Extended scaled particle theory for dilute solutions of arbitrary shaped solutes: applications to solvation free energies of alkanes, alcohols, and proteins

A Doctoral Thesis
Submitted to
Osaka University

Masayuki Irisa

1995

Yanagida bio–motron project, ERATO JRDC,
2-4-14 Senba-Higashi, Mino, Osaka 562, Japan

E-mail address: irisa@yanagida.jrdc.go.jp

Table of contents

General introduction

Proteins consist of only 20 kinds of amino acids but each protein has a unique 3-dimensional structure. This 3-dimensional structure is determined in thermodynamic equilibrium law. The solvation free energy is one of the key to understanding the stabilities of proteins in water. For the purpose of calculating solvation free energies, theories and simulations are used and checked by comparing calculated results with the experimental data of small molecules. But direct application of the techniques, which are originally used for small molecules, to biomolecules is very difficult due to the overwhelmingly large configurational space to which solvent molecules can access. In spite of the remarkable success of the molecular simulation with computers for the past few decades, there is not yet feasible method to determine the solvation free energy of the biomolecules. Scaled particle theory (SPT) [1] is one of the liquid theories used for calculating the solvation free energies and is widely used by solution chemists because of its simplicity and broad validity [2]. The difficulties of applying the original SPT to the biomolecule are a) consideration of the shape of the molecule, b) application to both polar and ionic solutes. Original SPT can only deal with spherical and convex molecules [3], the shape of which is different from molecules having multiple interaction sites. We extended SPT in order to apply to molecules having the shape of fused spheres.

In Chapter 1 of this thesis I review three methods concerning the solvation free energy calculation. I then explain the extension of the SPT for molecules with equal radii in Chapter 2, and the combination of the extended SPT with the Poisson-Boltzmann equation in Chapter 3 (the extension of the SPT with unequal radii is also explained). Applications of the theory to proteins are discussed in Chapter 4, and the mathematical basis of the extended SPT, a key of our success, is explained in Chapter 5. In

Chapter 6, our conclusions and some preliminary results of the application of our new method are presented, and the objective for future work is discussed.

References

1. H. Reiss, H. L. Fish, and J. L. Lebowitz, *J. Chem. Phys.* **31**, 369 (1959)
2. R. A. Pierotti, *Chem. Rev.* **76**, 717 (1965)
3. R. M. Gibbons, *Mol. Phys.* **17**, 81 (1969)

Chapter 1

Review of three methods relating with the calculation of the solvation free energy

Empirical method using the accessible surface area of the atom

One of the most widely used methods for calculating solvation free energies for biomolecules an the empirical method developed by Ooi et al.[1], in which the accessible surface area of the atoms in the molecule is used. The solvation free energy of the solute molecule is assumed to be,

$$\Delta\overline{g}_s^{\circ}(g \rightarrow m) = \sum_i ASA_i p_i \quad (1)$$

where ASA_i is the accessible surface area of the atom i, p_i is solvation free energy per accessible surface area of atom i. The parameters, p_i, of typical atoms are determined based on data from experiments on the solubilities of small molecules. This simple method predicts good values for the solvation free energies of biomolecules unless the electrostatic contribution is large in the solvation free energy.

Molecular dynamics simulations

The molecular dynamics simulation is also powerful for the solvation free energy calculation. The free energy perturbation method [2, 3] is well known for this purpose. In this method, the relative value of the solvation free energy is calculated using the gradual deformation of the solute molecule in the molecular dynamics simulation. For example, the difference between the solvation free energies of the glycine and the alanine can be calculated by deforming the glycine to the alanine in the simulation. The effects of water configurations around the solute molecule must be sufficiently averaged in every step in the course of the deformation of the

solute. However the application to estimate the large difference of the solvation free energy is limited because of the cpu-time required. Both contributions of hydrophobic and electrostatic parts to the solvation free energy are considered.

Poisson-Boltzmann equation

The estimation of the electrostatic contribution to the solvation free energy is very difficult. The classical method, in which the solvent is regarded as the dielectric medium, has been applied to the polar and ionic solutes. In the method, the Poisson-Boltzmann equation is solved numerically, resulting in reasonable Born energy (electrostatic part of the solvation energy) even for protein as long as the boundary condition is properly adjusted [4-6]. There are three algorithms to solve the Poisson-Boltzmann equation, the finite difference method, the finite element method, and the boundary element method. Warwicker applied the finite difference method to the protein in water for the first time. The finite difference method is widely used because the numerical solvation is suitable for the vector/parallel computers.

References

1. T. Ooi, M. Oobatake, G. Némethy, and H. A. Sheraga, *Proc. Natl. Acad. Sci. USA* **84**, 3086 (1987)
2. Special issue. *Chemical Physics*, **158** No.2,3 (1991)
3. D. Frenkel, "Free energy computation and first order phase transitions: Molecular dynamics simulation of statistical mechanical systems, North Holland (1986)
4. J. Warwicker, and H. C. Watoson, *J. Mol. Biol.*, **157** 671 (1982)

5. H. Nakamura, *J. Phys. Soc. Jpn.*, **57** 3702 (1988)
6. T. Takahashi, H. Nakamura, A. Wada, *Biopolymers*, **32** 897 (1992)

Chapter 2

Extension of scaled particle theory for solutes with multi interaction sites: applications to hydration free energies of hydrocarbons

Abstract

A method based on the scaled particle theory is proposed for calculating the solvation free energy of a solute with an arbitrary shape in the dilute solution. The new approach employs our extended version of Richmond method [J. Mol. Biol. 178 (1984) 63] which enables us to analytically calculate the excluded volume and the accessible surface area of a molecule. The theory is applied to estimate the solvation free energy for a series of hydrocarbons in water. The calculated results exhibit a remarkable agreement with the corresponding experimental values which have been revised recently by Sharp and coworkers [Biochemistry 30 (1991) 9686].

Introduction

Theoretical determination of the solvation free energy is important to pursue the prediction problem of the tertiary structure of the protein. Furthermore, it is primarily important in the various sciences such as chemical, biological and material sciences for the understanding of the nature of the solvation in terms of physical quantities in molecular level. In spite of the remarkable success of the molecular simulation with computers for the past few decades [1], it is not yet feasible for the method to determine the solvation free energy mainly due to the overwhelmingly large configurational space to which solvent molecules can access. The integral equation approach has been very successful in studying the solution consisting of relatively small solutes, but its application to large molecules such as polymers and proteins is still under development [2]. An alternative empirical approach based on the solvent accessible surface area (ASA) has been widely employed among polymer and protein chemists, but its success largely depends on the experimental parameters assigned to the constituent atom groups [3]. This uncomfortable situation urged us to develop a new method extended from the scaled particle theory (SPT) [4] for the determination of the solvation free energy where solutes can take an arbitrary shape. The theory can give a rigorous description for a liquid system consisting of hard spheres [5]. The applicability of the method to a wider class of systems such as aqueous solutions has already been demonstrated for the solubility of small nonpolar molecules in water by Pierotti [6]. An application of the method to the solvation of protein has been made first by Ben–Naim et al. focussing on the hydrophobic effect, but protein was modeled by a sphere which apparently oversimplifies characteristics of the solute [7].

Although there have been significant efforts devoted to extend the theory to cover non–spherical molecules, the mathematical problems associated have confined the extension within rather limited class of molecules with a convex shape [8,9]. The solvation free energy is a reversible work required to dissolve a solute molecule into solvent. For the solvent consisting of hard spheres, the dissolving procedure is a physical process to scale up a cavity from which the center of solvent molecules are excluded. Then, the crucial step to find the reversible work is to get an analytical expression for the volume of the cavity and its derivatives with respect to the scaling parameter. The mathematical problem associated with the procedure is non–trivial for the solute molecules with an arbitrary shape including a non–convex body.

In this report, we propose an approach for calculating the reversible work required to create a cavity with an arbitrary shape in the solvent system consisting of hard spheres based on the Richmond method which was originally developed to calculate an accessible surface area of protein [10]. Although there is no essential difficulty for applying our method to the complicated molecules such as proteins, we present an preliminary application to the solubility of a series of hydrocarbons in water in this report. The problem belongs to rather classical thermodynamics in physical chemistry [11], but has been reinvestigated recently by Sharp and coworkers particularly in connection with the hydrophobic effect in protein solutions [12,13].

I. Scaled Particle Theory for a solute with an arbitrary shape

We consider a system consisting of a solute molecule with an arbitrary shape and solvent molecules with a spherical shape. The solute molecule is modeled as a rigid body formed of fused hard spheres with

different radii and the solvent molecules are modeled as hard spheres with the same radius. The problem is to find an expression for the reversible work required to dissolve the solute in the solvent. As mentioned, the dissolving procedure can be viewed as a scaling–up–process of a cavity made in the solvent. The volume of the cavity is an excluded volume of the scaling solute (see Fig. 1a).

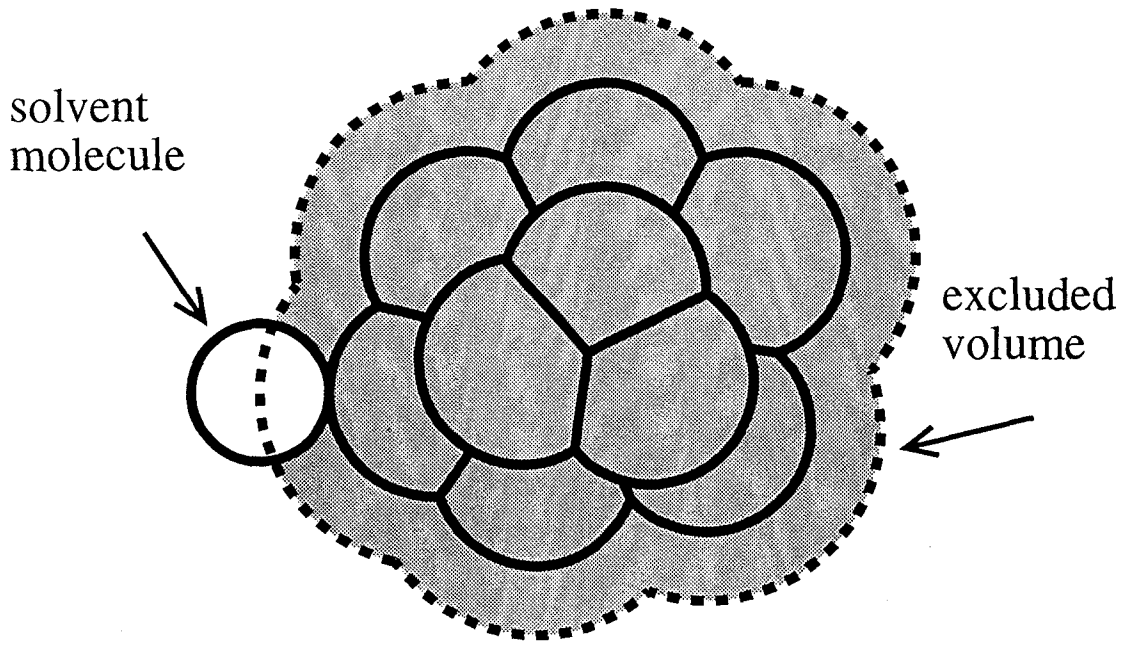


Fig. 1a The solute and solvent molecules in the extended scaled particle theory.A schematic illustrating of the solute volume excluded by solvents.

Then with the standard argument in SPT, the reversible work ($\bar{G}_c$) for creating the cavity with an arbitrary shape can be written as [8,14],

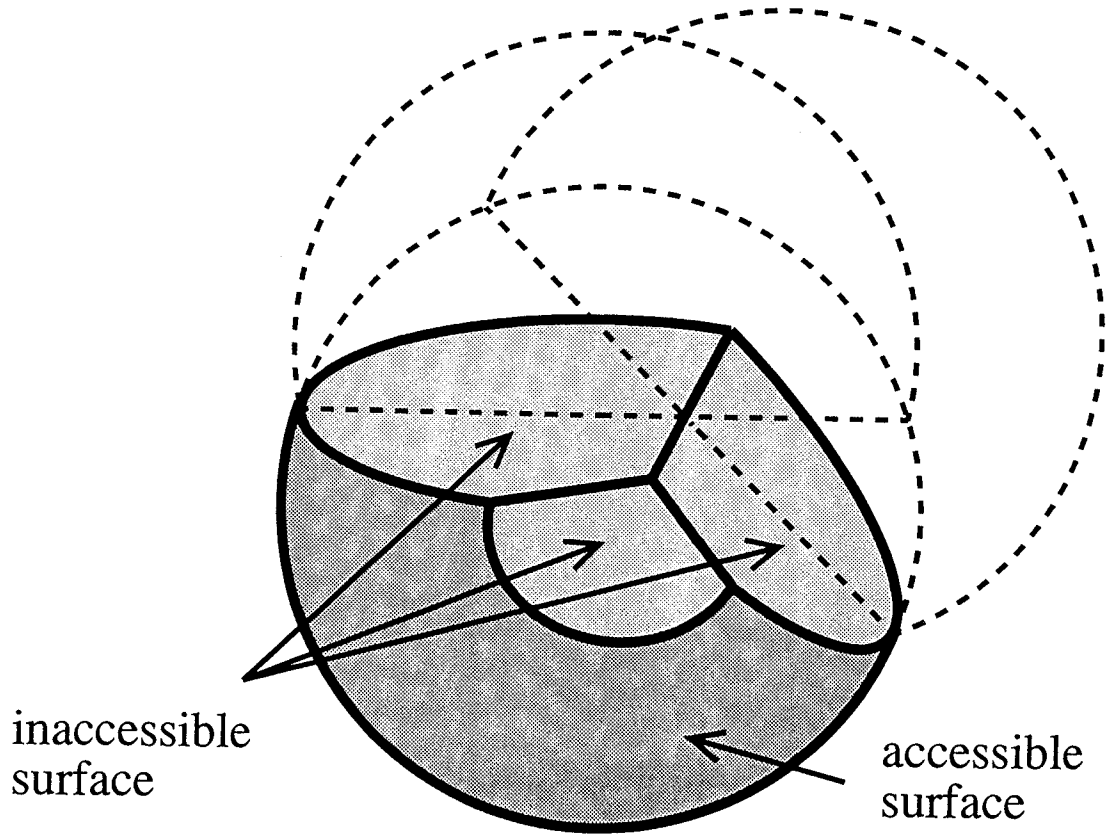


Fig. 1b The solute and solvent molecules in the extended scaled particle theory. Accessible and inaccessible surface of a component of fused spheres.

$$\bar{G}_c = N_A k_B T \left\{ -\ln(1-\rho V_C(0)) + \rho \int_0^1 G(\lambda)\left(\frac{\partial V_C(\lambda)}{\partial \lambda}\right) d\lambda \right\} \quad , \quad (1)$$

where N_A is the Avogadro number, k_B is the Boltzmann constant, T is the absolute temperature, ρ is the number density of the solvent, λ is the scaling parameter ($0 \le \lambda \le 1$), $V_c(\lambda)$ is the volume of the cavity, and $G(\lambda)$ is

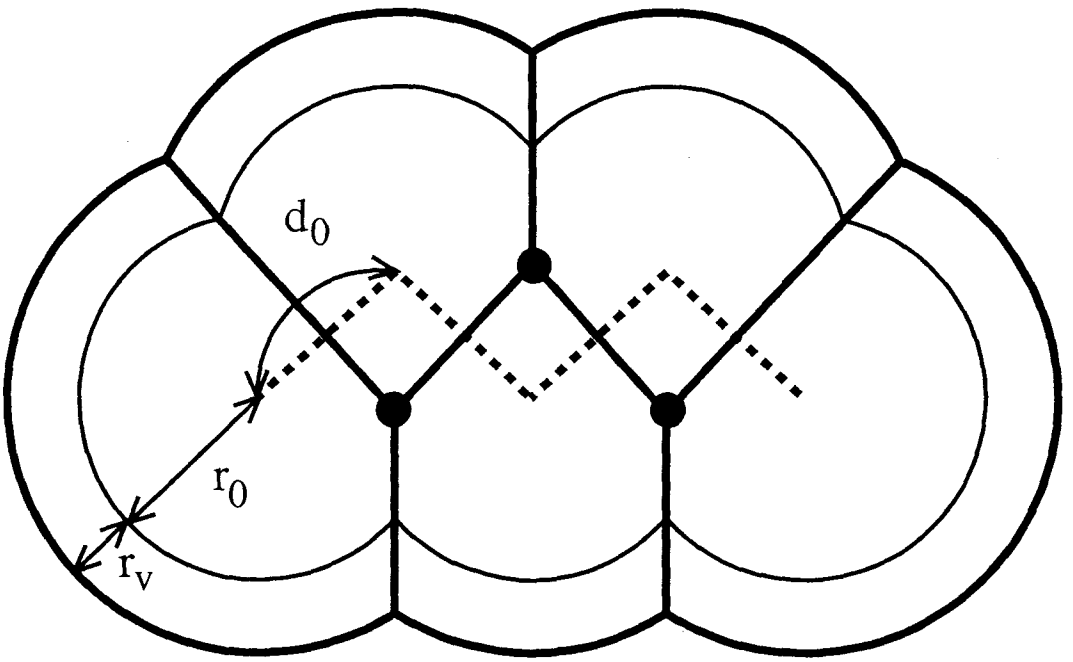


Fig. 2 Models of hydrocarbons, normal–alkane in the trans-conformation, as fused hard spheres. The symbol ● shows the location of the exposed vertices.

the distribution function of solvent centers at surface of the cavity. The first term in Eq.(1) is the work required to dissolve the material point as the seed of the solute molecule into the pure solvent. The second term represents the work required scaling–up the solute molecule from λ=0 to λ=1. The meaning of the second term is rather obvious since it is a product of the pressure of solvent exerted on the surface of the cavity, $k_BT\rho\bar{G}(\lambda)$, and the infinitesimal increment of the volume of the cavity. The integral can be evaluated analytically to lead

where P is the macroscopic pressure of the solvent, which is to be applied

$$\rho\int_0^1 G(\lambda)\left(\frac{\partial V_c(\lambda)}{\partial\lambda}\right)d\lambda = A(V_c(1)-V_c(0)) + B\left(\frac{\partial V_c}{\partial\lambda}\right)_{\lambda=0} + \frac{C}{2}\left(\frac{\partial^2 V_c}{\partial\lambda^2}\right)_{\lambda=0} \quad , \quad (2)$$

$$A = \frac{P}{k_B T} \quad , \quad B = \frac{\rho}{1-\rho V_c(0)} - A \quad , \quad C = \frac{\rho^2}{(1-\rho V_c(0))^2} \frac{\left(\frac{\partial V_c}{\partial\lambda}\right)^2_{\lambda=0}}{\left(\frac{\partial^2 V_c}{\partial\lambda^2}\right)_{\lambda=0}} + B \quad , \quad (3)$$

at the surface of the cavity with macroscopic size ($\lambda=\infty$). The derivatives of $V_c(\lambda)$ at $\lambda=0$ are the limit when the positive λ approaches 0. If the solute is a convex molecule, these equations are exactly the same as those of Boublík [14]. The V_c can be calculated for the solute with virtually any shape as follows [15],

$$V_c(\lambda) = \sum_a \left[\frac{1}{3} R_a S_a^A(\lambda) + \sum_b \frac{1}{3} F_{ab} S_{ab}^I(\lambda)\right] \quad , \quad (4)$$

$$F_{ab} = \frac{D_{ab}^2 + R_a^2 - R_b^2}{2D_{ab}} \quad , \quad R_a = r_a\lambda + r_v \quad , \quad D_{ab} = d_{ab}\lambda \quad , (5)$$

where a runs through all atoms in the solute molecule, b runs through atoms with which the atom a intersects to form inaccessible surfaces, $S_a^A(\lambda)$ and $S_{ab}^I(\lambda)$ are accessible and inaccessible surface area of the scaling atom a (see Fig. 1b), r_a is the radius of the atom a, r_v is the radius of the solvent molecule, and d_{ab} is the distance between atom a and b. The analytical equation of S_a^A, as the function of only atomic radii and interatomic distances, is derived by employing the Richmond method and that of S_{ab}^I is derived by us [16]. The analytical equations of the first and second derivatives of $V(\lambda)$ are obtained from $V(\lambda)$. More detailed description of the theory is presented elsewhere [4]. Here we give an expression for the case to be used in the next section.

The fused spheres with equal radii in Fig. 2 are regarded as a model of normal alkane in the trans-conformation.

$$V_c(\lambda)=nT_1+\sum_{(a,b)}T_2(a,b)+\sum_{(a,b,c)}T_3(a,b,c) \quad , \quad (6)$$

$$T_1=\frac{2}{3}\pi R^3(1+\delta_{1n}) \quad , \quad T_2(a,b)=\frac{1}{3}\pi F_{ab}(2R^2+J_{ab}^2) \quad ,$$

$$T_3(a,b,c)=\frac{1}{6}\left[-F_{ab}(2R^2+J_{ab}^2)\Phi(a,b,c)+R^3\Omega(a,b,c)+F_{ab}L(a,b,c)\sqrt{J_{ab}^2-L(a,b,c)^2}\right] \quad , \quad (7)$$

$$F_{ab}=\frac{D_{ab}}{2} \quad , \quad J_{ab}=\sqrt{R^2-F_{ab}^2} \quad , \quad L(a,b,c)=\frac{-F_{ab}\cos\theta_{abc}+F_{ac}}{\sin\theta_{abc}} \quad ,$$

$$\Omega(a,b,c)=-\cos^{-1}\left(\frac{R^2\cos\theta_{abc}-F_{ab}F_{ac}}{J_{ab}J_{ac}}\right) \quad , \quad \Phi(a,b,c)=\cos^{-1}\left(\frac{L(a,b,c)}{J_{ab}}\right) \quad , \quad (8)$$

$$\cos\theta_{abc}=\frac{D_{ab}^2+D_{ac}^2-D_{bc}^2}{2D_{ab}D_{ac}} \quad , \quad \sin\theta_{abc}=\sqrt{1-\cos^2\theta_{abc}} \quad ,$$

$$R=r_0\lambda+r_v \quad , \quad D_{ab}=d_{ab}\lambda \quad , \quad (9)$$

where δ is the Kronecker's delta, r_0 are the radii of solute atoms, d_{ab} is the distance between atoms a and b , and n is the number of atoms in the solute molecule, the first summation extends over all pairs of atoms which make inaccessible surfaces (Fig. 1b), the second summation extends over all triplets of atoms which make exposed vertices (● in Fig. 2), ie., both end points of a line made of two intersecting inaccessible surfaces. The summations can be performed using the suffices of the solute atoms which are numbered sequentially from one terminal atom to the other,

$$\sum_{(a,b)} T_2(a,b) = \sum_{a=3}^{n} T_2(a,a-2) + \sum_{a=2}^{n} T_2(a,a-1) + \sum_{a=1}^{n-1} T_2(a,a+1) + \sum_{a=1}^{n-2} T_2(a,a+2) \quad , \tag{10}$$

$$\sum_{(a,b,c)} T_3(a,b,c) = 2\left(\sum_{a=3}^{n} T_3(a,a-2,a-1) + \sum_{a=2}^{n-1} T_3(a,a-1,a+1) + \sum_{a=1}^{n-2} T_3(a,a+1,a+2) \right.$$
$$\left. + \sum_{a=3}^{n} T_3(a,a-1,a-2) + \sum_{a=2}^{n-1} T_3(a,a+1,a-1) + \sum_{a=1}^{n-2} T_3(a,a+2,a+1) \right) \quad . \tag{11}$$

This expression of normal alkane in our extended SPT includes the case of the single sphere solute (n=1) studied in the original SPT and the case of the hard dumbbell solute (n=2) studied by Boublík[17].

II. Solvation Free Energy of Hydrocarbons

As an application of the theory sketched above, we study the solubility of hydrocarbons in water of which importance in chemical biological sciences has been well addressed [7,11]. Recently, Sharp and coworkers restudied the problem with casting a suspicion to the traditional interpretation of experimental values of the transfer free energy to extract the solvation free energy [12,13]. The authors claimed from a thermodynamic arguments that the traditional method to evaluate the solvation free energy does not fully account for the change in volume entropy, and they revised the interpretation of experimental data on this account.

We employed the Pierotti's recipe [6] in order to calculate the solvation free energy of a series of hydrocarbons in water. The dissolution of the solute is decoupled into two steps by a thought experiment: (1)

creating a cavity in solvent, which has a right size as well as a shape to accommodate the solute, (2) establishing the attractive interaction between the solute and solvent. Then the solvation free energy can be written as a sum of the two terms:

$$\bar{G}=\bar{G}_c+\bar{G}_i \quad , \quad (12)$$

where $\bar{G}_c$ and $\bar{G}_i$ correspond to steps (1) and (2) above, respectively. The free energy of cavity formation $\bar{G}_c$ can be calculated from the extended SPT described in section I. We employ the Pierotti's value for the effective hard–sphere radius (1.375 Å) for water [18] which is very similar to the value commonly used in the study of biopolymers (1.4 Å) [19]. To estimate the contribution from the attractive interaction $\bar{G}_i$, we also utilized Pierotti's formula. He assumed $\bar{G}_i$ is approximately equal to the corresponding enthalpy, $\bar{H}_i$, due to the attractive force between solute and solvent. Then $\bar{H}_i$ is estimated as an integrated value of the Lennard–Jones type interactions between the solute and solvent molecules. The electrostatic interactions between the solute and solvent molecules are neglected. A following step–wise function was assumed for the distribution function of solvent molecules around a solute molecule,

$$\begin{cases} g(x,y,z)=0 \quad \cdots \quad (\exists b;\ d_{bv}\le r_b+r_v) \\ g(x,y,z)=1 \quad \cdots \quad (\forall b;\ d_{bv}>r_b+r_v) \end{cases} \quad , \qquad d_{bv}=\sqrt{(x-x_b)^2+(y-y_b)^2+(z-z_b)^2} \quad (13)$$

where (x_b,y_b,z_b) is a Cartesian coordinate of a center of the solute atom b. In this study, we calculated the value of $\bar{H}_i$ in the same way as the Pierotti [20] using distribution function $g(x,y,z)$ to cover a series of hydrocarbons,

$$\bar{H}_i=\sum_a \bar{H}_a \quad , \quad \bar{H}_a=-N_A C_a \iiint \rho\, g(x,y,z)\left(\frac{1}{d_{av}^6}-\frac{\sigma_{av}^6}{d_{av}^{12}}\right)dxdydz \qquad (14)$$

$$C_a=4(\varepsilon_a\varepsilon_v)^{1/2}\sigma_{av}^6 \quad , \quad \sigma_{av}=(\sigma_a+\sigma_v)/2 \quad ,$$

where $\bar{H}_a$ is the enthalpy contribution from atom a, ε_v and σ_v are the Lennard-Jones parameters for water (as the solvent), ε_a and σ_a are the Lennard-Jones parameters for solute atom a.

The solvation free energy was calculated for a series of the normal–alkanes, C_nH_{2n+2} (n=1,..,10). The extended atom representation is employed so that the methylene, methyl, and methane groups are regarded as single spheres. The Lennard–Jones parameters for calculating are listed in table 1. The TIPS parameters determined by Jorgensen are employed for the extended atoms CH_2, CH_3, and CH_4 [21]. The parameters for water are those determined by Pierotti from the solubility data of inert gas in water. The model of normal-alkanes is shown in Fig. 2. A single value 1.99 Å, which is a half the Lennard-Jones diameter (σ) for CH_2, is assigned to r_0 and r_b in Eq.(9) and (13) for CH_2, CH_3, and CH_4. The values of d_0 is 1.535 Å, the C-C bond length in ethane. The values of C-C-C bond angle is 109.47°.

The calculated values of the solvation free energy for the normal–alkane are plotted against the number of carbon atoms as shown in Fig. 3.

