Realization of the structural fluctuation of biomolecules in solution: Generalized Langevin mode analysis

Masatake Sugita 1,2 | Fumio Hirata 3

1Department of Bioinformatics, College of Life Sciences, Ritsumeikan University, Kusatsu, Shiga, Japan
2Department of Computer Science, School of Computing, Tokyo Institute of Technology, Tokyo, Japan
3Department of Theoretical and Computational Molecular Science, Institute for Molecular Science, Okazaki, Japan

Corespondence
Fumio Hirata, Department of Theoretical and Computational Molecular Science, Institute for Molecular Science, Okazaki 444-8585, Japan.
Email: hirata@ims.ac.jp

Abstract
A new theoretical method, referred to as Generalized Langevin Mode Analysis (GLMA), is proposed to analyze the mode of structural fluctuations of a biomolecule in solution. The method combines the two theories in the statistical mechanics, or the Generalized Langevin theory and the RISM/3D-RISM theory, to calculate the second derivative, or the Hessian matrix, of the free energy surface of a biomolecule in aqueous solution, which consists of the intramolecular interaction among atoms in the biomolecule and the solvation free energy.

The method is applied to calculate the wave-number spectrum of an alanine dipeptide in water for which the optical heterodyne-detected Raman-induced spectroscopy (RIKES) spectrum is available to compare with. The theoretical analysis reproduced the main features of the experimental spectrum with respect to the peak positions of the four bands around ~90 cm⁻¹, ~240 cm⁻¹, ~370 cm⁻¹, and 400 cm⁻¹, observed in the experimental spectrum, in spite that the physics involved in the two spectrum was not exactly the same: the experimental spectrum includes the contributions from the dipeptide and the water molecules interacting with the solute, while the theoretical one is just concerned with the solute molecule, influenced by solvation.

Two major discrepancies between the theoretical and experimental spectra, one in the band intensity around ~100 cm⁻¹, and the other in the peak positions around ~370 cm⁻¹, are discussed in terms of the fluctuation mode of water molecules interacting with the dipeptide, which is not taken explicitly into account in the theoretical analysis.

KEYWORDS
alanine dipeptide, aqueous solution, generalized Langevin mode, Hessian matrix, RIKES, RISM/3D-RISM theory, solvation free energy, structural fluctuation

1 INTRODUCTION

Function of protein is closely related to its structural fluctuation in the solution phase. For instance, a ligand molecule cannot access to the binding pocket without fluctuation of the protein since the binding mode between the protein and its ligand is mostly very tight. It is also known that the change of the fluctuation originated from a ligand binding at the distant position from the active site affect to the enzymatic activity. It indicates that the catalytic process of the enzyme is closely related to the inherent fluctuation of the protein. Since the fluctuation related to function of a protein is regarded as a long-period collective-mode, it is essential to elucidate the fluctuations with a long time scale at the atomic resolution for understanding the functional expression mechanism of the protein.

The dynamic behavior of proteins is experimentally analyzed by the spectroscopic methods such as NMR, transient grating spectroscopy, optical Kerr effect (OKE) spectroscopy. These methods can reveal the characteristics of the fluctuation of the
proteins. However, the resolution of the methods is not fine enough to understand the fluctuation in atomic detail. Therefore, it is a common maneuver to utilize a theoretical method such as molecular dynamics simulation (MD) in order to analyze fluctuations of the molecule in atomic resolution.\textsuperscript{[13–23]}

The most popular and effective method for analyzing fluctuations of a biomolecule computationally is the quasi-harmonic analysis, or the principal component analysis, of the variance–covariance matrix of the trajectory obtained from the MD simulation. However, there are certain limitations even in the method based on MD. For instance, in order to analyze detailed aspects of fluctuations, sampling of the trajectory should be at least several times longer than the relaxation time of the fluctuation. However, a usual MD for proteins is able to sample several tens of microseconds under infinite dilution conditions. It is obvious that a MD simulation cannot catch such fluctuations in detail, as its time constant exceeding tens of microseconds. Furthermore, sampling of the configuration space becomes increasingly difficult when the solvent consists of several chemical components in addition to water, such as electrolytes and denaturants.

There has been another method practiced to calculate the free energy surface and the spectral density of a molecule in solution.\textsuperscript{[22,23]} The method first calculates the density of states (DOS) from the velocity auto-correlation function of atoms in molecules, based on the Wiener–Khintchine theorem.\textsuperscript{[24]} Then, DOS is transformed into the vibrational partition-function as well as the free energy by means of the standard statistical-mechanics procedure. In a glance, the method seems to be promising for the calculation of the free energy surface. However, the method shares the same problem with the other methods based on the molecular-dynamics simulation, which is the sampling of the phase space concerning solvent molecules.

Another popular method for analyzing the structural fluctuation of biomolecules is the Normal Mode Analysis (NMA).\textsuperscript{[25,26]} The method gives the second derivative of the potential energy of a biomolecule, or Hessian, which represents the structural fluctuation of a molecule in terms of the frequency spectrum. The analysis provides the wide range of frequency spectrum of a molecule, ranging from the bond stretching-mode to that corresponding to the collective mode. Therefore, it is free from the concern about the length of trajectory of a MD simulation. However, this analysis has a fatal drawback. The analysis is performed in the environment of vacuum. So, the method in principle evaluates fluctuations in vacuum. The structural fluctuation, and dynamics of a protein in solution phase is completely different from those in vacuum. The method will give an erroneous picture for the structural fluctuation of protein.

