Supramolecular Chemistry
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713649759

Implementation of the solvent effect in molecular mechanics. 1. Model development and analytical algorithm for the solvent-accessible surface area

Valentin Gogonea[a]; Eiji Ōsawa[a]
[a] Contribution from the Department of Knowledge-Based Information Engineering, Toyohashi University of Technology, Toyohashi, Japan

To cite this Article Gogonea, Valentin and Ōsawa, Eiji(1994) 'Implementation of the solvent effect in molecular mechanics. 1. Model development and analytical algorithm for the solvent-accessible surface area', Supramolecular Chemistry, 3: 4, 303 — 317
To link to this Article: DOI: 10.1080/10610279408034930
URL: http://dx.doi.org/10.1080/10610279408034930

SUPRAMOLECULAR CHEMISTRY, Vol. 3, pp. 303–317
Reprints available directly from the publisher
Photocopying permitted by license only

Implementation of the solvent effect in molecular mechanics. 1. Model development and analytical algorithm for the solvent-accessible surface area

VALENTIN GOGONEA and EIJI ŌSAWA*

Contribution from the Department of Knowledge-Based Information Engineering, Toyohashi University of Technology, Tempaku-cho, Toyohashi 441, Japan

(Received July 28, 1992)

A short review is given on the treatment of the solvent effect based on the continuum medium theory. A new modification is proposed which extends Still's recent breakthrough. Essential points of our modification are (1) separation of cavity and intrinsic terms and (2) relating the cavity term to the molecular volume. Several technical problems that occurred in the algorithm of solvent-accessible molecular surface-area calculations and that concerning the discontinuity in the surface-area derivatives are discussed. Results of critical tests on the extensively revised algorithm are described. The most crucial situation arises when the function for calculating surface-area must be switched, in the course of geometry optimization, from the two-spheres intersection case to three-spheres case. A test using a simple three-spheres model indicates that the switching of the function will not be hazardous.

I. INTRODUCTION

The success of molecular mechanics in dealing with energetics, structures and dynamics of molecules in the gaseous state[1] has led to early attempts to extend the scheme to solutions: Allinger and coworkers[2] and Meyer[3] used the reaction field theory to develop a solvent effect term by taking into account the dipole and quadrupole contributions. Recently, Still and coworkers[4] introduced contributions from hydrophobic interactions by means of the solvent-accessible surface area of the solute and have used a generalized Born equation instead of reaction field theory for evaluating electrostatic solvent-solute interactions. Still's method is, however, designed for aqueous solutions and contains several empirical parameters in order to accelerate computation.[5] In view of its potential in molecular simulation,[6] it seems worthwhile to expand Still's methodology to the general solution. The paper, the first part of our series, describes the general scheme and some technical improvements in the analytical calculations of solvent accessible surface area.

II. MOLECULAR AND SOLVENT-ACCESSIBLE SURFACES

The concept of *molecular surface* emerged as a practical tool in studying properties of the molecule considered as a rigid body, especially in molecular graphics. It is defined as the surface which includes the *essential part* of the electronic cloud around a molecule. Such a surface is generally constructed by covering a sphere over each atom in the molecule, the sphere having the van der Waals radius of the atom,[7,8] and its center being located at the atomic nuclear position. Heavy overlap among spheres leaves a geometrically complex but closed free surface consisting of a collection of partial spheres (Fig 1a). Such a molecular surface resembles closely to the space-filling model, thus providing an analogy with macroscopic objects, and has helped to develop techniques of molecular graphics analysis.

The idea of solvent-accessible surface area originates from the scaled particle theory for liquids and solutions,[9] where it was defined as the surface of a spherical cavity obtained by excluding all the solvent hard spheres whose centers lie within a certain radius measured from the center of a solute molecule (Fig 1b). This idea developed into the solvent-accessible surface by including the concept of molecular surface.

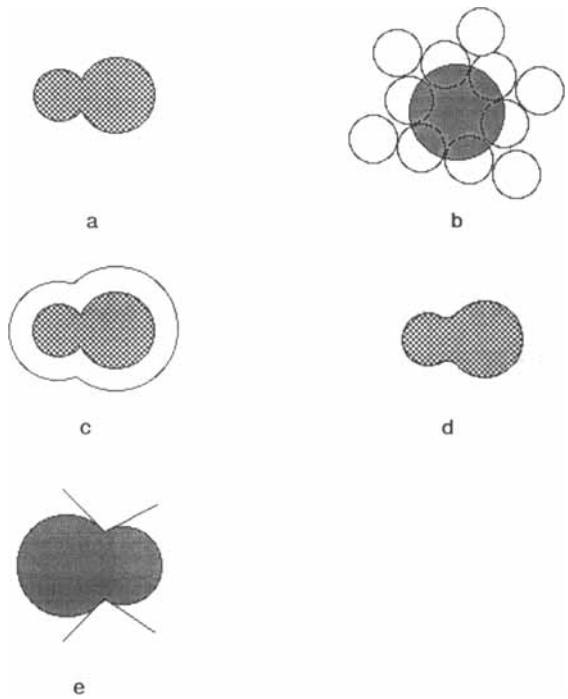


Figure 1 Various representations of solvent surfaces. (**a**) van der Waals surface, (**b**) cavity created by displacing solvent molecules from a spherical space in the liquid, (**c**) solvent-accessible surface, (**d**) solvent-rolled surface, (**e**) closed boundary made by intersection of two circles, the sudden change in tangent direction in the points where the two circles intersect each other marks the discontinuity in the first derivative.

At the moment, two definitions of the solvent-accessible surface are in use. The first one, introduced by Lee and Richards,[10] pertains to the surface of overlapping atomic spheres, each having as a radius the sum of the van der Waals radius of the atom and radius of a spherical model of the solvent (Fig 1c). This surface, also called *cavity surface*,[11] is a measure of the number of solvent molecules in the first solvation layer.

The second definition, also introduced by Richards,[12] corresponds to the surface drawn by a solvent hard sphere rolling over the solute molecule (Fig 1d). This surface consists of contact and reentrant surface.[13] The contact surface arises from that part of the van der Waals surface of atoms accessible to the solvent, while the reentrant surface has negative curvature and comes from the inward part of the solvent sphere when it is simultaneously in contact with more than one atom of the solute molecule.

Although the second definition gives a smooth surface and should be closer to the *real* molecular surface the solvent can access, the first one proved to be much easier to compute and as a consequence is widely used. The algorithm presented here is designed to compute the first type of surface area.

In order to compute the solvent-accessible surface area, a variety of algorithms of numerical, analytical or semianalytical type have been devised. Among them, the point-by-point scanning method[14] is widely used due to its simplicity.[15] Numerical algorithms for surface area computation are simple to program, but computationally expensive. For example, in the point-by-point scanning, a very fine mesh (0.005 Å) is needed in order to obtain acceptable accuracy. Besides, the optimization of a target function converges much slower when gradients obtained by numerical differentiation are used rather than by the analytical method.

The analytical computation of the surface area consisting of overlapping hard spheres proved a difficult task. Wodak and Janin[16] proposed an approximate analytical method based on a two-body distance function and corrective terms for multiple overlapping. The first exact solution to the analytical computation of the surface area of a fused hard spheres system was given by Connolly.[13] Shortly after him, Richmond proposed a similar algorithm for the computation of both area and its first derivative.[17] Later, Gibson and Scheraga[18] proposed analytical algorithms for computation of both surface area and volume of fused hard spheres. Connolly's algorithm is the most elaborate and computes the area obtained by rolling a probe sphere on the van der Waals surface. Here we used Richmond's method of analytical calculating molecular surface area.

A. Remark on Still's approach to solvent-accessible surface

Before describing our results, we briefly examine Still's approach of calculating the solvent-accessible surface area. He and his coworkers developed Wodak and Janin's approximate algorithm into a parameterized method.[19] Their main aim was to obtain an approximate surface area function which best fits the exact surface area and at the same time has continuous derivatives. The method is restricted to the computation of the surface area of overlapped hard spheres in water, for which parameters had been obtained by the least-squares fitting procedure, but it is highly effective in terms of computer time.

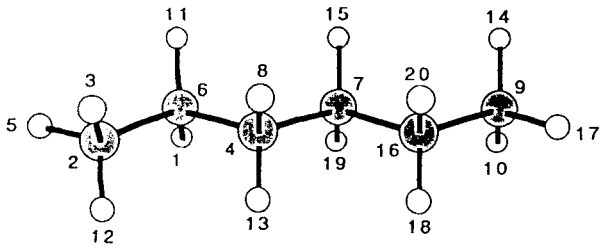
However, these authors state that the performance of their method is in general unpredictable. We carried out extensive tests but present here one example which seems to indicate a source of problem. The surface areas of all 12 unique conformers of *n*-hexane (**1**)[20] obtained by our modification (*vide infra*) of Richmond's method are compared with those obtained by Still's method in Table 1.