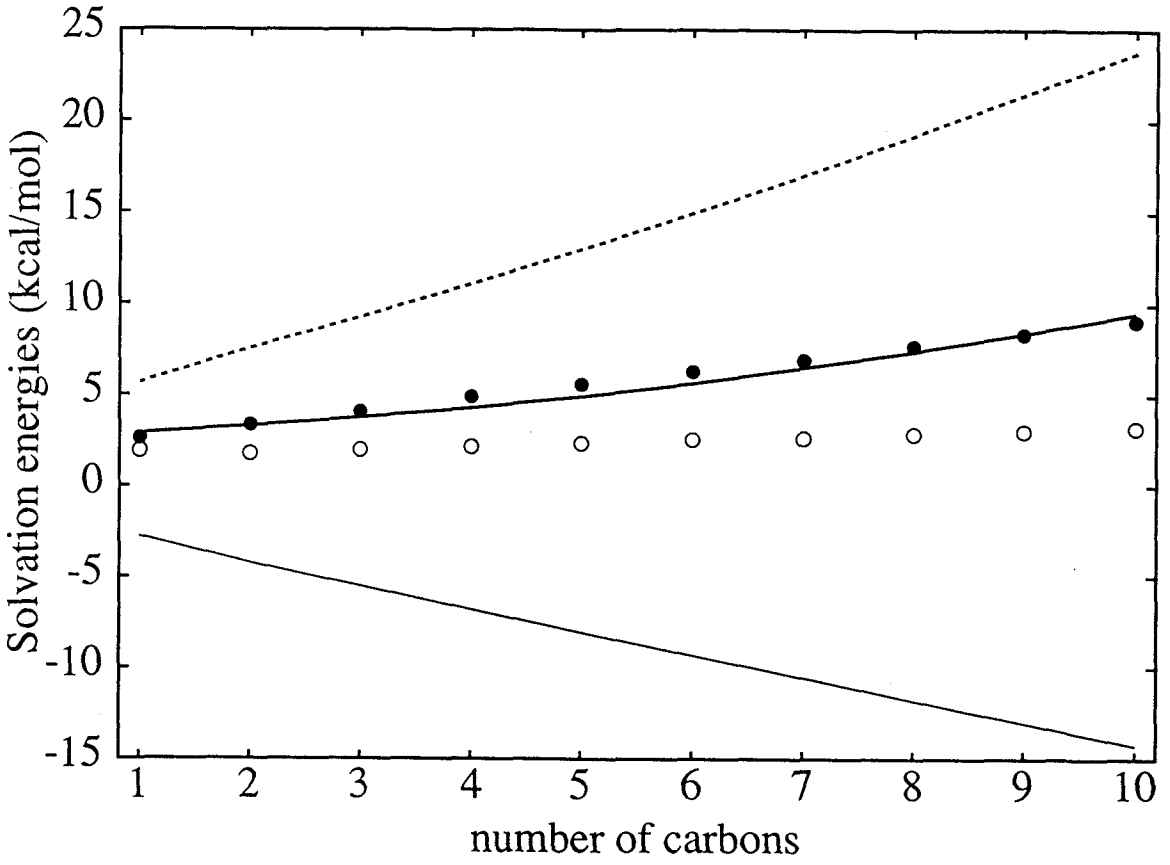


Fig. 3 Plot of the calculated values of the solvation free energy for the normal–alkane against the number of carbon atoms. The thick solid, dashed and thin solid lines represent $\bar{G}$, $\bar{G}_c$ and $\bar{G}_i$, respectively. The experimental values of $\bar{G}$ corrected by Sharp et al. (●) and uncorrected (○) are also plotted.

The corresponding experimental values corrected by Sharp et al. [13] are also plotted in the same figure along with the uncorrected values [22]. The way used to derive the solvation free energy of methane (n=1) is identical to that of Pierotti [12] except for the effective hard–sphere radius of methane. The theoretical results shows almost perfect agreement with the experimental results interpreted by Sharp et. al. within the experimental accuracy in spite of the neglect of the multiplicity of the conformation of

the solute molecule.

Concluding remarks

We have presented a method of calculating the solvation free energy of a solute with an arbitrary shape based on the scaled particle theory. The excluded volume of the solute molecule and its derivatives with respect to a scaling parameter, which are keys to the theory, have been analytically obtained using our extended version of Richmond method. The theory has been applied to the solvation free energy of a series of hydrocarbons in water. The results have shown a remarkable agreement with the experimental values which have been recently interpreted by Sharp and coworkers. The scaled particle theory has been believed to account for the hydrophobic effect reasonably well. Our results seem to have added another evidence to the confidence.

In this chapter, we studied small molecules with non–spherical shape. However, there is no principal difficulty to extend the method to cover more general class of solute molecules including proteins as long as the free energy for the cavity formation is concerned. The major difficulty expected for such extension will be the treatment of the electrostatic interactions which contribute to the solvation fee energy to a large extent for the ionic molecules. The treatment based on the Poisson–Boltzmann equation has been developed to account for the electrostatic effect [23], although it does not account for the hydrophobic effect. A hybrid between SPT and the Poisson–Boltzmann approach will complete the theory for the solvation free energy of biopolymers.

References

[1] See the articles appeared in the special issue of *Chemical Physics*, vol. 158, 1991.

[2] A. Kitao, F. Hirata, and N. Gō, in preparation.

[3] T. Ooi, M. Oobatake, G. Némethy, and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA*, **84** 3086 (1987)

[4] M. Irisa, K. Nagayama, and F. Hirata, in preparation.

[5] H. Reiss, H. L. Fisch, and J. L. Lebowitz, *J. Chem. Phys.* **31** 369 (1959)

[6] R. A. Pierotti, *Chem. Rev.* **76** 717 (1965)

[7] A. Ben-Naim, K. L. Ting, R. L. Jernigan, *Biopolymers*, **28** 1309 (1989)

[8] R. M. Gibbons, *Mol. Phys.* **17** 81 (1969)

[9] T. Boublík, *Adv. Chem. Ser.* **204** 173 (1983)

[10] T. J. Richmond, *J. Mol. Biol.* **178** 63 (1984)

[11] C. H. Tanford, "The Hydrophobic Effect: Formation of Micelles and Biological Membranes", Wiley, New York, (1980).

[12] K. A. Sharp, A. Nicholls, R. M. Fine, and B. Honig, *Science* **252** 106 (1991)

[13] K. A. Sharp, A. Nicholls, A., R. Friedman, and B. Honig, *Biochemistry* **30** 9686 (1991)

[14] T. Boublík, *Mol. Phys.* **27** 1415 (1974)

[15] L. R. Dodd and D. N. Theodorou, *Mol. Phys.* **72** (1991)

[16] M. Irisa, F. Hirata, and K. Nagayama, in preparation.

[17] T. Boublík, *Mol. Phys.* **44** 1369 (1981)

[18] R. A. Pierotti, J. Phys. Chem. 69 (1965) 281.

[19] F. M. Richards, *Annu. Rev. Biophys. Bioeng.* **6** 151 (1977)

[20] R. A. Pierotti, *J. Phys. Chem.* **67** 1840 (1963)

[21] W. L. Jorgensen, *J. Am. Chem. Soc.* **103** 335 (1981)

[22] A. Ben-Naim and Y. Marcus, *J. Chem. Phys.* **81**, 2016 (1984)

[23] N. Rogers, *Prog. Biophys. Molec. Biol.* **48** 37 (1986)

Table 1

Lennard–Jones parameters of water and extended atoms

Compd	σ,Å	ε/k_B,K
H_2O	2.75	85.3
CH_4	3.73	147.9
CH_3	3.86	91.1
CH_2	3.98	57.5

Chapter 3

Combination of the extended scaled particle theory and Poisson-Boltzmann equation: applications to hydration free energies alkanes and alcohols

Abstract

A hybrid approach of the extended scaled particle theory (SPT) and the Poisson–Boltzmann (PB) equation is proposed for calculating the solvation free energy of the solute with partial charges in dilute aqueous solution. The applicability of this method is demonstrated by taking a series of the normal alcohols as an example along with normal alkanes. The solvation free energy of normal alkanes which has been studied previously based on a rather crude model is recalculated using more elaborate model. The electrostatic contribution to the free energy for the polar solute is calculated by the PB equation. In order to take into account the complicated boundary condition associated with the shape of molecular surface, the PB equation is solved numerically using a finite difference method. A superposition of hydrophobic and the electrostatic contributions to the solvation free energy exhibits a reasonable agreement with corresponding experimental data for the polar solute. Possibility of further applications of the method is discussed.

Introduction

Increasing attention has been paid to the theoretical determination of the solvation free energy from a variety of fields including the stability analysis of protein and the chemical reaction rate in the condensed phase. In spite of the remarkable success of the molecular simulation with computers for the past few decades [1], it is still one of the most difficult tasks for the method to determine the solvation free energy mainly due to the overwhelmingly large configurational space to which solvent molecules can access. The integral equation approach has been very successful in studying the solution consisting of relatively small solutes, but its application to large molecules such as polymers and proteins is still under development [2]. An alternative empirical approach based on the solvent accessible surface area (ASA) has been widely employed among polymer and protein chemists, of which success has been marked by the prediction of the cold denaturation of proteins [3]. The ASA method has been applied even to a quantum-chemical calculation to investigate the solvent effect upon the electronic structure of a molecule in condensed phase. However, the success of the method largely depends on experimental parameters assigned to the constituent atom groups [4]. This uncomfortable situation urged us to develop a new method [5] extended from the scaled particle theory (SPT) for the determination of the solvation free energy where the solute can take an arbitrary shape. The theory gives a rigorous description for a liquid system consisting of hard spheres [6]. The applicability of the method to a wider class of systems such as aqueous solutions has already been demonstrated for the solubility of small nonpolar molecules in water by Pierotti [7]. Recently, the physical origin of the anomalous decrease in entropy in the solution process of solute from gas to aqueous environment

was investigated by Soda [8] based on an analysis of the free energy of the cavity formation calculated from SPT. The author has attributed the large decrease in entropy upon the insertion of a molecule into solvent to the decrease in volume available to solvent molecules. Although there have been significant efforts devoted to extend the theory to cover non–spherical molecules, the mathematical problems associated have confined the extension within rather limited class of molecules with a convex shape [9,10]. In the preceding chapter [5], we have proposed an extended version of SPT based on the Richmond method [11] of calculating the surface area and the volume of complex molecules, which is applicable to solute with an arbitrary shape including that with non-convex shape. In the chapter 2, we only outlined the method and reported the solvation free energies of normal alkanes as fused spheres with equal radii. Results for revised model of normal alkanes in which the fused spheres have unequal radii are reported in the this chapter.

The main objective of the present chapter is to extend the method further to be able to include polar and ionized solutes. A crucial problem to be solved in calculating the solvation free energy for those solutes is how to evaluate the electrostatic contribution or the Born energy [12]. Fully statistical-mechanical treatment of the problem has again been limited to relatively small molecules at this stage. A possible alternative is to employ the Poisson-Boltzmann (PB) equation which is based on the continuum solvent model [12,13]. The equation is known to give reasonable results for the Born energy even for protein provided the boundary condition is properly adjusted [13-15]. We employ the method in order to evaluate the Born energy of polar molecules. The boundary problem representing the complicated shape of molecular surface is solved numerically using the finite difference method [14,15]. A notorious drawback of the PB approach has been its inability to take account for the hydrophobic part of the free

energy. On the other hand, the Pierotti method based on SPT just takes care of the hydrophobic contribution in terms of the cavity formation and the attractive interaction between the solute and the solvent. Therefore, a hybrid between the two methods gives entire solvation free energy of polar and ionized solute working complementarily each other. We demonstrate the applicability of the method by taking a series of the normal alcohols as an example.

Method

In the following two sections, we give rather instructive description of the extended SPT. The details of the derivation of the equation of the volume of the fused spheres used in the extended SPT is given elsewhere [16]. In the section III, some details of the numerical calculations is given including the PB procedure.

We employ the Pierotti's recipe [7] to calculate the solvation free energy, $\Delta\bar{g}_s^o$(g→m), of a series of normal alcohols in water as well as normal alkanes. The dissolution process of the solute into dilute solution is decoupled into three steps by a thought experiment: (I) creating a cavity in solvent, which has a right size as well as a shape to accommodate the solute molecule; $\bar{g}_C$, (II) establishing the Lennard-Jones type interaction between the solute molecule and solvent; $\bar{g}_I$, (III) coupling the electrostatic interaction between the solute and the solvent molecules; $\bar{g}_B$. Then the solvation free energy can be written as a sum of the three terms:

$$\Delta\bar{g}_s^o(g\rightarrow m)=\bar{g}_C+\bar{g}_I+\bar{g}_B \quad (1)$$

I. Extended Scaled Particle Theory

We consider the rigid body system consisting of one solute molecule with an arbitrary shape and solvent molecules with a spherical shape. The solute molecule is modeled as the fused hard spheres with different radii and the solvent molecules are modeled as hard spheres with the same radius. This represents the system of aqueous solution of an

organic molecule. All hydrogen atoms in the system are regarded as being absorbed into the united atoms. The problem is to find an expression for the reversible work required to dissolve the solute in the solvent. In this section, at first step SPT is explained in the manner based on the work of Gibbons, R. [9] which is the first attempt to apply SPT to the non-spherical particle system. SPT is the theory to calculate $\bar{g}_C$ by scaling up a solute molecule in the solvent,

$$\bar{g}_c = W(\lambda = 1) \quad (2)$$

where λ is the scaling parameter varying from 0 (material point) to 1 (real size), $W(\lambda)$ is the work required to dissolve the scaling solute particle to the system. The theory assumes that $W(\lambda)$ can be represented for all positive values of λ by the following equation,

$$W(\lambda) = A + B\lambda + \frac{1}{2}C\lambda^2 + PV_c(\lambda) \qquad \lambda \geq 0 \quad (3)$$

where $V_C(\lambda)$ is the excluded volume of the scaling solute.where P is the macroscopic pressure of the solvent. The term PV_c is introduced since at very large values of λ the leading term in $W(\lambda)$ must be PV_c. This term is negligibly small at λ=1 under the pressure of one atmosphere. The cavity of the scaling molecule has non-zero volume even at λ=0. The equation of $W(\lambda)$ for negative values of λ is known ,

$$W(\lambda) = -k_B T \ln(1 - \rho V_C(\lambda)) \quad . \quad (4)$$

The coefficients A, B, and C are determined using the continuity up to second derivatives at λ=0. The explicit expression of A, B and C are following;

$$A=-k_BT\ln(1-\rho V_c(0)) \quad , \quad B=-k_BT\frac{1}{1-\rho V_c(0)}\rho\left(\frac{\partial V_c}{\partial\lambda}\right)_{\lambda=0} \quad ,$$

$$C=-k_BT\frac{1}{1-\rho V_c(0)}\rho\left(\frac{\partial^2 V_c}{\partial\lambda^2}\right)_{\lambda=0}+k_BT\frac{1}{(1-\rho V_c(0))^2}\rho^2\left(\left(\frac{\partial V_c}{\partial\lambda}\right)_{\lambda=0}\right)^2 \quad , \tag{5}$$

where k_B is the Boltzmann constant, T is the absolute temperature, ρ is the number density of the solvent, the derivatives of $V_c(\lambda)$ at $\lambda=0$ are the limit when the positive λ approaches 0. This expression in our extended SPT includes the case of the single sphere solute studied in the original SPT and the case of the hard dumbbell solute studied by Boublík [17]. If the solute is a convex molecule, these equations are the same as those of Gibbons [9] and Boublík [10]. Validity of the scaled particle theory for nonspherical particles has been demonstrated by comparing the compressibility factors with those of Monte Carlo simulation [16,18].

The reversible work for creating a cavity with an arbitrary shape can be written as follows [10],

$$W(\lambda)=k_BT\left[-\ln(1-\rho V_c(0))+\rho\int_0^1\Gamma(\lambda)\left(\frac{\partial V_c(\lambda)}{\partial\lambda}\right)d\lambda\right] \tag{6}$$

$\Gamma(\lambda)$ is the distribution function of solvent centers at the surface of the cavity. The first term in Eq. (6) is the work required to dissolve the material point as the seed of the solute molecule into the pure solvent. The second term represents the work required for scaling up the solute molecule from $\lambda=0$ to $\lambda=1$. The meaning of the second term is rather obvious since it is a product of the pressure of solvent exerted on the surface of the cavity, $k_BT\rho\Gamma(\lambda)$, and the infinitesimal increment of the volume of the cavity.

The second step is to make analytical formulations of excluded

volume, $V_C(\lambda)$, and its first and second derivatives. Because in this work a solvent molecule is regarded as a single sphere, the excluded volume of the solute molecule is exactly the same as a volume of fused spheres in which the value of the radius of each sphere is the radius of the atom forming the solute molecule plus the radius of the solvent molecule. The analytical formulation of the volume of the fused spheres with unequal radii is explained in the appendix. The topological nature of the fused spheres are essential to calculate the volume. Because in most of the cases, topological nature is not changed near λ=0 in the course of the scaling, the derivatives of the excluded volume are analytically derived using topological nature at λ=1.

II. Lennard-Jones interaction of the solute molecule with the solvent.

To estimate the contribution from the Lennard-Jones interaction $\bar{g}_l$, we also utilized Pierotti's method [19]. Because the solvent entropy change caused by coupling the Lennard-Jones type interaction is thought to be small, $\bar{g}_l$ is approximately equal to the corresponding enthalpy, $\bar{h}_l$, due to the attractive force between solute and solvent. Then $\bar{h}_l$ is estimated as an integrated value of the Lennard–Jones type interactions between the solute and solvent molecules. A following step–wise function was assumed for the distribution function of solvent molecules around a solute molecule,

$$\begin{cases} \eta(x,y,z)=0 \cdots (\exists\ b;\ d_{bv} \le r_b + r_v) \\ \eta(x,y,z)=1 \cdots (\forall\ b;\ d_{bv} > r_b + r_v) \end{cases}, \quad d_{bv}=\left\{(x-x_b)^2+(y-y_b)^2+(z-z_b)^2\right\}^{1/2} \tag{7}$$

where r_a and r_b are the radius of atom b and the solvent molecule, respectively, (x_b,y_b,z_b) are Cartesian coordinates of a center of the solute atom b. In this study, we calculate the value of $\bar{h}_l$ in the same way as

Pierotti has done [20] using distribution function $\eta(x,y,z)$ to cover a multi-site molecules.

$$\bar{h}_I=\sum_a \bar{h}_{Ia} \quad , \quad \bar{H}_{Ia}=-C_a \iiint \rho\eta(x,y,z)\left(\frac{1}{d_{av}^6}-\frac{\sigma_{av}^6}{d_{av}^{12}}\right)dxdydz \qquad (8)$$

$$C_a=4(\varepsilon_a\varepsilon_v)^{1/2}\sigma_{av}^6 \quad , \quad \sigma_{av}=(\sigma_a+\sigma_v)/2 \quad ,$$

where $\bar{h}_{Ia}$ is the enthalpy contribution from atom a, ε_v and σ_v are the Lennard-Jones parameters for water (as the solvent), ε_a and σ_a are the Lennard-Jones parameters for solute atom a. The calculation is performed numerically using equally spaced 3-d grids of which size is 0.02 Å. When the topology of the solute molecule is very simple, analytical calculation of $\bar{h}_a$ can be done using expression of accessible surface area of the solute molecule. We calculated the values of $\bar{h}_a$ for methane, ethane, and methanol analytically in order to check the corresponding numerical results.

III. Electrostatic interaction between the solute molecule and the solvent

To evaluate the hydrophilic part of the free energy change due to the electrostatic interaction (Born energy), we numerically solve the Poisson-Boltzmann equation of the solute-solvent system using a finite difference method [13-15] based on the macroscopic dielectric model [21]. We used essentially the same algorithm as that of Warwicker et al. [13] but modified the algorithm in a few points in the treatment of the boundary conditions [14,15]. The system is divided into two regions: solute region and solvent region. In this case, the dielectric constants of the solute molecule and the water are 1 and 78.4, respectively. The boundary between these dielectric media is just the surface of the solute molecule except hydrogen atoms bound to heteroatoms (see Fig. 1).

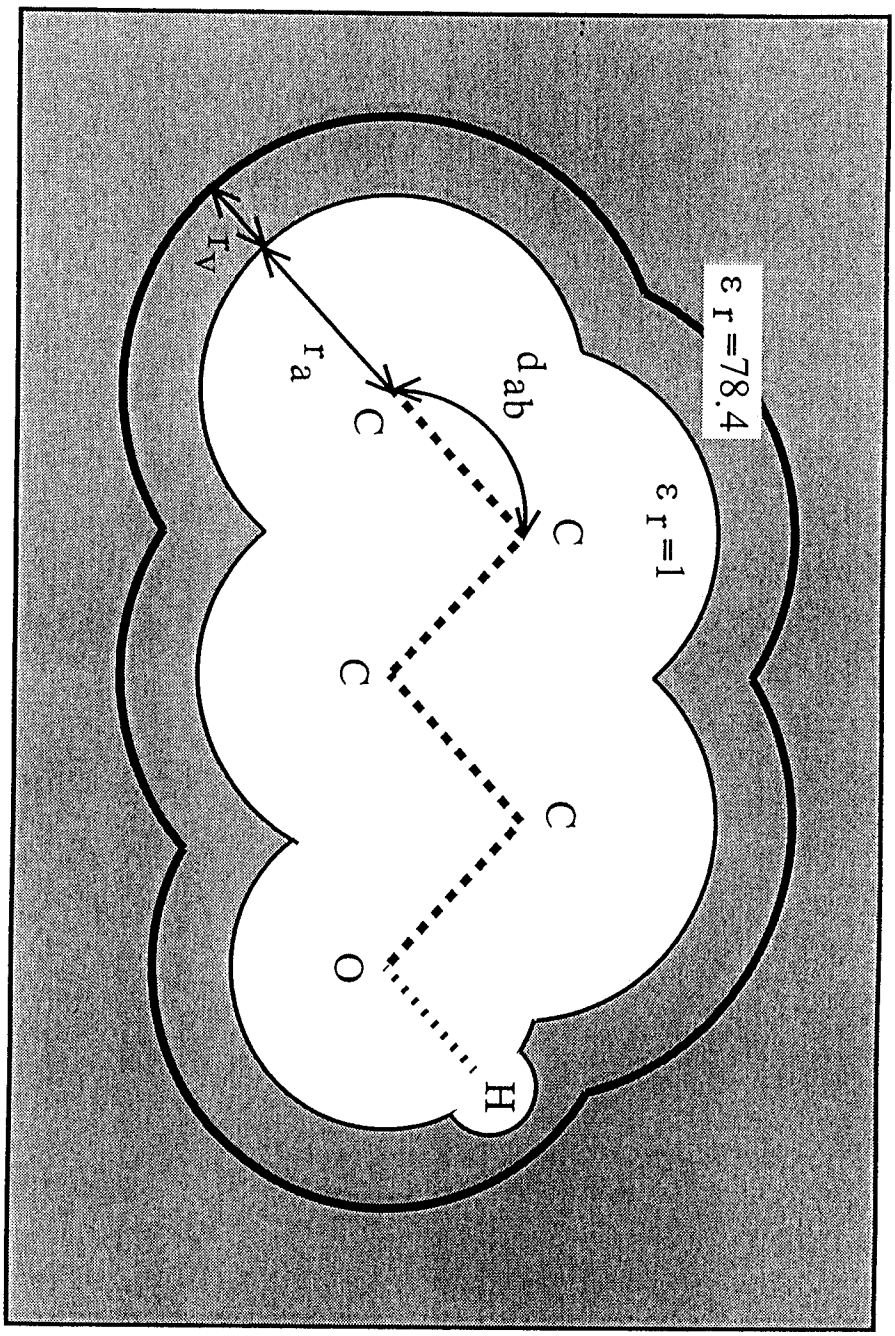


Fig. 1 Models of organic molecule, normal–alcohol in the trans-conformation, as fused hard spheres. The boundary of the excluded volume is represented by the thick solid lines. The radius of the solvent molecule is r_V. The dielectric medium as the solvent (water) is represented by the shaded region.

For hydrogen atoms bound to heteroatoms, which have partial charges, we use the dielectric boundary radius of 1.0 Å in order to avoid a situation in which the hydrogen charge comes too close to water [22]. To see the effect of the grid size, we performed the numerical calculation using different grid sizes, and found that the effect was negligible in case of 0.25 Å with which the present work were carried out. The Debye screening constant κ is set to zero corresponding to the condition of infinite dilution. Our program which used in this work to solve the Poisson-Boltzmann equation will be registered in JCPE (Japan chemistry program exchange) in this year from which anyone can get our program in the similar way as QCPE.

Energy and structure parameters

As has been mentioned already, solute hydrogen atoms are not treated explicitly, but are treated as absorbed into united atoms. Each radius r of the hard sphere making fused spheres is a half the Lennard-Jones diameter σ of the united atom. The Lennard–Jones parameters used in section II and electrostatic partial charges used in section III are listed in table 1. The TIPS parameters determined by Jorgensen are employed for the extended atoms CH_2, CH_3, CH_4, O in ROH, and H in ROH [23]. The parameters for water are those determined by Pierotti from the solubility data of inert gas in water [20]. The effective hard-sphere radius (1.375 Å) for water is similar to those commonly used in the study of biopolymers (1.4 Å) [24]. The structure parameters of normal-alkanes and alcohols are the same as those in the Jorgensen's paper [23].

Results

The calculated values of the solvation free energy for the normal–alkane in the trans-conformation are plotted against the number of carbon atoms as shown in Fig. 2. There have been an unsettled controversy concerning the experimental data for the solvation free energy at infinite dilution, with which the theoretical results should be compared. The experimental data can be abstracted from the solubility data according to the following equation,

$$\Delta\bar{g}_s^o(g \rightarrow m) = -k_B T \ln\left(\frac{\rho_s^g}{\rho_s^m}\right) + \chi \quad (9)$$

where ρ_s^g and ρ_s^m are the number densities of the solute molecules in the gas phase and in the mixture, respectively. The second term in Eq. (9) is a correction proposed by Sharp et al., [25, 26] which somehow takes account for the volume difference between solute and solvent, and it has the same expression with what appears in the Flory-Huggins theory [27, 28],

$$\chi = -k_B T\left(1 - \frac{v_s}{v_v}\right) \quad (10)$$

where v_s and v_v are the partial molar volumes of the solute s and solvent v, respectively. It is the unsolved question whether the correction is necessary or not. Since it is not the purpose of this article to solve the question, we leave it to future studies and plot both data in Fig. 2.

In our calculation, the method for deriving the solvation free energy of methane (n=1) is identical to that of Pierotti [20] except for the effective hard–sphere radius of methane. The theoretical results shows fairly good agreement with the both experimental results with and without Sharp's

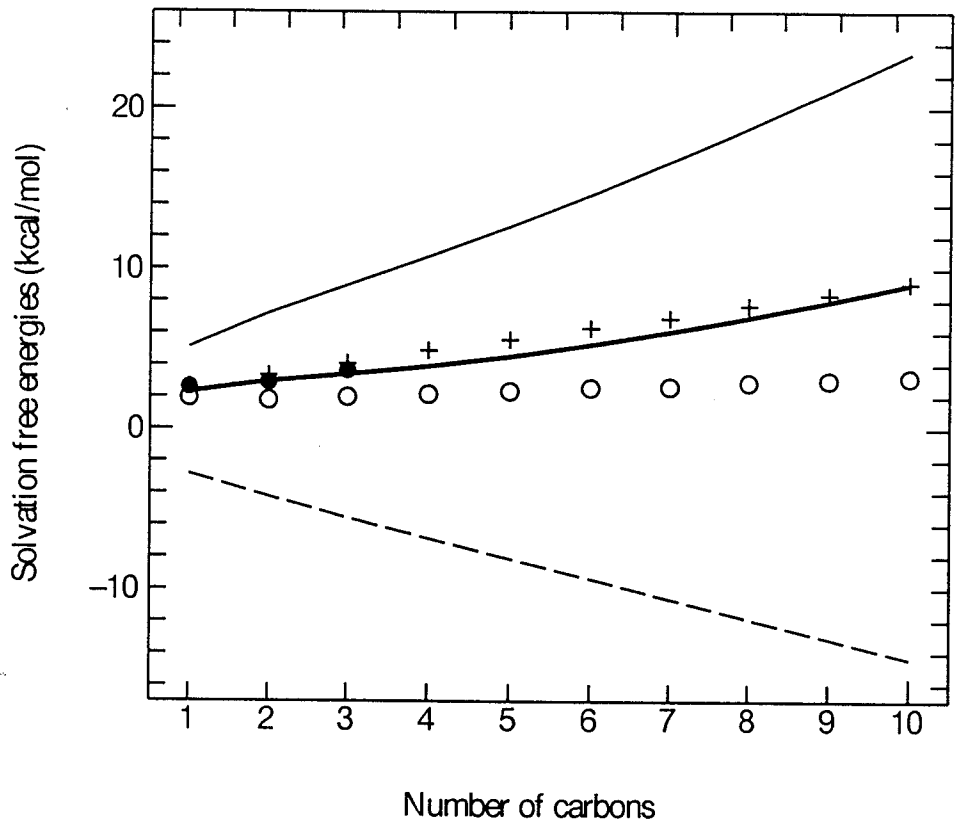


Fig. 2 Plot of the calculated values of the solvation free energy for the normal–alkane against the number of carbon atoms. The thick solid, thin solid and dashed lines represent $\bar{g}$, $\bar{g}_C$ and $\bar{g}_I$, respectively. The experimental values of $\bar{G}$ interpreted by Sharp et al. (● partial molar volume is used for v_v, + molar volume is used for v_v) and without correction term (○) are also plotted.

correction, although better agreement was obtained with that using Sharp's correction.

The calculated values of the solvation free energy for normal-alcohol in the trans-conformation are plotted against the number of carbon atoms as shown in Fig. 3. The corresponding experimental values are also plotted in the same figure. The theoretical results shows good agreement with the both experimental values with and without Sharp's correction. The agreement is better again with the experimental results corrected by Sharp et. al. Our results for the Born energies of methanol and ethanol are also in good agreement with those obtained from the free energy perturbation method, -7.1 and -7.4, and with those calculated from the generalized Born equation, -7.1 and -6.4 [22].

