

# Calculation of nuclear magnetic resonance order parameters in proteins by normal mode analysis

Shinji Sunada and Nobuhiro Go<sup>a)</sup>

*Department of Chemistry, Graduate School of Science, Kyoto University, Kyoto 606-01, Japan*

Patrice Koehl

*UPR 9003 du CNRS, ESBS, Boulevard Sebastien Brant, 67400 Illkirch Graffenstaden, France*

(Received 6 September 1995; accepted 15 December 1995)

An analytic formula is developed for calculating the generalized NMR order parameters in a protein from the normal mode analysis (NMA). The generalized order parameter,  $\bar{S}^2$ , is given as thermal ensemble average of a Taylor series in powers of  $\Delta$ , the displacement of internuclear vector from its mean. Henry and Szabo developed a method to calculate the ensemble average based on the NMA carried out in the Cartesian coordinate space (CCS). However, atomic motions in each individual CCS normal modes are linear in the three-dimensional space, which may cause interatomic distances even between covalently bonded atoms to change significantly. In this situation Henry and Szabo proposed to use a special formula for  $\bar{S}^2$  which includes a trick to compensate such changes. We showed that by carrying out the NMA in the dihedral angle space (DAS) and by interpreting each DAS normal mode into curved atomic motions,  $\bar{S}^2$  can be calculated reliably for spin pairs separated up to about 10 intervening covalent bonds by a natural formula without any trick. © 1996 American Institute of Physics. [S0021-9606(96)02311-6]

## I. INTRODUCTION

Nuclear magnetic resonance (NMR) has become a powerful experimental method to determine three-dimensional structures of proteins in solution. This power of NMR was recognized when a set of interproton distance constraints from NOE, which are individually rather crude, was realized to have strong enough information to determine three-dimensional structures of proteins.<sup>1,2</sup> Thus, in the current methods of structure determination, distance constraints are first deduced from experimental NMR data, and then a number of three-dimensional structures which satisfy this set of distance constraints are computationally obtained. Even though these methods have great practical importance for determining solution structures of proteins, they are unsatisfactory from the theoretical point of view.

At first these methods are based on an implicit assumption that a single microscopic rigid structure satisfies a set of NMR data. In practice one calculates a number of such single structures, which satisfy a given set of experimental distance constraints equally well. However, it must be clearly recognized that each of such single structures, not an ensemble of them, is assumed to satisfy the experimental data. One tends to feel that conformational variety among an ensemble of such computationally obtained structures reflects the real conformational flexibility. However, there is no firm theoretical basis for this feeling.

The reality is that protein conformation is not rigid, but quite flexible. And an ensemble of fluctuating conformations, not a rigid single conformation, determines experimental

NMR data. An ideal method of protein structure determination from NMR data should be based on this fact. Such an ideal method would enable us to determine not only an average structure but also an extent of conformational fluctuations around it on a theoretically sound basis.

The method employed in the structure determination by x-ray crystallography is better in this respect. There, protein structure is described in terms of an average conformation and fluctuation around it. Both of them, an average and fluctuation, are described by theoretical models with a number of adjustable parameters. By determining values of such adjustable parameters against experimental Bragg intensity data, one gets information about an average and fluctuation around it separately.

An important point in this structure determination is a balance between the number of bits of information in the experimental input data and that in values of output adjustable parameters. If the number of adjustable parameters is too large, the theoretical model may overadjust to the experimental data. An extreme of models with too many adjustable parameters is found in a once-popular treatment in which experimental data (such as Bragg intensities) are added as pseudopotential energy terms to usual potential energy to carry out molecular dynamics (MD) simulation. Because the Newton's equation of motion can find solution for a very wide range of potential energy function, the number of adjustable parameter is essentially infinite in the models of such treatment.<sup>3</sup> This would manifest, for example, in the case of x-ray crystallographic refinement by the fact that *R*-factor can be rendered as small as desired by carrying out MD simulation long enough. This is because molecular motions which perfectly satisfy even error-containing experimental data can be created as a solution of the Newton's equation. In order that results of an analysis faithfully reflect

<sup>a)</sup>To whom correspondence should be addressed: Department of Chemistry, Faculty of Science, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto 606-01, Japan. Tel: 81-75-753-4017; fax: 81-75-711-6083; electronic mail: go@qchem.kuchem.kyoto-u.ac.jp

the reality, the number of bits of information must be balanced between input and output.

Not only the number of adjustable parameters in a theoretical model but of course reality of the model is important. A model of the dynamic structure of proteins employed in the conventional x-ray crystallographic refinement is that each atom in a protein molecule fluctuates independently of each other and isotropically in space. This model, even though operationally easy to handle, is far from real atomic motions observed for example in MD simulations. The model of dynamical structure has been replaced by a realistic one in the recently developed normal mode refinement (NM-REF) method without increasing the number of adjustable parameters.<sup>4–6</sup> In this new model conformational fluctuations are assumed to take place in “an important conformational subspace” spanned by a relatively small number of low-frequency normal mode vectors. This method has been successfully applied for human lysozyme<sup>6</sup> and its mutant.<sup>7</sup> The success of the application of the normal mode concept for protein x-ray crystallography prompted similar efforts to interpret NMR data. Most of such efforts have been made along the line based on the “model-free” approach proposed by Lipari and Szabo.<sup>8,9</sup> They were concerned with NMR relaxation parameters, such as  $T_1$ ,  $T_2$ , and NOE which are given in terms of spectrum density function  $J_{\alpha\beta}(\omega)$  for each pair of spins,  $\alpha$  and  $\beta$ . They gave an explicit expression of  $J_{\alpha\beta}(\omega)$  in terms of four quantities,  $R$ ,  $S^2$ ,  $\tau_M$ , and  $\tau_e$ , where  $R$  is an average distance between the two spins,  $\tau_M$  is the correlation time of the overall tumbling motion of a protein molecule,  $\tau_e$  is the correlation time of internal motion of the molecule, and  $S^2$  is a parameter pertaining directional fluctuation of interspin vector due to the internal motion. This expression has a desirable character for developing a refined method of structure determination from NMR experimental data, because experimental observables are expressed in terms of quantities pertaining to an average structure and an extent of fluctuations around it.