An insightful work which may solve the problem has been published by Kim and Hirata in 2013.\textsuperscript{[27]} They have derived a generalized Langevin equation which represents the dynamics of a solute molecule conjugated with the density fluctuation of solvent. The most important point of the equation is that the shape of the equation is essentially a Langevin equation of a harmonic oscillator immersed in solvent or water. So, it has a general appearance similar to the Langevin equation treated by Wang and Uhlenbeck in 1945, and by Lamm and Szabo later in the context of Langevin mode analysis (LMA).\textsuperscript{[28,29]} However, there are two essential differences between the new and old theories. The first of those is concerned with solvent in which the harmonic oscillator is immersed. The solvent in the old theories is a sort of structureless “gel”, while that in the new theory is an assembly of molecules. The important terms of the equation, such as the friction and the random force, have a microscopic expression that can be realized with the aid of another generalized Langevin equation of fluid under a force exerted by the protein. Another difference from the old theories is in its restoring force proportional to the displacement of atoms from their equilibrium position. In the old theories, the proportional constant, or the Hessian, is the second derivative of the potential energy with respect to the atomic coordinate, while it is the second derivative of the free energy in the new theory, that includes the solvent induced force. The Hessian concerning the solvent induced force may be obtained from the RISM/3D-RISM theory as the second derivative of the solvation free energy. So, if one drops the memory and random force terms from the equation, it is formally identical to that of a harmonic oscillator employed in the normal mode analysis, but with the Hessian that includes solvent-induced force. The isomorphism between the two theories gives us a new possibility to analyze the structural fluctuation of a biomolecule in solution without being concerned with the notorious problem inherent to MD, or the sampling of configuration space.

Now, the problem of analyzing the structural fluctuation of a biomolecule in solution is reduced to a calculation of the Hessian including solvent-induced force. According to the Kim-Hirata theory, the Hessian can be obtained in principle from the RISM/3D-RISM theory as the second derivative of the free energy with respect to the atomic coordinates of the molecule. However, it is a non-trivial numerical problem to calculate the Hessian matrix, the number of elements of which amounts to a square of the number of atoms consisting the biomolecule: for a typical protein including $\sim 10^5$ atoms, the number of matrix elements becomes $\sim 10^{10}$.

In the present study, we propose a new method for evaluating the Hessian matrix based on the RISM/3D-RISM theory. In the method, we prepare a trajectory of a biomolecule in water by running a MD simulation in which each atom is driven by the interatomic forces as a derivative of the free energy surface with respect to the atomic coordinates.\textsuperscript{[29]} The free energy surface is evaluated by means of the RISM/3D-RISM method. The Hessian matrix is obtained from the trajectory taking the second derivative of the free energy surface with respect to the atomic coordinate of the solute at each snapshot, and taking the average over the entire trajectory.

The method is applied to a rather small biomolecule, alanine dipeptide, in water to calculate the wave number spectrum of the molecule. The result is compared with the experimental data of the spectrum around the tera Hertz region observed by means of optical heterodyne-detected Raman-induced spectroscopy (RIKES).\textsuperscript{[11,12]}

## 2. Theory

In the present study, we apply the generalized Langevin equation (GLE) of a protein in water at the infinite dilution, derived by Kim and
Hirata, to calculate the structural fluctuation. The equation has been derived by projecting all the variables in the phase space, concerning both the protein and water, onto the four dynamic variables: the density field and its conjugated momentum of water, and the spatial and momentum coordinates of atoms of a single protein molecule. It should be noted that the canonical ensemble average for the projection is taken over the entire coordinates, spatial and momentum, of both the protein and water. The projection gives rise essentially to two equations, one for a protein molecule, and the other for the density field of water molecules, which are conjugated each other. In the present study, we are just concerned with the GLE for a protein, which reads

\[
M_\alpha \frac{d^2 \Delta R_\alpha(t)}{dt^2} = - \sum_\beta k_{\alpha\beta} \Delta R_\beta(t) = \sum_\beta \gamma_{\alpha\beta}(t-s) \cdot \frac{P_{\alpha}(s)}{M_\alpha} + \mathbf{W}_\alpha(t) \tag{1}
\]

where \( \Delta R_\alpha(t) = R_\alpha(t) - \langle R_\alpha \rangle \) denotes the displacement of an atom \( \alpha \) of the protein from its equilibrium position \( \langle R_\alpha \rangle \). It is worthwhile to note that the ensemble average \( \langle \cdots \rangle \) is taken over the entire phase space including both the protein and water. The second and third terms in the right-hand side in Equation (1) are the friction and random forces acting on the atom \( \alpha \) of the biomolecule as in the standard GLE, which play crucial roles in the dynamics or relaxation from a fluctuated structure to the equilibrium one. We ignore those two terms in this study in which we are just interested in the mode and magnitude of the fluctuation.