Although the results from the two algorithms agree well for the stable conformers (upper entries), the error

Table 1 Solvent accessible surface area (Å^2) of twelve unique conformers of n-hexane as computed by both exact and approximate methods

Conformer[a]	*AA*[b]	*EA*[c]	*Difference, %*[d]
AAA	314.64	313.559	0.3
AAG	303.16	309.352	−2.0
AGA	303.88	309.497	−1.8
AGG	286.56	303.101	−5.4
GAG	291.21	304.929	−4.5
GAG′	294.84	305.413	−3.4
GGG	273.81	296.768	−7.7
AGG′	284.86	302.110	−5.7
AGG′	287.15	302.616	−5.1
GGG′	267.50	294.165	−9.0
GGG′	277.31	298.606	−7.1
GG′G	266.95	293.796	−9.1

[a] A = anti, G = gauche, G′ = −gauche, **G** = distorted G (90°), G′ = distorted G′ (−90°). Conformers are listed in the order of decreasing stability. See: Ref. 17a.
[b] Approximate area computed with Still's method.
[c] Exact area computed with our modified algorithm based on Richmond's method.
[d] $[(AA - EA)/EA] \times 100$.



I

in the approximate method (mostly negative) increases for the less stable conformers (lower entries). The more strained conformers have high internal crowding, hence high degrees of overlapping among van der Waals surfaces of atoms. Apparently, errors due to the increased overlapping must have accumulated.

III. DEVELOPMENT OF MODEL

A. Elaboration of hydrophobic term

The continumm medium model of solvation, which we follow in this work, expresses the solvation free energy G_{sol} as the sum of energy terms which correspond to the physical processes taking place when a solute molecule passes from gas to solution:

$$G_{sol} = G_{cav} + G_{vdW} + G_{pol} \quad (1)$$

G_{cav} is the free energy required to create a cavity in the solvent to accommodate the solute, and G_{vdW} the van der Waals interaction energy (repulsion and dispersion) between solute and solvent. The sum of these two terms are traditionally termed as the hydrophobic interaction term since most of the past solvation studies have been directed to the aqueous solution. G_{pol} is the electrostatic interaction energy between solute and solvent (also called polarization energy), and is usually evaluated by means of the reaction field theory.[21]

In the solvation model proposed by Still and coworkers,[4] the combined cavity and van der Waals terms were evaluated in terms of the solvent-accessible surface area. These two terms are, however opposite in sign, and almost cancel each other. For polar solutes this sum is only a fraction of solvation energy, but will overcome the polarization term for non-polar solutes. Contrary to the generally held idea, G_{vdW} sometimes exceeds the polarization terms for polar solutes. For example, in the case of N-methylformamide, G_{pol} and G_{vdW} are 3.23 and 4.47 kcal/mol,[22] respectively. For these reasons, we chose to compute the cavity and van der Waals terms separately and with at least the same accuracy as in the evaluation of solvation free energy itself.

B. Evaluation of cavity free energy G_{cav}

The idea that a spherical cavity of suitable size should be created into the solvent in order to accommodate the solute proved to be useful in explaining the solubility of gases in liquids.[9] One way to evaluate G_{cav} is to relate it with the solvent surface tension and cavity size:

$$G_{cav} = 4\pi a \gamma_\infty \quad (2)$$

where a is the cavity radius and γ_∞ surface tension of solvent liquid.[23] However, the use of macroscopic surface tension in evaluating G_{cav} of a cavity of microscopic size and the difficulty of determining the cavity radius make eqn. (2) less useful for such a model to be effective in conformational analysis.

An alternative in evaluating the free energy for creating a cavity is to use scaled particle theory (SPT) which defines G_{cav} as the work required to exclude the centers of solvent molecules from a certain region in space (cavity, Fig 1b):[9]

$$\frac{G_{cav}}{kT} = -\ln(1-y) + \frac{3y}{1-y}\xi + \left[\frac{3y}{1-y} + \frac{9}{2}\left(\frac{y}{1-y}\right)^2\right]\xi^2 + \frac{yP}{\rho_s kT}\xi^3 \quad (3)$$

where $y = \pi\rho_s \sigma_s^3/6$ is the reduced number density, $\xi = \sigma/\sigma_s$, σ and σ_s are hard sphere diameters of the solute and solvent, respectively, ρ_s the number density of the solvent ($\rho_s = N_A/V_m$, N_A = Avogadro's number, V_m = solvent molar volume), k Boltzmann constant, T

absolute temperature and P pressure. The above expression for the evaluation of cavity energy, although very successful for hard sphere liquids, proved difficult to apply to real solutions due to the impossibility to account for the real shape of the solute.[24]

Finally, we recall here a suggestion of Oakenfull and Fenwick[25] who applied relation 4 for the volume changes associated with the creation of a cavity:

$$G_{cav} = -\frac{1}{\beta_T} V_{cav} + c \tag{4}$$

where β_T is the isothermal compressibility, V_{cav} cavity volume and c an integration constant.

We combine here the concepts of SPT theory with the above suggestion made by Oakenfull and Fenwick in evaluating G_{cav} and consider the cavity free energy as being *directly related* with the cavity volume, where the latter was defined as the volume enclosed in solvent accessible surface area. We are backed in our choice in evaluating G_{cav} by a recent molecular dynamic study made by Postma *et al.*[26] Postma and coworkers performed a molecular dynamic simulation to create cavities in water using simple point charge (SPC) water molcules and obtained the free energies for five cavities of different sizes. Using their numerical results we find an excellent correlation between the cavity volume (computed from thermal radius of cavity) and the cavity free energy G_{cav} (Fig 2a). This correlation can be fit to eqn. (5):

$$G_{cav} = \zeta V_{cav} + \zeta_0 \tag{5}$$

with a correlation coefficient 1.000 for $\zeta = 0.0426$ and $\zeta_0 = -0.1173$.

We should note that the lack of similar molecular dynamics simulations for other solvents, and consequently the impossibility to obtain cavity free energy parameters for other solvents, temporarily limits the applicability of this model to water solutions. Nonetheless, extension to other solvent systems should by no means be impossible (*vide infra*).

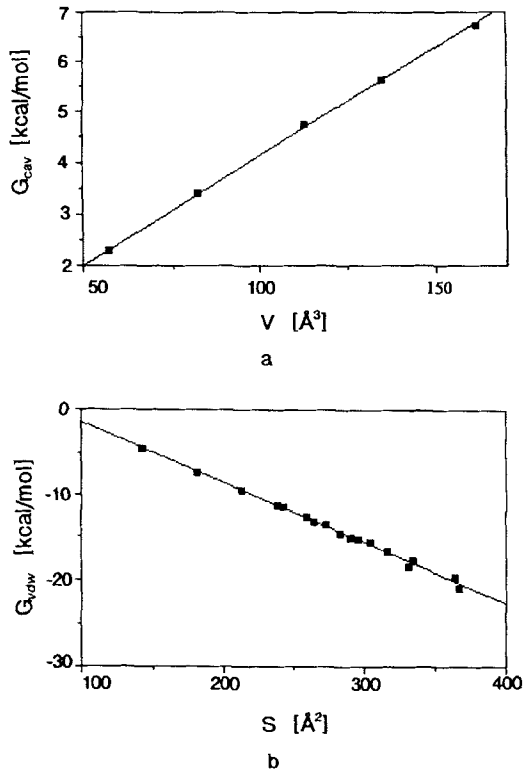


Figure 2 (**a**) Correlation of G_{cav} obtained from MD simulation with cavity volume; the dots on the correlation line mark the five cavities simulated by Postma *et al.*, (**b**) correlation of G_{vdw} with solvent-accessible surface area; the dots correspond to a series of seventeen linear and branched alkanes (from methane to 2,2,5-Me_3-hexane) for which G_{vdw} was evaluated from hydration free energies.

C. Evaluation of van der Waals interaction energy G_{vdW}

A common way to evaluate solute-solvent van der Waals interaction energy G_{vdW} is to use the Lennard-Jones potential[27] which contains both the repulsion and attraction between solute and the surrounding solvent. A good starting point in evaluating G_{vdW} seems to be the expression of van der Waals interaction energy for a solution of hard spheres:[27]

$$G_{vdW} = 4\pi\rho_s kT \int_{\sigma_0}^{\infty} u_{LJ} g(r) r^2 \, dr \tag{6}$$

where $\sigma_0 = (\sigma + \sigma_s)/2$, is solute-solvent distance at $u_{LJ} = 0$, $g(r)$ is the radial distribution function, r is the distance between the two interacting particles and u_{LJ} a potential function of the Lennard-Jones type containing r as the only variable. The integration is performed on the whole volume around the hard sphere solute of diameter σ.