In the case of normal mode refinement method of x-ray protein crystallography, the Debye–Waller factor, a quantity reflecting conformational fluctuation, is expressed in terms of a relatively small number of variances and covariances of low-frequency normal mode variables. In the line of such efforts, we are concerned in this paper, an expression of  $S^2$  in terms of normal mode variables. In fact, Henry and Szabo addressed this problem even before the success of NM-REF. Our approach is based heavily on the method they developed. However, their approach is based on the normal mode analysis carried out in the Cartesian coordinate space (CCS). Motions expressed by individual CCS normal modes are linear in space by definition. Such linear motions bring forth to internuclear distance changes even for a pair of spins of atoms covalently bonded to each other. Henry and Szabo realized that such artificial distance changes cause some troubles in the calculation of  $S^2$ . To solve this problem they invented a clever trick of using a special formula for  $S^2$ , which is designed for not being influenced by the artificial distance changes occurring in the CCS normal modes.

In this paper we are at first concerned with the same

problem Henry and Szabo faced. We will show in this paper that by carrying out the normal mode analysis not in the Cartesian coordinate space but in the dihedral angle space (DAS), and by interpreting linear motions of normal modes in DAS into curved atomic motions in the Cartesian coordinate space, the problem of artificial distance changes can be avoided and no special trick is needed for calculating  $S^2$ .

The developed mathematically natural method is then applied for calculation of  $S^2$  in a small globular protein BPTI, and the results are compared with numerically obtained simulated values. The results of comparison indicates that for pairs of spins separated by up to about ten intervening covalent bonds, values of  $S^2$  can be well reproduced by a formula involving thermal mean-square amplitudes of DAS normal mode variables. This fact promises us a hopeful path to development of a theoretically natural new method of protein structure determination from NMR data. For pairs of spins separated by a larger number of intervening covalent bonds, formula developed in this paper becomes more approximate.

## II. THEORY

### A. The model-free approach

The relaxation due to dipole–dipole interaction between a pair of nuclei can be described by the correlation function<sup>8,10,11</sup>

$$C_m(t) = \left\langle \frac{D_{m0}^{(2)*}[\Omega_{LF}(0)]D_{m0}^{(2)}[\Omega_{LF}(t)]}{r^3(0)r^3(t)} \right\rangle \quad (1)$$

$$= \int \int d\Omega d\Omega' dr dr' \frac{D_{m0}^{(2)*}(\Omega)}{r^3} p_{\text{eq}}(\Omega, r) \\ \times \frac{D_{m0}^{(2)}(\Omega')}{r'^3} p_{\text{eq}}(\Omega' r'; t | \Omega r; 0), \quad (2)$$

where  $D_{mn}^{(2)}(\Omega)$  is the Wigner rotation matrix element, the polar angles  $\Omega_{LF}$  specify the orientation of the unit vector  $\boldsymbol{\mu}_{LF}$  connecting the two nuclei, for example  $\text{C}^{13}$  and H, in the laboratory coordinate system,  $r$  is the distance between the two nuclei,  $p_{\text{eq}}(\Omega, r)$  is the equilibrium probability distribution, and  $p_{\text{eq}}(\Omega' r'; t | \Omega r; 0)$  is the conditional probability that, when the polar angles and the distance are  $\Omega$  and  $r$  at time 0, they are  $\Omega'$  and  $r'$  at time  $t$ . Note that, when  $n=0$ ,  $D_{mn}^{(2)}(\Omega)$  depends only on the polar angles, not on all three of the Euler angles.

The direct and cross relaxation rates are given in terms of the spectrum density  $J_m(\omega)$  defined by<sup>12,13</sup>

$$J_m(\omega) = 2 \int_0^\infty \cos(\omega t) C_m(t) dt. \quad (3)$$

When internal motions are assumed to be decoupled from the overall motion of the molecule and also the overall motion is assumed to be isotropic, the correlation function of Eq. (1) ceases to depend on suffix  $m$ , and (hence by dropping suffix  $m$ ) is given by<sup>10</sup>

$$C(t) = \frac{1}{5} \exp\left(-\frac{t}{\tau_M}\right) C_{\text{in}}(t), \quad (4)$$

$$C_{\text{in}}(t) = \sum_{n=-2}^2 \left\langle \frac{D_{n0}^{(2)*}[\Omega_{\text{MF}}(0)] D_{n0}^{(2)}[\Omega_{\text{MF}}(t)]}{r^3(0)r^3(t)} \right\rangle, \quad (5)$$

where  $\Omega_{\text{MF}}$  are the polar angles in the frame attached to the molecule (M-frame). Equation (5) is a general expression for the internal correlation function (hence suffix in) which characterizes NMR relaxation due to the internal motion in a protein.

In the model-free approach proposed by Lipari and Szabo,<sup>8,9</sup> the time course of the internal correlation function, Eq. (5), is assumed to have the following form:

$$C_{\text{in}}(t) = \tilde{S}^2 + (\langle r^{-6} \rangle - \tilde{S}^2) e^{-t/\tau_e}, \quad (6)$$

where  $\tilde{S}^2$  is the value of  $C_{\text{in}}(t)$  for  $t = \infty$  and is called the generalized order parameter,  $\langle r^{-6} \rangle$  is the value of  $C_{\text{in}}(t)$  for  $t = 0$ , and  $\tau_e$  is the effective correlation time. Thus, in the model-free approach, NMR experimental data are interpreted in terms of the three parameters,  $\tau_M$ ,  $\tau_e$ , and  $\tilde{S}^2$ .