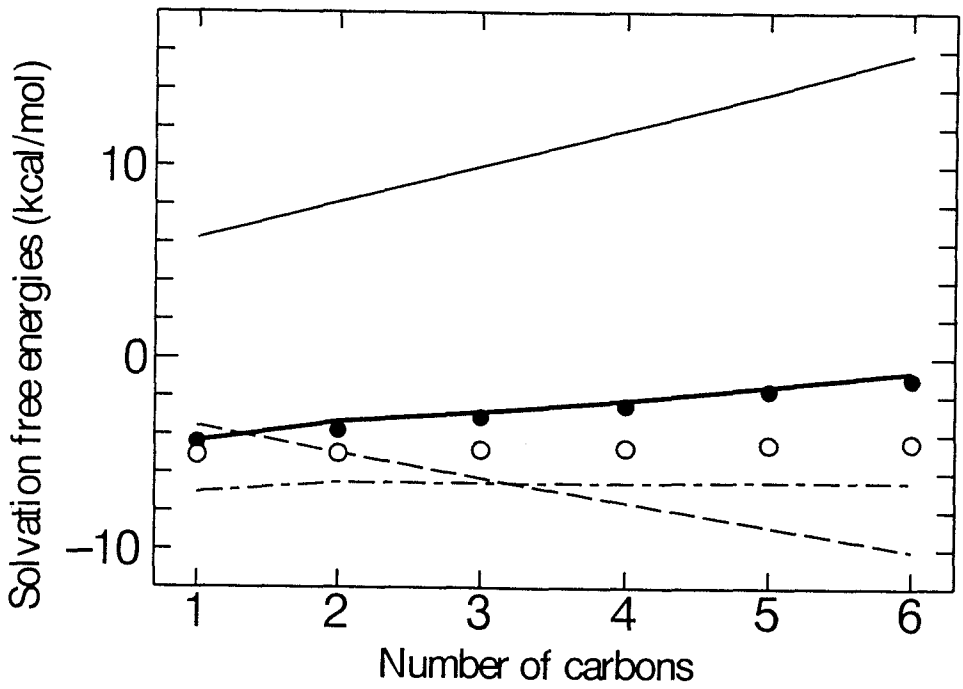


Fig. 3 Plot of the calculated values of the solvation free energy for the normal–alcohol against the number of carbon atoms. The thick solid, thin solid, dashed, and dashed-dot lines represent $\bar{g}$, $\bar{g}_C$, $\bar{g}_I$, and $\bar{g}_B$ respectively. The experimental values of $\bar{G}$ interpreted by Sharp et al. (● partial molar volume is used for v_v) and without correction term (○) are also plotted.

Discussion

It is not surprising why SPT works so well for liquids which consist of molecules interacting via short range potential such as the Lennard-Jones interaction, since thermodynamic quantities of those liquids are largely determined by the repulsive part of the potential. However, why the theory is successful also for aqueous system has been a puzzle since Pierotti published the famous article concerning the solubility of inert gases in water, because the liquid structure of water is largely determined by hydrogen-bonds between a pair of molecules and the hydrogen-bonding network and its modification due to the solute perturbation are believed to play an essential role in determining the solvation free energy. The modern computer simulation which of course incorporates such a highly non-central interaction seems to have been giving reasonable, if not perfect, agreement with SPT as long as the solvation free energy is concerned. It is therefore worthwhile to express our opinion for this problem even though the discussion is largely speculative in this stage.

According to SPT, the solvation free energy in water is characterized by the large positive contribution from the work required to make a cavity. The work is proportional to the local pressure which in turn is proportional to the number density of solvent as indicated in Eq. (6). It is therefore reasonable to attribute the large energy for the cavity formation to the large number density of water. Then, next question will be what distinguishes water from other solvent. In another word, what is the anomaly of water? The fact that water maintains liquid state at such high density will be the anomaly of the solvent, which can be possible only due to the strong attractive electrostatic interaction (hydrogen bond). Recently, Wakai and Nakahara presented interesting experimental results (personal communication) which concern the pressure dependence of the virial for a variety of liquid.

They have found the virial at room temperature for water stays essentially constant over a pressure range from 1 MPa to 300 MPa, while CH_3CN, $CHCl_3$ and C_6H_6 exhibit large dependence on the pressure. The results suggest that water is highly compressed liquid already at standard state due to the harsh balance between the repulsive and attractive forces. The low compressibility of water compared to other liquids support the view. Although SPT does not explicitly take account for the structural aspect of water, it reflects the most important nature of water implicitly through the number density which is anomalously large compared to other solvent.

Our results for both normal alkane and alcohols apparently support the interpretation proposed by Sharp et. al. However, the results should not be regarded conclusive considering various approximations with which the theory is associated at this stage. The approximation include (1) making solvent distribution uniform beyond the distance of closest approach in the calculation of $\bar{g}_l$, (2) using continuum dielectrics to evaluate the electrostatic contribution, and (3) neglecting contribution from the conformational isomers which possibly plays a significant role in solution. Further extensive study is needed to elucidate significance of those approximations. Rather encouraging for the theory is the fact that the theoretical results show reasonable agreement with either of the two experimental results after cancellation of the large positive contribution from the cavity formation and the attractive contribution due to the Lennard-Jones as well as the electrostatic interaction. The resolution of the theory must be sufficient for most of applications which concern a qualitative nature of the solvation process.

In fact, our theory reproduces qualitatively the difference in the solubility, between two typical solutes, hydrocarbons and alcohols, which respectively represent those solutes with low and high solubilities. It should be noted that the present theory is free from any experimental parameters

except for the dielectric constant of solvent which is employed for the evaluation of the electrostatic contribution.

For an actual application of the theory to a phenomenon in which only difference of the solvation free energy matters, the problem concerning the two experimental data becomes even less significant, because the error associated with the correction will be largely canceled out anyway. One example of such possible applications will be the thermal and cold denaturation of protein. A study has been carried by Ooi and Oobatake [3] based on the method of the accessible surface area with empirical energy parameters for the atom groups. They have calculated the free energy difference between the native state of the protein and one of the possible extended states (a denatured state), and could have successfully predicted the thermal as well as cold denaturation. The cold denaturation occurs due to the unfavorable change in the entropic contribution as temperature decreases, of which essential part is the free energy of the cavity formation. It is almost obvious that each part of the decomposed free energy in the present treatment shows different behavior with the conformation of protein and with the temperature change. If so, it is of great interest to calculate the free energy difference between the two conformations based on the method presented here, and to see how each part of the decomposition behaves as temperature changes.

Conclusions

In this chapter, we have presented a new method to evaluate the solvation free energy of polar and non-polar solute by coupling the extended scaled particle theory and the Poisson-Boltzmann equation. Considering the rather crude approximation employed for the model for solute and solvent, the theoretical results show a reasonable agreement with the experimental data. Although there remains few questions in the quantitative aspect in the method, we conclude that the theory can be applicable to more interesting classes of solvation phenomena including the thermal and cold denaturation of protein.

Appendix

Analytical formulation of the volume of the fused spheres with unequal radii

The body of fused spheres can be divided into the pieces of bodies belonging to the single spheres by the Dodd's definition. The volume of each piece, V_a, can be calculated from [29],

$$V_c(\lambda)=\sum_a V_a(\lambda) \quad , \quad V_a=\frac{1}{3}R_a S_a^A+\sum_{LST2'(a)}\frac{1}{3}F_{ab}S_{ab}^I \quad , \quad F_{ab}=\frac{D_{ab}^2+R_a^2-R_b^2}{2D_{ab}} \quad ,$$

(A1)

where a runs through all spheres considered, $LST2'(a)$ runs through spheres with which the atom a intersects to form inaccessible surfaces (see Fig. 4a), S_a^A and S_{ab}^I are accessible and inaccessible surface area of the sphere a, R_a is the radius of the sphere a, and D_{ab} is the distance between the sphere a and b. The piece belonging to each sphere has the intersection circles, exposed vertices, buried vertices which are important topological elements used soon (figure 4a). Eq. (A1) shows that the each piece can be divided into the smaller pieces consisted of cone-pyramids and spherical sectors [29] (see Fig. 4b).

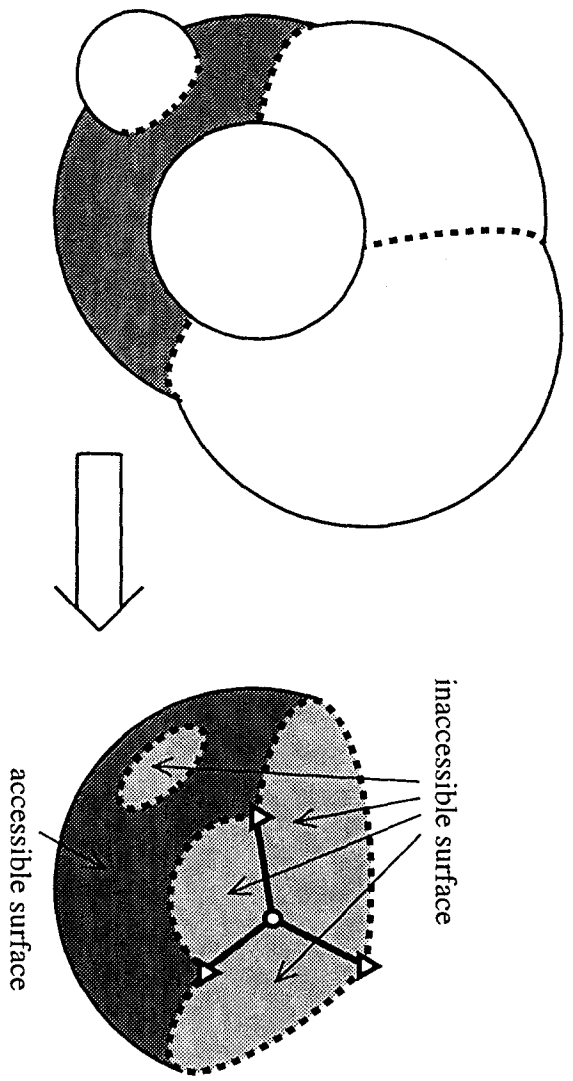


Fig. 4a Accessible and inaccessible surface of a component of fused spheres. The thick dashed line represents intersection circles (including one isolated intersection circle). The symbols ○ and △ show the locations of the exposed and buried vertices respectively.

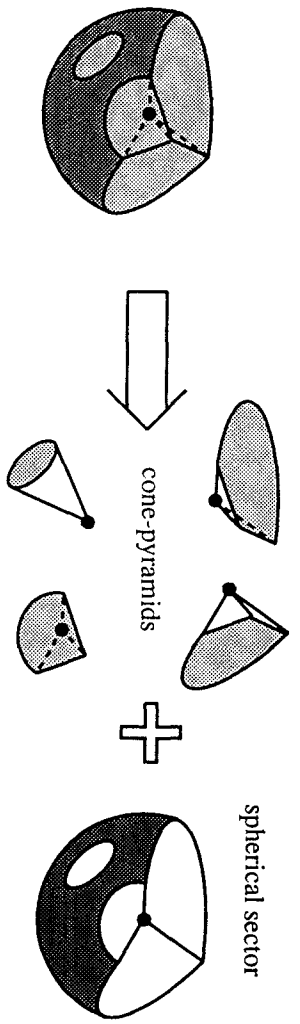


Fig. 4b Cone-pyramids and spherical sectors. The symbol ● shows the location of the center of the sphere (for the case of the left body, the center of the sphere is located inside the body).

The elegant analytical equation of V_a as the function of only sphere radii and inter-sphere distances, was derived by employing the Richmond method and that of S_{ab}^{I} is derived by one of us [16]. We originally started from the Eq. (A1), but after rearranging a lot of equations paying attention to the topologies of the pieces, we finally got the elegant equation. The only topological information needed to get V_a are what follows,

1) Target sphere, *a*, is isolated (having volume of 4/3 πR_a^3) or not.

2) The list of spheres (LST2(a)) making isolated intersecting circles with the sphere *a*.
   The component (b) of this list (LST2(a)={b}) is the sphere index which makes the isolated intersecting circle with sphere i.

3) The list of pairs of spheres(LST3(a)) making the exposed vertices of the sphere *a*.
   The component ((b,c)) of this list (LST3(a)={(b,c)}, b<c) contains the indices of the two spheres which make the exposed vertex with sphere i.

4) The list of triplets of spheres (LST4(a)) making the buried vertex of the sphere *a*.
   The component ((b,c,d)) of this list (LST4(a)={(b,c,d)}, b<c<d) contains the indices of the three spheres which make the buried vertex with the sphere i.

These topological information 1)-4) are obtained by the analytical calculation. From these information we can make the following lists.

0') The Euler-Poincaré characteristic, $\varkappa_a$, of the accessible surface of sphere a.

2') The list of spheres appeared in 2), 3), and 4). This is exactly the same as the list LST2'(a).

The volume, V_a, can be calculated for the fused spheres with virtually any shape as follows:

$$V_a = T_1(a) + \sum_{LST2'(a)} T_2(a,b) + \sum_{LST3(a)} \{T_{3a}(a,b,c) + T_{3b}(a,b,c) + T_{3b}(a,c,b)\} + \sum_{LST4(a)} \{T_4(a,b,c,d) + T_4(a,c,d,b) + T_4(a,d,b,c)\} \quad , \tag{A2}$$

$$T_1(a) = \frac{2}{3}\pi R_a^3 \chi_a^A \quad , \quad T_2(a,b) = \frac{2}{3}\pi R_a^2 F_{ab} + \frac{1}{3}\pi F_{ab} J_{ab}^2 \quad , \quad T_{3a}(a,b,c) = \frac{1}{3} R_a^3 \Omega(a,b,c$$

$$T_{3b}(a,b,c) = -\frac{1}{3} R_a^2 F_{ab} \Phi(a,b,c) + \frac{1}{6} F_{ab}\left[-J_{ab}^2 \Phi(a,b,c) + L(a,b,c)\left\{J_{ab}^2 - L(a,b,c)^2\right\}^{1/2}\right]$$

$$T_4(a,b,c,d) = -\frac{1}{3} R_a^2 F_{ab} \Psi(a,b,c,d) + \frac{1}{6} F_{ab}\left[-J_{ab}^2 \Psi(a,b,c,d) + M(a,b,c,d)\right] \quad ,$$

$$F_{ab}=\frac{D_{ab}^2+R_a^2-R_b^2}{2D_{ab}} \quad , \quad J_{ab}=\left\{R_a^2-F_{ab}^2\right\}^{1/2} \quad , \quad L(a,b,c)=\frac{-F_{ab}\cos\theta_{abc}+F_{ac}}{\sin\theta_{abc}} \quad ,$$

$$M(a,b,c,d)=\frac{2L(a,b,c)L(a,b,d)-(L(a,b,c)^2+L(a,b,d)^2)\cos\Psi(a,b,c,d)}{\sin\Psi(a,b,c,d)}$$

$$\Omega(a,b,c)=-\cos^{-1}\left(\frac{R_a^2\cos\theta_{abc}-F_{ab}F_{ac}}{J_{ab}J_{ac}}\right) \quad , \quad \Phi(a,b,c)=\cos^{-1}\left(\frac{L(a,b,c)}{J_{ab}}\right) \quad ,$$

$$\cos\theta_{abc}=\frac{D_{ab}^2+D_{ac}^2-D_{bc}^2}{2D_{ab}D_{ac}} \quad , \quad \sin\theta_{abc}=\left\{1-\cos^2\theta_{abc}\right\}^{1/2} \quad ,$$

$$\cos\Psi(a,b,c,d)=\frac{D_{ac}^2\sin^2\theta_{abc}+D_{ad}^2\sin^2\theta_{abd}-D_{cd}^2+(D_{ad}\cos\theta_{abc}-D_{ac}\cos\theta_{abd})^2}{2D_{ac}D_{ad}\sin\theta_{abc}\sin\theta_{abd}} \quad ,$$

$$\sin\Psi(a,b,c,d)=\left\{1-\cos^2\Psi(a,b,c,d)\right\}^{1/2} \quad ,$$

It is easily understood from our equation that the topological nature, ie. exposed and buried vertices, are important to calculate the excluded volume. The detail of the analytical calculation of the volume, including the analytical equation of the accessible surface area in the same manner of the volume, is explained elsewhere [16].

In our extended scaled particle theory, the excluded volume of the scaling solute molecule is calculated by setting the radius and the distance as follows,

$$R_a=r_a\lambda+r_v \quad , \quad D_{ab}=d_{ab}\lambda \quad \text{(A3)}$$

where r_a and r_b are the radius of the solute atom a and the solvent molecule, respectively, d_{ab} is the distance between atom a and b, λ is the scaling parameter. First and second derivatives of the excluded volume by the scaling parameter are also derived analytically from Eq. (A3), because our expression of the excluded volume is the function of only atomic radii

and interatom distances.

Table 1

Lennard–Jones parameters of water and extended atoms

Compd	σ,Å	ε/k_B,K	q,electrons
H_2O	2.75	85.3	0
CH_4	3.73	148.0	0
CH_3	3.86	91.2	*
CH_2	3.98	57.5	*
O in ROH	3.08	88.0	-0.685
H in ROH	0	0	0.40

* partial charges chosen to achieve neutrality of monomers in the same manner in TIPS.

References

(1) Special issue., *Chemical Physics*, **158**, No.2,3 (1991)

(2) Kitao, A., Hirata, F., and Gō, N., *Chem. Phys.*, **158**, 447 (1991)

(3) Ooi, T., and Oobatake, M., *J. Biochem.*, **103**, 114 (1988)

(4) Ooi, T., Oobatake, M., Némethy, and G., Sheraga, H.A., *Proc. Natl. Acad. Sci. USA*, **84**, 3086 (1987)

(5) Irisa, M., Nagayama, K., and Hirata, F., *Chem. Phys. Letters*, **207**, 430 (1993)

(6) Reiss, H., Fish, H.L., and Lebowitz, J.L., *J. Chem. Phys.*, **31**, 369 (1959)

(7) Pierotti, R.A., *Chem. Rev.*, **76**, 717 (1965).

(8) Soda, K., *J. Phys. Soc. Jpn.*, **62**, 1782 (1993)

(9) Gibbons, R.M., *Mol. Phys.*, **17**, 81 (1969).

(10) Boublík, T., *Mol. Phys.*, **27**, 1415 (1974)

(11) Richmond, T.J., *J. Mol. Biol.*, **178**, 63 (1984)

(12) Rogers, N.K., *Progr. Biophys. Mol. Biol.*, **48**, 37 (1986)

(13) Warwicker, J. and Watoson, H.C, *J. Mol. Biol.*, **157**, 671 (1982)

(14) Nakamura, H., *J. Phys. Soc. Jpn.*, **57**, 3702 (1988)

(15) Takahashi, T., Nakamura, H., and Wada, A., *Biopolymers*, **32**, 897 (1992)

(16) Irisa, M., in preparation.

(17) Boublík, T., *Mol. Phys.*, **44**, 1369 (1981)

(18) Boublík, T., *Adv. Chem. Ser.*, **204**, 173 (1983)

(19) Pierotti, R.A., *J. Phys. Chem.*, **67**, 1840 (1963)

(20) Pierotti, R.A., *J. Phys. Chem.*, **69**, 281 (1965)

(21) Tanford, C, and Kirkwood, J.G., *J. Am. Chem. Soc.*, **79**, 5333 (1957)

(22) Still, W.C., Tempczyk, A., Hawley, R.C., and Hendrickson, T., *J. Am. Chem. Soc.*, **112**, 6127 (1990)

(23) Jorgensen, W.L., *J. Am. Chem. Soc.*, **103**, 335 (1981)

(24) Richards, F.M., *Ann. Rev. Biophys. Bioeng.*, **6**, 151. (1977)

(25) Sharp, K.A., Nicholls, A., Friedman, R., and Honig, B., *Biochemistry*, **30**, 9686 (1991)

(26) Sharp, K.A., Nicholls, A., Fine, R.M., and Honig, B., *Science,*, **252**, 106 (1991)

(27) Flory, P.J., *J. Chem. Phys.*, **9**, 660 (1941)

(28) Huggins, M.L., *J. Chem. Phys.*, **9**, 440. (1941)

(29) Dodd, L.R., and Theodorou, D.N., *Mol. Phys.*, **72**, 1313 (1991)

Chapter 4

Application to the proteins: hydration free energies of avian pancreatic polypeptide and actin

Abstract

A hybrid approach combining the extended scaled particle theory (SPT) and the Poisson-Boltzmann (PB) equation for the solvation free energy of non-polar and polar solutes has been proposed by us. This new method is applied for the calculation of the hydration free energy of two proteins, avian pancreatic polypeptide (36 residues) and actin (372 residues). In the hydrophobic part of the solvation free energies of these proteins, cavity formation in the water and attractive force between the solute and the solvent compensate each other. Born energy is much larger than hydrophobic terms in this hydration free energy, because hydrophilic residues are ionized in the water. Temperature dependences of the hydrophobic part of the hydration free energies of the avian pancreatic polypeptide in the native and extended conformations are also calculated. The extended conformation is a model for the conformation of denatured proteins. The native conformation is destabilized at low temperature in terms of the hydrophobic effects. This fact shows that the hydrophobic hydration can be the origin of cold denaturation. This work is the first step for free energy difference calculation in stability analysis of the proteins.

Introduction

Increasing attention has been paid to the theoretical determination of the solvation free energy from a variety of fields including the stability analysis of protein. Solvation free energies of molecules, especially large

ones, are very difficult to calculate due to the overwhelmingly large configurational space. This uncomfortable situation urged us to develop a new method extended from the scaled particle theory (SPT) [1] for the determination of the solvation free energy where the solute can take an arbitrary shape [2]. Furthermore a hybrid approach in which the extended SPT and Poisson-Boltzmann equations work complementarily was proposed for the solvation free energy of the non-polar and polar solutes by us. We have demonstrated the applicability of this new method by taking a series of the normal alcohols as an example as well as normal alkanes in other place [3]. In this chapter, we explain the first attempt for the application of our new method to protein hydration.

Method

In the preceding chapters [2,3], we explained precisely about our method. Here, we explain the method briefly and add more information relating to the application for protein. We employed the Pierotti's recipe [1] to calculate the solvation free energy, $\Delta\bar{g}_s^o(g\rightarrow m)$. The dissolution process of the solute is decoupled into three steps by a thought experiment: (I) creating a cavity in solvent, which has a right size as well as a shape to accommodate the solute molecule; $\bar{g}_C$, (II) establishing the Lennard-Jones type interaction between the solute molecule and solvent; $\bar{g}_I$, (III) establishing the electrostatic interaction between the solute molecule and the solvent molecules; $\bar{g}_B$. Then the solvation free energy can be written as a sum of the three terms:

$$\Delta\bar{g}_s^o(g\rightarrow m)=\bar{g}_C+\bar{g}_I+\bar{g}_B \quad (1)$$

I. Extended Scaled Particle Theory

We consider the rigid body system consisting of one solute molecule with an arbitrary shape and solvent molecules with a spherical shape. The solute molecule is modeled as the fused hard spheres with different radii and the solvent molecules are modeled as hard spheres with the same radius. This corresponds to the system of aqueous solution of an organic molecule.

The reversible work for creating the cavity with an arbitrary shape can be written as follows [4]

$$\bar{g}_C = k_B T\left[-\ln(1-\rho V_c(0)) + \rho\int_0^1 \Gamma(\lambda)\left(\frac{\partial V_c(\lambda)}{\partial\lambda}\right)d\lambda\right] \quad (2)$$

where $V_C(\lambda)$ is the excluded volume of the scaling solute, $\Gamma(\lambda)$ is the distribution function of solvent centers at the surface of the cavity. The first term in equ. (2) is the work required to dissolve the material point as the seed of the solute molecule into the pure solvent. The second term represents the work required scaling up the solute molecule from λ=0 to λ=1. The meaning of the second term is rather obvious since it is a product of the pressure of solvent exerted on the surface of the cavity, $k_B T\rho\Gamma(\lambda)$, and the infinitesimal increment of the volume of the cavity. The difficulty appeared in the calculation of equ. (2), is we must make analytical formulations of excluded volume, $V_C(\lambda)$, and first and second derivatives of it. Because in this work the solvent molecule are regarded as a single sphere, the excluded volume of the solute molecule is exactly the same as a volume of fused spheres in which the value of the radius of each sphere is the radius of the atom consisting the solute molecule plus the radius of the solvent molecule (Fig. 1). The analytical formulation of the volume of the fused spheres with unequal radii made by us, is explained in the

elsewhere [5].

II. Lennard-Jones interaction of the solute molecule with the solvent.

To estimate the contribution from the Lennard-Jones interaction $\bar{g}_l$, we also utilized Pierotti's formula [6]. Because the solvent entropy change caused by the establishment of the Lennard-Jones type interaction is thought to be small, $\bar{g}_l$ is approximately equal to the corresponding enthalpy, $\bar{h}_l$, due to the attractive force between solute and solvent. Then $\bar{h}_l$ is estimated as an integrated value of the Lennard-Jones type interactions between the solute and solvent molecules. In the calculation, individual contribution from each atom in the protein is calculated first. Because it is better to calculate difference of enthalpy of the solute atom between in analytically known hypothetical situation (alone in the solvent) and in real situation (surrounded by other atoms in the solute). Numerical integration with grid size of 0.08 Å is done for each atom in the protein.

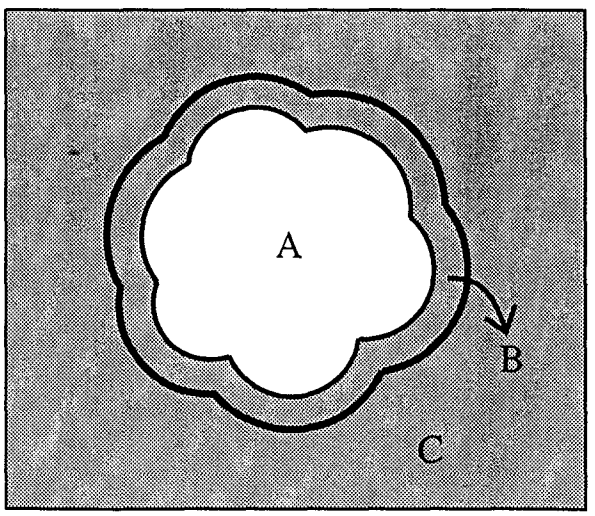


Fig.1 Schematic drawing of the system. A+B represents the excluded volume. Solute region (A) has the low dielecrtic constant and solvent region (B+C) has the high dielectric constant.

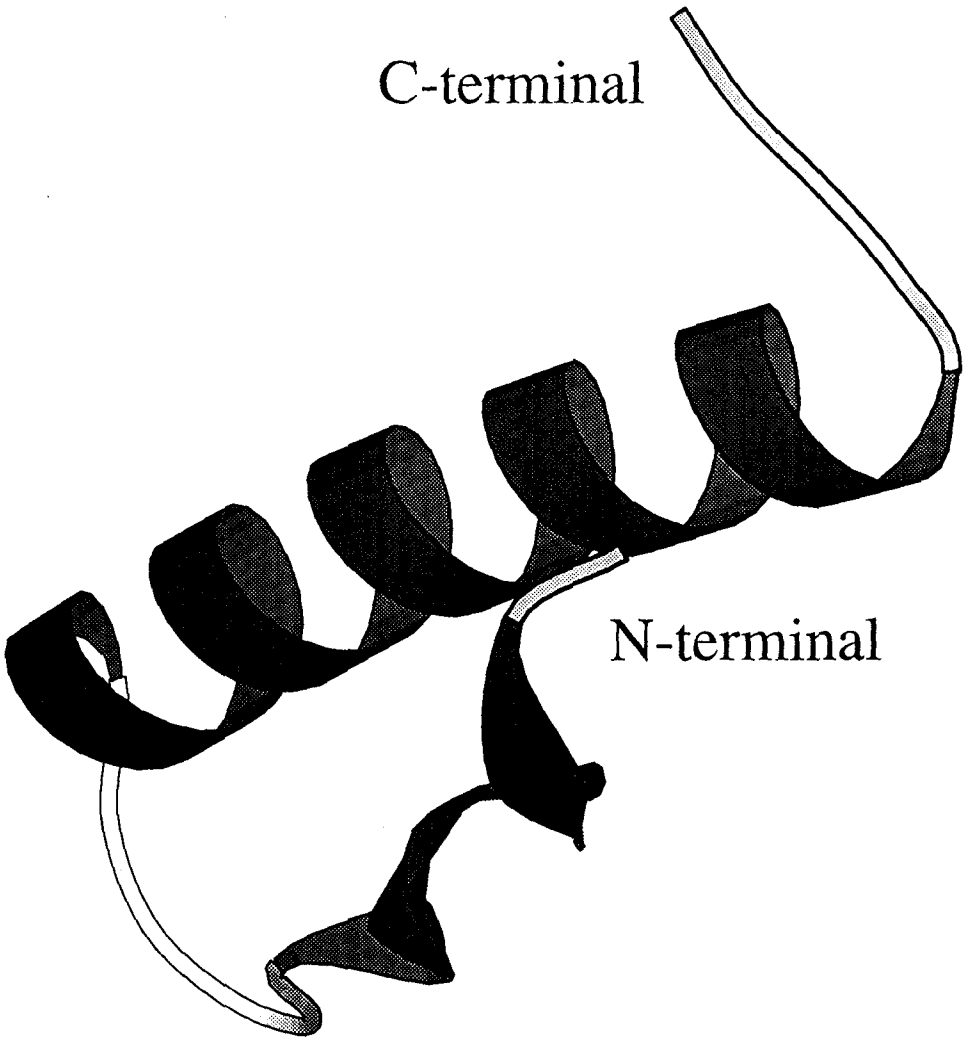


Fig.2 Avian pancreatic polypeptide (1PPT) [11] taken from protein data bank in BNL.

