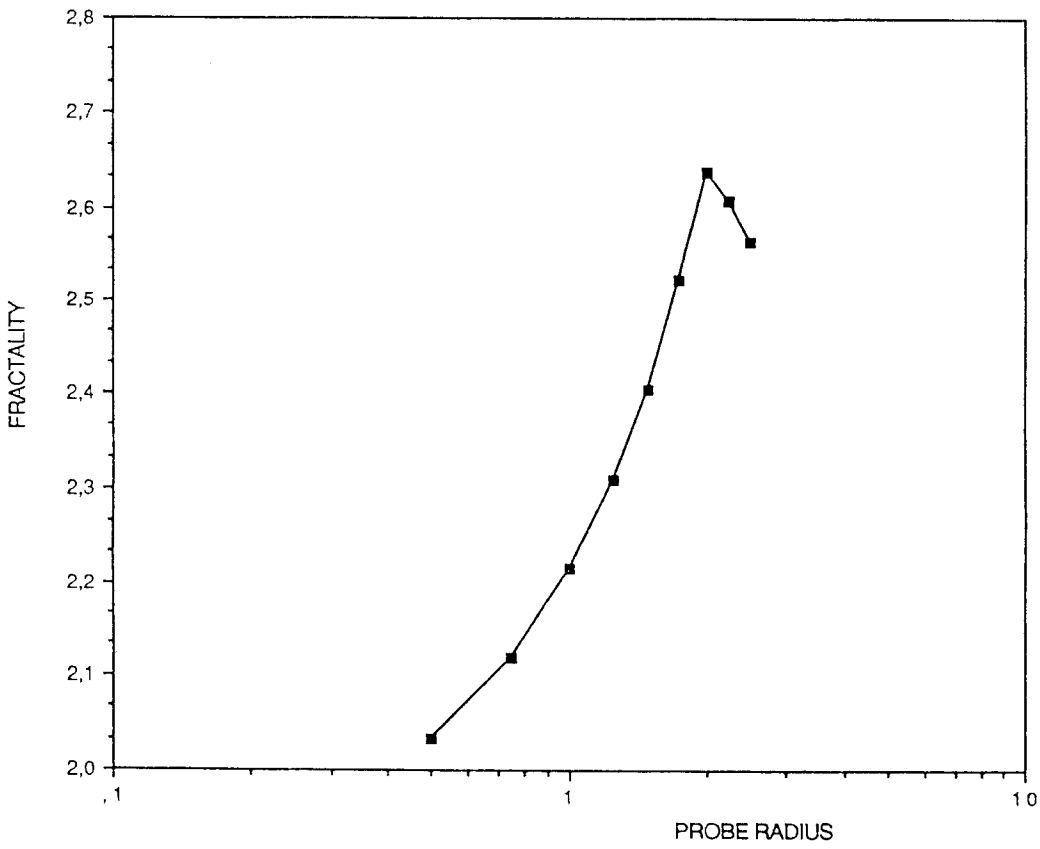


Figure 13. Index of fractality for the surface of RBP versus the probe radius in Å.

pally the radius of the probe sphere, and FRADIO which controls the creation process of new spheres. The latter parameter can take values between 0 and 1.0.

References

1. I. Langmuir, *Colloyd. Symp. Monogr.*, **3,** 3 (1925).
2. R.B. Hermann, *J. Phys. Chem.*, **76,** 2754 (1972).
3. S.C. Valvani, S.H. Yalkowsky, G.L. Amidon, *J. Phys. Chem.*, **80,** 829 (1976).
4. R.A. Pierotti, *Chem. Rew.*, **76,** 717 (1976).
5. S. Miertus, E. Scrocco, and J. Tomasi, *Chem. Phys.*, **55,** 117 (1981).
6. G. Floris and J. Tomasi, *J. Comp. Chem.*, **10,** 616 (1989).
7. L. Pauling, *Nature,* **161,** 707 (1948).
8. B. Linder, *Adv. Chem. Phys.*, **12,** 225 (1967).
9. B.J. Costa Cabral, D. Rinaldi, and J.L. Rivail, *C.R. Acad. Sc. Paris,* **298,** 675 (1984).
10. A. Horta, I. Fernandez-Pierola, *Macromolecules,* **14,** 1519 (1981).
11. A.Y. Meyer, *J. Chem. Soc. Perkin II,* 1161 (1985).
12. M.L. Connolly, *J. Appl. Cryst.*, **16,** 548 (1983).
13. M.L. Connolly, *Molecular Surface Program,* QCPE 429.
14. M.L.J. Drummond, *J. Chem. Phys.*, **88,** 5021 (1988).
15. H.R. Karfunkel and V. Eyraud, *J. Comp. Chem.*, **10,** 628 (1989).
16. B. Lee and F.M. Richards, *J. Mol. Biol.*, **55,** 379 (1971).
17. F.M. Richards, *Ann Rev. Biophys. Bioeng.*, **6,** 151 (1977).
18. J.L. Pacsual-Ahuir, E. Silla, J. Tomasi, and R. Bonaccorsi, *J. Comp. Chem.*, **8,** 778 (1987).
19. J.L. Pascual-Ahuir and E. Silla, Quantum Chemistry—Basic Aspects, Actual Trends, R. Carbó, Ed., *Studies in Physical and Theoretical Chemistry,* Elsevier, New York, 1989, Vol. 82, p. 597.
20. J.L. Pascual-Ahuir, E. Silla, J. Tomasi, and R. Bonaccorsi, *GEPOL QCPE Program 554* (1988).
21. O. Tapia, H. Eklund, and C.I. Bränden, in *Esteric aspects of Biomolecular Interactions,* G. Naray-Szabo and K. Simun, Eds., CRC Press, Boca Raton, FL, 1986.
22. O. Tapia and F. Colonna-Cesari, personal communication.
23. G.A. Arteca and P.G. Mezey, *J. Phys. Chem.*, **93,** 4746 (1989).
24. J.A. Pople and D.L. Beveridge, *Approximate Molecular Orbital Theory,* McGraw-Hill, New York, 1970.
25. M. Lewis and D.C. Rees, *Science,* **230,** 1163 (1985).
26. P. Pfeizer, U. Welz, and H. Wippermann, *Chem. Phys. Letters,* **113,** 535 (1985).
27. J. Åquist and O. Tapia, *J. Mol. Graph.*, **5,** 30 (1987).
28. J. Åquist, P. Sandblom, T.A. Jones, M.E. Newcomer, W.F. van Gunsteren, and O. Tapia, *J. Mol. Biol.*, **192,** 593 (1986).
29. A. Meyer, *Chem. Soc. Rev.*, **15,** 449 (1986).