In cases of relaxation between covalently bonded two nuclei, for example,  $\text{C}^\alpha\text{-H}^\alpha$  and  $\text{N-H}^N$ , internuclear distance  $r$  can be assumed to be constant  $r = R$ . In such cases it is convenient to define the following orientational (hence suffix  $o$ ) correlation function:

$$C_{\text{in},o}(t) = R^6 \times C_{\text{in}}(t) = \sum_{n=-2}^2 \langle D_{n0}^{(2)*}[\Omega_{\text{MF}}(0)] \times D_{n0}^{(2)}[\Omega_{\text{MF}}(t)] \rangle. \quad (7)$$

In the model-free approach the orientational correlation function, Eq. (7), is approximated to have the following form:

$$C_{\text{in},o}(t) = S^2 + (1 - S^2) e^{-t/\tau_e}. \quad (8)$$

If (a) overall motion is considerably slower than the internal motion ( $\tau_e \ll \tau_M$ ), (b) the internal motion is in the extreme narrowing limit [ $(\tau_e \omega)^2 \ll 1$ ] and (c)  $\tau_e$  is very small, the spectrum density is given by

$$J(\omega) = \frac{2}{5R^6} \left[ \frac{S^2 \tau_M}{1 + (\omega \tau_M)^2} \right]. \quad (9)$$

Equation (9) includes only  $S^2$  as the effect of internal motions. An expression corresponding to Eq. (9) for cases of variable internuclear distances can be derived from Eq. (6) as

$$J(\omega) = \frac{2}{5} \left[ \frac{\tilde{S}^2 \tau_M}{1 + (\omega \tau_M)^2} \right]. \quad (10)$$

These expressions are usually used in so-called ‘‘back calculation’’ in the structure refinement procedure of biomolecules.<sup>14</sup>

When the expression of Eq. (10) is valid, the dynamic correction factor defined by the ratio ( $\gamma$ ) of the (direct or cross) relaxation rate constants for mobile and rigid proton pairs, which is given by a linear sum of  $J(\omega)$  at a few different  $\omega$ 's, is given by

$$\gamma = \frac{\Gamma_{\text{dyn}}}{\Gamma_{\text{stat}}} = R^6 \tilde{S}^2, \quad (11)$$

where  $\Gamma_{\text{dyn}}$  corresponds to the dynamically averaged rate and  $\Gamma_{\text{stat}}$  is the rate calculated from the static structure (the energy minimized structure in this study). If no internal motion exists, the ratio  $\gamma$  is 1. When the internuclear distance is constant  $r = R$  as in the case of  $\text{C}^\alpha\text{-H}^\alpha$  spin pairs, this ratio reduces to the generalized order parameter  $S^2$ . In this work we treat cases of the fast internal motion for which NMR relaxation can be treated based on the Lipari–Szabo model.

## B. Expressions for generalized order parameter

Henry and Szabo developed a formalism for calculation of the order parameter based on NMA.<sup>15</sup> We start our treatment from their formalism. Henry and Szabo used a  $3 \times 3$  Cartesian tensor that had been previously introduced by Sykora *et al.*,<sup>16</sup>

$$\Phi_{oij} = \frac{r_i r_j}{r^2}, \quad (12)$$

where  $r_i$  (or  $r_j$ ) is M-frame  $x$ ,  $y$ , or  $z$  component of the vector  $\mathbf{r}$  which connects the two nuclei. Using this tensor the internal correlation function of Eq. (7) can be transformed to<sup>15</sup>

$$C_{\text{in},o}(t) = \frac{3}{2} \text{tr}(\Phi_o(t)\Phi_o(0)) - \frac{1}{2} \text{tr}(\Phi_o(t)) \cdot \text{tr}(\Phi_o(0)). \quad (13)$$

The generalized order parameter  $S^2$  is the value of  $C_{\text{in},o}(t)$  at  $t = \infty$ ,

$$S^2 = C_{\text{in},o}(\infty) = \frac{3}{2} \text{tr}(\langle \Phi_o \rangle^2) - \frac{1}{2} (\text{tr} \langle \Phi_o \rangle)^2. \quad (14)$$

Similarly the generalized order parameter  $\tilde{S}^2$  is expressed by

$$\tilde{S}^2 = C_{\text{in}}(\infty) = \frac{3}{2} \text{tr}(\langle \Phi \rangle^2) - \frac{1}{2} (\text{tr} \langle \Phi \rangle)^2, \quad (15)$$

where

$$\Phi_{ij} = \frac{1}{r^3} \Phi_{oij} = \frac{1}{r^3} \frac{r_i r_j}{r^2}. \quad (16)$$

Note that the quantities  $(1/r^3)$  and  $(r_i r_j / r^2)$  in Eq. (16) are influenced by distance fluctuation and orientational motion, respectively.

Following Sykora *et al.*,<sup>16</sup> the vector  $\mathbf{r}$  at one instantaneous time is written by

$$\mathbf{r} = \mathbf{R} + \Delta, \quad (17)$$

where  $\mathbf{R}$  is the mean position of vector  $\mathbf{r}$  and  $\Delta$  is an instantaneous displacement vector. Then Eq. (12) becomes

$$\begin{aligned} \Phi_{oij} &= \frac{(R_i + \Delta_i)(R_j + \Delta_j)}{(\mathbf{R} + \Delta)(\mathbf{R} + \Delta)} \\ &= \frac{[\zeta_j + (\Delta_j/R)][\zeta_j + (\Delta_j/R)]}{1 + 2(\sum_k \zeta_k \Delta_k)/R + (\sum_k \Delta_k \Delta_k)/R^2}, \end{aligned} \quad (18)$$

where  $i$  and  $j$  indicate  $x$ ,  $y$ , or  $z$ ,  $R$  is the norm of the vector  $\mathbf{R}$ , and  $\zeta_i = R_i/R$ . The right-hand side of Eq. (18) can be expanded into a Taylor series up to second order in  $\Delta/R$  when  $\Delta_i/R \ll 1$  is assumed. Then  $\langle \Phi_{oij} \rangle$  becomes

$$\begin{aligned} \langle \Phi_{oij} \rangle = & \zeta_i \zeta_j + \frac{1}{R} \left( \zeta_i \langle \Delta_j \rangle + \zeta_j \langle \Delta_i \rangle - 2 \zeta_i \zeta_j \sum_k \zeta_k \langle \Delta_k \rangle \right) \\ & + \frac{1}{R^2} \left[ \langle \Delta_i \Delta_j \rangle - \zeta_i \zeta_j \sum_k \langle \Delta_k \Delta_k \rangle \right. \\ & + 4 \zeta_i \zeta_j \sum_{k,l} \zeta_k \zeta_l \langle \Delta_k \Delta_l \rangle - 2 \left( \zeta_i \sum_k \zeta_k \langle \Delta_j \Delta_k \rangle \right. \\ & \left. \left. + \zeta_j \sum_k \zeta_k \langle \Delta_i \Delta_k \rangle \right) \right]. \end{aligned} \quad (19)$$

