Journal of Molecular Liquids 217 (2016) 23–28

Contents lists available at ScienceDirect

Journal of Molecular Liquids

journal homepage: www.elsevier.com/locate/molliq

Multi-scale dynamics simulation of protein based on the generalized Langevin equation combined with 3D-RISM theory

Fumio Hirata [a,*], Bongsoo Kim [b]

[a] College of Life Sciences, Ritsumeikan University, and Molecular Design Frontier Co. Ltd., Kusatsu, Shiga 525–8577, Japan
[b] Department of Physics and Institute for Soft and Bio Sciences, Changwon National University, Changwon 641–773, South Korea

ARTICLE INFO

Article history:
Received 8 June 2015
Accepted 14 July 2015
Available online 18 August 2015

Keywords:
Multiscale dynamics
Protein structure
Generalized Langevin theory
3D-RISM theory
Principal component analysis

ABSTRACT

A theory to realize a multiscale dynamics of protein is proposed based on the generalized Langevin equation combined with the 3D-RISM theory. The idea of normal mode analysis (NMA) is adopted to decouple the dynamics into different modes having respective time scales. In order to decouple the modes, the variance–covariance matrix concerning the structural fluctuation of protein in solution is diagonalized, instead of the Hessian matrix of a harmonic oscillator in vacuum. An algorithm to estimate the friction coefficient exerted on each atom of protein due to solvent is also proposed.

1. Introduction

Structural dynamics of protein plays a crucial role in the expression of their intrinsic functions. The best example is a process of the signal transduction due to protein, which in general consists of a series of the molecular recognitions, chemical reactions, and mass transportations for propagating a signal from upstream to downstream of the process. Each process of the signal-transduction should be precisely controlled in time, otherwise the system malfunctions due to failure of controlling the signal traffic, and the living body will become sick or dead. For example, a group of proteins called "ion channel" has a device to permeate an ion from one side of membrane to the other [1]. The conduction rate is determined by various factors related to structure of the protein, but the most important one is the mechanism called "gating," which plays a role similar to a "valve" in the hydrodynamic system. The opening and closing of a gate are often regulated by a structural fluctuation of the protein, such as a camera-iris-like motion found in the potassium channel [2].

There exist the two essential requirements for describing the structural dynamics of biomolecules theoretically in a sensible manner. Firstly, the structure and dynamics of protein should be described in atomic resolution. The reason is because all biological functions performed by protein are expressed in atomic scale. This requirement denies the popular approach so-called "coarse-grained model" for protein, in which an amino-acid residue is represented by a "big sphere," for examples [3–6]. Coarse graining amino-acid residues in pore of an ion channel, for example, will ruin the dynamics entirely, since such an ion is interacting with *atoms* of residues consisting the pore, as well as with water molecules in general [7]. Secondly, the protein dynamics should be conjugated or coupled in atomic scale with that of solvent or solution in which the protein is immersed. The second requirement turns down the conventional treatments of protein dynamics based on the phenomenological Langevin theory in which solvent dynamics is decoupled from that of solute and is represented by a phenomenological transport coefficient [8,9].

Recently, the authors have presented a theory which meets the two requirements stated above, based on the generalized Langevin theory combined with the 3D-RISM theory [10,11] (see also Appendix A). The theory consists of two equations of describing the time evolution of protein structure and the solvent density, which are correlated with each other: the deviation or fluctuation of atomic positions from its equilibrium states, and the collective density field of solvent. These coupled equations take a general form of the Langevin equation, having the friction and random force terms. In the theory, solvent dynamics is coarse-grained in the level of the collective density of an atom of solvent, but still retains atomic resolution. Two important features of the protein dynamics are revealed by the theory. Firstly, the structural dynamics of protein is taking place on the free energy surface of protein including the solvation free energy, not just on the molecular–mechanical (MM) energy surface of protein. Secondly, the structural fluctuation of protein has characteristics of the Gaussian process with the variance–covariance matrix concerning the displacement of atomic coordinates from its equilibrium position. The theory has been employed to conceptually clarify the structural fluctuation of protein induced by thermodynamic perturbations [12]. Although many important aspects of the conjugated

* Corresponding author.

http://dx.doi.org/10.1016/j.molliq.2015.07.037
0167-7322/

dynamics of protein and solvent have been revealed by the theory, the proposed equations have not been solved numerically yet due to the lack of recipe of solving the equations.