III.Electrostatic interaction between the solute molecule and the solvent

To evaluate the polar part of the free energy change due to the electrostatic interaction (Born energy), $\bar{g}_B$, we numerically solved the Poisson-Boltzmann equation of the solute-solvent system [7] based on the macroscopic dielectric model. Here, only the essence is mentioned. The system is divided into two regions: solute region and solvent region (Fig. 1). In this case, the dielectric constants of the solute molecule and the water are 1 and 78.4, respectively. The boundary between these dielectric media is just the surface of the solute molecule except hydrogen atoms bound to heteroatoms. For hydrogens bound to heteroatoms, which have charges, we used the dielectric boundary radius of 1.0 Å not to locate the charges too much close to the solvent having high dielectric constant [8]. In the numerical calculation of this work, the grid size is 0.5 Å and κ is set to zero. We treated the both terminals and hydrophilic residues of protein as ionized both in water and in vacuum .

Energy parameter and structure

The OPLS [9] parameters are employed for the united atoms and partial charges in protein. The parameters for water are those determined by Pierotti [10].

Avian pancreatic polypeptide:

The crystallographic structure of avian pancreatic polypeptide (1PPT) [11] is taken from protein data bank in BNL (Fig. 2). For electrostatic calculation, we add 67 hydrogen atoms, which are bound to heteroatoms, to crystallographic structure having 301 heavy atoms. In this structure, N- and C- terminals are positively and negatively ionized respectively and 3 residues (arginine) are positively ionized and 5 residues (asparagine and glutamine) are

Fig. 3 Actin (1ATN) [12] taken from protein data bank in BNL.

negatively ionized. Net charge of this polypeptide is -2 electron unit.

Actin:

The crystallographic structure of actin (1ATN) [12] is taken from protein data bank in BNL (Fig. 3). For electrostatic calculation, we add 628 hydrogen atoms, which are bound to heteroatoms, to crystallographic structure having 2907 heavy atoms. In this structure, N- and C- terminals are positively and negatively ionized respectively and 36 residues (arginine and lysine) are positively ionized and 50 residues (asparagine and glutamine) are negatively ionized. Net charge of this protein is -14 electron unit.

Temperature dependence of the solvation free energy

The calculation of the temperature dependence of the solvation free energy is a straight-forward application of our extended SPT. The only parameter affected by the temperature is the number density of the solvent in SPT. The solvation free energy at any temperature can be derived analytically by using the number density of the water. The number density of the water is determined as a 4-th polynomial function of temperature to reproduce the temperature dependence of the bulk water experiments of the number density. The parameters of the water molecule in SPT is constant and can be used for over the entire temperature range.

Results and Discussion

Avian pancreatic polypeptide:

The accessible surface area and excluded volume of this protein

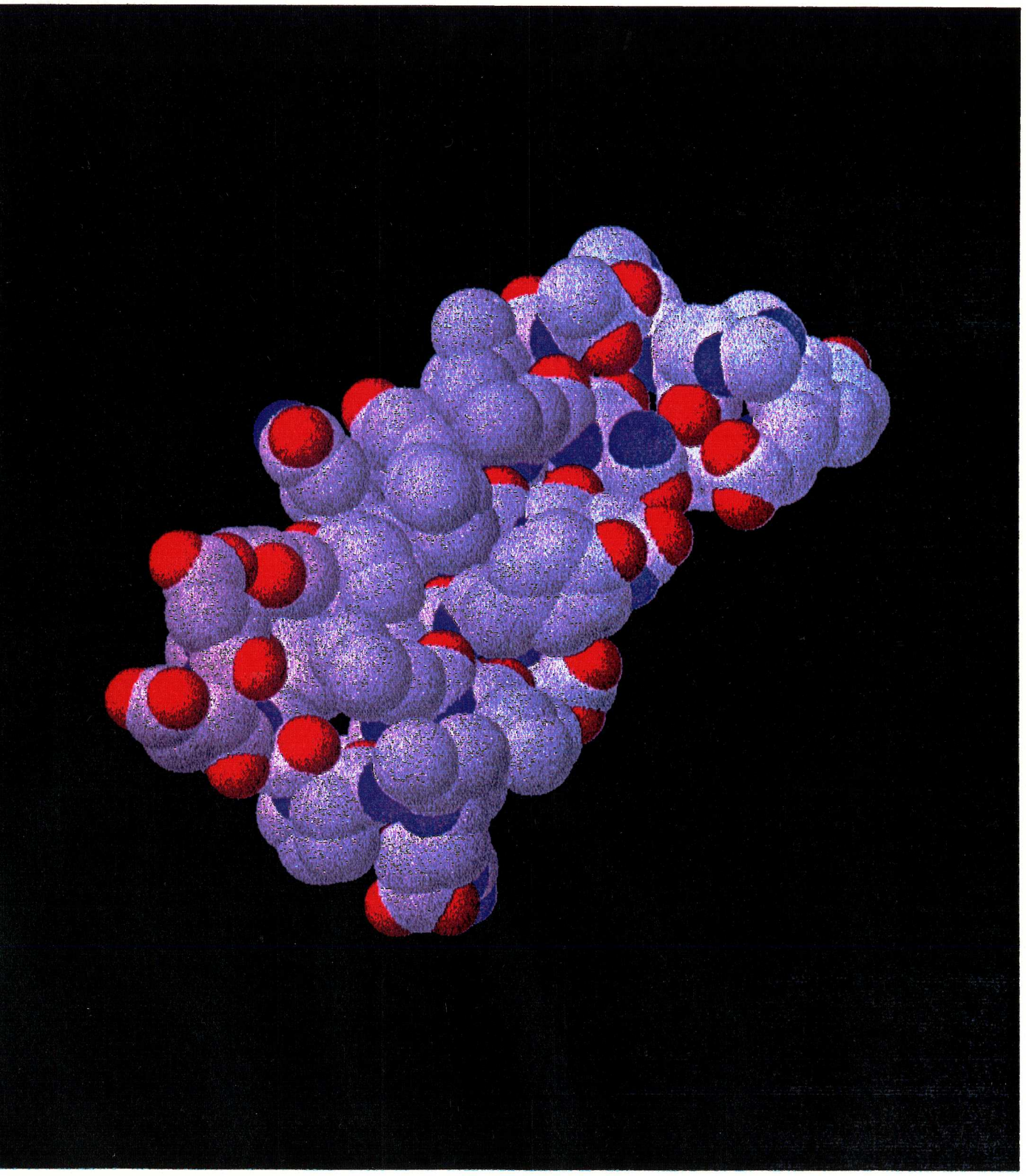


Fig. 4 Native conformation of avian pancreatic polypeptide.

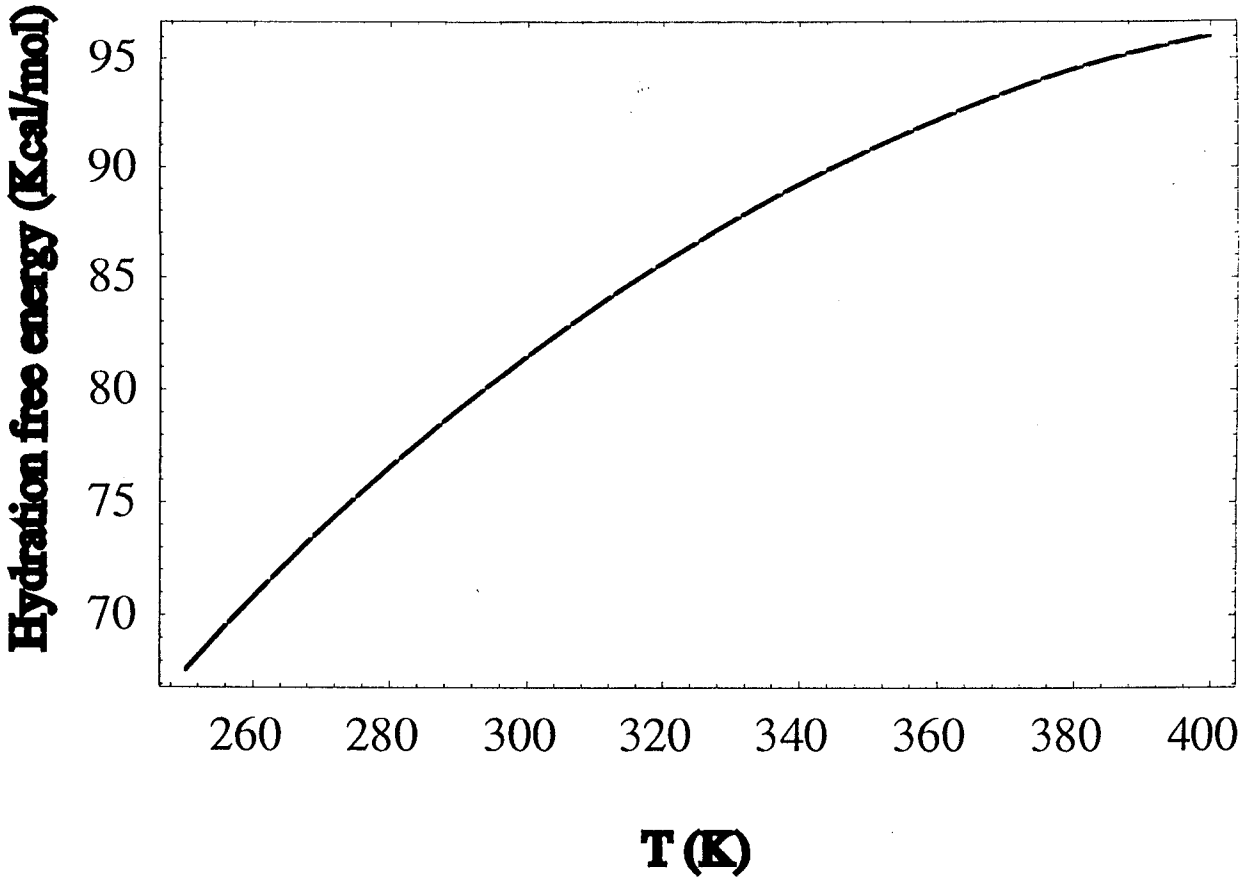


Fig. 5 Temperature dependence of the hydrophobic part of the hydration free energy of avian pancreatic polypeptide (native conformation)

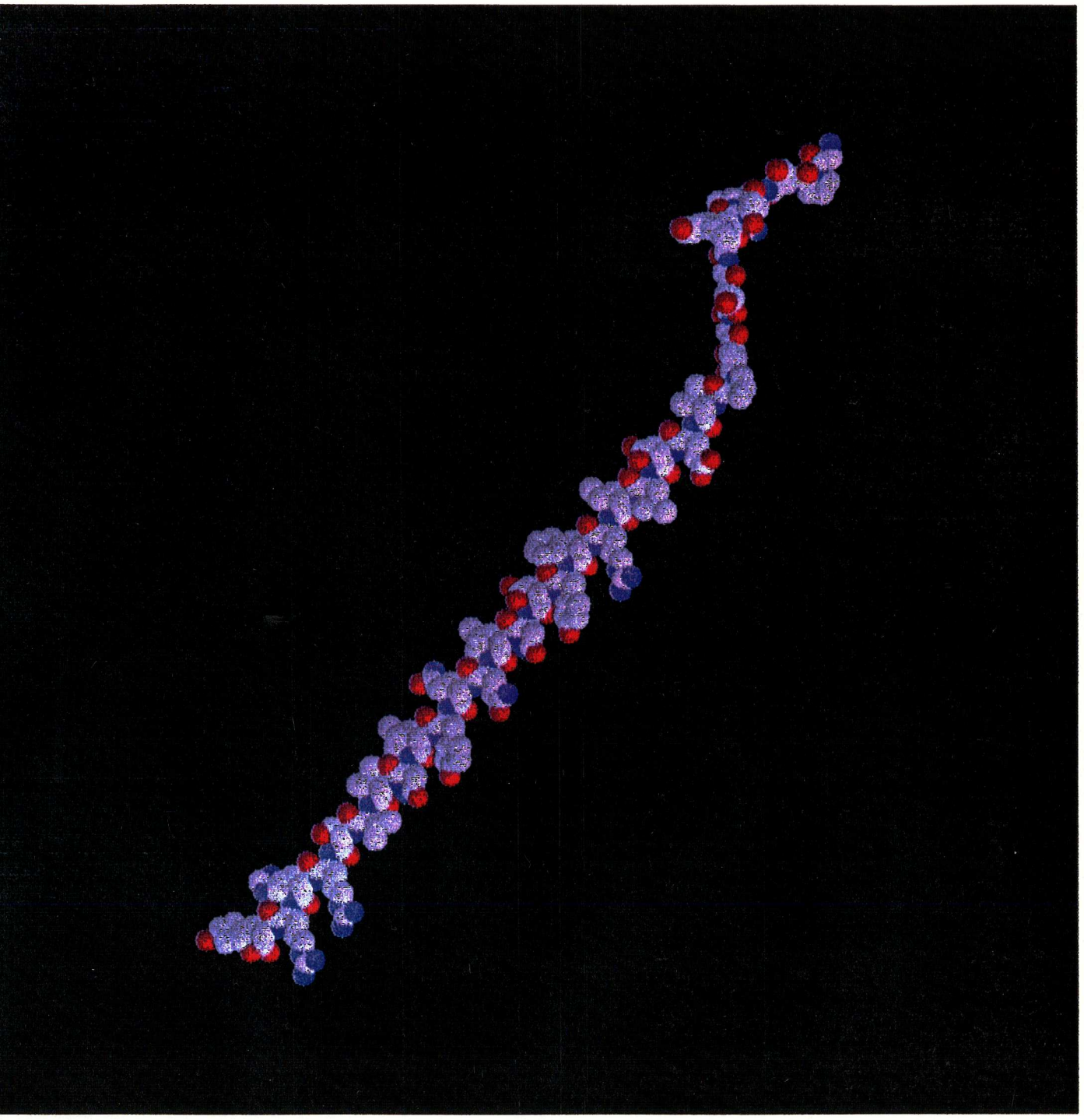


Fig. 6 Extended conformation of avian pancreatic polypeptide. All dihedral angles are set to be degree of 180 except for ϕ dihedral angles in proline residues. The accessible surface area and the excluded volume are 5073 $Å^2$ and 9618 $Å^3$ respectively.

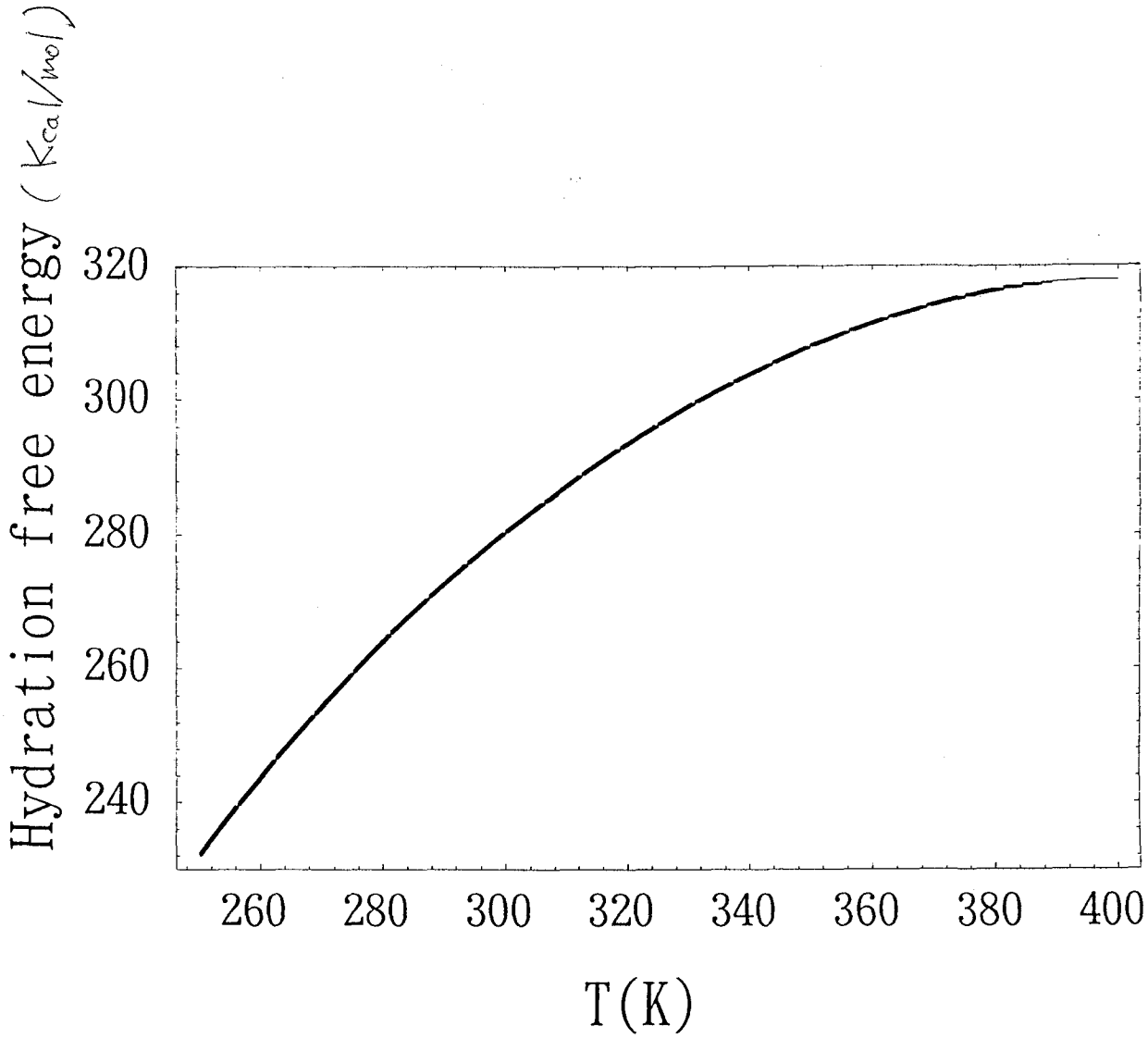


Fig.7 Temperature dependence of the hydrophobic part of the hydration free energy of avian pancreatic polypeptide (extended conformation)

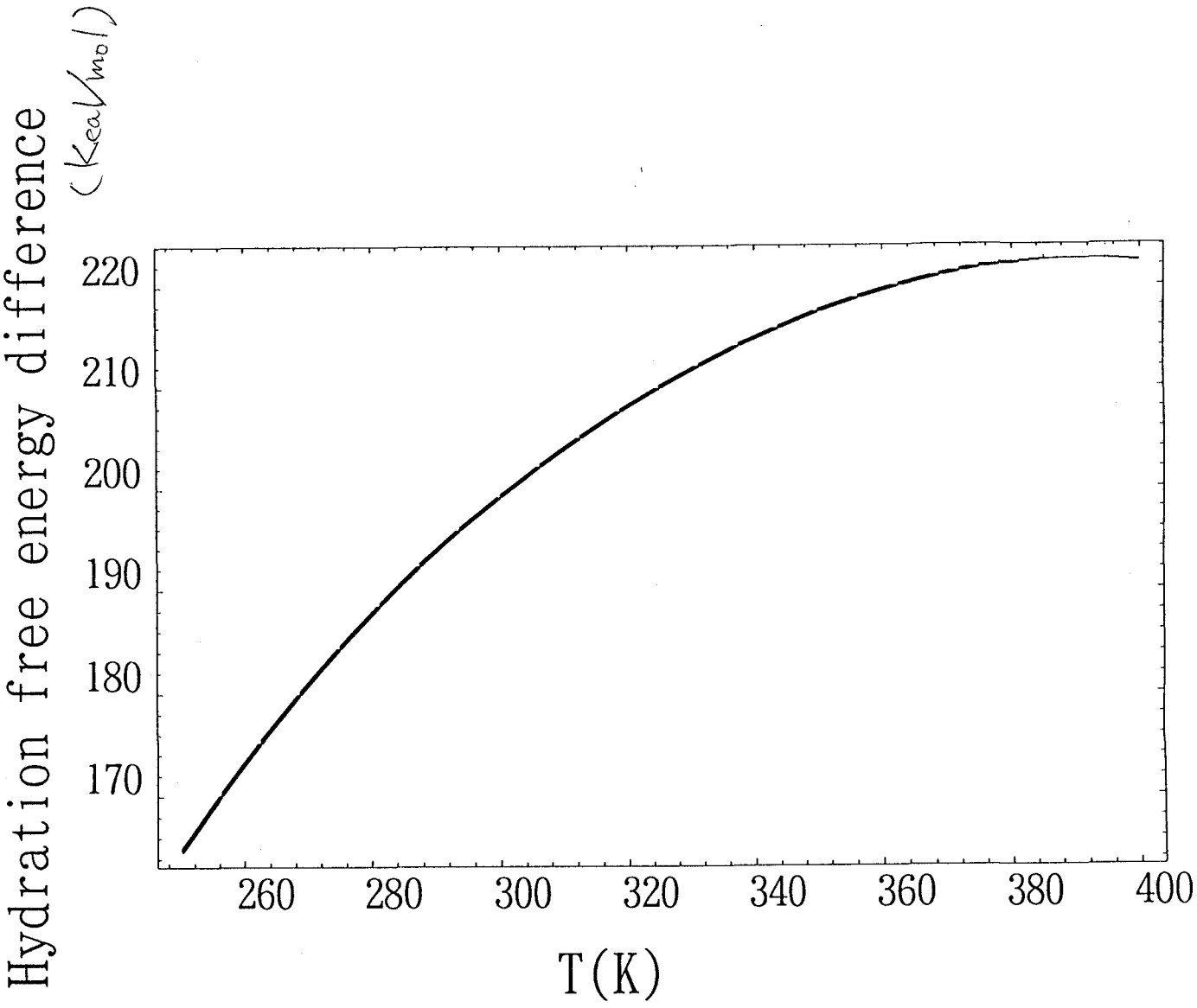


Fig. 8 Temperature dependence of the hydrophobic part of the hydration free energy difference. The plotted values are derived by subtracting the hydrophobic part of the hydration free energies of the native conformation from those of the extended conformation at the corresponding temperatures.

are 3336.8 Å^2 and 8524.1 Å^3, respectively. However this protein has two α-helicies, most of the heavy atoms in the protein are accessible from water molecules (only three heavy atoms are not accessible).

The calculated values of the solvation free energy for this protein are follows; $\bar{g}_C$, $\bar{g}_I$, $\bar{g}_B$ and $\Delta\bar{g}_s^o$(g→m) are 239.8 kcal/mol, –142.5 kcal/mol, -1612.4 kcal/mol and -1515.1 kcal/mol, respectively.

The temperature dependence of the hydrophobic part of the hydration free energy ($\bar{g}_C$ and $\bar{g}_I$) of this protein in the native conformation (Fig. 4) is calculated (Fig. 5). The temperature dependence of the hydrophobic part of the hydration free energy ($\bar{g}_C$ and $\bar{g}_I$) of this protein in the extended conformation (Fig. 6) is also calculated (Fig. 7). The extended conformation is a model for the conformation of the denatured protein. The plotted values in Fig. 8 are derived by subtracting the hydrophobic part of the hydration free energies of the native conformation from those of the extended conformation at the corresponding temperature. This result shows that the hydrophobic part of the hydration can be the origin of the denaturation at low temperature, which is well known as cold denaturation.

Actin:

The accessible surface area and excluded volume of this protein are 16735.5 Å^2 and 69959.3 Å^3, respectively.

The calculated values of the solvation free energy for this protein are follows; $\bar{g}_C$, $\bar{g}_I$, $\bar{g}_B$ and $\Delta\bar{g}_s^o$(g→m) are 777.5 kcal/mol, –741.4 kcal/mol, -1314.5 kcal/mol and -1278.4 kcal/mol, respectively (See Fig. 9, 10 and 11).

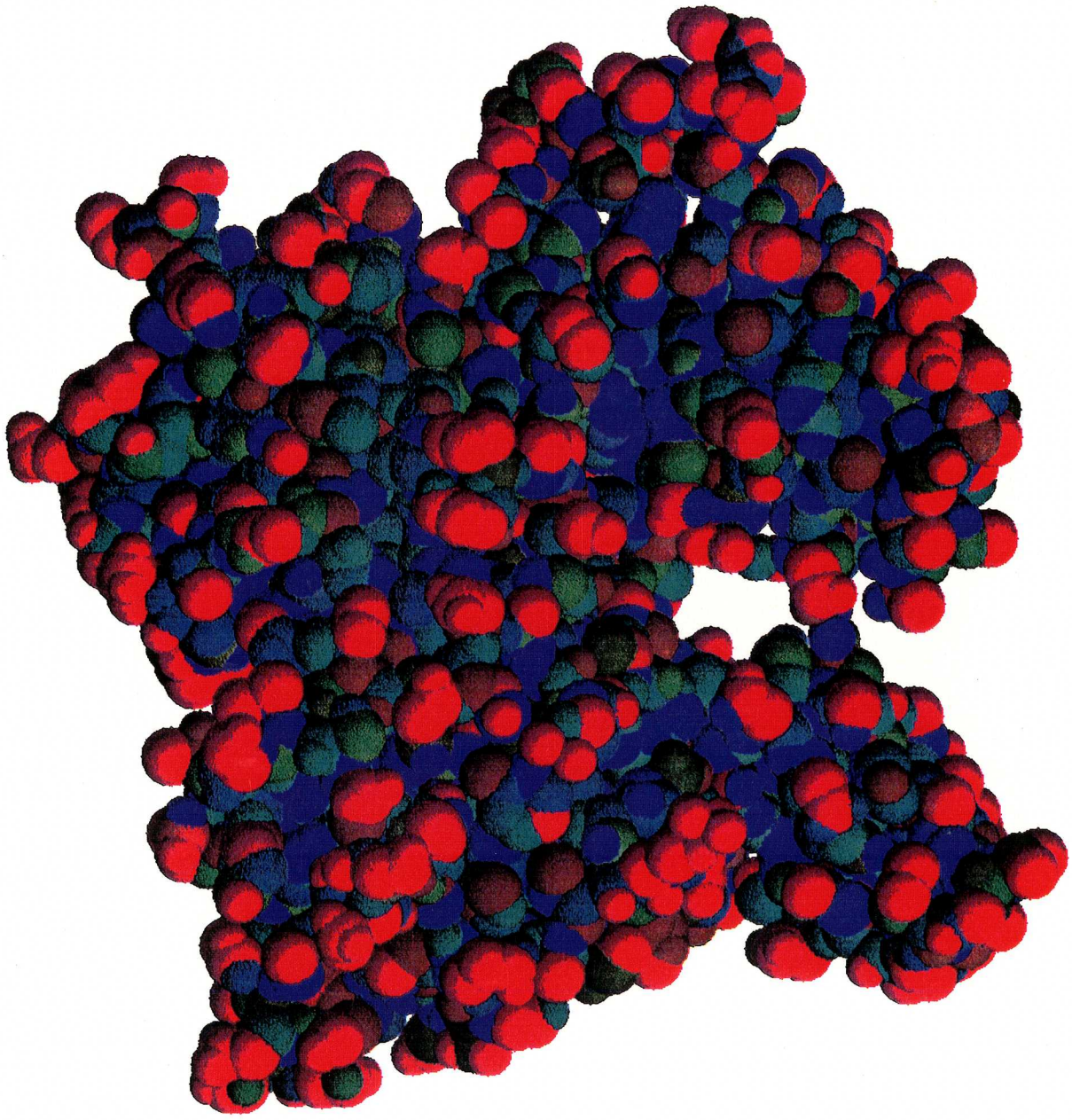


Fig. 9 Contribution of each atom to the first derivative of the excluded volume of actin. Blue, green and red represent the positive, zero and negative values, respectively.