The right-hand side of Eq. (16) can also be expanded by using

$$r = R + \frac{(\mathbf{R} \cdot \Delta)}{R} + \frac{1}{2} \frac{(\Delta \cdot \Delta)}{R} - \frac{1}{2} \frac{(\mathbf{R} \cdot \Delta)^2}{R^3} + \dots \quad (20)$$

to have the second order expression for the ensemble average of the tensor  $\Phi_{ij}$ ,

$$\begin{aligned} \langle \Phi_{s.o.ij} \rangle = & \frac{1}{R^3} \left\{ \zeta_i \zeta_j + \frac{1}{R} \left( \zeta_i \langle \Delta_j \rangle + \zeta_j \langle \Delta_i \rangle \right. \right. \\ & \left. \left. - 5 \zeta_i \zeta_j \sum_k \zeta_k \langle \Delta_k \rangle \right) + \frac{1}{R^2} \left[ \langle \Delta_i \Delta_j \rangle \right. \right. \\ & \left. \left. - \frac{5}{2} \zeta_i \zeta_j \sum_k \langle \Delta_k^2 \rangle + \frac{35}{2} \zeta_i \zeta_j \sum_{k,l} \zeta_k \zeta_l \langle \Delta_k \Delta_l \rangle \right. \right. \\ & \left. \left. - 5 \left( \zeta_i \sum_k \zeta_k \langle \Delta_j \Delta_k \rangle + \zeta_j \sum_k \zeta_k \langle \Delta_i \Delta_k \rangle \right) \right] \right\}. \end{aligned} \quad (21)$$

Up to this point derivations of equations are conceptually simple. Complication starts at this point. Henry and Szabo noted that, when values of  $\langle \Delta_i \rangle = 0$ , which are to be obtained by Cartesian coordinate NMA, are used, Eq. (21) leads to unphysical results. Their reasoning is essentially based on recognition that Eq. (20), by the use of which Eq. (21) was derived, leads to an unphysical result  $\langle r \rangle \neq R$  when applied for a case of fixed internuclear distance of  $r = R$ . This strange phenomenon that the mathematically sound expression of Eq. (20) yields an unphysical result reveals that the basic assumption underlying the Cartesian coordinate NMA is incompatible with the second order expansion of Eq. (20). In fact, in the normal mode analysis in the Cartesian coordinate space, all atomic motions in each normal mode are linear in space, thus inevitably involving bond length changes even for covalently bonded atom pairs.

In order to avoid the unphysical results noted by Henry and Szabo, it appears reasonable to carry out NMA in the internal coordinate space. Due to the spirit of NMA, each normal mode is a linear motion in the internal coordinate space, but, when translated into motions in the Cartesian

coordinate space, it should be a curved one.<sup>17</sup> In this paper we will show that, by calculating such curved motions accurately up to the second order, mathematically consistent and physically sound results can be obtained.

However, Henry and Szabo tried still to use the result of the Cartesian coordinate NMA. In order to avoid the unphysical results they proposed (instead of improving the treatment of NMA) to use instead of Eq. (20) the first order expansion of  $r$ ,

$$r = R + \frac{(\mathbf{R} \cdot \Delta)}{R} \quad (22)$$

in the factor of  $(1/r^3)$  in Eq. (16) to obtain a second order expansion expression of  $\Phi_{ij}$ . Thus, the result is

$$\begin{aligned} \langle \Phi_{f.o.ij} \rangle = & \frac{1}{R^3} \left\{ \zeta_i \zeta_j + \frac{1}{R} \left( \zeta_i \langle \Delta_j \rangle + \zeta_j \langle \Delta_i \rangle \right. \right. \\ & \left. \left. - 5 \zeta_i \zeta_j \sum_k \zeta_k \langle \Delta_k \rangle \right) + \frac{1}{R^2} \left[ \langle \Delta_i \Delta_j \rangle \right. \right. \\ & \left. \left. - \zeta_i \zeta_j \sum_k \langle \Delta_k^2 \rangle + 16 \zeta_i \zeta_j \sum_{k,l} \zeta_k \zeta_l \langle \Delta_k \Delta_l \rangle \right. \right. \\ & \left. \left. - 5 \left( \zeta_i \sum_k \zeta_k \langle \Delta_j \Delta_k \rangle + \zeta_j \sum_k \zeta_k \langle \Delta_i \Delta_k \rangle \right) \right] \right\}, \end{aligned} \quad (23)$$

where the subscript f.o. emphasizes the use of first order expansion of  $r$ . Use of a mathematical trick in derivation of Eq. (23) must be clearly noted.

Equations (21) and (23) are now at our hand to calculate the order parameters  $\tilde{S}^2$  by NMA.

### C. Normal mode analysis in the dihedral angle space

For the reason discussed above, we will now consider NMA in internal coordinate space (ICS). In this paper we will carry out NMA in a subspace of ICS, i.e., the dihedral angle space (DAS) in which bond lengths and bond angles are treated as kept fixed.

In simulation in DAS, it is necessary to remove six degrees of freedom for translation and rotation of the molecule. This procedure is attained by requiring that atomic coordinates  $\mathbf{y}_a$  always give the minimum of the sum of mass-weighted square deviations,

$$E = \sum_a m_a (\mathbf{y}_a^0 - \mathbf{y}_a)^2 \quad (24)$$

from position vectors  $(\mathbf{y}_a^0)$  of a reference conformation.<sup>18-20</sup> Here  $m_a$  is mass of atom  $a$ , and the summation is taken over all atoms in the molecule.