A difficult problem for applying the new theory to actual dynamics of protein in solution by means of the numerical integration concerns the "time step." If one uses the small time step for solving the equation, i.e., the time step typical to the all-atom-MD-simulation, or femtosecond, the benefit of using the GLT/3D-RISM equation will not be so great with respect to the computation time, although we still gain a lot concerning physics of the process under concern. On the other hand, if one uses a large time step typical to the collective motion such as the hinge-bending motion, the dynamics will quickly fails due to the divergence in energy and force. Therefore, it is desirable to device a method to decouple the modes from fast to slow, and assign appropriate time steps depending on the mode. In the present paper, we propose a method to decouple the dynamical modes based on the idea of normal mode analysis (NMA) [13,14].

The paper is devoted to the memory of Dr. K. Arakawa who guided F. Hirata, one of the co-authors, to the statistical mechanics of liquid and solution.

2. Multi-scale dynamic algorithm for solving the GLT/3D-RISM equation

In the ordinary NMA carried out in vacuum, one defines the normal mode vector as a linear combination of the real-space displacement–vector of atoms [13,14]. The matrix to transform the vectors is called the "Hessian" or the variance–covariance matrix, which is nothing but the force constant matrix in the coupled oscillator. By diagonalizing the variance–covariance matrix, one obtains the normal mode vector and the normal mode variables. So, one can assign an appropriate time step or "amplitude" to each mode in the space of normal mode. Then, by carrying out the inverse transformation from the normal mode vector to the real-space vector, one can recover the displacement of atoms in the real space. Since the normal mode vectors are completely orthogonal due to the diagonalization of Hessian, no concern is necessary for violation of physics in real space, such as overlap of atoms in protein, after the inverse transformation. It is the decoupling of modes that brings us a benefit in the dynamics simulation.

However, NMA in vacuum cannot be applied to the actual dynamics of protein in solutions. Firstly, the protein structures taken from the protein data bank (PDB) to minimize the energy are close to the equilibrium structures in *aqueous solutions*, but not in *vacuum*. The structures are those of local energy minimum, which are in largely fluctuated states from the equilibrium structure in vacuum. Secondly, due to the high-energy state, it quickly relaxes to the equilibrium state in vacuum by any finite displacement of atoms. The last statement is concerned with the essential requirement of NMA in which a Taylor expansion of the potential energy with respect to the atomic displacement should be truncated at the second order for the harmonic analysis, or the diagonalization of the Hessian matrix.

Here, unlike NMA in vacuum, we consider the dynamics of protein that is taking place in a global minimum of the free energy surface in *solution*. The protein restores its equilibrium structure eventually after any fluctuation around the minimum, small or large, unless the equilibrium conditions such as temperature and pressure are changed. In such a case, we have already derived a generalized Langevin equation for the structural dynamics of protein, which is combined with the 3D-RISM/RISM equation (A brief review of the theory is provided in the Appendix A.) [11]

$$M_\alpha \frac{d^2\Delta\mathbf{R}_\alpha(t)}{dt^2} = -k_BT\sum_\beta \left(\mathbf{L}^{-1}\right)_{\alpha\beta}\cdot\Delta\mathbf{R}_\beta(t) - \int_0^t ds \sum_\beta \Gamma_{\alpha\beta}(t-s)\cdot\frac{\mathbf{P}_\beta(s)}{M_\beta} + \mathbf{W}_\alpha(t). \quad (1)$$