Equation 6 could be transformed from a volume integral into a surface integral by means of Gauss theorem.[28,29] In this case we obtain:

$$G_{vdW} = \rho_s kT \sum_{i=1}^{N} \int_0^{\pi} \int_0^{2\pi} \int_{\sigma_i}^{\infty} \times u_{LJ}(r_p) g(r_p,\theta,\varphi) r_p^2 \sin\varphi \, dr_p \, d\theta \, d\varphi \tag{7}$$

where N is the number of atoms in the solute molecule and σ_i equals $r_i + (\sigma_s/2)$ where r_i is van der Waals radius of atom i. The integration is performed for all points in the polar coodinates r_p, θ, and φ which lie on exposed surface area of each atom of radius r_p. The disadvantage of this approach in evaluating G_{vdW} is that the surface integral has to be calculated by numerical integration which is very costly in computer

time when the optimization of solute geometry is required.

An alternative to eqn. (7) for evaluating G_{vdW} would be to treat every atom as a hard sphere solute and to evaluate its van der Waals interaction energy with the surrounding by means of eqn. (6). Finally, G_{vdW} for the polyatomic solute is computed from that part of interaction energies which correspond to the exposed surface area (solvent accessible surface area) of every atom:

$$G_{vdW} = \rho_s kT \sum_{i=1}^{N} \frac{S_i}{r_i^2} \int_{\sigma_i}^{\infty} u_{LJ}(r) g(r) r^2 \, dr \qquad (8)$$

where S_i is the exposed surface area of atom i with van der Waals radius r_i. Although this approach is less accurate than the eqn. (7), it has the advantage that the volume integral can be evaluated for every type of atoms *only once* and used as a parameter for any compound. Besides, this approach makes it possible to extend the solvent effect model to general solution.

In the evaluation of the integral of eqn. (8) the radial distribution function $g(r)$ could be taken as unity (homogenous approximation) or a simple functional form which meets boundary conditions (contact value, neighbour peaks) can be used in order to include partial information concerning solvent distribution around solute.

In a crude approximation, the whole integral can be substituted with a parameter determined empirically (*vide infra*). This option reduces the eqn. (8) to:

$$G_{vdW} = \sum_{i=1}^{N} v_i S_i + v_0 \qquad (9)$$

where v_i and v_0 are empirically determined parameters.

Floris and coworkers[30] computed G_{vdW} for hydrocarbons interacting with water using a model based on eqn. (7). Although, they did not compute the cavity term in order to compare their sum with the experimental solvation energies, we wish to remark that they found a very good correlation between the computed G_{vdW} and van der Waals surface area of hydrated hydrocarbons and that their finding supports the idea of evaluating G_{vdW} by means of eqn. (9).

We confirmed this linear relation between G_{vdW} and surface area by evaluating G_{vdW} for hydrated hydrocarbons by subtracting cavity energy from hydration free energy:[31]

$$G_{vdW} = G_{sol} - G_{cav} \qquad (10)$$

The cavity free energies G_{cav} were evaluated by means of eqn. (5) while the polarization term was set to zero. Figure 2b shows indeed a good correlation given by eqn. (9) between G_{vdW} computed with eqn. (10) and solvent-accessible surface area with a correlation coefficient of 0.996. Some of the earlier approaches have used a surface tension parameter for the whole surface area,[4] and others have used different surface tension parameters for every type of atoms.[5] Taking into account that different type of atoms have different Lennard-Jones potential parameters we attempted to determine parameters in eqn. (9) separately for C and H atoms. Results are satisfactory: $v_C = -0.1179$, $v_H = -0.0671$, $v_0 = 6.6463$ with the correlation coefficient of 0.9982. Parameters for heteroatoms could be determined in a similar way after adding the polarization term.

Using eqn. (9) in the evaluation of van der Waals interaction free energy first of all requires a high accuracy in the computation of atomic surface area, which will be discussed in some detail later in this paper.

D. Evaluation of polarization free energy G_{pol}

The polarization free energy is commonly evaluated by means of reaction field theory[21] or generalized Born (GB) method.[32] In generalized Born approach, the polarization free energy of a solute embedded in a spherical cavity surrounded by a solvent of ε dielectric constant is given as:

$$G_{pol} = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right)\left(\sum_{i=1}^{N} \frac{q_i^2}{a_i} + 2 \sum_{i=1}^{N-1} \sum_{j>1}^{N} \frac{q_i q_j}{d_{ij}}\right) \qquad (11)$$

where q_i and q_j are partial charges of solute atoms i and j and d_{ij} the distance between these atoms. For non-spherical molecules the cavity radius a_i of atom i is usually taken as the sum of van der Waals radius r_i and solvent radius. The use of GB approach would be straightforward provided that the cavity radii (a) could be determined in a simple way. In a recent implementation[4] of GB method, the cavity radii were evaluated in a rather sophisticated way (with a finite difference method) which allows making the radius a_i a function of molecular shape, but considerably increases the computing time.[5]

There is still another disadvantage of GB method: it does not take into account the polarizability of the solute. In order to cope with this problem, we choose to evaluate the polarization free energy G_{pol} by means of reaction field theory. Applying Onsager's idea that the solute molecule will be polarized by its own reaction field created in the surrounding dielectric, the polarization energy due to solute dipole is:

$$G_{dip} = -\frac{\kappa}{1 - 1\kappa} \frac{\mu^2}{a^3} \qquad (12)$$

where μ is the solute dipole moment, $\kappa = (\varepsilon - 1)/(2\varepsilon + 1)$, $1 = 2\alpha/a^3$, α being the molecular polarizability

of the solute. This quantity is usually taken as a function of refractive index n_d, namely $\alpha = 2(n_d^2 - 1)/(n_d^2 + 2)$.

Abraham and Bretschneider[33] have proposed to use the quadrupole term along with dipole term in order to account for the polarization energy of polar solute molecules with symmetrical charge distribution (with zero dipole moment):

$$G_{qdp} = -\frac{3\kappa}{5-\kappa}h \tag{13}$$

$$h = \frac{3}{2a^5}\sum_{i,j=x,y,z}^{i\neq j}[4Q_{ii}^2 + 3(Q_{ij} + Q_{ji})^2 - 4Q_{ii}Q_{jj}] \tag{14}$$

where Q_{ii} and Q_{ij} are axial and perpendicular components of quadrupole moment, respectively. Abraham and Bretschneider also took into account of dipole-dipole and dipole-quadrupole interactions between solvent and solute in order to incorporate Onsager's theory in highly polar media.[33] Thus, the expression for dipole-dipole and dipole-quadrupole interaction free energy G_{DQ} obtained is:

$$G_{DQ} = -bf[1 - \exp(-bf/16RT)] \tag{15}$$

where,

$$f = [(\varepsilon - 2)(\varepsilon + 1)/\varepsilon]^{\frac{1}{2}} \tag{16}$$

$$b = \frac{3}{a^3}\left[\frac{2V_mRT}{\pi}\right]^{\frac{1}{2}}\left[\mu^2 + \frac{3Z}{2a^2}\right]^{\frac{1}{2}} \tag{17}$$

and R is the gas constant and Z is the summation of eqn. (14) which is sometimes called the solute quadrupole moment.

Finally, the polarization free energy we used in the development of this solvent effect model is given as:

$$G_{pol} = G_{dip} + G_{qdp} + G_{DQ} \tag{18}$$

IV. SOME TECHNICAL PROBLEMS AND THEIR SOLUTIONS

A. Incorporation of solvent effect term into molecular mechanics geometry optimization

If the solvent effect is to be incorporated into molecular mechanics force field through a *linear approximation*, the total energy E of a solute molecule in solution will be given as the sum of the steric energy due to the gas state E_g and the steric energy due to solvent effect E_{sol}:

$$E = E_g + E_{sol} \tag{19}$$

Consequently, the gradient (∇) and hessian (Δ) matrices of the total steric energy with respect to the freedom of atomic movements will have linear expressions too:

$$\nabla E = \nabla E_g + \nabla E_{sol} \tag{20}$$

$$\Delta E = \Delta E_g + \Delta E_{sol} \tag{21}$$

The steric energy due to the solvent force field could be expressed as:

$$E_{sol} = E_{cav} + E_{vdW} + E_{pol} \tag{22}$$

where steric energy terms E_{cav}, E_{vdW} and E_{pol} have similar expression as those given in eqns (5), (9) and (18), but they refer to individual molecules, while G_{cav}, G_{vdW} and G_{pol} are statistical averaged quantities. Thus, the evaluation of the effect of the solvent requires the obtaining of the first and the second order derivatives of these energy terms.