A second order formula,

$$\Delta \mathbf{y}_a = \sum_{\alpha} \mathbf{k}_{a\alpha} \Delta \theta_{\alpha} + \frac{1}{2} \sum_{\alpha, \beta} \mathbf{L}_{a\alpha\beta} \Delta \theta_{\alpha} \Delta \theta_{\beta} \quad (25)$$

for conversion from a linear motion in DAS to a curved motion in Cartesian coordinate space (CCS) has been de-

rived recently by requiring the condition of Eq. (24).<sup>17</sup> Summation is taken over variable dihedral angles in the molecule. Explicit analytic expressions for the coefficients are given in Ref. 17. If the second order terms in Eq. (25) are neglected, then we obtain the formula for linear conversion,

$$\langle \Delta y_{ai} \rangle = 0 \quad (26)$$

and

$$\langle \Delta y_{ai} \Delta y_{bj} \rangle = \sum_{\alpha, \beta} k_{ai\alpha} k_{bj\beta} \langle \Delta \theta_\alpha \Delta \theta_\beta \rangle. \quad (27)$$

Here both  $a$  and  $b$  refer to atoms in the molecule and  $i, j$  are  $x, y,$  or  $z$  component. Use of these formula will yield the unphysical results noted by Henry and Szabo, which will be shown later. When second order terms are retained, we have

$$\langle \Delta y_{ai} \rangle = \frac{1}{2} \sum_{\alpha, \beta} L_{ai\alpha\beta} \langle \Delta \theta_\alpha \Delta \theta_\beta \rangle, \quad (28)$$

$$\langle \Delta y_{ai} \Delta y_{bj} \rangle = \sum_{\alpha, \beta} k_{ai\alpha} k_{bj\beta} \langle \Delta \theta_\alpha \Delta \theta_\beta \rangle + \frac{1}{4} \sum_{\alpha, \beta, \gamma, \delta} L_{ai\alpha\beta} L_{bj\gamma\delta} \langle \Delta \theta_\alpha \Delta \theta_\beta \Delta \theta_\gamma \Delta \theta_\delta \rangle. \quad (29)$$

Equation (28) shows shifts of mean position due to curved motion. A precise formula for Eq. (29) should also have fourth order terms involving  $k_{ai\alpha} \Delta \theta_\alpha$  and (not yet obtained) third order terms in the expansion of Eq. (25). Previous numerical calculation indicated that such fourth order terms can be neglected numerically.

### III. METHODS

The normal mode analysis (NMA) was carried out for a globular protein BPTI following the standard procedure.<sup>21</sup> The coordinate set of atoms in BPTI was taken from 4PTI (Ref. 22) in Protein Data Bank (PDB).<sup>23</sup> After regularization of the coordinates, conformational energy was minimized with the empirical parameter set ECEPP.<sup>24</sup> Then, NMA was carried out and quantities necessary to calculate NMR order parameters, i.e., eigenvalues, eigenvectors,  $K$ -matrix, and  $L$ -matrix, were obtained. We also have carried out a Monte Carlo conformation sampling in DAS by assuming a harmonic potential energy surface around the minima for which NMA is carried out. Order parameters  $S^2$  and  $\tilde{S}^2$  are calculated directly, i.e., without using the Taylor expansion by using sampled 450 000 conformations of BPTI. In both NMA and Monte Carlo simulation, thermal averages were taken at  $T=300$  K.

### IV. RESULTS

Order parameters  $S^2$  and  $\tilde{S}^2$  are calculated for various cases by using different approximations. It is convenient to discuss separately the two cases of constant internuclear distances and variable internuclear distances. When internuclear distance is constant  $r=R$ , we should physically expect the following relation:

$$S^2 = \tilde{S}^2 \times R^6. \quad (30)$$

TABLE I. Plots in Figs. 1–6.

Conversion from DAS to CCS	$S^2$ Eq. (19)	$S^2$ Eq. (12)	$\tilde{S}_{f.o.}^2 \times R^6$ Eq. (23)	$\tilde{S}_{s.o.}^2 \times R^6$ Eq. (21)	$\tilde{S}^2 \times R^6$ Eq. (16)
Linear Eqs. (26) and (27)	a		b	c	
Curved (2nd order) Eqs. (28) and (29)	d		e	f	
Precise numerical Calculation (MC result)		g			h

Figure 1: a,b,c    Figure 4: b,f,h  
Figure 2: d,e,f    Figure 5: b,f,h  
Figure 3: a,d,g    Figure 6: b,f

Even when internuclear distance is variable, the quantity  $\tilde{S}^2 \times R^6$  has the meaning of the dynamic correction factor of Eq. (11), where  $R$  is the internuclear distance in an assumed rigid conformation of the molecule. For this reason, when we present results of  $\tilde{S}^2$ , we will give  $\tilde{S}^2 \times R^6$ , where  $R$  is either a constant value of internuclear distance or its value at minimum energy conformation (MEC) for variable internuclear distance. Various approximations examined are summarized in Table I, and will be described in detail below.

#### A. C $^\alpha$ –H $^\alpha$ spin pairs: Cases of constant nuclear distance

Let us start with examination of the discussion and the proposal of Henry and Szabo.<sup>15</sup> They proposed a method of calculating  $S^2$  or  $\tilde{S}^2 \times R^6$  by the Cartesian coordinate NMA. The Cartesian coordinate NMA corresponds to using the linear relations of Eqs. (26) and (27). For the calculation of  $\tilde{S}^2 \times R^6$ , they proposed to use a special expression of Eq. (23) instead of more natural Eq. (21). To examine this point, we plotted calculated results of  $S^2$ ,  $\tilde{S}_{f.o.}^2 \times R^6$  and  $\tilde{S}_{s.o.}^2 \times R^6$  for C $^\alpha$ –H $^\alpha$  spin pairs in Fig. 1. We see that Eq. (30) does not hold, when the natural expression of Eq. (21) is used to calculate  $\tilde{S}^2$  (the unphysical result), but does hold, when the special formula of Eq. (23) is used instead. This result may appear to indicate that the use of Eq. (23) cures the problem

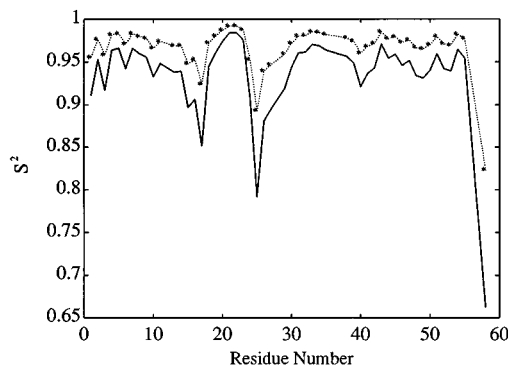


FIG. 1.  $S^2$  and  $\tilde{S}^2 \times R^6$  for C $^\alpha$ –H $^\alpha$  spin pairs in nonglycyl residues of BPTI calculated by using the linear relations of Eqs. (26) and (27). Orientational order parameter  $S^2$  (\*),  $\tilde{S}_{f.o.}^2 \times R^6$  (dotted line), and  $\tilde{S}_{s.o.}^2 \times R^6$  (solid line).