The first term in the right-hand-side is the restoring force acting on the atom \( \alpha \) from the other atoms represented by \( \beta \), which is proportional to the displacement of the atoms from their equilibrium position \( \langle R_\beta \rangle \), or the structural fluctuation. With those physics of GLE in mind, Equation (1) is adapted specially to the analysis of structural fluctuation in the form,

\[
M_\alpha \frac{d^2 \Delta R_\alpha(t)}{dt^2} = - \sum_\beta k_{\alpha\beta} \Delta R_\beta(t) \tag{2}
\]

where \( k_{\alpha\beta} \) is the force-constant matrix, or the Hessian matrix, of the restoring force, and the following expression was derived by Kim and Hirata.

\[
k_{\alpha\beta} = k_B T \langle \Delta R_\alpha \Delta R_\beta \rangle^{-1} \tag{3}
\]

Equation (2) is formally identical to the equation of a harmonic oscillator, so that we are able to use all the numerical recipe developed for the normal mode analysis (NMA) with a difference of physics involved in the two. Let us refer to the new method as “Generalized Langevin Mode Analysis (GLMA).”

The Equation (2) implies that the fluctuation of protein atoms is taking place on the energy surface which has a quadratic form with respect to the atomic coordinates, that is,

\[
F(\Delta R_1, \Delta R_2, \cdots, \Delta R_N) = \sum_\alpha \sum_\beta \Delta R_\alpha A_{\alpha\beta} \Delta R_\beta \tag{4}
\]

Kim and Hirata have made a physical ansatz that the energy should be the free energy consisting of the potential energy and the solvation free energy, not just a potential energy among the protein atoms, that is,

\[
F(\Delta R_1, \Delta R_2, \cdots, \Delta R_N) = U(\Delta R_1, \Delta R_2, \cdots, \Delta R_N) + \Delta \mu(\Delta R_1, \Delta R_2, \cdots, \Delta R_N) \tag{5}
\]

where \( U \) and \( \Delta \mu \) are the potential energy and the solvation free energy. The ansatz is a logical consequence derived from Equation (1) that concerns all the coordinates in the phase space including those of water molecules, not just those of protein. With the Equation (3), the probability distribution of the displacement of atoms from the equilibrium state, or the structural fluctuation, is written as,

\[
p(\Delta R_1, \Delta R_2, \cdots, \Delta R_N) = \sqrt{\frac{|k|}{(2\pi)^n}} \exp \left[ -\frac{1}{2} \sum_\alpha \sum_\beta \Delta R_\alpha k_{\alpha\beta} \Delta R_\beta \right] \tag{6}
\]

The variance-covariace matrix is defined by

\[
\langle \Delta R_\alpha \Delta R_\beta \rangle = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Delta R_\alpha \Delta R_\beta p(\Delta R_1, \Delta R_2, \cdots, \Delta R_N) d\Delta R_1 d\Delta R_2 \cdots d\Delta R_N. \tag{7}
\]

The Hessian matrix can be obtained from Equation (4) as the second derivative of the free energy surface,

\[
k_{\alpha\beta} = \frac{\partial^2 F}{\partial \Delta R_\alpha \partial \Delta R_\beta} \tag{8}
\]

So, with the definition of the Hessian matrix by Equation (8), the mathematical isomorphism between NMA and GLMA was established. In the case of NMA, the free energy \( F(\Delta R_1, \Delta R_2, \cdots, \Delta R_N) \) should be replaced just by the potential energy \( U(\Delta R_1, \Delta R_2, \cdots, \Delta R_N) \).

The mathematical task of NMA or GLMA is to perform a principal-axis (component) analysis, or to solve an eigen value problem, of the Hessian matrix (Equation (8)) or the variance-covariance matrix defined by Equation (7), which are related through Equation (3). However, there is a big stumbling block before we are able to perform the principal-axis analysis, that is how to realize the Hessian or variance-covariance matrix.

### 2.1 Determining the Hessian matrix from the 3D-RISM/MD simulation

From Equation (5) and (8), the Hessian matrix of the free energy surface is split into two terms,
\[
\begin{align*}
\kappa_{ji} & = \frac{\partial^2 F}{\partial \mathbf{R}_j \partial \mathbf{R}_i} \\
& = \frac{\partial^2 U}{\partial \mathbf{R}_j \partial \mathbf{R}_i} + \frac{\partial^2 \Delta \mu}{\partial \mathbf{R}_j \partial \mathbf{R}_i}.
\end{align*}
\]

(9)

The first term is the second derivative of the potential energy among atoms in the biomolecule, the calculation of which can be performed with a routine implemented in a standard program of MD simulation, such as AMBER. However, it is not a trivial problem to calculate the second term, because it is concerned with the solvation free energy of the biomolecule in solution. If one tries to do it with the MD simulation, it will be a disaster. Even to determine the solvation free energy of a single conformation of a biomolecule by a MD simulation will require a huge amount of computation time. However, it is relatively an easy task for the RISM/3D-RISM formalism thanks to a tactics proposed by Yu and Karplus, which enables to derive the derivatives of the correlation functions along the iteration process for solving the integral equations.\[^{30}\]

The procedure begins with the 3D-RISM formula for the force acting on an atom of protein from the solvent molecules, derived by Yoshida and Hirata, which requires the derivative of the solvation free energy with respect to protein atoms.\[^{31,32}\]

\[
\frac{\partial \Delta \mu}{\partial \mathbf{R}_j} = \sum_i \rho_j \left( \frac{\partial^2 \Delta \mu}{\partial \mathbf{R}_j \partial \mathbf{R}_i} \right) \delta_j^{(\alpha)}(r) dr,
\]

(10)

where \(\rho_j\) denotes the density of the \(j\)-th atom in the solvent molecule, \(\delta_j^{(\alpha)}(r)\) the interaction energy between the solute atom \(\alpha\) and the solvent atom \(j\), \(\delta_j^{(\alpha)}(r)\) the spatial distribution of the solute atom \(j\) around the protein. It should be noted that the both are an implicit function of the coordinates \(\mathbf{R}_j\) of the solute. The relation can be understood as a classical analogue of the Hellmann-Feynman theorem in the quantum mechanics if one regards \(\mathbf{R}_j\), \(\delta_j^{(\alpha)}(r)\), and \(\delta_j^{(\alpha)}(r)\) as the coordinate of a nucleus, the electronic energy, and the probability distribution of the electron which is a square of the electronic wavefunction.