We ignored here the effect of the solvent on the partition function of the solute and assumed that only the conformers population changes as passing from gas to solution. Consequently, the solvation free energy is evaluated by taking into account the low energy conformers obtained by mapping steric energy surface:[34]:

$$G_{sol} = \sum_{i=1}^{n_{cf}} p_i E_{sol}^i \tag{23}$$

where n_{cf} is the number of significantly populating conformers, p_i the population of conformer i ($\sum_{i=1}^{n_{cf}} p_i > 0.99$) and E_{sol}^i the solvent effect steric energy. Note that p_i is not available from the beginning and is determined by iteration using the gas state values as a starting value.

B. Continuity in the surface

Van der Waals *molecular surface*, when described as a function of Cartesian coordinates, has discontinuities in the first partial derivatives on the circles of intersection.[35] Figure 1e illustrates the case in two dimensions: the function describing a closed boundary has discontinuities in the first derivative at the intersection of the two circles.

However, unlike the molecular surface, the *surface area* of a body composed of intersecting two hard spheres has continuous derivatives. In this case, the surface area S is given as:

$$S = 2\pi\sum_{i=1}^{2} r_i(r_i + t_i) \tag{24}$$

$$t_i = (d^2 + r_i^2 - r_j^2)/2d \qquad i,j = 1,2; \quad i\neq j \tag{25}$$

For symbols, see Fig 3, d is the distance between the centres of spheres i and j. Setting the radii of both atoms to unity ($r_1 = r_2 = 1$) does not reduce the generality of the above relationships, but simplifies the expressions for the first and second order derivatives with respect to the Cartesian coordinates of spheres centres. It is also convenient here to use the 3 by N matrix notation of Cartesian coordinates for N atomic

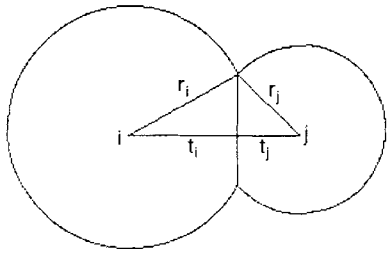


Figure 3 The definition of quantities which appear in eqns (24) and (25). These quantities are used to compute the surface area of a body made by intersecting two spheres.

centres:

$$\boldsymbol{x} = \begin{vmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ & \cdots & \\ x_i & y_i & z_i \\ & \cdots & \\ x_N & y_N & z_N \end{vmatrix}$$

where x_i, y_i and z_i are the Cartesian coordinates of atom i. Then, the derivatives of surface area can be written down as the following matrix notation:

$$\frac{\partial S}{d\boldsymbol{x}} = \frac{2\pi}{d}\boldsymbol{x} \tag{26}$$

$$\frac{\partial^2 S}{\partial \boldsymbol{x}^2} = \frac{2\pi}{d}\left(\eta - \frac{1}{d^2}\boldsymbol{x}^2\right) \tag{27}$$

where $\eta = 1$ if both the first and second differentiations are done with respect to the same coordinates, or $\eta = 0$ otherwise. Equations (26) and (27) show a discontinuity in the first and second order derivatives only for $d = 0$ (i.e. when the centres of two spheres coincide).

C. Gauss-Bonnet theorem

The calculations of solvent-accessible surface area S_i of sphere i in a system formed by overlapping spheres can be accomplished by means of the Gauss-Bonnet theorem:[36]

$$S_i = r_i^2\left[2\pi\chi + \sum_{\lambda=1}^{v}\Omega_{\lambda,\lambda+1} + \sum_{\lambda=1}^{v}\int_{C_\lambda} k_\lambda ds_\lambda\right] \tag{28}$$

where v is the number of regular curves C_λ of length s_λ (arc of intersection circle, Fig 4) which make a closed boundary around S_i, k_λ is the geodesic curvature of C_λ, $\Omega_{\lambda,\lambda+1}$ is the external angle between the vectors tangent to two connected arcs C_λ and $C_{\lambda+1}$ (Fig 4). The solvent-accessible surface area of a certain atom could be made of one or more separate pieces, each of them being enclosed in an independent closed boundary. Two less the number of these non-adjacent parts gives the Euler-Poincare chracteristic χ. The derivation of eqn. (28) as well as the details of the computation of all quantities appearing in this equation are described by Richmond.[17]

D. Richmond's algorithm

Computation of surface areas is straightforward as long as only two atomic spheres are involved in an intersection (two-spheres case; this is done by eqns (24) and (25)). Difficulty arises when more than two spheres overlap (when eqn. (28) applies). In a three-spheres case (i,j,k), two circles of intersection, one between i and j (hereafter designed as i/j) and the other between i and k (i/k), cross each other.

Computation of the solvent-accessible surface area of the higher-spheres case proceeds through the following steps:

(1) All of the atomic spheres which overlap with the sphere of atom i are found and sorted in the decreasing order of the radius of intersection circle with i. Assume that these spheres come from atoms $j, k, l, \ldots$.

(2) In order to simplify the computation, the molecule is rototranslated so that atom i is placed at the origin of the coordinate system used, and atom j is placed on the positive side of z axis.

[a] From among atoms $k, l, \ldots$, all those atomic spheres $m, n, o, \ldots$ which overlap both i and j are found. The plane of intersection circle between atoms i and m (i/m) will cut out a free arc on the intersection circle i/j (Fig 5a, the arc which borders the shaded area is inside atom m and the other arc is outside of atom m).

[b] Every intersection plane between atom i and other atom ($m, n, o, \ldots$ except atom j) will divide the intersection circle between i and j into two arcs. After considering all the intersection planes between atom i and atoms $m, n, o, \ldots$ the arcs which belong to i/j intersection plane and are not overlapped by other atoms are called free arcs (C_λ, Fig 5b, in

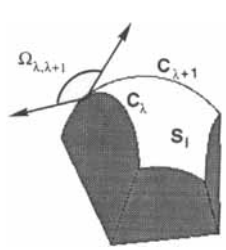


Figure 4 A piece of solvent-accessible surface S_i enclosed in a boundary of C_λ regular curves. $\Omega_{\lambda,\lambda+1}$ is the external angle between the tangents of connected curves. The external angle $\Omega_{\lambda,\lambda+1}$ along with C_λ curves are used in computing S_i by means of eqn. (28).

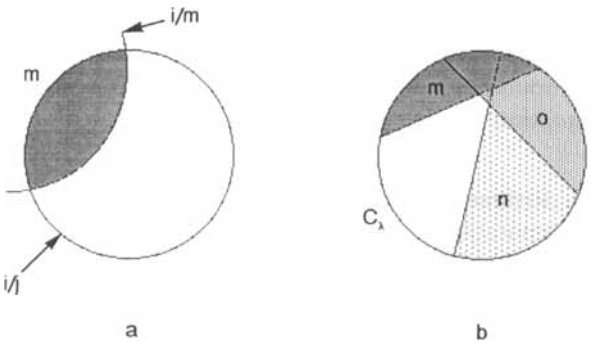


Figure 5 Definition of free arcs C_λ. (**a**) the arcs defined by cutting the i/j intersection circle with the i/m cutting plane, (**b**) the free arc C_λ which remains after taking into account all cutting planes on i/j intersection circle.

this case there is only one free arc). If there are more than one free arc, they belong to different closed boundries (*vide infra*).

[c] The external angle $\Omega_{\lambda,\lambda+1}$ is calculated at the two end-points of C_λ (Fig 4).

(3) The Euler-Poincare characteristic χ is obtained by a topological procedure. Every closed boundary makes a polygon of free arcs. The procedure checks the neighbourhood of every polygon and gives a *connectivity error* message if open boundaries are found.

(4) The total exposed area on atomic sphere i is obtained by summing up patches of free surfaces.

E. Cautionary note on numerical consequences

Care must be taken on the numerical aspects of the surface area computation by Richmond's algorithm. Errors arising from truncation in the computation of arc cosine function,[37] used in the calculation of arc length, are found sometime significant. These errors appear most often when spheres overlap in such a way that the cutting lines made by, for example, planes i/m and i/n on the i/j plane are close and parallel (Fig 6).

A *real* molecule with a *real* geometry was selected to illustrate how our modified algorithm manages to avoid the above shortcoming. Table 2 shows contrasting areas computed by using Richmond algorithm and by our modification for three atoms in n-heptane in all-anti conformation (**2**), Richmond algorithm fails in area computation of three atoms, C4, H8 and H13. The failure also appears at atoms C16, H18 and H20 which are symmetrically placed with respect to the middle of the carbon chain.