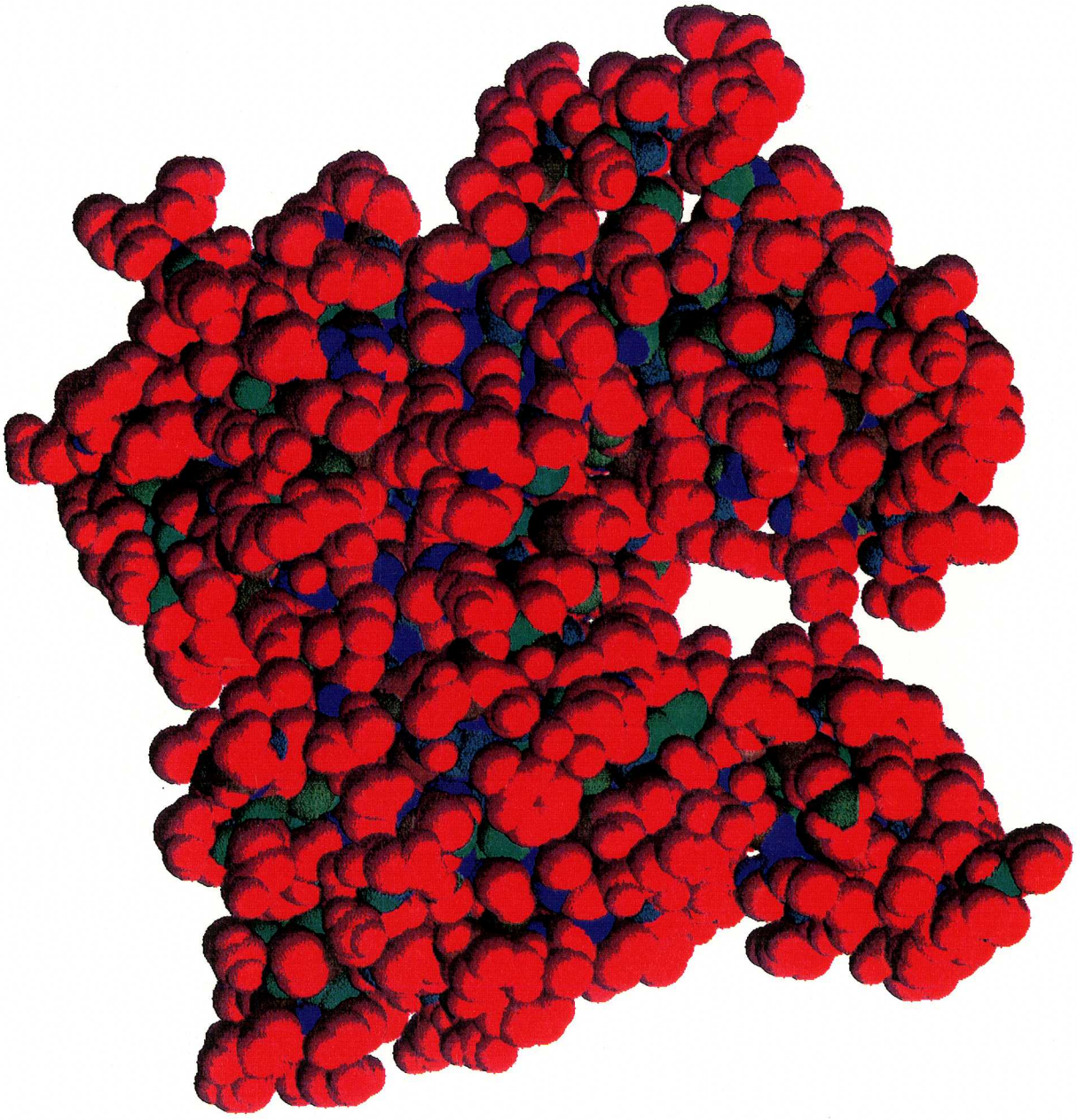


Fig. 10 Contribution of each atom to the second derivative of the excluded volume of actin. Blue, green and red represent the positive, zero and negative values, respectively.

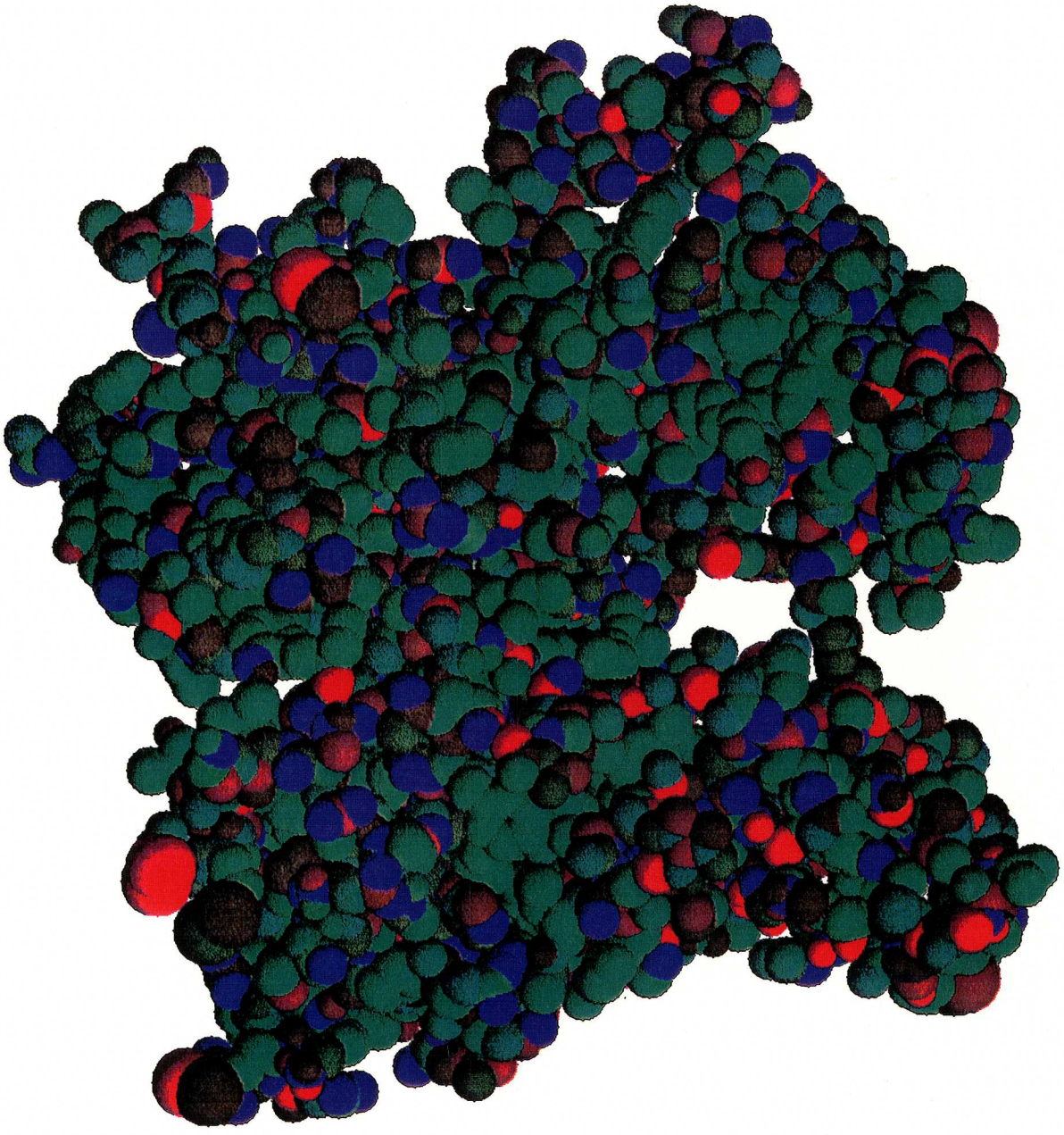


Fig. 11 Born energies of atoms are shown. Blue, green and red represent the positive, zero and negative values, respectively.

In the hydrophobic part of the solvation free energies of each protein, cavity formation in the water and attractive force between the solute and the solvent compensate each other. In the cavity formation of each protein, Born energy is much larger than other terms in this hydration free energy, because hydrophilic residues are ionized in the water. The Born energy calculated here is the preliminary one, so more investigation is needed. This method is suitable for the analysis of the solvation of the protein, because this can be used without any problems with the overwhelmingly large configurational space due to the solvent molecules. The most cpu-time consuming part is numerical integration of attractive forces between solute and solvent molecules. Analytical calculation of this part may be possible by using the analytical formulation of the excluded volume calculation. Although Born energy is dominant in the solvation free energy of this protein, hydrophobic term is practically important because our main purpose is an application of this method to the free energy difference calculation in the future work. For the thermal stability of the proteins, for example, temperature dependence of the free energy difference is required. To apply the empirical method using accessible surface area of the solute to such a problem, the experimental parameter sets at individual temperatures are needed. In our method, temperature dependence of the solvation free energy can be obtained by changing few parameters. It will be also good application of our method to investigate the effect of the solute shape to the solvation, because in our method we can treat molecules of arbitrary shape. Conformational change of proteins and the interface of the two proteins (e.g. motor proteins, actin and myosin) are good targets.

References

1 R. A. Pierotti, Chem. Rev. **76**, 717 (1965)

2 M. Irisa, K. Nagayama, and F. Hirata, Chem. Phys. Letters, **207**,

430 (1993)

3 M. Irisa, T. Takahashi, K. Nagayama, and F. Hirata, in submittion.

4 T. Boublík, Mol. Phys. **27**, 1415–1427 (1974)

5 M. Irisa, in submittion.

6 R. A. Pierotti, J. Phys. Chem. **67**, 1840 (1963)

7 N. Rogers, Progr. Biophys. Mol. Biol. **48**, 37 (1986)

8 W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, J.Am.Chem.Soc. **112,** 6127 (1990)

9 R. A. Pierotti, J. Phys. Chem. **69**, 281 (1965)

10 W. L. Jorgensen and J. T-Rives, J. Am. Chem. Soc. **110**, 1657 (1988)

11 T. L. Blundell, J. E. Pitts, I. J. Tickle, S. P. Wood, C.-W. Wu, Proc. Nat. Acad. Sci. USA **78**, 4175 (1981)

12 W. Kabsch, H. G. Mannherz, D. Suck, E. F. Pai, and K. C. Holmes, Naute. **347**, 37 (1990)

Chapter 5

Mathematical basis of extended scaled particle theory: analytical method for calculating volume of the fused spheres with different radii

Abstract

We propose a new algorithm for calculating volume of fused spheres. The new algorithm is superior to the old methods in two aspects. One is the separation of the algorithm into topological and analytical parts. This makes mathematical nature of the problem "what is minimum topological information required to calculate the volume of fused spheres?" clear. The number of conditional branches is drastically reduced. The remained conditional branches are those which are topologically essential. For practical use, only the list of the spheres forming isolated spheres, isolated sphere circles, exposed vertices, and buried vertices are required. The other aspect which make the new algorithm superior to the old ones is the analytical expression which is unrestricted to a choice of the coordinate frame. All equations are written only by radii of spheres and distances between spheres. This enables us to calculate derivatives of the excluded volume with respect to a scaling parameter which is used in the scaled particle theory of liquid.

Introduction

Analytical calculation for volume of fused spheres with different radii is a key for formulating liquid state theories of hard bodies, which includes the scaled particle theory of arbitrary shaped solutes [1,2] and the virial coefficients of the liquid consisting of hard spheres [3,4]. In this chapter we will propose a new algorithm of the analytical calculation for volume of fused spheres with different radii. The aim of our work is the volume calculation of the molecule [5], but our algorithm is effective for any volume calculation for the fused spheres. Some algorithms have been proposed for this purpose, but all those are not well elaborated in a sense that they involve too many conditional branches depending on configurations of the spheres. Most of the conditional branches used in the earlier works come from two points; 1) the machine (vector/parallel computer) dependent acceleration of the speed of the algorithm, 2) the topology of the fused spheres required for the analytical expression. In order to make the problems in the analytical calculation for volume of fused spheres clear, we focus on the point 2) in this work. In the discussion of this chapter, you would find the future strategy of point 1) become clearer by the study of point 2) in this work. The other aspect which make our new algorithm superior to the old ones is the analytical expression which is unrestricted to a choice of the coordinate frame. The analytical expression is applicable for the investigation of the molecular volume changes depending on the configurations in the theories/simulations where the internal coordinates of the atoms in the molecule (such as Van der Waals radii of the atoms and interatomic distances) are important. An application of the analytical expression of the volume of the fused spheres for a liquid theory is explained in the discussion in which the numerical/probabilistic calculation of the volume [5,6] is not suitable.

There are two problems to be solved in the analytical methods to calculate the volume of the fused spheres. The first is how to define the shape of the solid body, which can be done by giving the configurations and radii of the fused spheres. The shape determines topology of boundary made by intersecting spheres. The topology in turn determines conditional branches in the algorithm of the equation of volume. The solid body so defined can be viewed as a set of primary bodies with simpler geometry. It is the second problem how to express the volume of the simple objects by one analytical equation. Earlier studies concerning the volume calculation, which aim the application for the molecule, can be classified into two categories depending on the definition of the topology of boundary made by intersecting spheres: inclusion-exclusion algorithms by Kratky [7] and building block algorithm by Dodd et al. [8].

Basic concept in Kratky's inclusion-exclusion method [8] is analogous to the mathematical set theory, where volume of fused spheres corresponds to a union of sets. An instructive example is the case in which two spheres overlap. It is analogous to a union of two sets which can be derived from the individual sets and the intersection of the two sets. Extension of the method to more general cases in which more than three spheres become increasingly complicated. However, Kratky has proved that any intersection made by more than four spheres can be expressed by a combination of the intersection made by up to four spheres. This means the final expression for the volume of the fused spheres becomes a sum of terms corresponding to up to four spheres. However, the straight forward implementation of Kratky's method for a solid body with more than four spheres overlapped requires enormous amount of computation time. Sheraga and his coworkers have tried to bypass the problem using a "shortcut algorithm" which identifies the vanishing contributions concerning more than four spheres overlapped [9]. This was the first attempt to implement

Kratky's method for the calculation of the fused spheres. However, their calculation did not give the right answer to a test problem which involves more than five spheres overlapped due to an unidentified mistake in their algorithm [8].

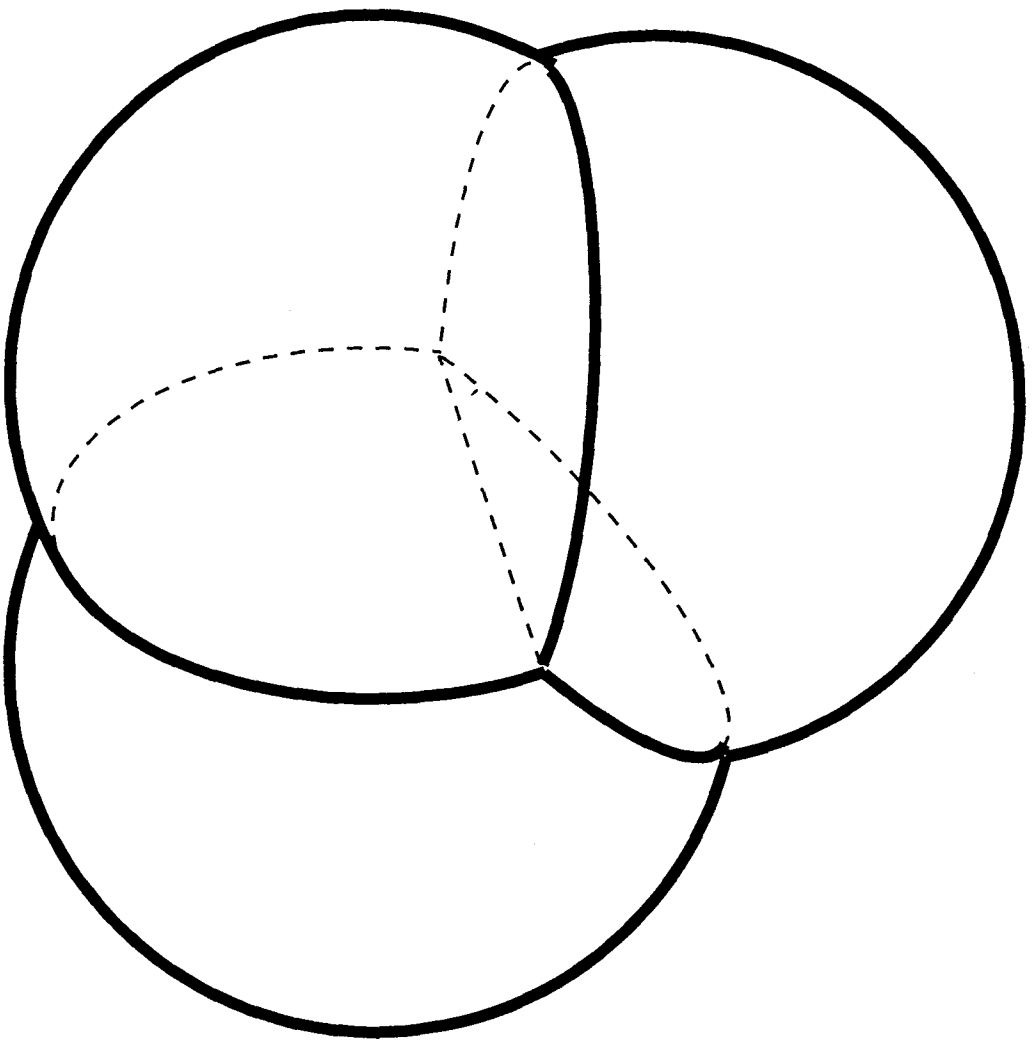


Fig. 1 Schematic drawing of partitions of fused spheres in which the partitions of pair of the spheres are flat planes.

Dodd's method [8] resembles constructing a house from building blocks. Fig. 1 illustrates how a solid body is constructed from "building blocks" each of which consists of a part of spherical surface and two flat planes. The building blocks themselves can be divided successively into pieces with simpler geometry. The essential idea due to Dodd is to make the division until the simplest geometry such as cones and pyramid is reached, which of course can be calculated analytically. A difficulty in the method arises from the definition of division which is not necessarily unique. An arbitrary choice of the division can give rise to a messy computational algorithm with many conditional branches. One of the problems, which is apparent in the Dodd division, is that the expression for the primary body with the simplest geometry requires the information of up to five spheres in contradiction with Kratky's finding [7]. The use of the Cartesian coordinates in his method, which is related to the five-sphere problem, is another disadvantage in such cases in which the derivatives of the volume with respect to the radii of spheres and inter-sphere distances are required. The application to the scaled particle theory is a typical example of such cases.

A crucial step in Dodd's method is how to calculate the spherical portion of the surface which makes the building block. An analytical method proposed by Richmond [10] can be employed for that purpose, which uses an elegant theorem due to Gauss and Bonnet. The Gauss-Bonnet theorem enables us to calculate area of the spherical surface bounded with connected arcs from the length of the arcs and the angles made by adjacent two arcs. Unfortunately, the expression of the accessible surface area proposed by Richmond is too difficult to understand due partly to the use of the Cartesian coordinates. Improvement which we could have achieved in this work for the algorithm of the volume calculation largely due to the reconstruction of the Richmond method by changing the Cartesian

coordinates to the internal coordinates (sphere radii and inter-sphere distances).

In the present chapter, we propose a new algorithm for calculating a solid body consisting of fused spheres. The method employs the internal coordinates instead of the Cartesian coordinates. The use of the internal coordinates facilitates the direct calculation for the volume of the building blocks of a solid body, an example of which appears in Fig. 1. Since further division of the building blocks as in the Dodd method is not necessary, the final expression for the volume is far more elegant compared to the earlier works and the conditional branches in the computational algorithm is drastically reduced. Moreover, any analytical term in our expression uses information up to four spheres in consistent with Kratky's finding [7]. The use of the internal coordinates makes some applications for the liquid state theories possible, in which the derivatives of the volume with respect to the internal coordinates are required.

The relation between our work and other works [11-15] using the power diagrams for the volume calculation in the computational geometry field is follows. Our new algorithm is explicitly separated into topological part and analytical parts. In this chapter, we conclude that the topological information required for the volume calculation of the fused spheres are topology of the building blocks. This building block exactly corresponds to the intersection of the sphere and the "cell" of the power diagram [11,13,16,17](which is a kind of Voronoi diagram). In our work, after the topology of the building blocks is known, the volume of the building block can be calculated at once by using "one" analytical equation without considering conditional branches which are caused by the further division of the building block. But all other works using power diagram divide the building block into pieces with simpler geometry [11,12,14,15]. Our work considers all possible configurations of the spheres which include the

singular cases appeared in the configurations of the molecules. The singular cases are a) one exposed vertex is formed by one spherical surface and more than two intersection planes, b) one buried vertex is formed by more than three intersection planes. For example, one of the case a) is the benzene molecule whose atoms are located on the vertices of the regular polygon. This difficulty in the volume calculation of the fused spheres has not been discussed explicitly in the works using the power diagrams for volume calculation. The singular cases corresponds to the singular cases in power diagrams. In these cases, one vertex of the cell in the power diagram is formed by more than three planes of the cells.

Method

In the definition we define the "building block" of a solid body and names of the topological elements. At first, we show the expression of the volume and the surface areas of the fused spheres. Then we will explain our algorithm by employing Richmond method [10]. After that we also show the expression of the volume and the surface area of the fused spheres with an special configuration which does not appear in the almost all molecules.

I. Definition of the division of the fused spheres (building block)

At first we define the building block precisely for later use. The definition was first proposed by Dodd and co-worker [8] in the molecular science field. A geometrical 3-dimensional body of fused spheres (B) composed of n spheres is expressed as following equation,

$$B=S_1^{(3)} \cup S_2^{(3)} \cup \cdots \cup S_{n-1}^{(3)} \cup S_n^{(3)} \quad (1)$$

where $S_i^{(3)}$ is a solid body of sphere i. This solid body B is also expressed by B_i (building block) as follows,

$$B=\bigcup_{i=1}^{n} B_i \quad , \quad B_i \cap B_j = \varnothing \quad (\text{if } i \neq j) \quad ,$$
$$B_i = S_i^{(3)} - \bigcup_{j(\neq i)}^{n} \left\{ \text{OutHalf}_i (S_i^{(3)} \cap S_j^{(3)}) \right\} \quad . \qquad (2)$$

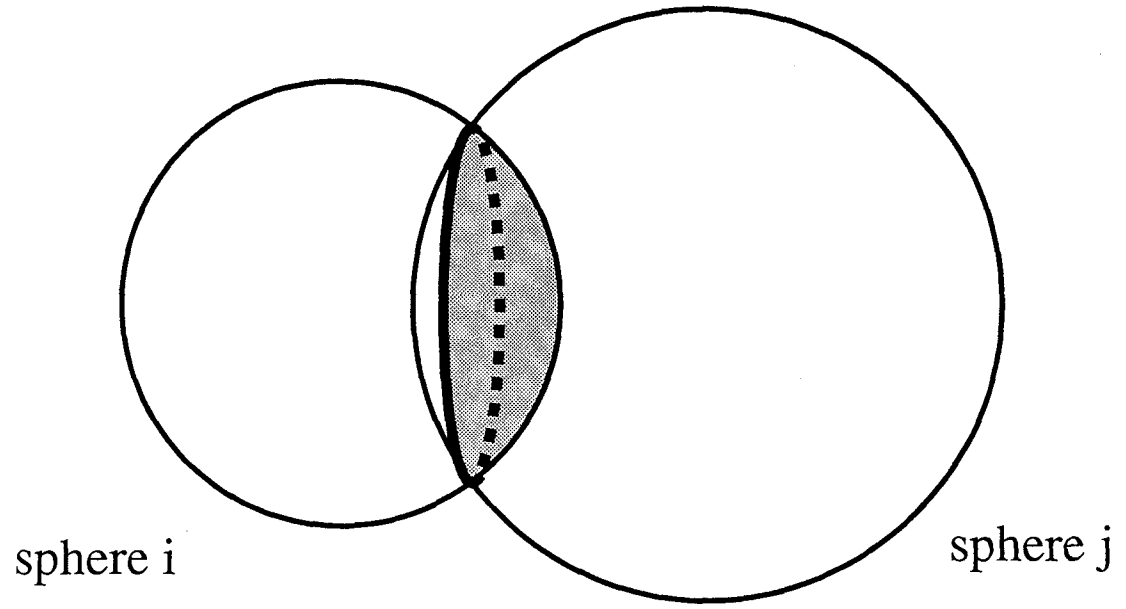


Fig. 3 Schematic drawing of the function $\mathrm{OutHalf}_i$. Shaded region is $\mathrm{OutHalf}_i(S_i^{(3)} \cap S_j^{(3)})$. Thick line represents sphere circle, $P_{ij}^{(1)}$. The partition plane $P_{ij}^{(2)}$ is the plane containing $P_{ij}^{(1)}$. $P_{ij}^{(2)}$ divides the whole space into sub spaces $P_{ij}^{(3)}$ and $\bar{P}_{ij}^{(3)}$.

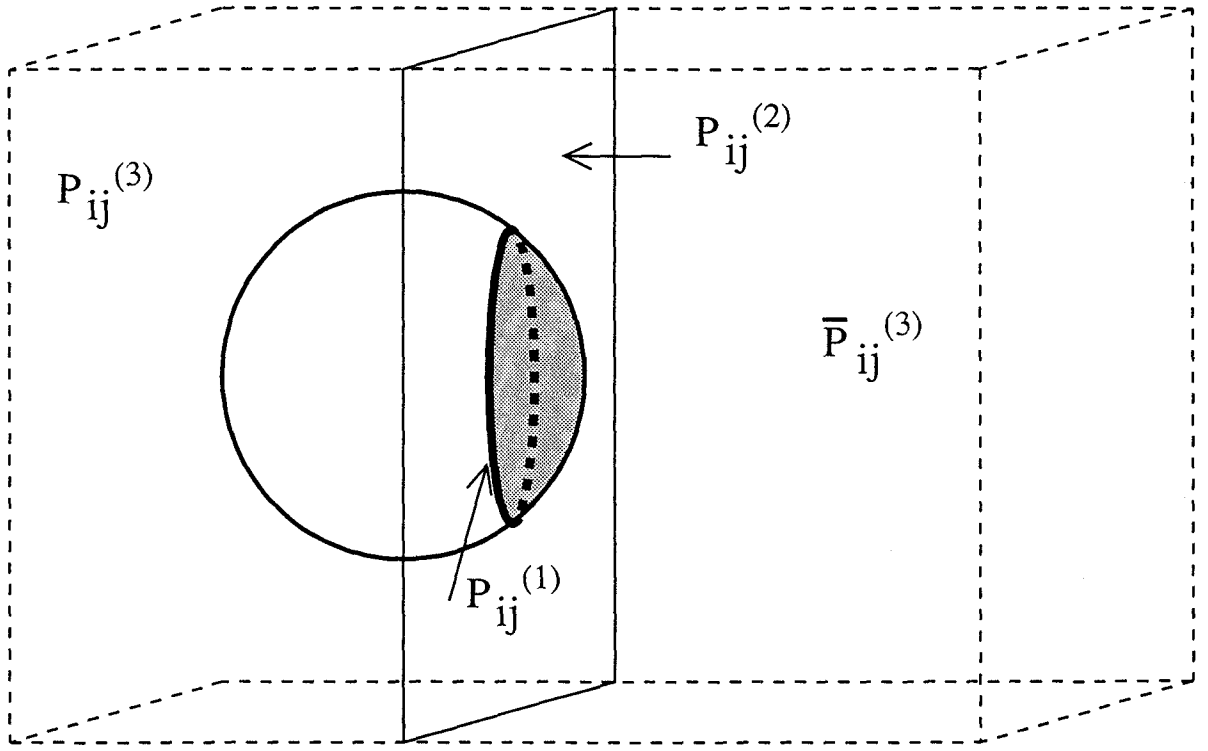


Fig. 2 Schematic drawing of the function $\mathrm{OutHalf}_i$. Shaded region is $\mathrm{OutHalf}_i(S_i^{(3)} \cap S_j^{(3)})$. Thick line represents sphere circle of the sphere i and j.

where $\mathrm{OutHalf}_i$ is a function which divides the intersection made by $S_i^{(3)}$ and $S_j^{(3)}$ into two pieces and chooses a piece formed by the sphere circle and the buried part of $S_i^{(2)}$, as is shown in Fig. 2. $S_i^{(2)}$ is the whole spherical surface of $S_i^{(3)}$. The function $\mathrm{OutHalf}_i$ is also defined by following equation.

$$\mathrm{OutHalf}_i(S_i^{(3)} \cap S_j^{(3)}) = \bar{P}_{ij}^{(3)} \cap S_i^{(3)} \quad (3)$$

where the set of $\bar{P}_{ij}^{(3)}$ is a space divided by a partition plane $P_{ij}^{(2)}$ and contains the buried part of $S_i^{(3)}$. $\bar{P}_{ij}^{(3)}$ is a composite set of $P_{ij}^{(3)}$ (see Fig. 3). The partition plane $P_{ij}^{(2)}$ is defined as the plane which contains the sphere circle $(P_{ij}^{(1)})$

$$P_{ij}^{(2)} \supset P_{ij}^{(1)} \quad (4)$$

The building block B_i exactly corresponds to the intersection of the sphere i and the "cell", which belongs to i, of the power diagram [11,13,16,17](which is a kind of Voronoi diagram) in the computational geometry.