The second derivative of \(\Delta \mu\) with respect to the protein coordinate is derived simply from Equation (10) as

\[
\frac{\partial^2 \Delta \mu}{\partial \mathbf{R}_j \partial \mathbf{R}_i} = \sum_i \rho_j \left( \frac{\partial^2 \Delta \mu}{\partial \mathbf{R}_j \partial \mathbf{R}_i} \right) \delta_j^{(\alpha)}(r) dr + \sum_i \rho_i \left( \frac{\partial^2 \Delta \mu}{\partial \mathbf{R}_i \partial \mathbf{R}_j} \right) \delta_j^{(\alpha)}(r) dr
\]

(11)

Equation (11) includes the derivative of the spatial distribution function which makes the calculation potentially non-trivial. However, it can be performed by the Yu-Karplus method\[^{29}\] along the course of iteration for solving the RISM/3D-RISM equation as follows.

Although the method is common to any closure to solve the equation, here, we just present the procedure corresponding to the Kovalenko-Hirata closure. Then, the RISM/3D-RISM equation consists of the two equations, which are written as

\[
\begin{align*}
\frac{\partial h_j^{\alpha\beta}(r)}{\partial \mathbf{R}_j} & = \sum_i \delta_j^{(\alpha\beta)}(r) \xi_i^{\alpha\beta}(r) dr' \\
& = \chi_j^{\alpha\beta} c_j^{\alpha\beta}(r)
\end{align*}
\]

(12)

\[
\begin{align*}
\frac{\partial h_j^{\alpha\beta}(r)}{\partial \mathbf{R}_j} & = \left\{ \begin{array}{ll}
\exp \left[ -u_j^{\alpha\beta}(r)/k_BT + \xi_j^{\alpha\beta} c_j^{\alpha\beta}(r) \right] & \text{for } -u_j^{\alpha\beta}(r)/k_BT + \xi_j^{\alpha\beta} c_j^{\alpha\beta}(r) > 0 \\
1 & \text{for } -u_j^{\alpha\beta}(r)/k_BT + \xi_j^{\alpha\beta} c_j^{\alpha\beta}(r) \leq 0
\end{array} \right.
\end{align*}
\]

(13)

\[
\begin{align*}
\xi_j^{\alpha\beta} c_j^{\alpha\beta}(r) & = -u_j^{\alpha\beta}(r)/k_BT + \exp \left[ -u_j^{\alpha\beta}(r)/k_BT + \xi_j^{\alpha\beta} c_j^{\alpha\beta}(r) \right]
\end{align*}
\]

(14)

where \(\xi_j^{\alpha\beta}(r)\) is the site-site pair correlation function of solvent that plays the susceptibility or response function to the perturbation \(c_j^{\alpha\beta}(r)\) from the solute molecule.\[^{33}\] The derivative of the correlation functions with respect to the atomic coordinate of protein can be written as

\[
\frac{\partial h_j^{\alpha\beta}(r)}{\partial \mathbf{R}_j} = \sum_i \chi_j^{\alpha\beta} \frac{\partial h_j^{\alpha\beta}(r)}{\partial \mathbf{R}_j}
\]

(15)

\[
\begin{align*}
\frac{\partial h_j^{\alpha\beta}(r)}{\partial \mathbf{R}_j} & = \left\{ \begin{array}{ll}
\frac{1}{k_BT} \frac{\partial^2 h_j^{\alpha\beta}(r)}{\partial \mathbf{R}_j^2} & \text{for } -u_j^{\alpha\beta}(r)/k_BT + \xi_j^{\alpha\beta} c_j^{\alpha\beta}(r) > 0 \\
\frac{1}{k_BT} \frac{\partial^2 h_j^{\alpha\beta}(r)}{\partial \mathbf{R}_j^2} & \text{for } -u_j^{\alpha\beta}(r)/k_BT + \xi_j^{\alpha\beta} c_j^{\alpha\beta}(r) \leq 0
\end{array} \right.
\end{align*}
\]

(16)

The derivatives can be calculated along the course of iteration to find the solutions for the correlation functions themselves.

In order to compare the theoretical results with experimental data of the wavenumber spectrum, the Hessian matrix is diagonalized to find the eigen value and vector for each snapshot of the simulation, and averaged over the trajectory to take account for the thermal fluctuation of solute conformation.

It will be worthwhile to make some note concerning the validity of the analysis proposed here. The generalized Langevin equation (Equation (1)) employed here assumes existence of some equilibrium structure (\(\mathbf{R}_e\)): for example, it will be the native structure in the case of a protein in the native condition. So, the following question may be naturally raised. How do you define the equilibrium structure without making any analysis? The answer to the question is that we do not have to use the information concerning the equilibrium structure for an analysis of Hessian. The Hessian matrix defined by Equation (9) has a local character, and it is determined at each point of the conformational space. If the free energy surface at a point of the conformational space is concave, Hessain or the second derivative is positive, and vice versa.