Let's consider the overlap between atomic spheres C4 and H8. Figure 7a shows the 4/8 intersection circle as being cut by other nine intersecting planes. The cutting lines made by 4/16 and 4/20 planes are parallel and close. The arcs cut out by these lines have almost identical end-points (10^{-8} radian apart).[38]

Our modified algorithm avoids the connectivity error by using the following set of conditions:

$$L^k_{i/j} < L^l_{i/j} \tag{29}$$

where $L^k_{i/j}$ is the length of the arc made by atom k on the i/j intersection circle (in this case, $i = 4$, j, k, l belongs to the set of indices $\{8, 16, 20\}$ with $j \neq k \neq l$). The algorithm swaps the end-point values when the arcs are found to be in the wrong sequence. For example, the end-point values are swapped between atoms 16 and 20 as shown in Fig 7b.[39]

In order to detect and treat these special cases, new quantities are needed such as the three-plane intersection

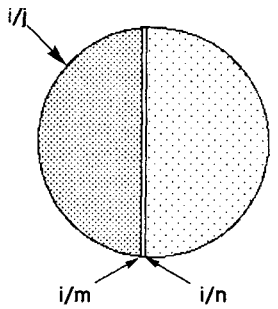
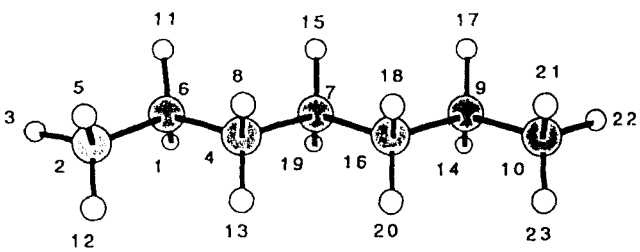


Figure 6 The case where the numerical errors appear in the calculation of solvent-accessible surface area with Richmond's algorithm. When the intersection circle planes i/m and i/n cut the intersection circle i/j, and the resultant cutting lines are close and parallel, numerical difficulty arises.

Table 2 Computed atomic surface area S_i (Å^2) for several atoms of n-heptane in the global minimum conformation **2**

Atom	*EP*[a]	S_i *(Richmond)*	S_i *(this work)*	*Difference* %
C4	–	5.312	1.616	228.7
H8	1	17.580	14.981	17.3
H13	1	21.277[b]	14.981[b]	42.0
S[c]		363.393	338.209	7.4

[a] Euler-Poincare characteristic.
[b] H13 should have the same S_i value as H8 due to symmetry. This premise is realized in the present modification.
[c] Solvent accessible surface area of the whole molecule (in Å^2).



2

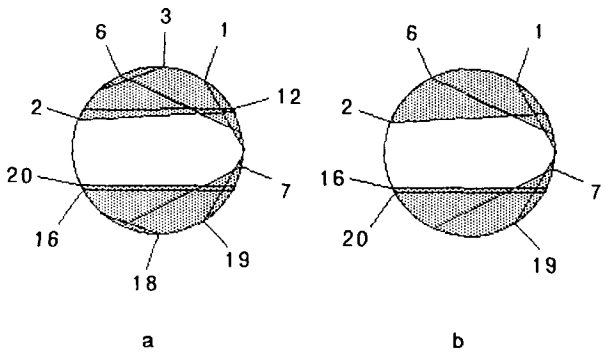


Figure 7 Intersection circle made by overlapping atoms C4 and H8 in **2**. Lines numbered k is the cutting line of the $4/k$ intersection circle plane with plane 4/8. (**a**) Richmond's algorithm, (**b**) modified algorithm.

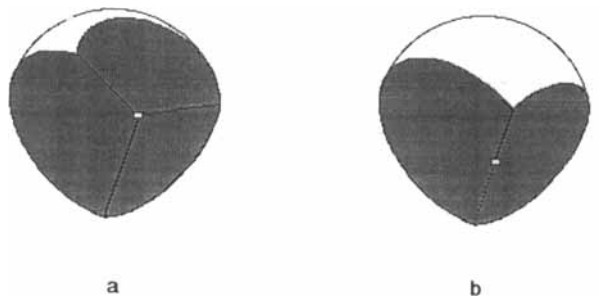


Figure 8 (**a**) Three planes intersection point, (**b**) the middle of the segment which connects the points made by three spheres intersection. These quantities are computed in order to check for cases where numerical difficulties appear and for further developments (volume computation).

points (Fig 8a) and the middle of the segment which connects the two points made by three-spheres intersection (Fig 8b).

New computer codes are added in order to: a) facilitate computation of the middle of the segment which connects the two points made by three-spheres intersection, b) avoid the potentially troublesome intersection cases as those in Fig 6, and c) avoid completely buried atoms, intersection planes and cutting planes which are buried by other planes.

Computationally expensive calculations of three-plane intersection points greatly affect the computer time, specially in large molecules. However, these new enhancements are useful also in analytical computation of the cavity volume and the first and second order derivatives of both surface area and volume. Thus, the middle of the segment and the three-plane intersection points (Fig 8) are also utilized in setting the integral limits for volume calculation.

Having proceeded to the point where solvent-accessible surface area can be accurately calculated by using Richmond's algorithm, we proceeded to deriving the first and second order derivatives of surface area.

V. ANALYTICAL FIRST AND SECOND ORDER DERIVATIVES OF MOLECULAR SURFACE AREA

The pioneeering work in obtaining the analytical first derivative for the surface area of overlapping spheres with unequal radii belongs to Richmond.[17] Our interest in extending this work to the analytical computation of volume and its derivatives imposed us the task of modifying the Richmond's original algorithm. Consequently, even the formulae that we use for the first derivatives of surface area somewhat differ from those proposed by Richmond. The main difference lies in the calculation of free arcs which contribute to the evaluation of the last term in eqn. (28).

In a hypothetical example, Fig 9 shows the free arc (the border of unshaded area) made by two cutting planes i/k and i/l on the intersection circle i/j. Using the quantities defined in Fig 9, we can write down the expression for the free arc length s_λ and its first derivatives as:

$$s_\lambda = \omega r_{i/j} \tag{30}$$

$$\omega = (\omega_l^0 - \omega_l) - (\omega_k^0 + \omega_k) \tag{31}$$

$$\frac{\partial \omega}{\partial x} = \frac{\partial(\omega_l^0 - \omega_k^0)}{\partial x} - \frac{\partial(\omega_k + \omega_l)}{\partial x} \tag{32}$$

where $r_{i/j}$ is the radius of i/j intersection circle. The angles ω_k^0 and ω_l^0 are related with the position of the centres of spheres k and l. ω_k and ω_l correspond to the end-points of the arcs cut out by intersection planes i/k and i/l on i/j intersection circle. The Richmond's method evaluates the free arc length using an expression similar to eqn. (31), but in calculating the derivatives only the second term of eqn. (32) is taken into account. Indeed, this last term has the largest contribution to the free arc derivative, but the first

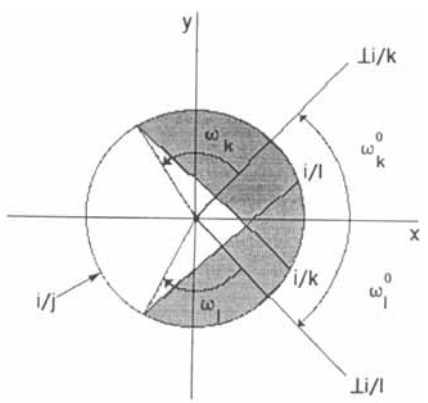


Figure 9 Model for computation of free arc length. $\perp i/k$ belongs to i/j intersection plane, passes through the centre of i/j intersection circle and is perpendicular on the line made by i/k cutting plane on i/j intersection circle. $\perp i/l$ is defined in a similar way.

term, although very small in many cases, increases its contribution to the total differential[40] of surface area ($\nabla S\delta x$, where ∇S is the first derivative matrix of S and δx is the increments in coordinates) in very complex systems of hard spheres.

In order to demonstrate this point, the total differential of the surface area was calculated using the first derivatives for a set of compounds involving a variety of structural features. Results are summarized in Tables 3 and 4. Note that, as the increment in coordinates (δx) decreases, the total differential should approach the increment in surface area (δS), and $\delta S - \nabla S\delta x$ should vanish as δx goes towards zero. Some of the compounds used for testing were generated using MACROMODEL atom types,[41] while the rest of them include all atoms explicitly. The last column of Table 3 lists the contribution (in %) of the derivatives of ω^0 angle to the total differential.

When this contribution doesn't vanish, the total differential of surface area computed by Richmond's original program (lower line in each entry) does not converge to δS as the step of integration decreases. When the computations were repeated after the derivatives due to ω^0 angle were removed in our program, it gave identical results for all compounds with Richmond's. Table 3 reveals that our modifications are valid, the total differential area always approaching to δS whenever δx is decreased.