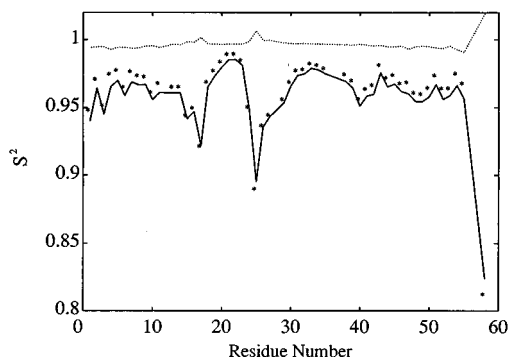


FIG. 2.  $S^2$  and  $\tilde{S}^2 \times R^6$  for  $C^\alpha-H^\alpha$  spin pairs in nonglycyl residues of BPTI calculated by using the curved relations of Eqs. (28) and (29). Orientational order parameter  $S^2$  (\*),  $\tilde{S}_{f.o.}^2 \times R^6$  (dotted line), and  $\tilde{S}_{s.o.}^2 \times R^6$  (solid line).

existing in the use of Cartesian coordinate NMA in Eq. (21). In fact, we can show the relation of Eq. (30) analytically for a case of constant  $r=R$ . In such a case we have  $\mathbf{R} \cdot \Delta = 0$ . When this relation is used in Eq. (16), we can derive  $\Phi_{oij} = R^3 \Phi_{f.o.ij}$ , hence  $S^2 = \tilde{S}_{f.o.}^2 \times R^6$ . This agreement does not of course prove validity of Eq. (23). As indicated in Sec. II, more essential problem exists in the use of the linear relations of Eqs. (26) and (27).

Next, we calculated  $S^2$  and  $\tilde{S}^2 \times R^6$  by using the curved relations of Eqs. (28) and (29). The results shown in Fig. 2 indicate that the relation of Eq. (30) is recovered when the natural relation of Eq. (21) is used together with the curved relations of Eqs. (28) and (29). This means that the “unphysical result” pointed out by Henry and Szabo does not have its root in Eq. (21), but in the use of Cartesian coordinate NMA result, as they have already noted in their paper.

We also checked that the second order expansion in  $\Delta/R$  in Eq. (19) is valid for these cases of  $C^\alpha-H^\alpha$  spin pairs. Figure 3 shows that the calculation of orientational order parameter  $S^2$  via Eq. (19) by NMA is sufficient to reproduce the results from the Monte Carlo calculation. As long as  $S^2$  is concerned, simple use of the linear relations of Eqs. (26) and (27) works for constant distance cases in a practical sense.

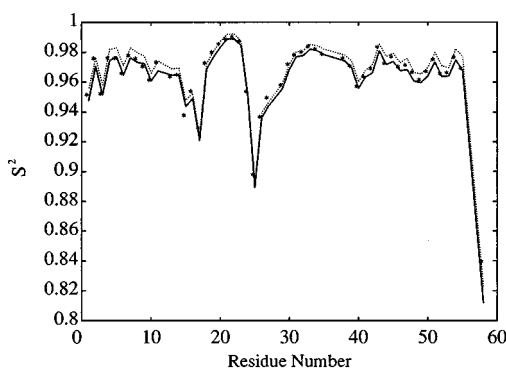


FIG. 3.  $S^2$  values for  $C^\alpha-H^\alpha$  spin pairs in nonglycyl residues of BPTI calculated by Eq. (19) and the linear relations, Eqs. (26) and (27) (dotted line), calculated by Eq. (19) and the curved relations, Eqs. (28) and (29) (solid line), and calculated directly by Eq. (12) and Monte Carlo simulation (\*).

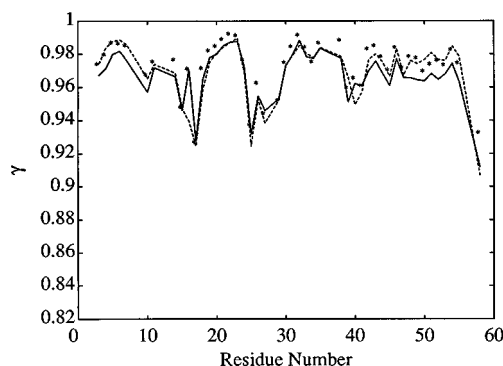


FIG. 4.  $\gamma = \tilde{S}^2 \times R^6$  for  $H^N(i)-H^\alpha(i)$  spin pairs calculated by Eq. (23) and the linear relations of Eqs. (26) and (27) as proposed by Henry and Szabo (broken line), calculated by Eq. (21) and the curved relations of Eqs. (28) and (29) (solid line), and calculated directly by Eq. (16) and Monte Carlo simulation (\*).