In Eq. (1), the Greek suffices specify atoms in protein. $\Delta\mathbf{R}_\alpha(t)$ is the deviation of atom α in protein from its equilibrium position, defined by, $\Delta\mathbf{R}_\alpha(t) \equiv \mathbf{R}_\alpha(t) - \langle\mathbf{R}_\alpha\rangle$, where $\langle\cdots\rangle$ means the statistical average at equilibrium, and $\mathbf{P}_\beta(t) = M_\beta\, d\mathbf{R}_\beta(t)/dt$. $\Gamma_{\alpha\beta}(t-s)$ and $\mathbf{W}_\alpha(t)$ are the friction coefficient and the random force, respectively, whose microscopic expressions are derived in Ref. [11]. The matrix **L** is the variance–covariance matrix concerning the structural fluctuation of atoms, defined by,

$$\mathbf{L} = \langle\Delta\mathbf{R}\Delta\mathbf{R}\rangle. \quad (2)$$

The essential feature of the Eq. (1) lies in the first term in the right-hand-side, that takes a form of the Hookian restoring force with $k_BT\mathbf{L}^{-1}$ as the *force constant* or the *Hessian*. Namely, the Eq. (1) takes the form of a Langevin harmonic oscillator. It is worthwhile to note that the harmonic nature of the equation is not originated from the truncation of Taylor expansion of the potential energy at the second order of the displacement as in NMA, but from the projection of all other variables in the phase space onto the dynamic variables including the atomic displacement of protein. The projection turned the dynamics from deterministic to stochastic or diffusive ones that are governed by the fluctuation–dissipation theorem. It is this feature that we can take advantage in order to decouple the modes of protein dynamics.

We adopt the idea of NMA in order to decouple the different modes of dynamics of protein in solution, but with some modifications, which we refer to as "Generalized Langevin Mode Analysis (GLMA)." Firstly, the *Hessian* or *force constant* in Eq. (1) is identified as the second derivative of the free energy surface including solvent effect, not of the mechanical potential energy surface, that is,

$$k_BT\left(\langle\Delta\mathbf{R}\Delta\mathbf{R}\rangle^{-1}\right)_{\alpha\beta} = \frac{\partial^2 F(\{\mathbf{R}\})}{\partial\mathbf{R}_\alpha\partial\mathbf{R}_\beta}, \quad (3)$$

where $F(\{\mathbf{R}\})$ is the free energy surface of the protein in solution, consisting of the interactions among atoms in protein, $U(\{\mathbf{R}\})$, and the solvation free energy $\Delta\mu(\{\mathbf{R}\})$ ($\{\mathbf{R}\}$ represents a set of coordinates of atoms in protein) [11]. Secondly, all the eigenvalues of the variance–covariance matrix should not necessarily be positive. Since our dynamics takes place in the global minimum of the free energy surface, the eigenvalues associated with the global structural change should be positive. However, those associated with the local structural change such as the exchange of a hydrogen-bond between an amino-acid residue and a water molecule, can be negative, which correspond to a transient state among local free energy minima. We assume that only few largest eigenvalues associated with the collective modes should be positive in order for the protein to stay in the global minimum. In that respect our approach is not just NMA in its narrow meaning, but rather a general principal-axis (or component) analysis on the quadratic free energy surface [15,16].

In the actual implementation of the theory to the GLT/3D-RISM simulation, an optimum size of time step should be carefully chosen to meet the following two requirements: firstly, the time step should be as large as possible in order to maximize the sampling of the free energy surface, and secondly, the first few eigenvalues should be positive during the simulation.

We follow the Verlet-type algorithm for the numerical integration of the GLT/3D-RISM equation. The Verlet-type algorithm reads,

$$\Delta\mathbf{R}_\alpha(t+\Delta t) = 2\Delta\mathbf{R}_\alpha(t) - \Delta\mathbf{R}_\alpha(t+\Delta t) + \frac{\partial^2\Delta\mathbf{R}_\alpha(t)}{\partial t^2}\Delta t^2. \quad (4)$$