Table 3 Comparison of the first order derivative of surface area (total differential) as computed by our modification (upper line) and by the original Richmond algorithm[a] (lower line)

		δx[c]					
Structure	NA[b]	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	$\delta\omega^{0}$[d]
Adamantane	26	7.50	0.75	0.08	0.01	0.00	0.0
		7.50	0.75	0.08	0.01	0.00	
Adamantane[e]	10	−9.02	−0.83	−0.08	−0.01	0.00	12.5
		4.18	12.64	13.46	13.56	13.57	
Bicyclo[2,2,1]heptane	19	−4.35	−0.41	0.04	0.00	0.00	0.0
		−4.35	−0.41	0.04	0.00	0.00	
Ribose	20	12.12	1.19	0.12	0.01	0.00	−1.4
		10.56	−0.22	−1.30	−1.41	−1.45	
Oxadecalin[e]	20	14.81	1.73	0.17	0.02	0.00	36.6
		83.06	62.23	60.13	59.83	73.69	
Arachidonic Acid[e]	23	−5.41	−0.51	−0.05	−0.01	0.00	0.6
		−4.78	0.08	0.55	0.60	0.66	
Glucopyranose	24	7.88	0.79	0.08	0.01	0.00	−5.4
		2.29	−4.63	−5.38	−5.45	−5.7	
Cholestanol[e]	29	24.96	2.5	0.25	0.03	0.00	−25.5
		−0.98	−23.10	−25.81	−26.25	−25.36	
Peniciline	30	9.83	0.74	0.07	0.01	0.00	−0.8
		8.89	−0.12	−0.79	−0.86	−0.75	
Porphyrin	38	10.28	1.06	0.11	0.01	0.00	0.3
		10.60	1.35	0.40	0.3	0.29	
Ala_6 Helix[e]	41	−261.1	−34.49	−2.74	−0.19	−0.02	−4.3
		−256.1	−39.83	−6.15	−1.50	−2.70	
Ala_6 Extended[e]	41	44.59	4.96	0.49	0.04	0.01	−33.1
		7.29	−28.40	−34.20	−30.60	−46.63	
Ala_8 Reverse Turn[e]	53	−26.49	−0.07	0.00	0.00	0.00	9.3
		−14.71	10.20	10.31	9.53	10.05	

[a] $(\delta S - \delta S^*)/\delta S^*$ in %. $\delta S = S - S_0$ is the difference in surface area between two consecutive iterations, $\delta S^* = \nabla S(x - x_0)$ is the total differential of surface area, ∇S is the first order derivatives matrix of surface area. S and S_0 are surface areas computed for x and x_0, where x is the matrix of atomic coordinates.

[b] Number of atoms (see also note[e]). van der Waals radii are taken from Gavezzotti, A.; *J. Am. Chem. Soc.* **1983**, *105*, 5220–5225, (except for that of sulfur which is taken from Ooi, T. *et al. Proc. Natl. Acad. Sci. USA* **1987**, *84*, 3087–3090.

[c] Integration increment in Å, coordinates are increased by δx for half of the atoms and decreased by δx for the rest.

[d] $\delta\omega^0 = \nabla\omega^0(x - x_0)$, *contribution of* ω^0 *in computing the total differential of surface area.* Unit is in %. $\nabla\omega^0$ is matrix of the first order derivatives of ω^0.

[e] These compounds are generated using MACROMODEL atom types, see Mohamadi, F.; Richards, N.G.J.; Guida, W.C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W.C.; *J. Comput. Chem.* **1990**, *11*, 440–467.

Table 4 Comparison of the second order derivative of surface area (quadratic approximation) as computed by full matrix approximation (upper line) and by block diagonal approximation (lower line)

		δx[b]			
Structure	*NA*[a]	10^{-1}	10^{-2}	10^{-3}	10^{-4}
Adamantane	26	−33.7[c]	0.0	0.0	0.0
		−65.7[d]	−3.9	−0.4	0.0
Adamantane[e]	10	−22.5	0.0	0.0	0.0
		135.0	−1.8	−0.2	0.0
Bicyclo[2,2,1]heptane	19	7.5	0.0	0.0	0.0
		−6.5	−0.9	−0.1	0.0
Ribose	20	−13.7	0.2	0.0	0.0
		−7.6	1.3	0.1	0.0
Oxadecalin[e]	20	−21.5	−2.2	0.0	0.0
		5.3	3.1	0.6	−0.3
Arachidonic Acid[e]	23	0.7	0.0	0.0	0.0
		−32.5	−1.8	−0.2	0.0
Glucopyranose	24	−30.5	0.0	0.0	0.0
		−82.7	−6.7	−0.7	−0.1
Cholestanol[e]	29	8.9	−0.1	0.0	0.0
		11.4	−0.5	0.0	0.0
Peniciline	30	−12.9	2.4	0.0	0.0
		−12.6	2.4	0.0	0.0
Porphyrin	38	−7.2	−0.2	0.0	0.0
		2.1	1.4	0.2	0.0
Ala_6 Helix[e]	41	−10.5	−0.1	0.0	0.0
		20.9	64.2	−8.6	−0.7
Ala_6 Extended[e]	41	−325.6	12.0	0.0	−0.2
		227.1	−340.2	8.4	0.5
Ala_8 Reverse Turn[e]	53	8.8	−26.3	−0.1	0.0
		−69.4	−36.8	−0.9	−0.1

[a] Number of atoms.
[b] Integration increment in Å, all coordinates for half of atoms were increased by δx and decreased by δx for the remaining atoms.
[c] In full matrix approximation, $(\delta S - \delta S^{**})/\delta S^{**}$ in %. $\delta S = S - S_0$ is the difference in surface area between two consecutive iterations, $\delta S^{**} = \nabla S(x - x_0) + 0.5\Delta S_{FM}(x - x)^2$, ∇S and ∇S_{FM} are first and second order derivatives matrices of surface area (FM = full matrix). S and S_0 are surface areas computed in x and x_0, where x is the atomic coordinates matrix.
[d] In block diagonal matrix approximation, $(\delta S - \delta S^{**})/\delta S^{**}$ in %. $\delta S^{**} = \nabla S(x - x_0) + 0.5\Delta S_{BD}(x - x_0)^2$, ∇S and ΔS_{BD} are first and second order derivatives matrices of surface area (BD = block diagonal matrix).
[e] Compounds generated with MACROMODEL atom types.

A. Test with a simple method of three-spheres case

A prerequisite of using analytical derivatives in Newton-Raphson optimizations is that they have to be continuous functions in the Cartesian coordinates. This usually happens with the potential functions of the molecular mechanics force field, but the analytical expression of surface area function (eqn. (28)), changes with the geometry of molecule. The 'transition' point, where the function of surface area changes, e.g. from a two-spheres function to a three-spheres function, could be discontinuous for the analytical derivatives.

The behaviour of the derivatives around these points is studied using a simple arrangement of three spheres intersecting each other, where the position of transition points could be anticipated. One such transition point appears when the surface area function switches from a two-spheres case function (intersection circles do not intersect each other) to a three-spheres case function (Fig 10a–c). Let's first consider the model depicted in Fig 10a. This system is essentially of the two-spheres intersection type and its surface area can be readily computed using eqns (24) and (25). As sphere A is moved to the right along the x coordinate, the function which describes the surface area will change when the x coordinate of sphere A reached a transition point where all three-spheres intersect simultaneously (Fig 10b). At this point the surface area function switches from two-spheres case function to three-spheres one.

Figure 10c shows a similar situation, although in this case, sphere C is rotated, starting from Fig 10a, around sphere A. Two transition points could be anticipated in this case. The first one occurs at $\phi = 2\pi/3$ and the second one occurs for $\phi = 0$, where ϕ is the angle defined as shown in Fig 10c.

The behaviour of derivatives around these transition points can be monitored by the error in the surface area evaluation between two consecutive iterations. For this purpose, we define here a target function $(\delta S - \delta S^{**})/\delta S^{**}$ where δS is the increment in solute surface area at two consecutive iterations. δS^{**} can also be evaluated analytically by using the quadratic approximation of a Taylor series, $\delta S^{**} = \nabla S \delta x + 0.5\Delta S \delta x^2$. If the surface area function is differentiable around the point where computations are performed, then the quadratic approximation (δS^{**}) is a good approximation of the change in surface area. Evidently, the lack of the differentiability in this point could give large differences between the quadratic approximations (δS^{**}) and the difference between the surface area (δS) at two consecutive iterations.

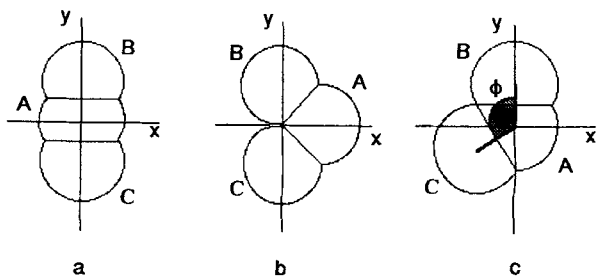


Figure 10 Arrangements of model spheres for checking the behaviour of the derivatives at transition points. (**a**) spheres arrangement for which surface area is computed with a two-spheres intersection case function, (**b**) transition point obtained by moving sphere A along x axis, (**c**) transition point obtained by moving sphere C around sphere A.