For later usage, we must give definitions for the topological elements (Fig. 4). We name the exposed spherical surfaces of the body B_i as "accessible surface", $B_i^{(2)A}$, and plane surfaces of the body B_i as "inaccessible surface", $B_{ij}^{(2)I}$,

$$B_i^{(2)A} = S_i^{(2)} \bigcap B_i \quad , \quad B_{ij}^{(2)I} = P_{ij}^{(2)} \bigcap B_i \quad , \quad (5)$$

The term "exposed vertex" and "buried vertex" are the vertex of B_i which forms boundary of $B_i^{(2)A}$ and the vertex of B_i which is surrounded by $B_{ij}^{(2)I}$ respectively. The term "A-A inner-angle" is the angle which is formed by the two adjacent arcs at the exposed vertex on the accessible surface, $B_i^{(2)A}$. The term "A-L inner-angle" is the angle which is formed by the one arc and one straight line at the exposed vertex on the inaccessible surface, $B_{ij}^{(2)I}$. The term "L-L inner-angle" is the angle which is formed by the two straight lines at the vertex on the inaccessible surface, $B_{ij}^{(2)I}$. The term "isolated sphere circle" is the sphere circle which has no exposed vertex intersected with another sphere circle.

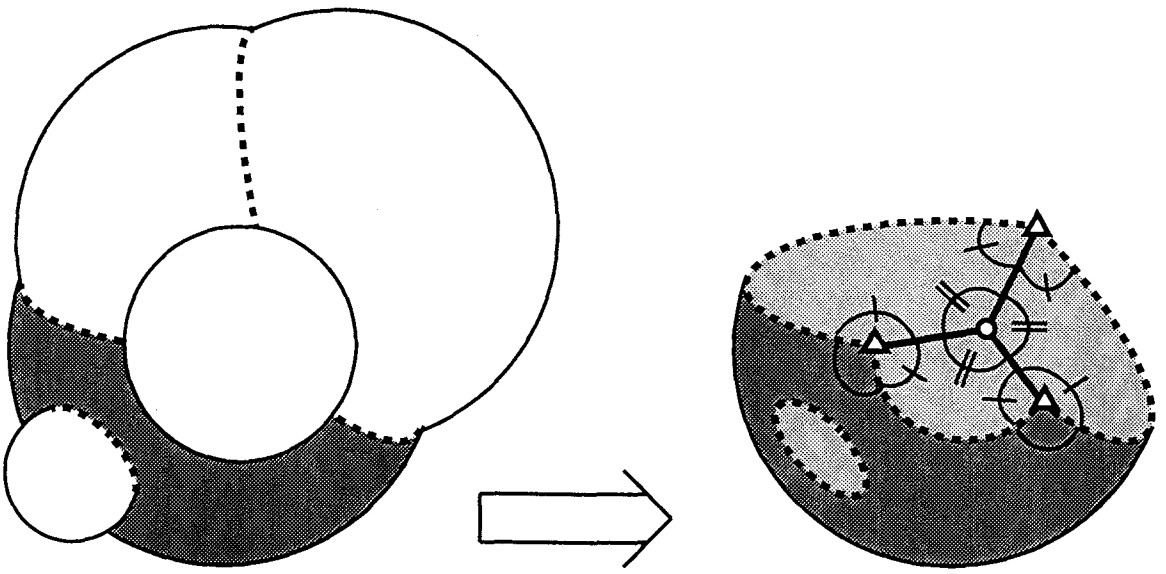


Fig. 4 Definitions for the topological elements. Thick and thin shaded surfaces are accessible and inaccessible surfaces, respectively. Exposed vertex and buried vertex are represented as open triangle and open circle respectively. A-A, A-L, and L-L inner-angles are represented as small arc, small arc with -, small arc with =, respectively.

II. Expression of the volume and surface area of the fused spheres in our method

When the fused spheres have ordinary configuration in which each exposed vertex on the fused spheres is made of three arcs and one straight line, and each buried vertex is made of four straight lines (Fig. 5), the topological information required for the analytical calculation of the volume of fused spheres is only as follows,

1) The list of isolated spheres (LST1).

2) The list of two spheres (LST2) making isolated intersecting circles. The component ((i,j)) of this list (LST2={(i,j)}, i<j) is the indices of the two spheres which make the isolated intersecting circle.

3) The list of three spheres (LST3) making exposed vertex on the accessible surface.
The component ((i,j,k)) of this list (LST3={(i,j,k)}, i<j<k) is the indices of the three spheres which make the exposed vertex.

4) The list of four spheres (LST4) making buried vertex on the inaccessible surface.
The component ((i,j,k,l)) of this list (LST4={(i,j,k,l}, i<j<k<l) is the indices of the four spheres which make the buried vertex.

Above topological information is obtained by the simple analytical calculation explained in the Appendix A. By using lists in 1), 2), 3), and 4), we make the following lists.

0') The Euler-Poincaré characteristics (χ_i^A) of the accessible surface of each sphere (i). The detail of the derivation of χ_i^A is explained in the Appendix B.

1') The list of spheres (LST1') which appeared in 1), 2), and 3).

2') The list of pairs of spheres (LST2') which appeared in 2), 3), and 4).

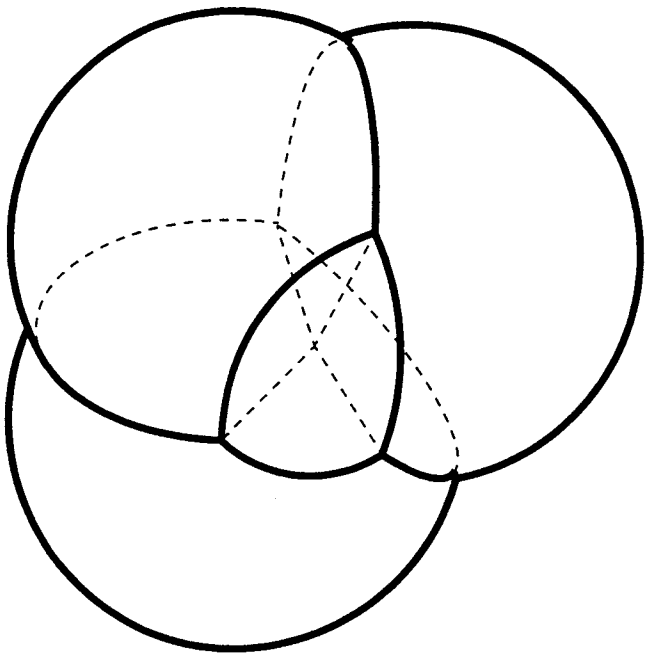


Fig. 5 Regular configuration in which one exposed vertex has three arc and one straight line, and one buried vertex has four straight lines.

The volume V and surface area A can be calculated for the fused spheres as follows:

$$
\begin{aligned}
V \;=& \sum_{LST1'} T_1(i) + \sum_{LST2'} \{T_2(i,j)+T_2(j,i)\} \\
&+ \sum_{LST3} \{T_{3a}(i,j,k)+T_{3a}(j,k,i)+T_{3a}(k,i,j) \\
&\qquad +T_{3b}(i,j,k)+T_{3b}(j,k,i)+T_{3b}(k,i,j) \\
&\qquad +T_{3b}(i,k,j)+T_{3b}(k,j,i)+T_{3b}(j,i,k)\} \\
&+ \sum_{LST4} \{T_4(i,j,k,l)+T_4(i,k,l,j)+T_4(i,l,j,k) \\
&\qquad +T_4(j,k,l,i)+T_4(j,l,i,k)+T_4(j,i,k,l) \\
&\qquad +T_4(k,l,i,j)+T_4(k,i,j,l)+T_4(k,j,l,i) \\
&\qquad +T_4(l,i,j,k)+T_4(l,j,k,i)+T_4(l,k,i,j)\} \quad ,
\end{aligned}
\tag{6}
$$

$$
\begin{aligned}
A \;=& \sum_{LST1'} T_{A1}(i) + \sum_{LST2'} \{T_{A2}(i,j)+T_{A2}(j,i)\} \\
&+ \sum_{LST3} \{T_{A3a}(i,j,k)+T_{A3a}(j,k,i)+T_{A3a}(k,i,j) \\
&\qquad +T_{A3b}(i,j,k)+T_{A3b}(j,k,i)+T_{A3b}(k,i,j) \\
&\qquad +T_{A3b}(i,k,j)+T_{A3b}(k,j,i)+T_{A3b}(j,i,k)\} \\
&+ \sum_{LST4} \{T_{A4}(i,j,k,l)+T_{A4}(i,k,l,j)+T_{A4}(i,l,j,k) \\
&\qquad +T_{A4}(j,k,l,i)+T_{A4}(j,l,i,k)+T_{A4}(j,i,k,l) \\
&\qquad +T_{A4}(k,l,i,j)+T_{A4}(k,i,j,l)+T_{A4}(k,j,l,i) \\
&\qquad +T_{A4}(l,i,j,k)+T_{A4}(l,j,k,i)+T_{A4}(l,k,i,j)\} \quad ,
\end{aligned}
\tag{7}
$$

$$T_1(i)=\frac{1}{3}\rho_i T_{A1}(i) \quad , \quad T_2(i,j)=\frac{1}{3}\rho_i T_{A2}(i,j)+T_{I2}(i,j) \quad ,$$

$$T_{3a}(i,j,k)=\frac{1}{3}\rho_i T_{A3a}(i,j,k) \quad , \quad T_{3b}(i,j,k)=\frac{1}{3}\rho_i T_{A3b}(i,j,k)+T_{I3}(i,j,k) \quad , \quad (8)$$

$$T_4(i,j,k,l)=\frac{1}{3}\rho_i T_{A4}(i,j,k,l)+T_{I4}(i,j,k,l)$$

$$T_{A1}(i) = 2\pi\rho_i^2\chi_i^A \quad , \quad T_{A2}(i,j) = 2\pi\rho_i g_{ij} \quad ,$$

$$T_{A3a}(i,j,k) = \rho_i^2\Omega(i,j,k) \quad , \quad T_{A3b}(i,j,k) = -\rho_i g_{ij}\phi(i,j,k) \quad , \quad (9)$$

$$T_{A4}(i,j,k,l) = -\rho_i g_{ij}\omega(i,j,k,l)$$

$$T_{I2}(i,j) = \frac{1}{3}g_{ij}\pi r_{ij}^2 \quad ,$$

$$T_{I3}(i,j,k) = \frac{1}{6}g_{ij}\left[-r_{ij}^2\phi(i,j,k)+h(i,j,k)\sqrt{r_{ij}^2-h(i,j,k)^2}\right] \quad , \quad (10)$$

$$T_{I4}(i,j,k,l) = \frac{1}{6}g_{ij}\left[-r_{ij}^2\omega(i,j,k,l)+e(i,j,k,l)\right]$$

$$g_{ij}=\frac{d_{ij}^2+\rho_i^2-\rho_j^2}{2d_{ij}} \quad , \quad r_{ij}=\sqrt{\rho_i^2-g_{ij}^2} \quad , \quad h(i,j,k)=\frac{-g_{ij}\cos\theta_{ijk}+g_{ik}}{\sin\theta_{ijk}} \quad ,$$

$$e(i,j,k,l)=\frac{2h(i,j,k)h(i,j,k)-(h(i,j,k)^2+h(i,j,k)^2)\cos\omega(i,j,k,l)}{\sin\omega(i,j,k,l)}$$

$$\Omega(i,j,k)=-\cos^{-1}\left(\frac{\rho_i^2\cos\theta_{ijk}-g_{ij}g_{ik}}{r_{ij}r_{ik}}\right) \quad , \quad \phi(i,j,k)=\cos^{-1}\left(\frac{h(i,j,k)}{r_{ij}}\right) \quad , \tag{11}$$

$$\cos\theta_{ijk}=\frac{d_{ij}^2+d_{ik}^2-d_{jk}^2}{2d_{ij}d_{ik}} \quad , \quad \sin\theta_{ijk}=\sqrt{1-\cos^2\theta_{ijk}} \quad ,$$

$$\cos\omega(i,j,k,l)=\frac{d_{ik}^2\sin^2\theta_{ijk}+d_{il}^2\sin^2\theta_{ijl}-d_{kl}^2+(d_{il}\cos\theta_{ijk}-d_{ik}\cos\theta_{ijl})^2}{2d_{ik}d_{il}\sin\theta_{ijk}\sin\theta_{ijl}} \quad ,$$

$$\sin\omega(i,j,k,l)=\sqrt{1-\cos^2\omega(i,j,k,l)} \quad .$$

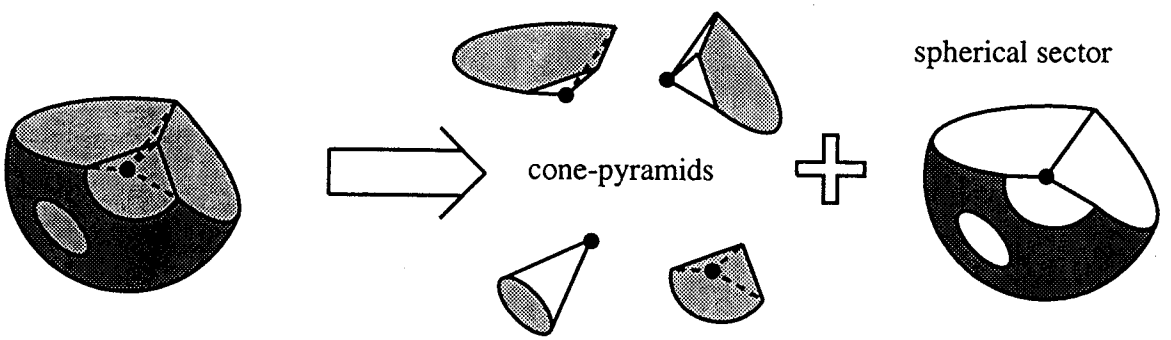


Fig. 6 Schematic drawing of the division of the sold body B_i into pieces of solid bodies. The pieces have two types of solid bodies, $Cntprj^{(3)}(o_i,B_i^{(2)A})$ (spherical sector) and $Cntprj^{(3)}(o_i,B_{ij}^{(2)I})$ (cone-pyramid). $B_i^{(2)A}$ and $B_{ij}^{(2)I}$ are accessible and inaccessible surface respectively. Center of the sphere is represented as closed circle.

where ρ_i and ρ_j are radii of sphere i and j respectively, d_{ij} is a distance between the sphere i and j. From now on, ρ indicates a radius of a sphere and d indicates a distance between two spheres.

III. Derivation of the volume expression

In order to derive our expression of the volume of the fused spheres, we will explain below a very simple method to calculate the volume or/and surface area of a geometrical 3-dimensional body. The method consists of hierarchial division of a geometrical 3-dimensional body into pieces of which geometrical description becomes less and less complex. But the final expression of the volume and surface area of the fused spheres described in the previous section is independent of the way of this hierarchical division.

The building block B_i can be decomposed into pieces of solid bodies (see Fig. 6),

$$B_i=\mathrm{Cntprj}^{(3)}(o_i,B_i^{(2)A})+\sum_j \mathrm{sign}^{(3)}(o_i,B_{ij}^{(2)I})\mathrm{Cntprj}^{(3)}(o_i,B_{ij}^{(2)I}) \quad ,$$

$$\mathrm{sign}^{(3)}(o_i,B_{ij}^{(2)I})=\begin{cases} 1 & (\text{if } o_i \in P_{ij}^{(3)}) \\ -1 & (\text{if } o_i \notin P_{ij}^{(3)}) \end{cases} \qquad (12)$$

where o_i is a center of the sphere i, $B_i^{(2)A}$ (accessible surface) represents one spherical plane or some spherical planes depending on the configuration of spheres, j run through all spheres making inaccessible surface in B_i, $\mathrm{Cntprj}^{(3)}(x, Y)$ is a function making a geometrical 3-dimensional body composed of Y (a spherical plane or spherical planes or a flat plane) and all straight lines from the point x to Y. The one of the component of the solid body $\mathrm{Cntprj}^{(3)}(o_i, B_i^{(2)A})$ and the solid body $\mathrm{Cntprj}^{(3)}(o_i, B_{ij}^{(2)I})$ were

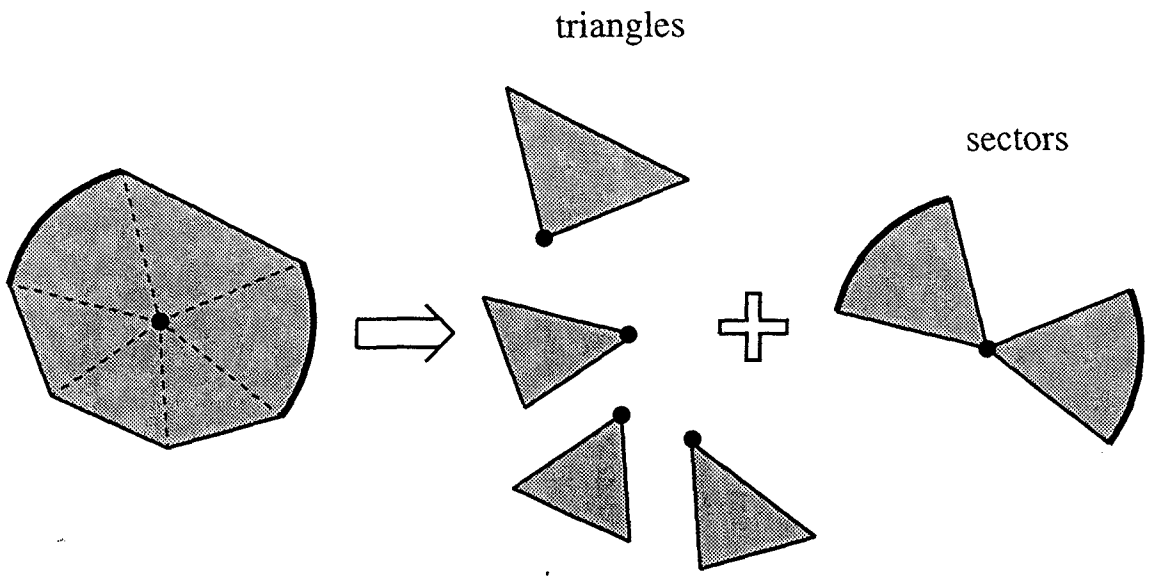


Fig. 7 Schematic drawing of the division of the inaccessible surface B_i into pieces of plane figures. The pieces have two types of plane figures, $Cntprj^{(2)}(o'_{ij},B_{ij}^{(1)A})$ (sector) and $Cntprj^{(2)}(o'_{ij},B_{ijk}^{(1)I})$ (triangle). $B_{ij}^{(1)A}$ and $B_{ijk}^{(1)I}$ are arc and straight line respectively. Center of the sphere circle $P_{ij}^{(1)}$ is represented as closed circle.

named as a spherical sector and a cone-pyramid respectively in Dodd's paper [8]. Both the spherical sector and the cone-pyramid have vertices at o_i.

The surface $B_{ij}^{(2)I}$ (inaccessible surface) can be decomposed into pieces of plane figures (see Fig. 7) in an analogous manner with Eq. (12),

$$B_{ij}^{(2)I}=\mathrm{Cntprj}^{(2)}(o_{ij}^{/},B_{ij}^{(1)A})+\sum_{k}\mathrm{sign}^{(2)}(o_{ij}^{/},B_{ijk}^{(1)I})\mathrm{Cntprj}^{(2)}(o_{ij}^{/},B_{ijk}^{(1)I})$$

$$B_{ij}^{(1)A}=S_i^{(2)}\cap B_{ij}^{(2)I}\quad,\quad B_{ijk}^{(1)I}=P_{ik}^{(2)}\cap B_{ij}^{(2)I}\quad, \tag{13}$$

$$\mathrm{sign}^{(2)}(o_{ij}^{/},B_{ijk}^{(1)I})=\begin{cases}1 & (\text{if } o_{ij}^{/}\in(P_{ik}^{(3)}\cap P_{ij}^{(2)}))\\ -1 & (\text{if } o_{ij}^{/}\notin(P_{ik}^{(3)}\cap P_{ij}^{(2)}))\end{cases}$$

where o'_{ij} is a center of the sphere circle $P_{ij}^{(1)}$, $B_{ij}^{(1)A}$ represents one arc or some arcs on $B_{ij}^{(2)A}$ depending on the configuration of spheres, k runs through all spheres making straight lines in the inaccessible surface $B_{ij}^{(2)I}$, $B_{ijk}^{(1)I}$ represents a straight line on $B_{ij}^{(2)I}$ formed by the sphere k, $\mathrm{Cntprj}^{(2)}(x', Y')$ is a function making a plane figure composed of Y' (an arc or arcs or a straight line) and all straight lines from the point x' to Y'. Components of the surface $\mathrm{Cntprj}^{(2)}(o'_{ij}, B_{ij}^{(1)A})$ and the surface $\mathrm{Cntprj}^{(2)}(o'_{ij}, B_{ijk}^{(1)I})$ are sectors and triangles respectively. Both a sector and a triangle have vertices at o'_{ij}.

When the plane figure $\mathrm{Cntprj}^{(2)}(o'_{ij}, B_{ij}^{(1)A})$ is expressed as the sum of some sectors,

$$\mathrm{Cntprj}^{(2)}(o_{ij}^{/},B_{ij}^{(1)A})=\sum_{\eta}\mathrm{Cntprj}^{(2)}(o_{ij}^{/},B_{ij,\eta}^{(1)A}) \tag{14}$$

$$B_{ij}^{(2)I}=\sum_{\eta}\mathrm{Cntprj}^{(2)}(o_{ij}^{/},B_{ij,\eta}^{(1)A})+\sum_{\mu}sign^{(2)}(o_{ij}^{/},B_{ij,\mu}^{(1)I})Cntprj^{(2)}(o_{ij}^{/},B_{ij,\mu}^{(1)I}) \tag{15}$$

where η and μ run through all arcs and straight lines in the boundary of

the inaccessible surface $B_{ij}^{(2)I}$ respectively, $B_{ij,\eta}^{(1)A}$ and $B_{ij,\mu}^{(1)I}$ are arc and straight line respectively. When straight line μ is formed by the inaccessible surface $B_{ij}^{(2)I}$ and sphere k, $B_{ij,\mu}^{(1)I}$ equal to $B_{ijk}^{(1)I}$.

III-1. Analytical part

The volume of B is calculated by,

$$V=\sum_i V_i \quad (16)$$

where V_i is the volume of B_i. In this section, we assume we know i) whether $B_i=\varnothing$ or not, ii) whether $B_{ij}^{(2)I}=\varnothing$ or not, iii) which spheres form exposed vertices, arcs, and lines, and their relative orientations, and iv) the Euler-Poincaré characteristic of the accessible surface of the sphere i. By the last of this section you would understand this assumption is too much information about the topology of the fused spheres for the volume calculation.

The volume of B_i which is not zero is calculated by the following equations,

$$V_i=\frac{1}{3}\rho_i S_i^A+\sum_j \frac{1}{3}g_{ij}S_{ij}^I \quad (allow\ g_{ij}<0) \quad (17)$$

$$g_{ij}=\frac{d_{ij}^2+\rho_i^2-\rho_j^2}{2d_{ij}} \quad (allow\ g_{ij}\leq 0) \quad (18)$$

where j is the sphere index which shares the inaccessible surface $B_{ij}^{(2)I}$ with sphere i ($B_{ij}^{(2)I}\neq\varnothing$), S_i^A and S_{ij}^I are the areas of $B_i^{(2)A}$ and $B_{ij}^{(2)I}$, respectively. The first and second terms in Eq.(17) represent volumes of $\mathrm{Cntprj}^{(3)}(o_i, B_i^{(2)A})$ and $\mathrm{Cntprj}^{(3)}(o_i, B_{ij}^{(2)I})$ with signs, respectively. By using g_{ij} which is a distance with the sign between the center of the sphere i and the partition plane $P_{ij}^{(2)}$, the conditional branch of $\mathrm{sign}^{(3)}(o_i, B_{ij}^{(2)I})$ is automatically selected.

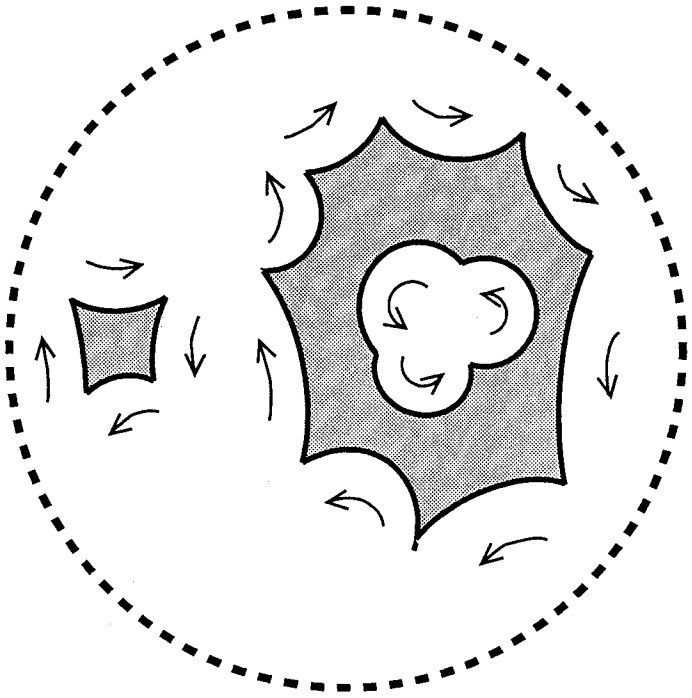


Fig. 8 Oriented arcs in boundaries of the accessible surface. Orientation of numbering of the arcs in the boundary is shown as arrows. Only accessible surfaces of solid body B_i are represented as shaded regions.

The accessible surface area of the sphere i (S_i^A) is calculated by the Richmond method which uses the Gauss-Bonnet theorem in the differential-topological geometry,

$$S_i^A=\rho_i^2\ [2\pi\chi_i^A+\sum_{\nu}(\sum_{\gamma}\Omega_{\gamma,\gamma+1}+\sum_{\gamma}k_\gamma s_\gamma)] \quad (19)$$

where ν runs through closed boundaries of the accessible surface $B_i^{(2)A}$ which has sometimes more than one boundary depending on the configuration of fused spheres, γ runs through arcs in the closed boundary ν, $\Omega_{\gamma,\gamma+1}$ is an angle formed from a tangent line to the arc γ, and a tangent line to the arc $\gamma+1$ at the exposed vertex, k_γ and s_γ are the geodesic curvature and length of arc γ, respectively, and χ_i^A is the Euler-Poincaré characteristic of $B_i^{(2)A}$. The numbering of arcs in the boundary is clockwise (Fig. 8) with respect to the accessible region. The orientation of the numbering affects only the sign of $\Omega_{\gamma,\gamma+1}$ (always negative in this method). Now let us focus on a closed boundary ν. In this boundary when arc γ is formed by sphere i and j, and arc $\gamma+1$ is formed by sphere i and k, $\Omega_{\gamma,\gamma+1}(i,j,k)$ is given as (Fig. 9),

$$\Omega_{\gamma,\gamma+1}(i,j,k)=-\cos^{-1}\left(\frac{\rho_i^2\cos\theta_{ijk}-g_{ij}g_{ik}}{r_{ij}r_{ik}}\right) \quad (20)$$

$$\cos\theta_{ijk}=\frac{d_{ij}^2+d_{ik}^2-d_{jk}^2}{2d_{ij}d_{ik}} \quad (21)$$

$$r_{ij}=\sqrt{\rho_i^2-g_{ij}^2}\ \ (r_{ij}\geq 0) \quad (22)$$

where r_{ij} is the radius of the sphere circle of $P_{ij}^{(1)}$, $\Omega_{\gamma,\gamma+1}$ and θ_{ijk} is limited as $0\leq\theta_{ijk}\leq\pi$. When arc γ is formed by sphere i and j, the geodesic curvature k_γ is given by the following equation,

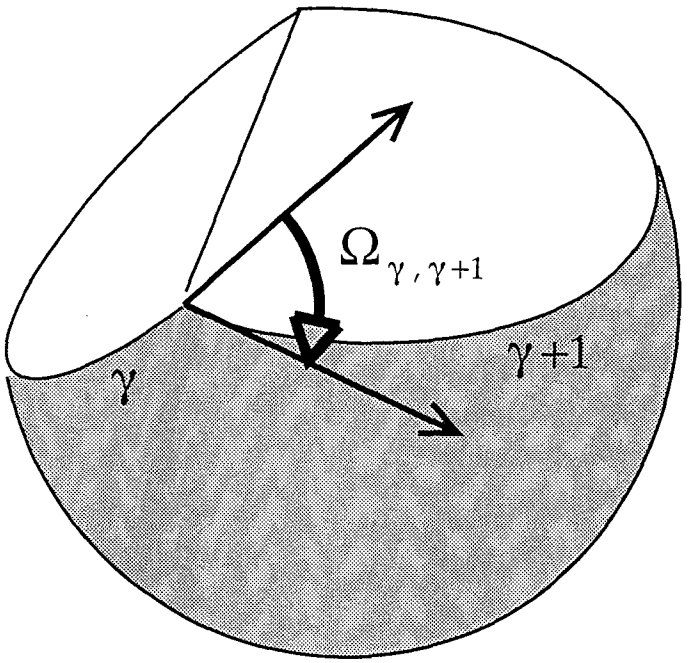


Fig. 9 Angle $\Omega_{\gamma,\gamma+1}$ formed by arc γ and $\gamma+1$. The tangent line to the arc γ and the tangent line to arc $\gamma+1$ form the angle $\Omega_{\gamma,\gamma+1}$.

$$k_\gamma = \frac{g_{ij}}{\rho_i r_{ij}} \quad (23)$$

The length of the arc γ, s_γ is given as,

$$s_\gamma = r_{ij}\Theta_\gamma \quad (24)$$

where Θ_γ is the center angle of the arc γ in the intersecting circle $P_{ij}^{(2)}$, this term will be discussed soon after in this section.