### B. H–H spin pairs: Cases of variable nuclear distance

In this section we consider two approaches to calculate  $\gamma = \tilde{S}^2 \times R^6$ , the linear relations of Eqs. (26) and (27) with Eq. (23) as proposed by Henry and Szabo and the curved relations of Eqs. (28) and (29) with Eq. (21). At first we consider the cases of  $H^N(i)-H^\alpha(i)$  and  $H^\alpha(i)-H^N(i+1)$  ( $i$ , residue number). Distances between these proton pairs are variable, because there are one and two variable dihedral angles in between, respectively. As shown in Fig. 4, both the linear relations of Eqs. (26) and (27) with Eq. (23) and the curved relations of Eqs. (28) and (29) with Eq. (21) can reproduce well the values of Monte Carlo simulation. However, in Fig. 5, we see non-negligible discrepancies between results from the linear relations and the numerical calculation, but the agreement between those from the curved relations and the numerical calculation is good. The difference between Fig. 4 and Fig. 5 may be attributed to more complex motion of the interspin vector in Fig. 5 due to one more variable dihedral angle. Generalizing this tendency, we expect that the values of  $\gamma$  calculated by the method of Henry and Szabo becomes worse as the number of intervening variable dihedral angles

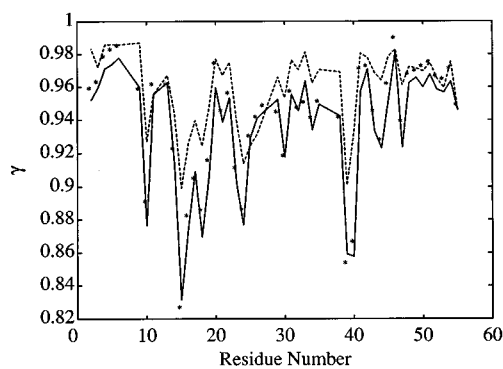


FIG. 5.  $\gamma = \tilde{S}^2 \times R^6$  for  $H^\alpha(i)-H^N(i+1)$  spin pairs calculated by Eq. (23) and the linear relations of Eqs. (26) and (27) as proposed by Henry and Szabo (broken line), calculated by Eq. (21) and the curved relations of Eqs. (28) and (29) (solid line), and calculated directly by Eq. (16) and Monte Carlo simulation (\*).

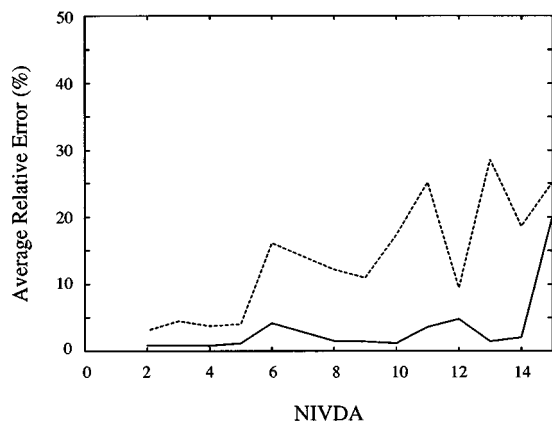


FIG. 6. Average error of the values of  $\gamma$  relative to the precise Monte Carlo numerical values. For  $\gamma$  calculated by Eq. (23) and the linear relations of Eqs. (26) and (27) as proposed by Henry and Szabo (broken line), calculated by Eq. (21) and the curved relations of Eqs. (28) and (29) (solid line). Results are shown for each case of NIVDA, the number of intervening variable dihedral angles (from 2 to 15).

(NIVDA) increases. To see also how our method of using the curved relations of Eqs. (28) and (29) together with Eq. (21) works for various spin pairs, values of  $\gamma$  are calculated for 304 proton-proton spin pairs which belong to different residues and whose internuclear distance is less than 3 Å in the minimum energy conformation. Figure 6 represents average errors (%) of the values of  $\gamma$  relative to the precise numerical values (MC results) in each NIVDA (from 2 to 15). We see that, for  $\text{NIVDA} \leq 14$ , our method works with average errors less than several percent, while the method of Henry and Szabo becomes increasingly worse as NIVDA becomes large. Figure 7 shows distribution of values of relative errors

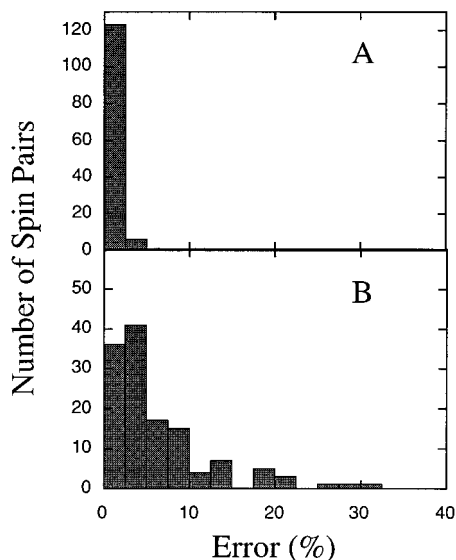


FIG. 7. Distribution of values of relative errors in 131 pairs in the range of proton-proton  $2 \leq \text{NIVDA} \leq 10$ . The results calculated by Eq. (21) and the curved relations of Eqs. (28) and (29) (a), and those calculated by Eq. (23) and the linear relations of Eqs. (26) and (27) as proposed by Henry and Szabo (b).

TABLE II. Average errors (%) of the value of  $\gamma$  relative to the precise numerical values for various ranges of NIVDA.<sup>a</sup>

NIVDA <sup>a</sup>	Number of spin pairs	Average errors (%)	
		<i>b</i> <sup>b</sup>	<i>f</i> <sup>b</sup>
2–10	131	6.31	1.10
11–50	70	42.3	13.7
51–100	71	43.2	15.8
100–	32	160	136

<sup>a</sup>The number of intervening variable dihedral angles.

<sup>b</sup>The same notation as in Table I.

in 131 proton-proton pairs whose NIVDA is between 2 and 10. We see that, for these cases of NIVDA, our method gives errors mostly less than a few percent. However, also our method gives large average errors for large NIVDA ( $\geq 15$ ) as shown in Table II. This table summarizes values of average errors for various ranges of NIVDA.

## V. CONCLUSIONS AND DISCUSSION

In this paper we studied validity of various formulas to calculate the generalized order parameters  $S^2$  (for cases of fixed internuclear distances) or  $\tilde{S}^2 \times R^6$  (for cases of variable internuclear distances) in terms of thermal mean-square fluctuations of normal mode variables.

Two types of normal mode analysis are considered, one carried out in the Cartesian coordinate space (CCS) and the other in the dihedral angle space (DAS). Atomic motions in each CCS normal mode is linear in space. This causes changes of internuclear distances even between spins of covalently bonded atoms. Hence, CCS normal modes give us a rather approximate picture of atomic motions in a protein. Each DAS normal mode is linear in the dihedral angle space. But, such linear motion can cause curved atomic motion in the Cartesian coordinate space. We developed two types of formulas to interpret linear normal mode motion in DAS to atomic motions in CCS, one which is linear in CCS [Eqs. (26) and (27)] and the other which is curved [Eqs. (28) and (29)]. If we use the formula for linear interpretation, the DAS normal mode analysis is essentially the same as the CCS normal mode analysis.