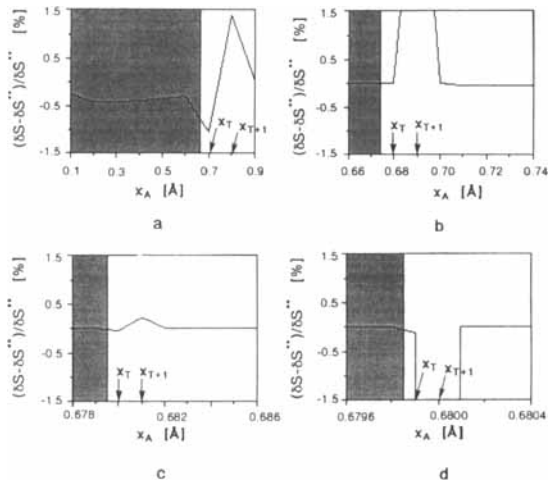


Figure 11 Computational tests for the arrangement of spheres shown in Fig 10b. δS is the difference in surface area between two consecutive iterations, δS^{**} is the quadratic approximation of surface area, x_T is the value of x_A at transition point, here the surface area is computed with a three-spheres intersection case function and the quadratic approximation δS^{**} is computed with the derivatives based on two-spheres intersection case function, x_{T+1} is the value of x_A which follows x_T. The interval scanned by x_A is gradually lowered in order to see if the discontinuity in derivatives is wide or sharp. (**a**) $0.1 < x_A < 0.9$, (**b**) $0.66 < x_A < 0.74$, (**c**) $0.678 < x_A < 0.686$, (**d**) $0.6796 < x_A < 0.6804$.

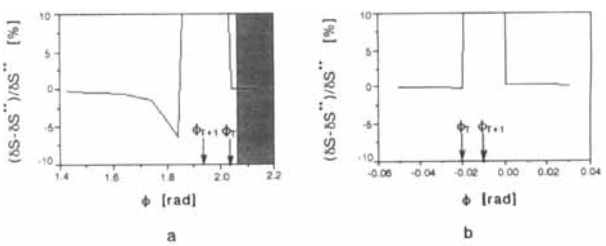


Figure 12 Computational tests for the arrangement of spheres shown in Fig 10c, δS and δS^{**} have the same meaning as in Fig 11. (**a**) ϕ_T is the value of ϕ at transition when the surface area is computed with a three-spheres intersection case function and the quadratic approximation δS^{**} based on two-spheres interaction case function. ϕ_{T+1} is the value which ϕ takes after ϕ_T, (**b**) in ϕ_T and ϕ_{T+1} both derivatives and surface area are computed with the three-spheres case function but the derivatives have a discontinuity when spheres C and B coincide (at $\phi = \phi_{T+1}$) due to asymptotic values of some quantities in eqn. (28). This leads to the disagreement between δS and quadratic approximation in ϕ_{T+1}.

Figure 11a–d shows computational results performed for the case presented in Fig 10b. The agreement between δS and δS^{**} is generally good. The darkened area corresponds to those cases where the computation of surface area is carried out using the eqns (24) and (25) (two-spheres case). In these cases, the only variable is the x-coordinate of the centre of sphere A (x_A). Similarly, the unshaded area corresponds to the case where eqn. (28) is used to compute the surface area (three-spheres case). The interval scanned by x_A was gradually decreased in going from Fig 10a to 10b in order to see if the discontinuity in derivatives is sharp or wide. A sharp discontinuity should diturb the optimization less than a wide one, if we could skip it by taking appropriate interval.

In Fig 11 we mark the value of x_A at transition with x_T. In this point the surface area is computed with a three-spheres case function and the quadratic approximation is based on derivatives calculated with a two-spheres case function at the previous iteration. When $x_A = x_T$, both derivatives are highly unstable (small changes in x coordinate will cause large changes in target function) due to some quantities in Gauss-Bonnet function which take asymptotic values (radii of intersection circles), and consequently the agreement at the next iteration (x_{T+1}) worsens.

Figure 12a–b show similar results obtained when sphere C is moved around sphere A (Fig 10c). In this case angle ϕ is used as the variable and two transition points are obtained. The first one is reached when ϕ is about $2\pi/3$ (Fig 12a). As in the previous example, the agreement between the increment in surface area (δS) and quadratic approximation (δS^{**}) worsens at the next step (ϕ_{T+1}) due to the same asymptotic change in some of the quantities appearing in eqn. (28). The second transition point is reached when ϕ is close to zero (C and B spheres coincide). Both before and after this point the same function (three-spheres case function, eqn. (28)) is used to calculate the derivatives but at $\phi = 0$ the derivatives are again unstable.

The increased discontinuities in the analytical derivatives of molecular surface area along with their complexity are among the major disadvantages of analytical gradients. The difficulty in locating the transition points on the steric energy surface for real molecules prevented us from providing adequate algorithm to deal with these situations. However, from above results, we deduce that these discontinuities should be of less practical importance for the optimization because the algorithm could manage to jump over or get out of them. In order to determine the effect of the discontinuities in the derivatives upon optimization we performed the minimization of surface area of the spheres arrangement shown in Fig 10b by allowing sphere A to move freely. The optimization proceeded smoothly and at the end we found sphere A positioned at the origin, identical with the arrangement having the lowest surface area, Fig 10a. Our conclusion is that the optimization algorithm manages to jump over the transition point due to the sharpness of the discontinuity. We also performed optimizations on hydrocarbons in water without any difficulty. It may also be noted that the optimization algorithms used in molecular mechanics set up an upper limit for the increment in atomic coordinates to about 0.2 Å and thus diminish the effect of the discontinuities in the derivatives of surface area on optimization.

VI. CONCLUSIONS

Analytical derivatives of surface area certainly will be of interest in such diverse fields of computational chemistry as drug design, molecular graphics, and protein docking algorithms. One way of using them in molecular mechanics optimization would be to add them at the end of the optimization before re-optimizing the geometry or to switch them on only in a later optimization step. These procedures will save a lot of computer time and we hope that it will still provide computational results with the required consistency.

Although a very crude approximation, the solvent effect model we have presented here is a starting point in a series of solvent effect models we wish to develop for molecular mechanics. Molecular surface area is a very common quantity but it has proved to be difficult to compute analytically. Besides, developing analytical derivatives proved to be a very challenging task. The main advantage of the solvent effect method we have proposed here is its simplicity along with its close dependence on molecular geometry which will make it useful in conformational analysis. The need to include the solvation effect into an optimization scheme made us avoid to use any of the integrals (eqns (7,8)) which would have given a more accurate treatment of the solvent-solute van der Waals interaction but also will dramatically increase the computing time.