3 | COMPUTATIONAL DETAILS

To calculate the Hessian matrix, \(k_{\alpha\beta}\), we first perform the molecular dynamics simulations based on the MTS-MD/OIN/GSFE/3D-RISM.
KH (MD/3D-RISM-KH) program\textsuperscript{[34]} implemented in the AMBER 18 software package.

In the MD/3D-RISM-KH simulation, we use the reference system propagator algorithm (RESPA) proposed by Omelyan and Kovalenko, in which the calculation of the solvent induced force is performed at every \textit{rismnrespa} steps, not at every step, of the MD simulation in order to save the computation time.\textsuperscript{[34]} So, the solvent induced force (Equation (10)) is calculated in every \textit{rismnrespa} steps. Furthermore, the mean force induced by the solvent is extrapolated in \textit{fcetre} steps. Thus, 3D-RISM calculations are performed every \textit{fcetre} steps \times \textit{rismnrespa} steps.

The ff14sb force field parameter is selected. Water molecules are represented by the modified SPC/E model. The cutoff radius of the solute-solvent interactions is set to 14 Å. The 3D-RISM-KH integral equations were discretized on a rectangular grid of resolution $\delta r = 0.5$ Å and converged to a relative root-mean-square residual tolerance of $\delta e = 10^{-4}$ by using the MDIIS accelerated numerical solver.

For the simulation of the alanine dipeptide, \textit{fcetre}, \textit{fcenbas}, and \textit{fcenbasis} is set to 10, 20, and 200, respectively. An 80 ns simulation was performed using these parameters. For the simulation of the Met-enkephalin, \textit{fcetre}, \textit{fcenbas}, and \textit{fcenbasis} is set to 10, 40, and 500, respectively. A 200 ns simulation was performed using these parameters.

4 | RESULTS AND DISCUSSIONS

4.1 | 3D-RISM/MD trajectory of the dipeptide projected onto dihedral-angle space

Depicted in Figure 1 are distributions of the MD trajectory over 200 ns projected onto the angular coordinate space spanned by the two dihedral angles $(\psi_1, \phi_2)$ illustrated in the figure. A major peak of the distribution is found around $\psi_1 = 150^\circ$ and $\phi_2 = -60^\circ$, marked by (ii) in Figure 1B, which corresponds roughly to the trans-gauche conformation. There is also a minor peak of the distribution around $\psi_1 = 150^\circ$ and $\phi_2 = -150^\circ$, which corresponds roughly to the trans-trans conformation.

4.2 | Spectrum from multiple snapshots

Shown in Figure 2A is the wavenumber spectrum of an alanine dipeptide in water, obtained from the Hessian matrix, $k_{\phi\psi}$, by diagonalizing the matrix and averaging over 1000 snapshots. The snapshots are taken from an 80 ns 3D-RISM/MD trajectory, evenly spaced. The structure of each snapshot was not minimized for the free energy. Shown in Figure 2B for comparison is the wavenumber spectrum calculated from the 100 snapshots, the free energy of which is minimized.

The spectrum depicted in Figure 2A can be discussed in the four different regions: (1) the region greater than $\sim$2700 cm$^{-1}$, (2) the region between $\sim$500 to $\sim$1800 cm$^{-1}$, (3) the region less than $\sim$500 cm$^{-1}$, (4) the region of negative wave number. The regions greater than $\sim$2700 cm$^{-1}$ is attributed apparently to stretching and bending of chemical-bonds related to hydrogen atoms. On the other hand, the spectrum between $\sim$500 to $\sim$1800 cm$^{-1}$ seems to be assigned to the local oscillations, stretching and bending, with which the heavy atoms such as C, O, N are concerned. Those vibrational modes are largely determined by the model of interatomic interaction-potential employed by the simulation, and they are not of much interest for the structural fluctuation of biomolecules. It is the regions (3) and (4) that is closely related to the structural fluctuation. The exploded view of the spectrum in that wave-number region is depicted in the right figure. So, we focus our attention on that region of the spectrum in the following.

Depicted in Figure 3 is the spectrum of the dipeptide in the wave-number less than 500 cm$^{-1}$, which is compared with a corresponding experimental data obtained by Klaas and his coworkers by means of optical heterodyne-detected Raman-induced spectroscopy (RIKES).\textsuperscript{[11,12]} Although there are apparent differences observed between the two results, there is a common feature in the two spectra. The two spectra have four peaks between the wave numbers, 0 to 500 cm$^{-1}$, which are relatively close each other, that is, $\sim$90, 250, 370 and 450 cm$^{-1}$ in the RIKES spectra, while $\sim$90, $\sim$240, $\sim$320, and $\sim$450 cm$^{-1}$ in the GLMA spectra.

There are marked differences between the two spectra in the following respects. (1) the negative frequency region observed in the GLMA result, which is absent in the experimental data; (2) small sub-peak around 0 cm$^{-1}$ seen in the RIKES result, which is absent in the GLMA spectra; (3) the large difference in the intensity between the two spectra, especially around 0 to 100 cm$^{-1}$ region; (4) relatively large difference, about 50 cm$^{-1}$, in the peak positions around 320 cm$^{-1}$.