The inaccessible surface area S_{ij}^{I} is calculated by the following equations (see Fig. 10),

$$S_{ij}^{I} = \sum_{\eta} \frac{1}{2} r_{ij} s_\eta + \sum_{\mu} \frac{1}{2} h_\mu t_\mu \quad (allow\ h_\mu < 0) \quad (25)$$

where s_η and t_μ are the lengths of $B_{ij,\eta}^{(1)A}$ and $B_{ij,\mu}^{(1)I}$, η and μ run through all arcs and straight lines in the boundary of the inaccessible surface $B_{ij}^{(2)I}$ respectively, and h_μ is the distance with sign between the center of the intersecting circle o_{ij}' and the line μ. The conditional branch of sign$^{(2)}$(o'_{ij}, $B_{ij,\mu}^{(1)I}$) is automatically selected. When a sphere k forms straight line μ on $B_{ij}^{(2)I}$,

$$h_\mu(i,j,k) = h_{ijk} = \frac{-g_{ij}\cos\theta_{ijk} + g_{ik}}{\sin\theta_{ijk}} \quad (allow\ h < 0) \quad (26)$$

$$\sin\theta_{ijk} = \sqrt{1 - \cos^2\theta_{ijk}} \quad (27)$$

The expression for h_{ijk} is derived in Appendix C. The first and second terms in Eq.(25) represent areas of sectors and triangles respectively. The numbering of straight

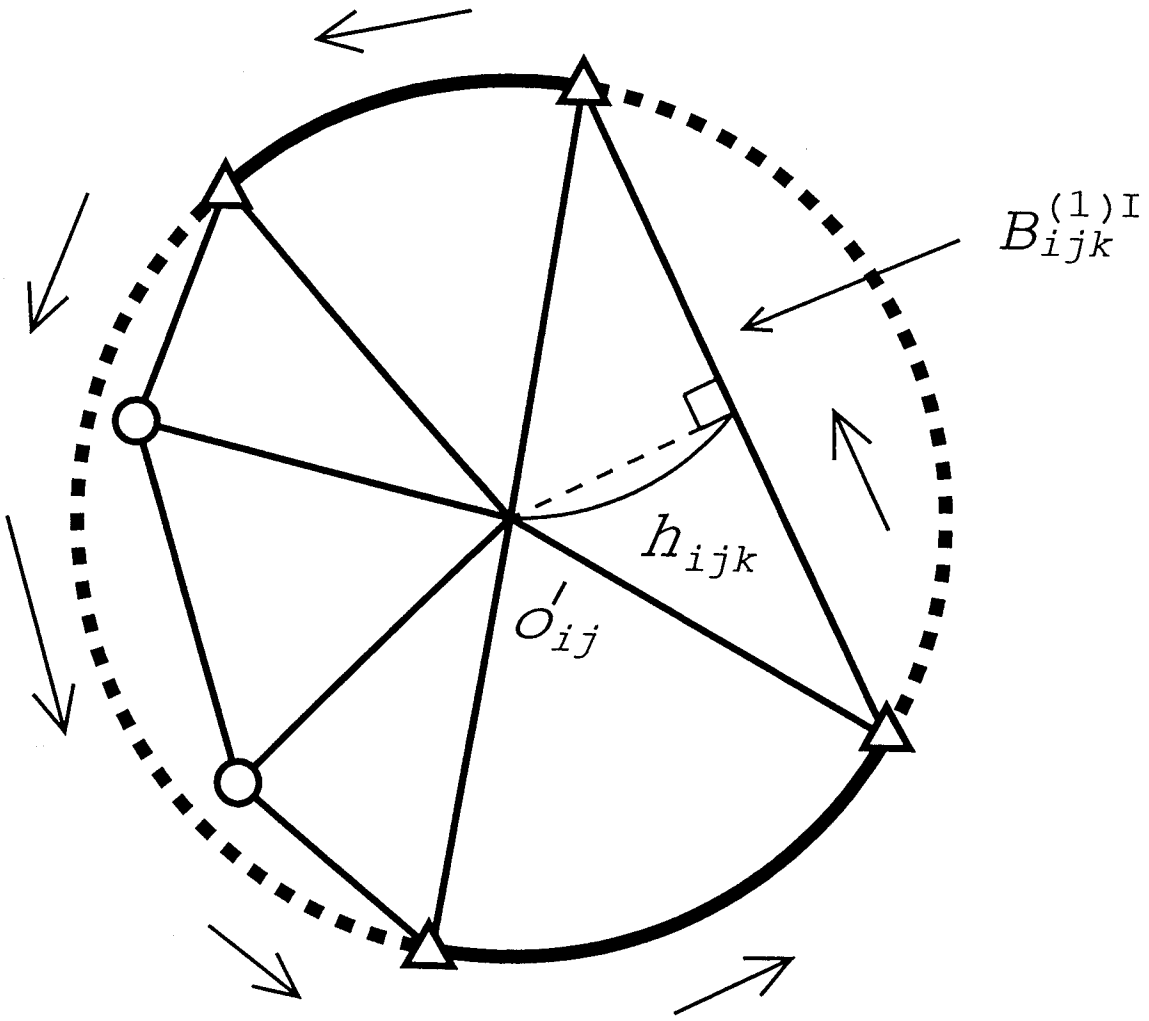


Fig. 10 Sectors and triangles in an inaccessible surface. Height with sign of the triangles are h_{ijk}. Exposed vertex and buried vertex are represented as open triangle and open circle respectively. Orientation of the numbering of the arcs and straight lines in the boundary is also shown as arrows.

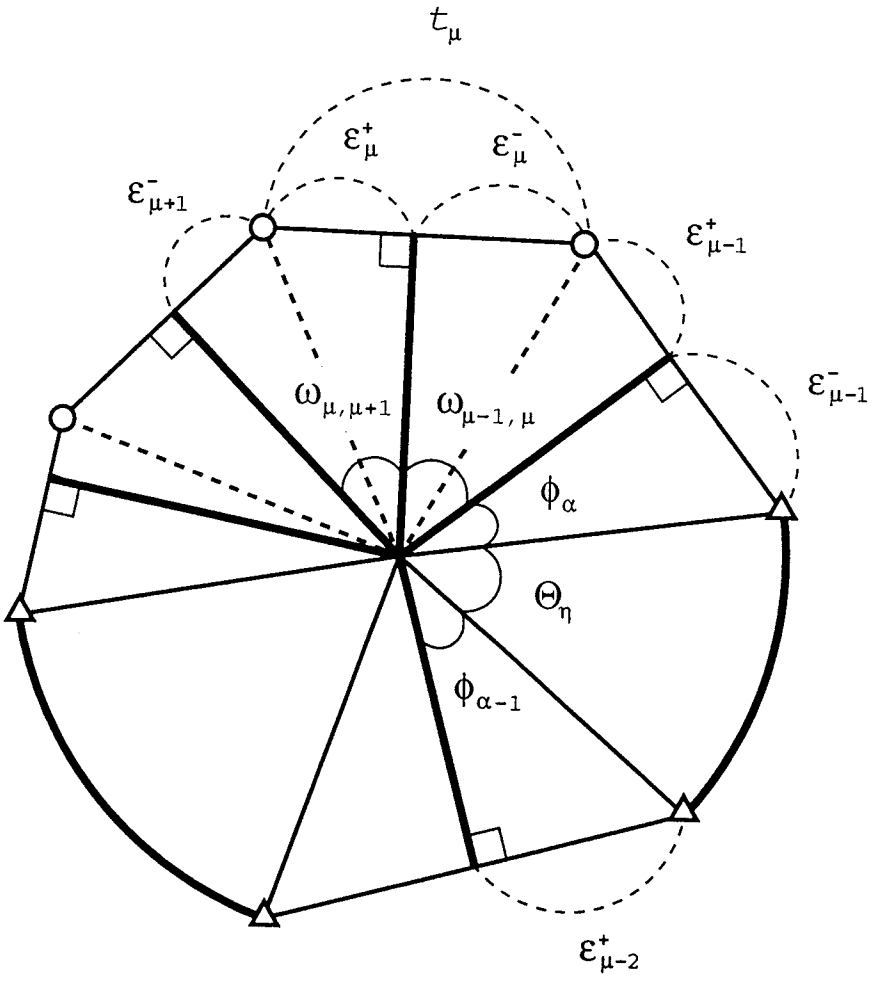


Fig. 11 Straight lines and Center angles of an inaccessible surface. The length of the line t_μ is divided into ε_μ^- and ε_μ^+. Negative value is allowed for ε_μ^- and ε_μ^+. The Solid line represents the line rectangle to straight line $B_{ijk}^{(1)I}$ from o'_{ij} center of the sphere circle $P_{ij}^{(1)}$. Center angle for the arc is represented as Θ_η. Center angle for the triangle neighboring the exposed vertex is represented as ϕ_α. Center angle for the quadrangle including the buried vertex is represented as $\omega_{\mu-1,\mu}$. Exposed vertex and buried vertex are represented as open triangle and open circle respectively.

lines in the boundary is anti-clock-wise. The straight line $B_{ij,\mu}^{(1)I}$ is decomposed into pieces, the length of the straight line, t_μ is given by following equations,

$$t_\mu=\varepsilon_\mu^-+\varepsilon_\mu^+ \quad (t_\mu\geq 0)\ (allow\ \varepsilon_\mu^-,\varepsilon_\mu^+<0) \quad (28)$$

where ε_μ^- and ε_μ^+ are defined as Fig. 11,

$$\varepsilon_\mu^-=\frac{h_{\mu-1}-h_\mu\cos\omega_{\mu-1,\mu}}{\sin\omega_{\mu-1,\mu}} \quad , \quad \varepsilon_\mu^+=\frac{h_{\mu+1}-h_\mu\cos\omega_{\mu,\mu+1}}{\sin\omega_{\mu,\mu+1}} \quad (29)$$

When straight lines μ-1 and μ in inaccessible surface $B_{ij}^{(2)I}$ are formed by sphere k and l respectively,

$$\cos\omega_{\mu-1,\mu}(i,j,k,l)=\frac{d_{ik}^2\sin^2\theta_{ijk}+d_{il}^2\sin^2\theta_{ijl}-d_{kl}^2+(d_{il}\cos\theta_{ijk}-d_{ik}\cos\theta_{ijl})^2}{2d_{ik}d_{il}\sin\theta_{ijk}\sin\theta_{ijl}} \quad (30)$$

$$(0\leq\omega_{\mu-1,\mu}\leq\pi)$$

where $\omega_{\mu-1,\mu}$ is the angle on $B_{ij}^{(2)I}$ between the projected vectors pointing from o' to sphere centers of k and l. Above equations of ε_μ^- and ε_μ^+ are valid for negative values of those variables. This is proved in Appendix C. When the line μ has an arc as its neighbor,

$$\varepsilon_\mu^-\ or\ \varepsilon_\mu^+=\sqrt{r_{ij}^2-h_\mu^2} \quad (\varepsilon_\mu^-\ or\ \varepsilon_\mu^+\geq 0) \quad (31)$$

The second term in Eq.(25) is described as the summation through the individual straight lines, $B_{ij,\mu}^{(1)I}$, of the inaccessible surface $B_{ij}^{(2)I}$. This summation can be replaced by the summation through the individual vertices of $B_{ij}^{(2)I}$. The calculation of the individual line length is not required for the inaccessible surface area calculation.

$$\sum_{\mu}\frac{1}{2}h_{\mu}t_{\mu}=\frac{1}{2}\sum_{\mu}h_{\mu}(\varepsilon_{\mu}^{-}+\varepsilon_{\mu}^{+})=\frac{1}{2}\{\sum_{\alpha}f_{\alpha}+\sum_{\beta}e_{\beta}\} \quad (32)$$

where α and β represent the exposed vertex on $B_{ij}^{(2)I}$ (which has one-to-one correspondence to A-L inner-angle) and the buried vertex on $B_{ij}^{(2)I}$ (which has one-to-one correspondence to L-L inner-angle), respectively.

When A-L inner-angle α is formed by straight line μ and the inaccessible surface $B_{ij}^{(2)I}$, and L-L inner-angle β is formed by the two adjacent straight lines μ-1 and μ and the inaccessible surface $B_{ij}^{(2)I}$,

$$f_{\alpha}=h_{\mu}\sqrt{r_{ij}^{2}-h_{\mu}^{2}} \quad (33)$$

$$e_{\beta}=\frac{2h_{\mu-1}h_{\mu}-(h_{\mu}^{2}+h_{\mu}^{2})\cos\omega_{\mu-1,\mu}}{\sin\omega_{\mu-1,\mu}} \quad (34)$$

$$\sin\omega_{\mu-1,\mu}=\sqrt{1-\cos^{2}\omega_{\mu-1,\mu}} \quad (35)$$

The first term in Eq.(25) is described as follows,

$$\sum_{\eta}\frac{1}{2}r_{ij}s_{\eta}=\frac{1}{2}r_{ij}^{2}\sum_{\eta}\Theta_{\eta} \quad (36)$$

It is the important point that the individual values of Θ_{η} is not required in the above equation. Finally following equation enables us to get required quantities, the sum of Θ_{η},

$$\sum_{\eta}\Theta_{\eta}=2\pi-\sum_{\alpha}\phi_{\alpha}-\sum_{\beta}\omega_{\beta} \quad (37)$$

where α runs through A-L inner-angles at the exposed vertices of the inaccessible surface $B_{ij}^{(2)I}$ and β runs through the L-L inner-angles at the vertices on the inaccessible surface $B_{ij}^{(2)I}$. When A-L inner-angle α is formed

by the straight line μ and inaccessible surface $B_{ij}^{(2)I}$,

$$\phi_\mu=\cos^{-1}\left(\frac{h_\mu}{r_{ij}}\right) \qquad (0\le\phi\le\pi) \quad . \quad (38)$$

Final expression of S_{ij}^{I} is as follows,

$$\begin{aligned} S_{ij}^{I} &= \frac{1}{2}r_{ij}^2\left(2\pi-\sum_\alpha \phi_\alpha-\sum_\beta \omega_\beta\right)+\frac{1}{2}\left(\sum_\alpha f_\alpha+\sum_\beta e_\beta\right) \\ &= \pi r_{ij}^2+\sum_\alpha\left(-\frac{1}{2}r_{ij}^2\phi_\alpha+\frac{1}{2}f_\alpha\right)+\sum_\beta\left(-\frac{1}{2}r_{ij}^2\omega_\beta+e_\beta\right) \end{aligned} \qquad (39)$$

In the above derivation of S_{ij}^{I}, individual length of arcs $(B_{ij,\eta}^{(1)A})$ and straight lines $(B_{ij,\mu}^{(1)I})$ are not necessary.

Let us return to the accessible surface area calculation. The summation of the last term in Eq.(19) is carried out through the all arcs on the closed boundary γ and then carried out through the all closed boundaries. This can be replaced as the summation, "at first sum through the all arcs on $B_{ij}^{(2)I}$ and then sum through the all inaccessible surfaces of $B_{ij}^{(2)I\mu}$, as follows,

$$\sum_\nu\sum_\gamma k_\gamma s_\gamma=\sum_\nu\sum_\gamma k_\gamma r_\gamma\Theta_\gamma=\sum_j\sum_\eta k_{ij}r_{ij}\Theta_\eta=\sum_j k_{ij}r_{ij}\left(2\pi-\sum_\alpha\phi_\alpha-\sum_\beta\omega_\beta\right) \qquad (40)$$

In this equation, summation through the boundaries of the accessible surfaces is changed to summation through the inaccessible surfaces. Then we can use Eq.(37). The lengths of the individual arcs are not required for the accessible surface area calculation in the volume calculation. The final expression of S_i^A is as follows,

$$
\begin{aligned}
S_i^A &= 2\pi\rho_i\chi_i^A + \sum_j \left\{ \sum_\alpha \frac{1}{2}\Omega_\alpha + k_{ij}r_{ij}\left(2\pi - \sum_\alpha \phi_\alpha - \sum_\beta \omega_\beta\right)\right\} \\
&= 2\pi\rho_i\chi_i^A + \sum_j \left\{ 2\pi k_{ij}r_{ij} + \sum_\alpha \left(\frac{1}{2}\Omega_\alpha - k_{ij}r_{ij}\phi_\alpha\right) + \sum_\beta \left(-k_{ij}r_{ij}\omega_\beta\right)\right\}
\end{aligned}
\tag{41}
$$

In the above derivation of S_i^A, individual length of arc ($B_{ij,\eta}^{(1)A}$) is not necessary.

III-2. Topological part

Our work considers all possible configurations of the spheres which include the singular cases appeared in the molecular science. The singular cases are a) one exposed vertex is formed by one spherical surface and more than two intersection planes (see Fig.12a), b) one buried vertex is formed by more than three intersection planes (see Fig.12b). For example, one of the case b) is that benzene molecule whose atoms are located on the vertices of the regular polygon.

The topological information required for the analytical calculation of the volume (V_i) and accessible surface area (A_i) of each sphere in the fused spheres are as follows,

I1) Whether the target sphere, i, is isolated (having volume of 4/3 $\pi\rho_i^3$ and χ_i^A=2) or not.

I2) A list of spheres (LSTI2(i)) making isolated intersecting circles with the sphere i.

The component (j) of the list (LSTI2(i)={j}) is the sphere index which makes the isolated intersecting circle with the sphere i.

I3a) A list of pairs of spheres (LSTI3a(i)) making the A-A inner-angles at exposed vertices of the sphere i.

The component ((j,k)) of the list (LSTI3a(i)={(j,k)}) consists of the indices of the two spheres which make the A-A inner-angle with the sphere i. Depending on the configuration a same pair appears twice in this list (e.g. the fused spheres made of three spheres like Fig. 1).

I3b) A list of pairs of spheres (LSTI3b(i)) making the A-L inner-angles at exposed vertices of the sphere i.

The component ((j,k)) of this list (LSTI3b(i)={(j,k)}) consists of the index of the sphere (j) making the inaccessible surface on which this A-L inner-angle is located, and the index of the sphere

(k) making the A-L inner-angle with the spheres i and j.

I4) A list of triplets of spheres (LSTI4(i)) making the L-L inner-angles at the exposed and buried vertices of the sphere i.
The component ((j,k,l)) of the list (LSTI4(i)={(j,k,l)}) consists of the index of the sphere (j) making the inaccessible surface on which this L-L inner-angle is located, and the indices of the two spheres (k and l) making the L-L inner-angle with the spheres i and j.

We do not need information of the orientation of the spheres in the pair in I3a), I3b), and the triplet in I4) relative to each other. Above topological information from I1) to I4) can be obtained by the simple analytical calculation explained in the Appendix A. From these information we make the following lists.

0') The Euler-Poincaré characteristics of the accessible surface of sphere i.

2') The list of spheres (LSTI2'(i)) appeared in 2), 3), and 4).

The list LSTI2' corresponds to the summation through the inaccessible surfaces, $B_{ij}^{(2)I}$, in B_i. The V_i and A_i can be calculated for the fused spheres by summing up all the contributions listed above.

$$V_i = T_1(i) + \sum_{LSTI2'(i)} T_2(i,j) + \sum_{LSTI3a(i)} T_{3a}(i,j,k) + \sum_{LSTI3b(i)} T_{3b}(i,j,k) + \sum_{LSTI4(i)} T_4(i,j,k,l) \quad , \tag{42}$$

$$A_i = \left[T_{A1}(i) + \sum_{LSTI2'(i)} T_{A2}(i,j) + \sum_{LSTI3a(i)} T_{A3a}(i,j,k) + \sum_{LSTI3b(i)} T_{A3b}(i,j,k) + \sum_{LSTI4(i)} T_{A4}(i,j,k,l) \right] , \quad (43)$$

$$T_1(i)=\frac{1}{3}\rho_i T_{A1}(i) \quad , \quad T_2(i,j)=\frac{1}{3}\rho_i T_{A2}(i,j)+T_{I2}(i,j) \quad ,$$

$$T_{3a}(i,j,k)=\frac{1}{3}\rho_i T_{A3a}(i,j,k) \quad , \quad T_{3b}(i,j,k)=\frac{1}{3}\rho_i T_{A3b}(i,j,k)+T_{I3}(i,j,k) \quad , \quad (44)$$

$$T_4(i,j,k,l)=\frac{1}{3}\rho_i T_{A4}(i,j,k,l)+T_{I4}(i,j,k,l)$$

This formulation contains the very rare case of the configuration of the spheres shown in Fig. 12.

In what follows, we show a case which is less general but contains most of the configurations of spheres in the practical use (for example the calculation of the excluded volume of a molecule represented by fused spheres etc.). If each exposed vertex is made of one A-A inner-angle and two A-L inner-angles, and each buried vertex is made of three L-L inner-angles, then the topological information needed for the analytical calculation of the volume of each sphere in the fused spheres is decreased. The topological information for such a case is as follows,

1) Whether the target sphere, i, is isolated (having volume of 4/3 $\pi\rho_i^3$ and $\chi_i^A=2$) or not.

2) A list of spheres (LST2(i)) making isolated intersecting circles with the sphere i.

   The component (j) of the list (LST2(i)={j}) is the sphere index

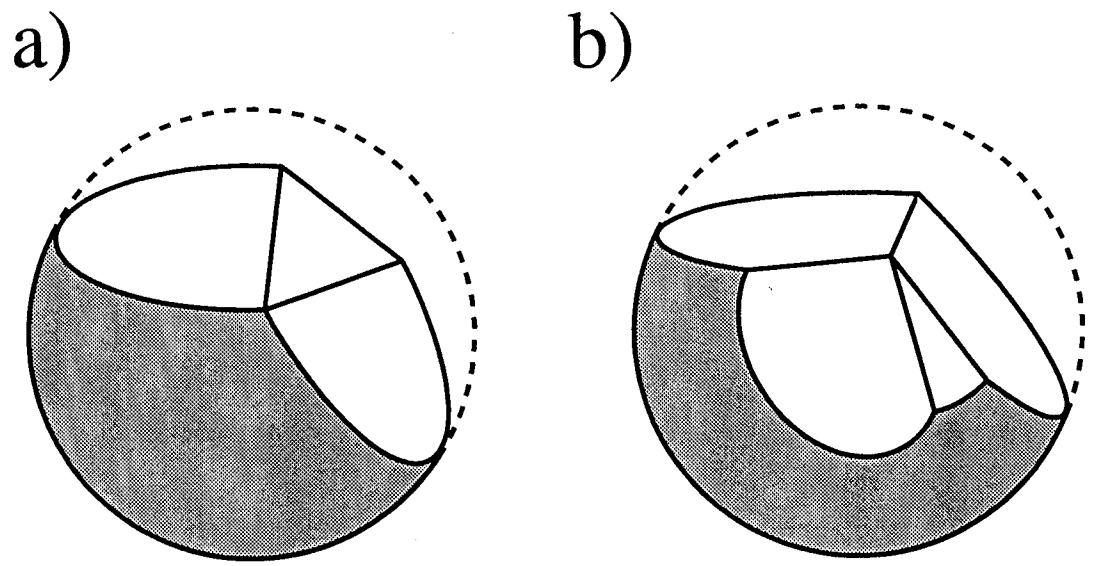


Fig. 12 Solid body B_i in special configuration of fused spheres a) The solid body B_i where an exposed vertex has L-L inner-angle. b) The solid body B_i where a buried vertex has more than three L-L inner-angle.

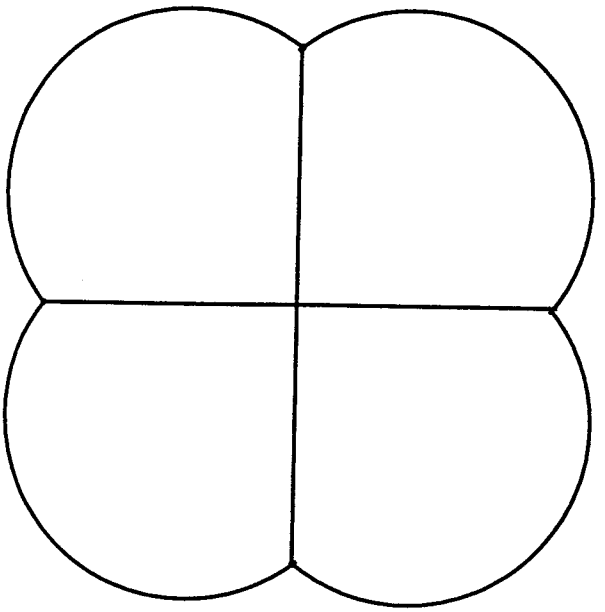


Fig. 13 Special configuration where sphere centers are located on the regular polygon.

which makes isolated intersecting circles with the sphere i.

3) A list of pairs of spheres(LST3(i)) making the exposed vertices of the sphere i.

The component ((j,k)) of the list (LST3(i)={(j,k)}, j<k) consists of the indices of the two spheres which make and exposed vertex with the sphere i.

4) The list of triplets of spheres (LST4(i)) making a buried vertex of the sphere i.

The component ((j,k,l)) of this list (LST4(i)={(j,k,l)}, j<k<l) consists of the indices of the three spheres which make a buried vertex with the sphere i.

Above topological information is obtained by the simple analytical calculation explained in the Appendix A. From these information we make the following lists.

0') The Euler-Poincaré characteristic of the accessible surface of sphere i.

2') The list of spheres (LST2'(i)) appeared in 2), 3), and 4).

The V_i and A_i can be calculated in a simpler way as,

$$V_i = T_1(i) + \sum_{LST2'(i)} T_2(i,j) + \sum_{LST3(i)} \{T_{3a}(i,j,k) + T_{3b}(i,j,k) + T_{3b}(i,k,j)\} + \sum_{LST4(i)} \{T_4(i,j,k,l) + T_4(i,k,l,j) + T_4(i,l,j,k)\} \quad , \tag{45}$$

$$A_i = T_{A1}(i) + \sum_{LST2'(i)} T_{A2}(i,j) + \sum_{LST3(i)} \{T_{A3a}(i,j,k) + T_{A3b}(i,j,k) + T_{A3b}(i,k,j)\} + \sum_{LST4(i)} \{T_{A4}(i,j,k,l) + T_{A4}(i,k,l,j) + T_{A4}(i,l,j,k)\} \quad , \tag{46}$$

Note that the special case like Fig. 13 (a configuration of which sphere centers are located on the regular polygon, is one of the example) can be treated by doing the topological search explained above for each B_i, because the vertex has different meaning for each B_i.

We can get the final expression described in the beginning of the section 1 by summing V_i and A_i and by doing the topological search for B_i at one time (in this case, the special case like Fig. 13 cannot be treated). If we assume all radii of the spheres are equal,

$$\sum_{LSTI'} T_{AI}(i)=2\pi\rho^2\sum_{LSTI'}\chi_i^A=2\pi\rho^2\chi \quad (47)$$

χ is the Euler-Poincaré characteristic of the whole body, B, of the fused spheres (two examples are shown in Fig. 14).

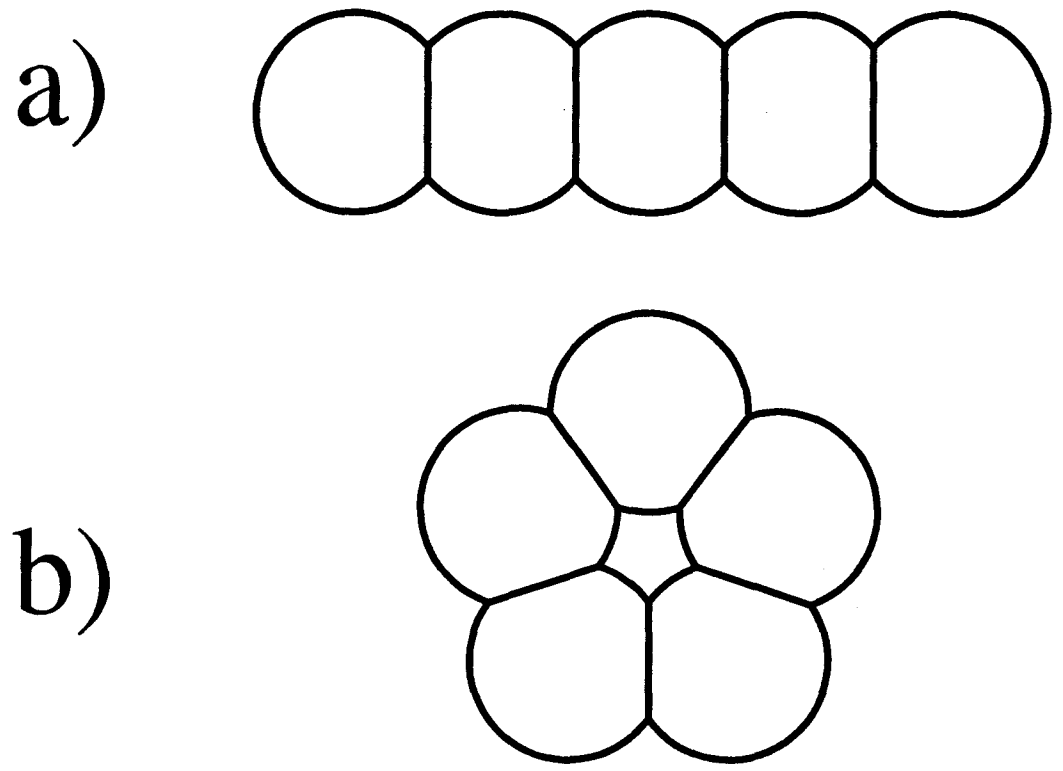


Fig. 14 Examples of the Euler-Poincaré characteristic of the fused spheres B. a) Straight chains which has isolated sphere circles and no vertex. In this case χ=2. b) Ring chains which has isolated sphere circles and no vertex. In this case χ=0.