REFERENCES

1 Burkert, U.; Allinger, N.L.; *Molecular Mechanics*, American Chemical Society, Washington, D.C., **1982**.
2 Dosen-Micovic, Lj.; Jeremic, D.; Allinger, N.L.; *J. Am. Chem. Soc.* **1983**, *105*, 1722.
3 Meyer, A.Y.; *J. Comput. Chem.* **1981**, *2*, 384.
4 Still, W.C.; Tempczyk, A.; Hawley, R.C.; Hendrickson, T.; *J. Am. Chem. Soc.* **1990**, *112*, 6127.
5 Cramer, C.C.; Truhlar, D.G.; *J. Am. Chem. Soc.* **1991**, *113*, 8305.
6 See, however: Smith, D.A.; Vijayakumar, S.; *Tetrahedron Lett.* **1991**, *32*, 3617.
7 Francl, M.M.; Hout, Jr. R.F.; Hehre, W.J.; *J. Am. Chem. Soc.* **1984**, *106*, 563.
8 Bondi, A.; *J. Phys. Chem.* **1964**, *68*, 441.
9 Pierotti, R.A.; *Chem. Rev.* **1976**, *76*, 717.
10 Lee, B.; Richards, F.M.; *J. Mol. Biol.* **1971**, *55*, 379.
11 Hermann, R.B.; *J. Phys. Chem.* **1972**, *76*, 2754.
12 Richards, F.M.; *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151.
13 Connolly, M.L.; *J. Appl. Cryst.* **1983**, *16*, 548.
14 Meyer, A.; *J. Chem. Soc. Perkin Trans. II* **1985**, 1161.
15 Wang, H.; Levintal, C.; *J. Comput. Chem.* **1991**, *7*, 868.
16 Wodak, S.J.; Janin, J.; *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 1736.
17 Richmond, T.J.; *J. Mol. Biol.* **1984**, *178*, 63.
18 Gibson, K.D.; Scheraga, H.A.; *Molec. Phys.* **1987**, *62*, 1247.
19 Hasel, W.; Hendrickson, T.F.; Still, W.C.; *Tetrahedron Comput. Methodol.* **1988**, *1*, 103.
20 Tsukuzi, S.; Schafer, L.; Goto, H.; Jemmis, E.D.; Hosoya, H.; Siam, K.; Tanabe, K.; Ōsawa, E.; *J. Am. Chem. Soc.* **1991**, *113*, 4665.
21 Botcher, C.; *Theory of electric polarization* Vol. 1, Elsevier, Amsterdam, **1973**.
22 Rinaldi, D.; Costa Cabral, B.J.; Rivail, J.L.; *Chem. Phys. Lett.* **1986**, *125*, 495.
23 For a glossary of symbols, see Appendix.
24 The hard sphere diameters were usually determined from solubility data and considered to be less sensitive to changes in molecular geometry.
25 Oakenfull, D.G.; Fenwick, D.E.; *J. Phys. Chem.* **1974**, *78*, 1759.
26 Postma, J.P.M.; Berendsen, H.J.C.; Haak, J.R.; *Faraday Symp. Chem. Soc.* **1982**, *17*, 55.
27 Ben-Naim, A.; *Water and Aqueous Solutions*, Plenum Press, New York, **1974**.
28 Hermann, R.B.; *J. Phys. Chem.* **1975**, *79*, 163.
29 Huron, M.J.; Claverie, P.; *J. Phys. Chem.* **1972**, *76*, 2123.
30 Floris, F.M.; Tomasi, J.; Pascual-Ahuir, J.L.; *J. Comput. Chem.* **1991**, *12*(7), 784.
31 Ben-Naim, A.; Marcus, Y.; *J. Chem. Phys.* **1984**, *81*(4), 2016.
32 Rashin, A.A.; Honig, B.; *J. Phys. Chem.* **1985**, *89*, 5588.
33 Abraham, R.J.; Bretschneider, E.; in *Internal Rotation in Molecules* (Orville-Thomas, W.J., ed.), Wiley, London, **1974**.
34 Ōsawa, E.; Gotō, H.; in *Computer Aided Innovation of New Materials* (Doyama, M. *et al.*, eds.), Elsevier Sci. Publ. B.V. (North-Holland), Amsterdam, **1991**, p. 411–416.
35 Mezei, P.G.; in *Reviews in Computational Chemistry* (Lipkowitz, K.B., Boyd, D.B., eds.), VCH Publishers, New York, **1990**.
36 Do Carmo, M.P.; *Differentil Geometry of Curves and Surfaces*, Prentice-Hall, Englewood Cliffs, **1976**.
37 This function is evaluated by a power series and its accuracy depends, among other factors, on the FORTRAN subroutine library used.
38 A similar case was found in n-hexane in the 4/8 intersection circle but in this case the arc end-points were found to be about 10^{-3} radian apart. In this case the uncertainty introduced by numerical computation do not affect the calculation.
39 Besides this correction, the modified algorithm elminates the cutting planes which are completely buried by other planes and are not needed for the evaluation of free arcs. Thus, in Fig 7b the planes due to atoms 3, 12 and 18 are eliminated.
40 Grossman, S.I.; *Multivariable Calculus, Linear Algebra, and Differential Equations*, Academic Press Inc., Orlando-Fl., **1986**.
41 Mohamadi, F.; Richards, N.G.J.; Guida, W.C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W.C.; *J. Comput. Chem.* **1990**, *11*, 440.

APPENDIX: SYMBOLS AND NOTATIONS

a	radius of a spherical cavity in solvent
a_i	radius of atom i used in the evaluation of G_{pol} in GB method, appears in eqn. (11)
A	label for sphere in the spheres arrangement of Fig 10
Å	Ångstrom unit
b	abbreviation of relation 17, appears in eqn. (15)
B	label for sphere in the spheres arrangement of Fig 10
c	integration constant, appears in eqn. (4)
C	label for sphere in the spheres arrangement of Fig 10
C_λ	piece of regular curve which borders the exposed surface area S_i of atom i
d	the distance between the centres of two overlapping spheres, appears in eqn. (25)

d_{ij} the distance between solute atom i and j, appears in eqn. (11)
E total steric energy in solution
E_{cav} steric energy due to creating a cavity in solvent
E_g solute steric energy in vapor phase
E_{pol} steric energy due to solute-solvent electrostatic interaction
E_{sol} steric energy due to solvent effect force field
E^i_{sol} solvent effect steric energy of solute conformer i
E_{vdW} steric energy due to solute-solvent interaction by van der Waals forces
f abbreviation of relation 16, appears in eqn. (15)
g the radial distribution function
G_{cav} free energy due to creating a cavity in solvent
G_{dip} polarization free energy due to solute dipole
G_{pol} polarization free energy due to solute-solvent electrostatic interaction
G_{qdp} polarization free energy due to solute quadrupole
G_{sol} solvation free energy
G_{vdW} free energy due to solute-solvent interaction by van der Waals forces
G_{DQ} polarization free energy due to solvent-solute dipole-dipole and dipole-quadrupole interactions
h abbreviation of relation 14, appears in eqn. (13)
i/j intersection circle made by overlapping spheres i and j
k Boltzmann constant
k_λ geodesic curvature of C_λ
l abbreviation of relation $2\alpha/a^3$
$L^k_{i/j}$ Length of the arc cut by atom k on i/j intersection circle
n_{cf} the number of significantly populating conformers
n_d solute refractive index
N number of atoms in solute molecule
N_A Avogadro's number
p_i free energy population of solute conformer i in solution
P the pressure of solute-solvent system
q_i partial charge of atom i
Q_{ii} axial quadrupole component of quadrupole moment
Q_{ij} perpendicular quadrupole component of quadrupole moment
r the distance between two interacting hard spheres
r_i van der Waals radius of atom i
$r_{i/j}$ the radius i/j intersection circle
r_p polar coordinate, appears in eqn. (7)
R gas constant
s_λ the length of the arc C_λ, appears in eqn. (28)
S solute solvent accessible surface area
S_i solvent accessible surface area of atom i
t_i abbreviation of relation 25, appears in eqn. (24)
T absolute temperature
u_{LJ} Lennard-Jones potential
v the number of regular curves C_λ which make a closed boundary
V_{cav} cavity volume, the volume enclosed in the solvent accessible surface area
V_m the molar volume of the solvent
x_A the coordinate x of sphere A, see Fig 10
x_i the Cartesian coordinate x of the center of atom i
x_T the value of x_A at transition, see Fig 11
x_{T+1} the value which x_A takes the next iteration after transition, see Fig 11
y reduced number density of the solvent, appears in eqn. (3)
y_i the Cartesian coordinate y of the centre of atom i
z_i the Cartesian coordinate z of the centre of atom i
Z abbreviation of summation in eqn. (14), appears in eqn. (17)

α solute polarizability
β_T isothermal compressibility
χ Euler-Poincare characteristic, appears in eqn. (28)
δS the increment in solute surface area
δS^* the total differential of surface area
δS^{**} the quadratic approximation of surface area
$\delta \boldsymbol{x}$ the increment in coordinates
ε solvent dielectric constant
ϕ the angle between the line which connects the centres of spheres A and B and the line which connects the centres of spheres A and C, see Fig 10
ϕ_T the value of ϕ at transition, see Fig 12
ϕ_{T+1} the value which ϕ takes the next iteration after transition, see Fig 12
γ_∞ macroscopic surface tension of the solvent
η switching parameter, appears in eqn. (27)
φ polar coordinate, appears in eqn. (7)
κ abbreviation for the relation $(\varepsilon - 1)/(2\varepsilon + 1)$
λ index in eqn. (28)
μ solute dipole moment
v_0, v_i empirically determined free energy parameters for solute-solvent interaction by van der Waals forces

θ polar coordinate, appears in eqn. (7)
ρ_s solvent number density, appears in eqn. (6)
σ solute hard sphere diameter
σ_0 solute-solvent distance at $u_{LJ} = 0$, $\sigma_0 = (\sigma + \sigma_s)/2$
σ_i the distance between solvent and atom i of the solute, $\sigma_i = r_i + \sigma_s/2$
σ_s solvent hard sphere diameter
ω the angle which corresponds to the free arc, appears in eqn. (30)
ω_k the angle which corresponds to the end-points of the arc cut out by sphere k on i/j intersection circle, appears in eqn. (31)
ω_k^0 this angle depends on the position of the centre of atom k, appears in eqn. (31)
ω_l the angle which corresponds to the end-points of the arc cut out by sphere l on i/j intersection circle, appears in eqn. (31)
ω_l^0 this angle depends on the position of the centre of atom l, appears in eqn. (31)
ξ the ratio between solute and solvent hard sphere diameters, appears in eqn. (3)
ζ, ζ_0 empirical free energy parameters for cavity creation, appear in eqn. (5)
$\Omega_{\lambda,\lambda+1}$ external angle between the tangents drawn from intersection point of C_λ and $C_{\lambda+1}$ regular curves
∇ gradient matrix
Δ hessian matrix