The spectrum at the negative frequency region in the theoretical result is apparently reflecting the representative points in the $\phi-\psi$ plot, which are assigned to the transient regions among the three stable conformations of the dipeptide molecule, depicted in Figure 1. These represent rare events, and there is no way to be observed by the spectrum in the real world.

In order to confirm if the structures contributing to the negative frequency are those in the transient states between two stable conformations or not, we have minimized the free energy of 100 snapshots evenly spaced from the trajectory. The spectrum calculated from the minimized structure is depicted in Figure 2B. As is expected, the spectrum at negative wave number region has disappeared. The result is not surprising, if the negative wave number region of spectrum is truly due to the contributions from the transient states, since the structures in the states fall down to a local minimum of the free-energy surface, where the curvature is positive.

The appearance of the spectrum at the negative frequency seems pathological in the sense that it does not agree with the experimental results of the spectroscopy. However, such a viewpoint may be one-sided, because the molecular structures that yield the spectrum at negative wave number have a well-defined physical meaning, that is, the transition state. The analysis of such a region of the "transition
Figure 1  The trajectory projected onto the dihedral angle (Ψ1,φ2) space: (A) the snapshots at every 80 steps, (B) the distribution of the trajectory. The molecular pictures depicted under the two figures (A) and (B) illustrate the conformations corresponding to the snapshots (i), (ii), and (iii) in fig. (B).

Figure 2  The wavenumber spectrum of an alanine-dipeptide in water: (A) the spectrum calculated from the 1000 snapshots, evenly spaced, without minimizing the free energy; (B) the spectrum calculated from the 100 snapshots, evenly spaced, after minimizing the free energy. The inset is exploded view of the spectrum in the range of wave number from –100 to 500 cm⁻¹.
Comparing the low frequency spectrum from GLMA
The modes of fluctuation corresponding to the peaks is indicat-
is the structure and the fluctuational mode of
are the radial distribution functions (RDF) of
is a manifestation of the hydrogen-bond
state” is of great interest from a view point of the kinetic analysis of a
chemical reaction. (In the present case, the chemical reaction is an
“isomerization”) Nevertheless, we just focus our attention on the
structural fluctuation of the biomolecule, projected on the real fre-
quency in the present study.

The second to fourth differences are mainly caused by the dis-
crepancy in the physics involved in the two analyses. According to the
authors of the experimental paper, the spectral data was obtained by
subtracting the intensity concerning pure solvent from overall spec-
trum including solute and solvent. Therefore, the intensity
depicted in the figure includes two contributions, one from the solute
and the other from water molecules that are interacting with the sol-
ute. The small subpeak seen around 0 cm\(^{-1}\) is likely to be assigned to
the diffusive motion of water molecules interacting with the dipeptide.
On the other hand, the theoretical analysis does not include any spec-
trum contributed by water, although the contribution from solute-
solvent interaction is included implicitly in the spectrum of the solute.
It may be the reason why the theoretical spectrum does not have the
intensity around 0 cm\(^{-1}\). The experimental spectra may also be con-
tributed from the translational and rotational diffusion of the dipeptide.
Such degrees of freedom are also removed from the theoretical
analysis.

The same reason is attributed to the large difference in the intensity
between the two spectra, especially those around -100 cm\(^{-1}\) region.
The region of the RIKES spectrum includes contributions from water mol-
ecules interacting with the solute. Those water molecules interacting with
the solute, especially via hydrogen-bond, are likely to be involved in oscil-
latory motions in lower frequency modes. So, it is suggested that the large
intensity around 100 cm\(^{-1}\) is assigned to the intermolecular oscillatory
motion of water molecules interacting with the solute.

The suggestion is supported partially by a simulation study carried out
for water molecules around a monoatomic solute, Na\(^+\), K\(^+\), Ne, A,
and Xe, to calculate the wavenumber spectrum of water. All the
spectra show large intensity between 0 to 300 cm\(^{-1}\), which are
assigned to the librational mode of water molecules.

The rather large discrepancy in the peak positions between the theo-
retical and the experimental spectra, \(\sim 320\) cm\(^{-1}\) vs. \(\sim 370\) cm\(^{-1}\), may
require a structural analysis of the mode of fluctuation.

Illustrated in Figure 4 is the structure and the fluctuational mode of
the dipeptide corresponding to the peak positions in the GLMA spec-
trum, which are obtained by diagonalizing the Hessian matrix. In the
figures, the direction and amplitude of the fluctuation of each atom
are illustrated by thick arrows.

As can be seen, all the modes carry a collective character, more or
less, in the sense that the oscillations extend over the entire molecule.
For example, the mode with the lowest frequency, or \(87\) cm\(^{-1}\), looks
like a “hinge-bending” motion around the C=O carbonyl bond, since
the central carbonyl group and the two terminal groups are oscillating
in the opposite phases. On the other hand, the mode at \(452\) cm\(^{-1}\)
seems to be more localized around the N-terminus group. It may be the
reason why the frequency is relatively high.

An interesting behavior is seen in the mode assigned to
319 cm\(^{-1}\), in which the carbonyl and the amide nitrogen are oscil-
lating in the opposite phase. The oscillation is indicative of a water
molecule bridging between the two atoms through a hydrogen-
bond. In order to clarify if it is the case or not, the solvation struc-
ture of the molecule in that fluctuation mode, or \(k = 319\) cm\(^{-1}\),
was analyzed.

