[2016 Report]

Statistical mechanics theory to control the structural fluctuation of protein in aqueous solution

Fumio Hirata (Toyota Physical and Chemical Institute)

The fellow has developed a statistical mechanics of liquids (RISM & 3D-RISM) to explore chemical processes in solution and in life phenomena, including chemical reactions and the structure-function relation of protein. However, an important problem is remained unsolved. It is the structural fluctuation and dynamics of the biomolecule in water. The structural fluctuation plays a crucial role when a protein expresses its function such as an enzyme and ion channels. It is also important in a process of drug design to analyse the structural fluctuation of a target protein. A purpose of the research is to develop a new theory to control the structural fluctuation of protein in aqueous solutions, combining the two theories in statistical mechanics, the generalized Langevin theory and the RISM/3D-RISM/ theory.

[**Introduction**]

Needless to say, protein is a main player for living bodies to maintain their life. Concerning the structural fluctuation of protein, a physical model has been *professed* persistently among biophysisists. It is a *harmonic oscillator model or normal mode analysis in vacuum.* [1,2] The model, however, is unable to describe a structural fluctuation of *living* protein in our body. Why? It is because such a *protein* will undergo a non-linear (prastic) deformation by a small perturbation. (If you doubt the statement, perform a molecular dynamics simulation of a protein in *vacuum* for, maybe, 100 pico seconds, starting from its X-ray structure.) A protein in a cell makes large conformational change when it is functioning. However, the protein restores its native structure after a perturbation associated with its activity is removed, on the contrary to the case of a portein in vacuum. For examples, an enzyme in a cell makes large structural change when it is working as a catalyst. However, it recovers its original structure upon completing its catalytic activity, and becomes ready for next reaction. (In fact, it is the reason why the molecule is called “biocatalyst.”) That is, the structural response of a protein at work to a perturbation is *linear* or *elastic deformation*. The

linearlity of the response is an essential requirement for a living body to maintain its life. Then, why is the structural fluctuation of *living* protein linear? The answer to the question lies in *water*. In order to understand the statement intuitively, let's look at Fig. 1. Illustrated conceptually in Fig. 1 is the structural response of a protein to applied pressure. Fig. 1(a) illustrates the situation in which pressure is applied to a protein in *vacuum*. It is readily imagined that the protein will be crushed by small pressure, and that the protein will not restore its original structure spontaneously. On the other hand, illustrated in Fig. 1(b) is a protein in water. In this case, the pressure (or perturbation) gives influence on the structure of protein through water. The structure will be deformed completely into a random coil if a large magnitude of pressure is applied. However, the native strcuture will be restored upon the pressure being removed, contrary to the case of protein in vacuum, as has been proven by Akasaka experimentaly by means of the pressure NMR. [3] Namely, *water* makes the structural response of protein to pressure to be *linear*.

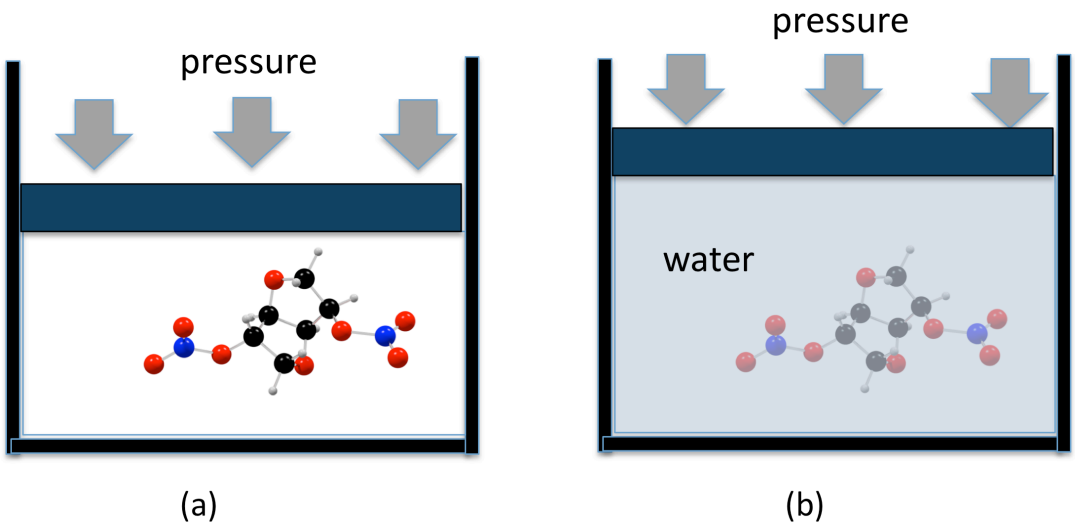

Fig. 1

Why does the structural flucuation of protein in water become linear? The answer to the question is that the structure is determined not only by the intramolecular interaction among atoms in protein, but also by the *solvation free energy*. For examples, the secondary structure of protein is supported by hydrogen-bonds among backone atoms. If the large pressure is applied to such a molecule in vacuun, extensive reorganization of

hydrogen-bonds will take place to make protein structure entangled, and the original structure will never restore after the presuure is removed. However, if the molecule is in water, the broken hydrogen-bonds due to the perturbation will be complemented by those with water molecules. Upon the pressure being removed, recombination of hydrogen-bonds will take place to recover the native structure.