Henry and Szabo have previously shown that for calculation of  $\tilde{S}^2 \times R^6$ , a special formula [Eq. (23)] must be used to obtain physically reasonable results. This formula was devised by using a trick to compensate the artificial distance changes caused by linear atomic motions in CCS normal modes. We have shown that, when we use DAS normal modes and interpretation of them into curved atomic motion in the Cartesian coordinate space, Eqs. (28) and (29), then a mathematically natural formula [Eq. (21)] gives physically reasonable results. We have further shown that this formula together with the curved DAS normal modes can be used to reliably calculate  $\tilde{S}^2 \times R^6$  for pairs of spins of atoms separated by up to 10 intervening covalent bonds, while the method proposed by Henry and Szabo gives numerically reasonable results only for cases of spin pairs separated by no or one intervening covalent bond.

Summarizing the above, we have attained our aim of developing a theoretically sound method of calculating the NMR order parameter  $\bar{S}^2 \times R^6$  by the normal mode analysis. This method enables us to calculate the order parameter reliably for spin pairs separated by up to about 10 intervening covalent bonds.

The above success has been attained only by carrying out the normal mode analysis in DAS, i.e., by a treatment in which bond lengths and bond angles are fixed. It should be stressed that, even though normal mode analyses in CCS and DAS are equivalent to each other within the linear range,<sup>25,26</sup> the DAS normal modes have more pieces of information about curved motions. This fact puts us in a paradoxical situation, because Brüschweiler<sup>27</sup> has shown that such high frequency motions involving changes of bond lengths and bond angles that must be treated quantum mechanically have significant effects on the order parameters. This situation suggests that we should do (1) the DAS normal mode analysis for low frequency modes in a protein and (2) the CCS normal mode analysis for high frequency modes. Because high frequency modes are localized in a relatively small number of atoms, we may be able to develop a formulation in which CCS normal mode analysis can be carried out not for a whole protein molecule but for small local fragments of the molecule.

#### ACKNOWLEDGMENTS

This work was started when P. Koehl stayed at Kyoto University as a short-term HFSP fellow in the summer, 1991. Computation has been done at Computer Centers of Kyoto University and of Institute for Molecular Science. This work has been supported by grants from Ministry of Education,

Japan, from the Mitsubishi Foundation and from International Human Frontier Science Program Organization.

- <sup>1</sup>W. Braun, C. Boesch, L. R. Brown, N. Go, and K. Wüthrich, *Biochem. Biophys. Acta* **667**, 377 (1981).
- <sup>2</sup>W. Braun and N. Go, *J. Mol. Biol.* **186**, 611 (1985).
- <sup>3</sup>A. T. Brünger and M. Nilges, *Q. Rev. Biophys.* **26**, 49 (1993).
- <sup>4</sup>A. Kidera and N. Go, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 3718 (1990).
- <sup>5</sup>A. Kidera and N. Go, *J. Mol. Biol.* **225**, 457 (1992).
- <sup>6</sup>A. Kidera, K. Inaka, M. Matsushima, and N. Go, *J. Mol. Biol.* **225**, 477 (1992).
- <sup>7</sup>A. Kidera, K. Inaka, M. Matsushima, and N. Go, *Protein Sci.* **3**, 92 (1994).
- <sup>8</sup>G. Lipari and A. Szabo, *J. Am. Chem. Soc.* **104**, 4546 (1982).
- <sup>9</sup>G. Lipari and A. Szabo, *J. Am. Chem. Soc.* **104**, 4559 (1982).
- <sup>10</sup>D. Wallach, *J. Chem. Phys.* **47**, 5258 (1967).
- <sup>11</sup>R. J. Wittebort and A. Szabo, *J. Chem. Phys.* **69**, 1722 (1978).
- <sup>12</sup>A. Abragam, *The Principles of Nuclear Magnetism* (Clarendon, Oxford, 1961).
- <sup>13</sup>M. Goldman, *Quantum Description of High-Resolution NMR in Liquids* (Clarendon, Oxford, 1988).
- <sup>14</sup>T. M. G. Koning, R. Boelens, G. A. van der Marel, J. H. van Boom, and R. Kaptein, *Biochem.* **30**, 3787 (1991).
- <sup>15</sup>E. R. Henry and A. Szabo, *J. Chem. Phys.* **82**, 4753 (1985).
- <sup>16</sup>S. Sykora, J. Vogt, H. Bosiger, and P. Diehl, *J. Magn. Reson.* **36**, 53 (1979).
- <sup>17</sup>S. Sunada and N. Go, *J. Comp. Chem.* **16**, 328 (1995).
- <sup>18</sup>C. Eckart, *Phys. Rev.* **47**, 552 (1935).
- <sup>19</sup>A. D. McLachlan, *J. Mol. Biol.* **128**, 49 (1979).
- <sup>20</sup>T. Noguti and N. Go, *J. Phys. Soc. Jpn.* **52**, 3283 (1983).
- <sup>21</sup>T. Nishikawa and N. Go, *Proteins* **2**, 308 (1987).
- <sup>22</sup>M. Marquart, J. Walter, J. Deisenhofer, W. Bode, and R. Huber, *Acta Crystallogr. Sect. B* **39**, 480 (1983).
- <sup>23</sup>F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, *J. Mol. Biol.* **112**, 535 (1977).
- <sup>24</sup>G. Némethy, M. S. Pottle, and H. A. Scheraga, *J. Phys. Chem.* **87**, 1883 (1983).
- <sup>25</sup>A. Kitao and N. Go, *J. Comp. Chem.* **12**, 359 (1991).
- <sup>26</sup>A. Kitao, S. Hayward, and N. Go, *Biophys. Chem.* **52**, 107 (1994).
- <sup>27</sup>R. Brüschweiler, *J. Am. Chem. Soc.* **114**, 5341 (1992).