Depicted in Figure 5 are the radial distribution functions (RDF) of
water molecules around the carbonyl oxygen and the amide nitrogen
of the dipeptide. The sharp peak at \(r \sim 1.8\) A in the O(peptide)-H (water) RDF in Figure 5C is a manifestation of the hydrogen-bond
between the carbonyl oxygen and the water hydrogen. The sharp
peak at \(r \sim 2.8\) A in the N(peptide)-O(water) RDF in Figure 5A is indic-
ative of a strong electrostatic-interaction between the amide nitrogen
and the water oxygen. The two sharp peaks in RDFs are strong evi-
dence of the existence of the water-bridge between the carbonyl
oxygen and the amide-nitrogen through those strong interactions.

The situation is illustrated by the cartoon at bottom right in Figure 5.
The analysis suggests that the water-bridge through the two strong
interactions may be the origin of the fluctuation mode assigned to 319 cm$^{-1}$.

Now, we are at the position to answer the question raised at the end of the last subsection, namely, the origin of the difference in the peak positions at 319 and 370 cm$^{-1}$, respectively, in the theoretical and experimental spectra. As we have already mentioned, the experimental spectrum includes both the contributions from the dipeptide and water molecules, while the theoretical one is concerned only with the interactions within the dipeptide, which include contributions from solvent implicitly. Therefore, the peak in the theoretical spectrum at 319 cm$^{-1}$ is originated mainly by the H(peptide)-O(peptide) interaction bridged by a water molecule. On the other hand, the peak in the experimental spectrum at 370 cm$^{-1}$ is a composite band consisting of the contributions from the H(peptide)-O(peptide) interaction and the water molecule bridging the two atoms in the peptide. The frequency of the mode of the water molecule may be higher than that of H(peptide)-O(peptide) interaction, because the water molecule is connected with the two atoms in the peptide through the two strong interactions.

5 | CONCLUDING REMARK

A new theoretical method, referred to as Generalized Langevin Mode Analysis (GLMA), was proposed to analyze the mode of structural fluctuations of a biomolecule in solution. The method combines the two theories in the statistical mechanics, or the Generalized Langevin theory and the RISM/3D-RISM theory, to calculate the second derivative, or the Hessian matrix, of the free energy surface of a biomolecule in solution, which consists of the intramolecular interaction among atoms in the biomolecule and the solvation free energy.

The method was applied to the alanine dipeptide in aqueous solution to calculate the wave number spectrum of the system, especially focusing on the collective modes of the fluctuation less than 500 cm$^{-1}$, for which an experimental data by means of optical heterodyne-detected Raman-induced spectroscopy (RIKES) is available. The theoretical analysis reproduced the main features of the experimental spectrum with respect to the peak positions of the four bands around 90, 240, 370, and 400 cm$^{-1}$, observed in the experimental spectrum, in spite that the physics involved in the two spectrum was not exactly the same: the experimental spectrum includes the contributions from the dipeptide and the water molecules interacting with the solute, while the theoretical one is just concerned with the solute molecule.

There are two major discrepancies between the two spectra: one in the band intensity around 100 cm$^{-1}$, and the other in the peak positions, that is, 320 cm$^{-1}$ in the theory, while 370 cm$^{-1}$ in the experiment. The large band intensity around 100 cm$^{-1}$ in the RIKES spectrum was attributed to the oscillatory motion of water molecules interacting with the dipeptide, which is not included in the spectrum from GLMA.
In order to examine the shift of the peak positions from \(\sim 320 \text{ cm}^{-1}\) in the GMLA spectrum to \(\sim 370 \text{ cm}^{-1}\) in the RIKES spectrum, the radial distribution function of water around the dipeptide, which is in the fluctuation mode around \(\sim 320 \text{ cm}^{-1}\), was analyzed. From the analysis, it was concluded that the shift of the peak position from \(\sim 320 \text{ cm}^{-1}\) in the GMLA spectrum to \(\sim 370 \text{ cm}^{-1}\) in the RIKES spectrum is attributed to the oscillatory motion of a water molecule which bridges between two atoms, N and O, in the dipeptide, through the hydrogen-bond between H(water) and O(peptide), and the electrostatic interaction between N(peptide) and O(water).

Two possible improvements of the theory are conceivable for getting better agreement with the RIKES data. One of those is to include the friction in the Hessian matrix in Equation (2), just as been proposed by Lamm and Szabo in their Langevin mode analysis.\(^{[29]}\) The effect may change some modes of fluctuation toward lower wavenumber side. Less the wavenumber, greater the effect may be, since such a collective mode with lower frequency involves more atoms to be exposed to water in motion.

The other point of the possible improvement concerns water molecules interacting with the peptide. In order to get better agreement with the RIKES data, the spectrum from water molecules themselves should be taken into account. Such a goal may be achieved by solving the generalized Langevin equation for water molecules interacting with the solute molecule. Such an equation has been already derived by Kim and Hirata, along with the generalized Langevin equation for a solute in water, which was employed in the present paper to analyze the structural fluctuation of the dipeptide.

The GLMA method developed here can be applied to explore the structural fluctuation of protein in solution without any further development in the theory, but with an assist of much greater computational power. It is our future plan to carry out such studies concerning the structural fluctuation of protein.