Implementation of the method

The excluded volumes of a pentane and a protein is calculated by using our program ESPER (Extended Scaled Particle theory PERformer). The values of the excluded volume and accessible surface area of a pentane are 353.8Å^3 and 256.3Å^2 for all-trans conformation, 344.6Å^3 and 247.3Å^2 for all-

gauche conformation, respectively. The values of the excluded volume and accessible surface area of a protein, avian pancreatic polypeptide (1PPT) [18], are 8524Å^3 and 3337Å^2, respectively. The excluded volumes calculated by using the numerical method with 3-dimensional mesh (the grid size is 0.02Å) are 353.8Å^3 for the all-trans pentane, 344.4Å^3 for the all-gauche pentane, and 8522Å^3 for 1PPT. In the calculation, the molecule is modeled as a fused spheres whose radius of the sphere is the van der Waals radius of the atom plus the radius of water molecule (1.375Å). The TIPS parameters [19] are employed for the conformation and the van der Waals radii of the pentane. The OPLS parameters [20] are employed for the van der Waals radii of 1PPT. The ESPER was originally intended for the calculation of first and second derivatives of the volumes in the scaled particle theory [1,2] and has not been well optimized for the acceleration of the speed of the calculation yet.

Discussion

In the previous sections, we could have derived an elegant expression for the volume of fused spheres. Our algorithm employs the method of Richmond [10] and the building block usage of Dodd [8], but the final expression for the fused spheres is more elegant than the old methods, because some analytical functions (g_{ij}, h_{μ}, e_{β}, etc.) appeared in the expression absorbs the conditional branches which otherwise is necessary to take account for the configuration of the spheres. Topological information required is explicitly separated.

One of the reason why we can express the volume by the spheres radii and inter-sphere distances, is as follows. In the methods of Richmond (surface area calculation) [10] and Dodd (surface area and volume calculation) [8], the topological information of the arcs and straight lines of the building blocks of the fused spheres is required, because the length of the arcs and straight lines are needed. In the topological search algorithm, it is easy to find the list of the spheres making vertices, but it is very difficult to find those of the spheres making the arcs (or straight lines). We can easily find one vertex (e.g. as an intersection point of partition planes explained in Appendix A). But in order to find the vertex located at the other edge of an arc (or straight line), we must search this among other vertices. Since Cartesian coordinate is used in such a search and in the calculation of the length of the arc (or straight line), the final analytical expression of the volume or surface area of both methods could not be expressed by spheres radii and inter-sphere distances. In our algorithm we do not need the information with respect to which vertices make a pair of edges for the individual arc (or straight line).

In exchange for the advantage described above, we cannot trace the boundaries of accessible surfaces in our method. According to the Gauss-

Bonnet theorem, we need the Euler-Poincaré characteristics of the accessible surfaces in the calculation of the accessible surface area of the fused spheres. To avoid direct calculation of the Euler-Poincaré characteristics, Richmond and Dodd used the trace of the closed paths of the arcs which surround the accessible surfaces in order to count the number of the isolated accessible surfaces and the holes of the surface. When one wants to get only the analytical expression of surface area (like Richmond method), this technique is natural. But in the analytical volume calculation, we can directly calculate Euler-Poincaré characteristics. In our method we directly count the number of faces, lines and vertices of the inaccessible surfaces in order to derive the Euler-Poincaré characteristics of the accessible surfaces (Appendix B). Therefore our method does not require the trace of the closed boundaries of accessible surface.

What is minimum topological information required to calculate the volume of fused spheres? Our method suggests that the answer is the topology of the building block, especially isolated sphere circles, exposed vertices, and buried vertices. These topological elements, especially buried vertices, depend on the local configuration of up to four spheres. From this fact, our final expression for the volume of the fused spheres becomes a sum of terms corresponding to up to four spheres in consistent with Kratky's finding. In the method of Dodd, length of the straight line, which is required in their algorithm, depends on the local configuration of up to five spheres. When we know nothing about the 3-dimensional configuration of the fused spheres, buried vertices of the building block enables you to image the shape of interior of the fused spheres. In the Kratky's method [7], topological information of the interior of the fused spheres is expressed as many intersecting regions made by spheres. Our method may be the "shortcut algorithm" Sheraga's group wanted to complete [8,9].

We found that our work and the work in the computational

geometry which uses power diagram complement each other. In the computational geometry, there exists the efficient search algorithm for the topology of the power diagrams [16,17]. The building block in our work exactly corresponds to the intersection of the sphere and the cell of the power diagram [11,13]. In our work, after the topology of the building blocks is known, the volume of the building block can be calculated at once by using "one" analytical equation without considering conditional branches which are caused by the further division of the building block. But all other works using power diagrams divide the building block into pieces with simpler geometry [11,12,14,15].. This further division of the building block is not necessary. The hybrid of our algorithm and power diagram method would be the best algorithm in the mathematical sense. But the problems of singular cases described in section 2.3.2 must be cleared, because these problems have not been explicitly discussed in the power diagram methods related to the volume calculation of the fused spheres.

The acceleration of the speed of the algorithm is the next problem in our work. The target of the acceleration is very clear, because the topological part and the analytical part are explicitly separated in our algorithm. The search of the topology of the building block must be accelerated, because the analytical part can easily be optimized in the use of the vector machine. The power diagram method mentioned before would be suitable for this purpose. But we cannot say which is faster the power diagram method or the vector/parallel machine dependent algorithms like the method of Dodd [8].

When we express the analytical volume of the fused spheres with equal radii, the analytical expression of volume of this fused spheres depends on the Euler-Poincaré characteristics of the whole fused spheres χ, not each sphere χ_i^A. This result is one of the good example how our method reflect the shape of the fused spheres well, and is very interesting

as well for the future use in the virial coefficient calculation in the liquid state theories.

Although the algorithm proposed here is general enough to be applied to many problems which concerns the volume and accessible surface area of a geometrical 3-dimensional body consisting of fused spheres, an application which takes advantage of the algorithm most is probably to the scaled particle theory (SPT) of liquids [1,2]. The scaled particle theory is a method to evaluate the thermodynamic work required to make a cavity in the system of hard spheres. If the cavity is regarded as a region from which solvent molecules are excluded, then the thermodynamic work is a part of the solvation free energy of a solute molecule. The scaled particle theory requires derivatives of the excluded volume with respect to a scaling parameter which linearly scales the cavity up from 0 to its final size as seen in

$$\bar{G}_c = N_A k_B T \left[-\ln[1-\rho V_c(0)] + \rho \int_0^1 G(\lambda) \left(\frac{\partial V_c(\lambda)}{\partial \lambda} \right) d\lambda \right] \quad (48)$$

where N_A, k_B, T have usual meanings, λ is the scaling parameter and $G(\lambda)$ is the radial distribution function of solvent molecules around solute at contact. Our new method explained above enables us to take derivatives of the excluded volume with respect to the scaling parameter. This was indeed a key to the extension of the scaled particle theory to an arbitrary shaped solute molecule. We have succeeded in the calculation of the solvation free energy of the aqueous solution of the nonpolar (normal-alkanes, from methane to dekane) [1] and polar solutes (normal-alcohols, from methanol to pentanol) [2] by using this new algorithm. We can get the perfect agreement of this calculated result with the experimental data. This means the scaled particle theory can be applied to the larger molecules such as proteins.

Appendix

When the molecule is small enough so that its surface (accessible and inaccessible) can be displayed by the computer graphics. You may be able to find the topological information to be used in the volume calculation by visual inspection. But when the molecule is large such as protein, you must get the information by the calculation. An practical method to obtain the topological information is explained in the Appendix A. This method has been implemented in the program ESPER which calculates the solvation free energy of the molecule by the extended scaled particle theory [1,2].

In the Appendix B, the brief explanation and derivation of the Euler-Poincaré characteristic of the accessible surface is presented. In the Appendix C, the proof of the equations of h_{ijk}, ε_{μ}^{-}, and ε_{μ}^{+} are given.

APPENDIX A: ALGORITHM USED IN THE SEARCH OF TOPOLOGY OF THE FUSED SPHERES

Loop for single spheres

1. List the isolated spheres (condition A1), completely buried spheres (condition A2), and others (stored in list B).

Loop for pairs of spheres in list B

2. List the isolated sphere circles (condition B1) and completely buried sphere circles (condition B2), and others (stored in list C).

Here, all vertices of the fused spheres are not searched at once, but vertices relating to each sphere are searched step by step.

Loop for single spheres, represented sphere i, in list B

Loop for pairs of partition planes of sphere i in list C.

3-1. List the vertices each of which is formed by two partition planes and the target sphere i. Vertices are stored in list D.

Loop for vertices in list D and the partition plane of sphere i in list C.

3-2. Check if each vertex is remained or not by using the partition planes. Remained ones are exposed vertices.

Loop for triplet of partition planes of sphere i in list C.

4-1. List the vertices each of which is formed by three partition planes. Vertices are stored in list E.

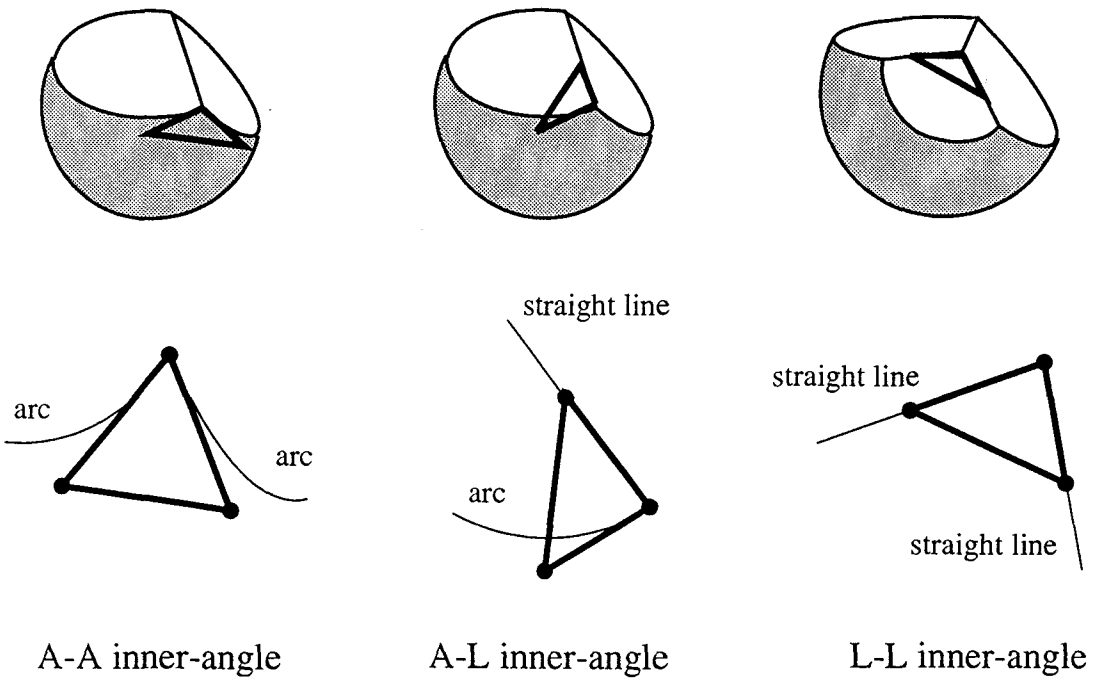
Loop for vertices in list E and the partition plane of sphere i in list C.

4-2. Check each vertex is remained or not by using the partition planes. Remained ones are buried vertices.

The exposed vertex corresponds one A-A inner-angle and two A-L

inner-angles of focused sphere i. The buried vertex corresponds to three L-L inner-angles of focused sphere i.

In most of the cases, the conditions listed above is enough for the topology search. If the very rare case exists, in which some exposed and buried vertices are located at a same position (see Fig. 12),



A-A inner-angle A-L inner-angle L-L inner-angle

Fig. 15 Schematic drawing of the triangles of A-A, A-L, L-L inner angles, which are used in the search of the remained inner-angles for very rare case when two or more vertices are located at the same position. The triangles have vertices on the straight line or the tangent line to the arc which make inner-angles. If an inner-angle is remained, any of the vertices of the triangle of the inner-angle are not cut out by any partition plane which makes vertex located at the same position with the vertex of the inner-angle.

5-1. List the exposed and buried vertices which are located at the same positions. Partition planes relating with these vertices located at the same position are listed in F.

5-2. List all inner-angles considered to be possible as A-A, A-L, and L-L inner-angles located at the same position. Make a triangle which is made of each (A-A or A-L or L-L) inner-angle (see Fig. 15).

5-3. Check if each vertex of the triangle is remained or not by using the partition planes in list F. When all three of the vertex of the triangle is not cut out by any partition plane, the (A-A or A-L or L-L) inner-angle corresponds to this triangle is remained.

Conditions used above algorithm and the procedure which checks if the vertex is remained or not by using partition planes are as follows,

A1) If any sphere j(≠i) satisfies the condition described below, the sphere *i* is an isolated sphere.

$$\forall j;\ g_{ij} \geq \rho_i\ (d_{ij} \neq 0)\quad and\quad \rho_i \geq \rho_j (d_{ij}=0)\quad and\quad i<j\ (d_{ij}=0\ ,\ \rho_i=\rho_j) \tag{49}$$

A2) If a sphere j(≠i) satisfies the condition described below, the sphere *i* is completely buried.

$$\exists j;\ g_{ij} \leq \rho_i\ (d_{ij} \neq 0)\quad or\quad \rho_i<\rho_j (d_{ij}=0)\quad or\quad i>j\ (d_{ij}=0\ ,\ \rho_i=\rho_j) \tag{50}$$

B1) The sphere circle made by pair of spheres (i,j) is isolated, when any other pair of spheres belonging to B satisfies the condition described below,

$$\begin{aligned} \forall k;\ h_{ijk} > r_{ij}\ (\sin\theta_{ijk} \neq 0)\quad and\quad g_{ij} \leq g_{ik}\ (\sin\theta_{ijk}=0) \\ and\quad d_{ij}<d_{ik}\ (\sin\theta_{ijk}=0,\ g_{ij}=g_{ik}) \end{aligned} \tag{51}$$

B2) The sphere circle made by pair of spheres (i,j) is completely buried, when another pair of spheres (i,k) belonging to B satisfies the condition described below,

$$\exists k;\ h_{ijk}<-r_{ij}\ (\sin\theta_{ijk}\neq 0)\quad or\quad g_{ij}>g_{ik}\ (\sin\theta_{ijk}=0)\quad or\quad d_{ij}>d_{ik}\ (\sin\theta_{ijk}=0,\ g_{ij}=g_{ik}) \tag{52}$$

The procedure of checking if the vertex is remained or not

A vertex of sphere i is not remained when one of the partition planes of sphere i satisfies the following condition which uses the Cartesian coordinate frame having the origin at the center of the sphere i.

$$g_{ij}>n_x(x-x_0)+n_y(y-y_0)+n_z(z-z_0) \tag{53}$$

where (n_x, n_y, n_z) is the normal vector of the partition plane $P_{ij}^{(2)}$ pointing from sphere i to j, (x, y, z) is the coordinate of the vertex.

APPENDIX B: EULER-POINCARÉ CHARACTERISTIC OF THE ACCESSIBLE SURFACE OF A SPHERE

Euler-Poincaré characteristic of the accessible surface of sphere i, χ_i^A, is calculated as follows,

$$\chi_i^A = c_i^A + n_i^A - d_i^A \tag{54}$$

where c_i^A, n_i^A, d_i^A are the numbers of vertices on the accessible surface, regions of closed boundaries on the accessible surface (an isolated accessible region is +1, a hole region is -1), and arcs on the accessible surface, respectively. Some examples are shown in Fig. 16. Since it is difficult to check if the region is accessible or hole, we calculated in alternative way.

If $B_i\neq\varnothing$, Euler-Poincaré characteristic of the accessible surface can

be calculated as follows,

$$\chi_i^A = 2 - \left(c_i + n_i^I - \sum_j c_{ij} \right) \quad (55)$$

$$c_i = a_i + b_i \ , \quad c_{ij} = q_{ij} + \frac{q'_{ij}}{2} \quad (56)$$

$$q_{ij} = \frac{a_{ij}}{2} \ , \quad q'_{ij} = a_{ij} + b_{ij} - q_{ij} \quad (57)$$

where c_i are the total number of vertices in the body B_i, a_i and b_i are the numbers of the exposed and buried vertices in the body B_i respectively, a_{ij} and b_{ij} are the numbers of A-L and L-L inner-angles in the inaccessible surface $B_{ij}^{(2)I}$, respectively, c_{ij} is the numbers of arcs and lines in the inaccessible surface $B_{ij}^{(2)I}$. If $B_i = \varnothing$, χ_i^A is equal zero.

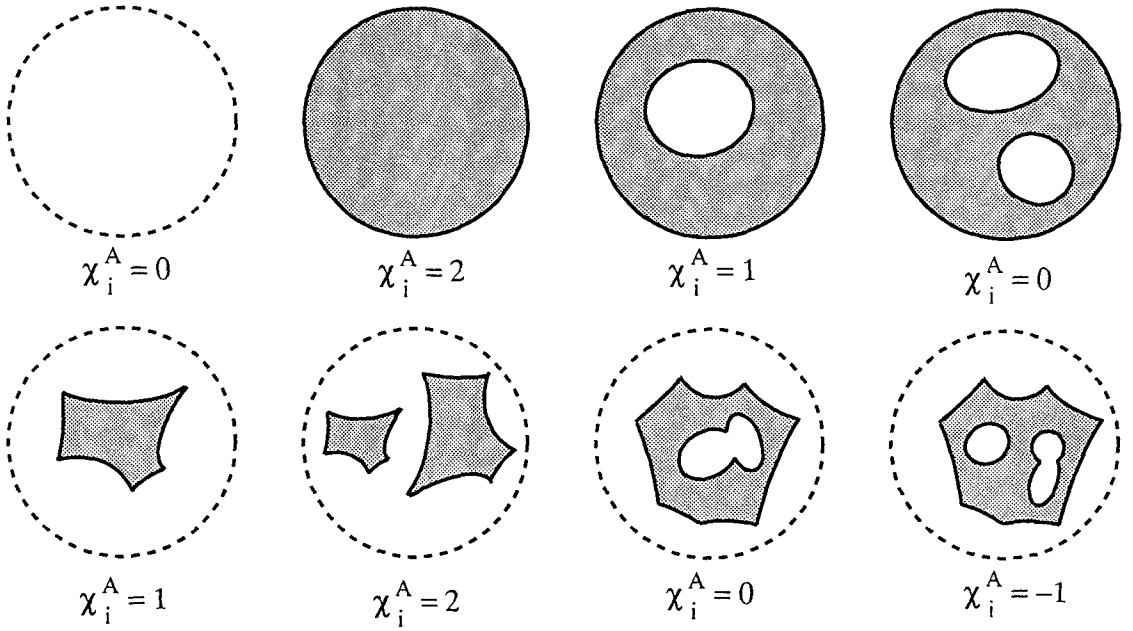


Fig. 16 Schematic drawing of the accessible surface. Shaded region represents accessible surface. Inaccessible surfaces are not shown. Euler-Poincaré characteristic of the accessible surfaces are listed below each figure.

APPENDIX C: DERIVATION OF THE EQUATIONS OF h_{ijk}, ε_{μ}^{-}, AND ε_{μ}^{+}

At first we derive the equation for h_{ijk} but you will find ε_{μ}^{-} and ε_{μ}^{+} can be derived in analogous way with h_{ijk}. When two sphere circles of a sphere i have two intersection points Q, the projection of the sphere circles onto the plane on which centers of three spheres i, j, and k are located, is represented as Fig. 17. The sphere i and angle θ_{ijk} are represented as a circle and a small arc. In the Fig.17 the centers of spheres j and k are located at the direction of y-axis, and the direction of thick arrow, respectively. The value of h_{ijk} corresponds to the x component of the position of Q, even if the value of h_{ijk} is negative. The position of Q is derived by solving these two equations which are also valid for negative values of g_{ij} and g_{ik}. Second, the derivation of ε_{μ}^{-} for which the same

$$\begin{cases} y = g_{ij} \\ \cos\theta_{ijk}(y - g_{ik}\cos\theta_{ijk}) = -\sin\theta_{ijk}(x - g_{ik}\sin\theta_{ijk}) \end{cases} \quad (58)$$

figure as Fig.17 can be used by replacing the meaning of the geometrical elements is following. In this explanation, straight lines μ-1 and μ in inaccessible surface $B_{ij}^{(2)l}$ are formed by spheres k and l, respectively. When three partition plane of sphere i have a intersection point Q, the projection of the two partition planes onto the partition plane made by the spheres i and j is represented as Fig.17. In the figure the sphere circle made by i and j is represented as a circle. An angle $\omega_{\mu-1,\mu}(i,j,k,l)$ is represented as a small arc. The projected positions of the centers of sphere k and l are located at the direction of thick arrow, and the direction of y-axis, respectively. The value of ε_{μ}^{-} corresponds to the x component of the position of Q, even if the value of ε_{μ}^{-} is negative.

$$\begin{cases} y=h_{ijk} \\ \cos\omega_{\mu-1,\mu}\cdot(y-h_{ijl}\cos\omega_{\mu-1,\mu})=-\sin\omega_{\mu-1,\mu}\cdot(x-h_{ijl}\sin\omega_{\mu-1,\mu}) \end{cases} \tag{59}$$

We can derive ε_{μ}^{+} in the same way described above.

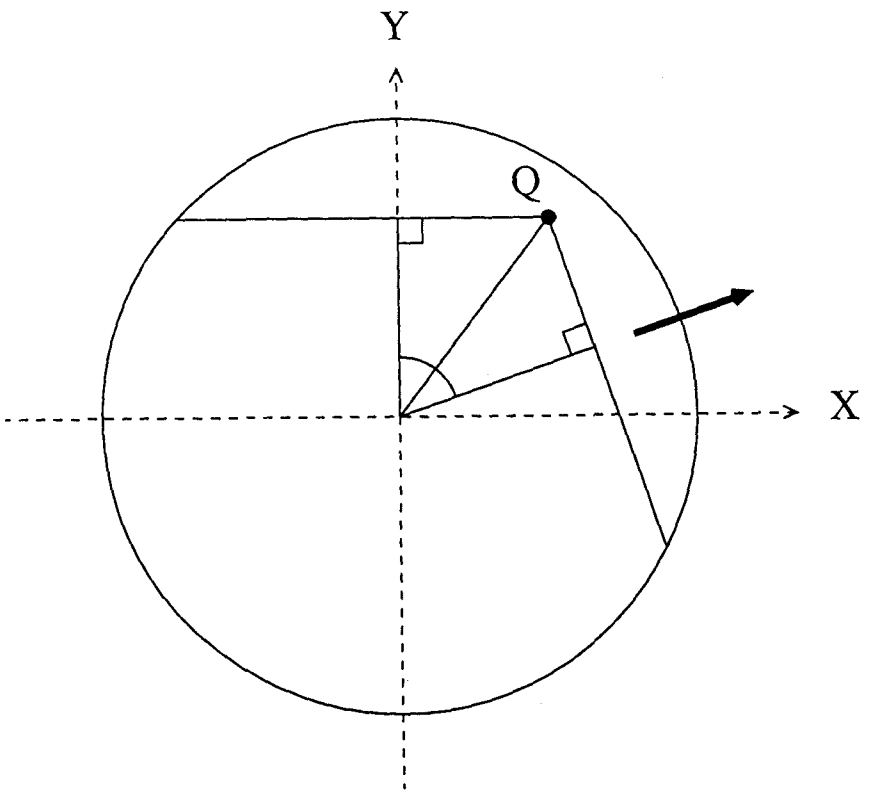


Fig. 17 This figure has different meaning depending on the function, h_{ijk}, ε_{μ}^{-}, or ε_{μ}^{+} for which this figure is used in the derivation, because theses three functions can be derived in analogous way (see detail in the text).

References

1. M. Irisa, K. Nagayama, and F. Hirata, *Chem. Phys. Letters* **207**, 430 (1993).

2. M. Irisa, K. Nagayama, and F. Hirata, *Mol. Phys.* in press.

3. F. H. Ree and W. G. Hoover, *J. Chem. Phys.* **40** 939 (1964).

4. F. H. Ree and W. G. Hoover, *J. Chem. Phys.* **46** 4181 (1967).

5. F. M. Richards, *Ann. Rev. Biophys. Bioeng.* **6**, 151 (1977).

6. Paul G. Spirakis, *Proc. 7th Annual Symposium on Theoretical Aspects of Computer Science, Lecture Notes in Computer Science* **182**, 328 (1985).

7. K. W. Kratky, *J. Statist. Phys.* **25**, 619 (1981).

8. L. R. Dodd and D. N. Theodorou, *Mol. Phys.* **72**, 1313 (1991).

9. K. D. Gibson and H. A. Sheraga, *Mol. Phys.* **62**, 1247 (1987).

10. T. J. Richmond, *J. Mol. Biol.* **178**, 63 (1984).

11. F. Aurenhammer, *SIAM J. Comput.* **16**, 78 (1987).

12. D. Avis, B. K. Bhattacharya, and H. Imai, *The Visual Computer* **3**, 323 (1988).

13. H. Edelsbrunner, *Discrete Comput. Geom.* **13**, 415 (1995).

14. H. Edelsbrunner and P. Fu, *National Center for Supercompuring Applications, Tech. Rep.* 010 (1993).

15. H. Edelsbrunner, M. Facello, P. Fu, and J. Liang, *Proceedings of the 28th Annual Hawaii International Conference on System Science* 256 (1995).

16. F. Aurenhammer, *ACM Computing Surveys* **23**, 345 (1991).

17. H. Edelsbrunner, *EATCS Monographs on Theoretical Computer Science*, **10**, Springer-Verla (1987)
"Algorithms in combinatorial geometry".

18. T. L. Blundell, J. E. Pitts, I. J. Tickle, S. P. Wood, C.-W. Wu, *Proc. Nat. Acad. Sci. USA* **78**, 4175 (1981).

19. W. L. Jorgensen, *J. Am. Chem. Soc.* **103**, 335 (1981).

20. W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.* **110**, 1657 (1988).

Chapter 6

Conclusions

I developed the new method, the extended SPT with PB equation, for calculating the solvation free energy of biomolecules. This method was successfully applied to alkanes, alcohols and proteins. Our new method is superior to the previous ones in three aspects. 1) In the calculation of the solvation free energy, our new method requires less than 1/100 of the cpu-time for the molecular dynamics simulation. 2) This method is suitable for the analysis of the solvation of the protein, because it can be used without any problems on the overwhelmingly large configurational space due to the solvent molecules. 3) Temperature dependence of the solvation free energy can be obtained by changing only few parameters. To apply the empirical method using accessible surface area of the solute to the thermal stability, experimental parameter sets at individual temperatures are needed.

It has been very difficult to consider the effect of water in studying the large molecules. Hydration free energy can be used as the effective potential representing the effects of water in long time scale where the effect of configurational changes of water is averaged. Conformational change of the protein and the interface of the two proteins (e.g. motor proteins, actin and myosin) are good targets.

Acknowledgements

I would like to thank Professor Toshio Yanagida, University of Osaka, Professor Kuniaki Nagayama, University of Tokyo, and Professor Fumio Hirata, Institute of Molecular Science, for their helpful discussions and encouragements in order to complete this work, Dr. Takuya Takahasi, University of Tokyo, for his kind help in constructing the hybrid of the extended SPT and the Poisson-Boltzmann equation, Dr. Haruki Nakamura and Dr. Akinori Kidera, Biomolecule Engineering Research Institute (BERI), Makio Tokunaga, Yanagida biomotron project ERATO JRDC, for their kind suggestions in the process of the application of our method on the protein, Dr. Toshio Yamazaki, University of Osaka, for his kind help in the proof of the analytical calculation of the volume of fused spheres, all members in four groups, Yanagida biomotron project ERATO JRDC, Laboratory of Professor Yanagida in University of Osaka, Department of Bioinfomatics in BERI, and Nagayama Biometorology Laboratory in JEOL, for their kind encouragements.