**ACKNOWLEDGMENT**

The paper is contributed to the special issue of the journal, which is published for the memory of Sihyun Ham who was a collaborator and a good friend of Fumio Hirata.

The authors gratefully acknowledge Klaas Wynne for providing the RIKES data. They are also grateful to Hideaki Shirotá for bringing Wynne's paper to our attention. This work was partly supported by JSPS KAKENHI Grant Number 30737523. The part of the work was carried out while FH was a fellow of Toyota Physical and Chemical Research Institute (Toyota Riken).

**DATA AVAILABILITY STATEMENT**

Data openly available in a public repository that issues datasets with DOIs

**ORCID**

Masatake Sugita https://orcid.org/0000-0002-6298-4055

Fumio Hirata https://orcid.org/0000-0002-9430-0754

**REFERENCES**

[1] K. Gekko, N. Obu, J. Li, J. C. Lee, Biochemistry 2004, 43, 3844.
[2] K. Gekko, T. Kamiyama, E. Ohmoe, K. Katayanagi, J. Biochem. 2000, 128, 21.
[3] G. Bhabha, J. Lee, D. C. Eikert, J. Gam, I. A. Wilson, H. J. Dyson, S. J. Benkovic, P. E. Wright, Science 2011, 332, 234.
[4] K. Akasaka, Chem. Rev. 2006, 106, 1814.
[5] K. Kuroi, K. Okajima, M. Ikeuchi, S. Tokutomi, M. Terazima, Proc. Natl. Acad. Sci. 2014, 111(41), 14764.
[6] K. Tanaka, Y. Nakasone, K. Okajima, M. Ikeuchi, S. Tokutomi, M. Terazima, J. Mol. Biol. 2009, 386(5), 1290.
[7] J. Monod, J. Wymann, J. P. Changeux, J. Mol. Biol. 1965, 12(1), 88.
[8] D. E. Koshland, G. Némethy, D. Filmer, Biochemistry 1966, 5(1), 365.
[9] B. Srivasan, F. Forouhar, A. Shukla, C. Sampangi, S. Kulkarni, M. Abashidze, J. Seetharaman, S. Lew, L. Mao, T. B. Acton, et al., FEBS J. 2014, 281, 1613.
[10] M. A. Cuendet, H. Weinstein, M. V. LeVine, J. Chem. Theory Comput. 2016, 12, 5758.
[11] G. Giraud, K. Wynne, J. Am. Chem. Soc. 2002, 124(41), 12110.
[12] G. Giraud, J. Karolin, K. Wynne, Biophys. J. 2003, 1–11, 1903.
[13] A. Kitao, F. Hirata, N. Go, Chem. Phys. 1991, 158, 447.
[14] A. Kitao, S. Hayward, N. Go, Proteins 1998, 33, 496.
[15] Kitao, A.; Go, N. Investigating protein dynamics in collective coordinate space. 1999, 1–6.
[16] N. Karasawa, A. Mitsutake, H. Takano, J. Chem. Phys. 2019, 150, 084113.
[17] A. E. García, Phys. Rev. Lett. 2011, 68(17), 2696.
[18] C. R. Schwantes, V. S. Pande, J. Chem. Theory Comput. 2013, 9, 2000.
[19] T. Mori, S. Saito, J. Chem. Phys. 2015, 142, 135101.
[20] G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, F. Noé, J. Chem. Phys. 2013, 139, 015102.
[21] S. Sakuraba, Y. Joti, A. Kitao, J. Chem. Phys. 2010, 133, 185102.
[22] P. H. Berens, D. H. J. Mackay, G. M. White, K. R. Wilson, J. Chem. Phys. 1983, 79, 2375.
[23] L. D. M. Peters, J. C. Dietschreit, J. Kussmann, C. Orschenufeld, J. Chem. Phys. 2019, 150, 194111.
[24] Statistical mechanics, McQuarrie, D. A., Harper & Row publishers, 1976.
[25] Harmonic Dynamics of Proteins, Proc. Natl. Acad. Sci. 1983, 80, 6571.
[26] N. Go, T. Nishikawa, Proc. Natl. Acad. Sci. 1983, 80, 3696.
[27] B. Kim, F. Hirata, J. Chem. Phys. 2013, 138(5), 54108.
[28] M. C. Wang, Uhlenbeck., Rev. Mod. Phys. 1945, 17, 323.
[29] G. Lamm, A. Szabo, J. Chem. Phys. 1986, 85, 7334.
[30] H.-A. Yu, M. Karplus, J. Chem. Phys. 1988, 89, 2366.
[31] T. Miyata, F. Hirata, J. Comput. Chem. 2007, 29, 871.
[32] N. Yoshida, F. Hirata, J. Comput. Chem. 2006, 27(4), 453.
[33] Hirata, F., Exploring Life phenomena with statistical mechanics of molecular liquids, CRC press, 2021.
[34] I. Omelyan, A. Kovalenko, J. Chem. Theory Comput. 2015, 11(4), 1875.
[35] R. A. X. Persson, V. Pattini, A. Singh, S. M. Kast, K. Heyden, J. Chem. Theory Comput. 2017, 13, 4467.

**How to cite this article:** M. Sugita, F. Hirata, J. Comput. Chem. 2023, 44(10), 1031. https://doi.org/10.1002/jcc.27062