We have derived a generalized Langevin equation to describe the structural dynamics of a protein in water. The equation suggests strongly that the structural fluctuation of protein is *linear*. If the strutural fluctuation is linear, the *linear response theory* should be valid. Therefore, we have proposed a linear reponse theory to describe a structural response to a thermodynamic (static) perturbation. In the last year, the theory was further extended to a structural response to a time dependent perturbation, an example of which is the photo-induced structural dynamics of a protein. Those works are briefly reviewed in the present report.

I. Theory of the structural fluctuation of protein [B. Kim & F. Hirata, *J. Chem. Phys.,* 138, 054108 (2012)**]**

The research starts from the following equation that describes the strutcural fluctuation of protein in aqueous solutions. (The equation was derived by us in 2012.)

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

where $\Delta\mathbf{R}_\alpha(t)$ represents the displacement of atom α of protein from its equilibrium position, that is, "fluctuation." The equation has a form of the Langevin equation: the first, second and third terms in the right hand side, are a restoring force proportional to the displacement, a frictional force proportinal to the velocity, and a random force exerted by solvent, respectively. If one ignores the second and third terms, it reduces to an equation which looks like a "harmonic oscillator." The most intriguing aspect of the equation lies in the first term which has a restoring force proportional to the displacement. If it is the case in a harmonic oscillator in "vacuum," the restoring force is derived from the Taylor expansion of the potential energy with respect to the atom position, by ignoring the terms higher than the second order, based on an assumption of

"small oscillation." On the contrary, our equation does not include such an assumption. The restoring force was natually derived as a result of the ensemble average of the mechanical variables over the entire phase-space including both protein and solvent. And, the "force constant" gets the following expression in terms of the variance-covariance (or dsipersion) matrix ($\langle \Delta\mathbf{R}\Delta\mathbf{R} \rangle$) of the fluctuation.

$$k_B T\mathbf{L}^{-1} \equiv k_B T \langle \Delta\mathbf{R}\Delta\mathbf{R} \rangle^{-1} \tag{2}$$

The ensemble average in the variance-covariance matirx involves not only the coordinates of protein but also those of water molecules. From an analogy to the case of a harmonic oscillator in vacuum, we gave the following definition to the varinace-covariance matrix.

$$k_B T \left(\mathbf{L}^{-1}\right)_{\alpha\beta} = \frac{\partial^2 F(\{\mathbf{R}\})}{\partial\Delta\mathbf{R}_\alpha \partial\Delta\mathbf{R}_\beta} \tag{3}$$

In the equation, $F(\{\mathbf{R}\})$ is the free energy surface of protein, that can be defined by the following equation;

$$F(\{\mathbf{R}\}) = U(\{\mathbf{R}\}) + \Delta\mu(\{\mathbf{R}\}) \tag{4}$$

where $U(\{\mathbf{R}\})$ and $\Delta\mu(\{\mathbf{R}\})$ are the interaction energy among protein atoms and the solvation free energy, respectively. Considering that the equilibrium structure of a protein is a minimum point of the free energy surface, the definition for the *force constant* is rational. The free energy surface can be obtained from the RISM and 3D-RISM theories, to which the author has made significant contributions. Therefore, the *force constant* of the free energy surface can be calculated from Eq. (3).

II. Linear response theory

If the force constant in Eq. (1) is expressed in terms of the free energy by Eq. (3), the free energy should be expressed reversely in the quadratic function of $\Delta\mathbf{R}$ as

$$F(\{\Delta\mathbf{R}\}) = \frac{1}{2} k_B T \sum_{\alpha,\beta} \Delta\mathbf{R}_\alpha \cdot \left(\mathbf{L}^{-1}\right)_{\alpha,\beta} \cdot \Delta\mathbf{R}_\beta . \tag{5}$$

Suppose a perturbation is applied to the system. The structure of protein as well as its free energy should be changed due to the perturbation. The structural change can be predicted based on the variational principle as follows. Firstly, we add a energy change, proportinal to the perturbation (**f**), to the free energy surface.

$$F\left(\{\Delta\mathbf{R}\}\right)=\frac{1}{2}k_BT\sum_{\alpha,\beta}\Delta\mathbf{R}_\alpha\cdot\left(L^{-1}\right)_{\alpha,\beta}\cdot\Delta\mathbf{R}_\beta-\sum_\alpha\Delta\mathbf{R}_\alpha\cdot\mathbf{f}_\alpha \qquad (6)$$

Then, apply the variational priciple to Eq. (6) to find a linear responce expression describing structural change induced by the perturabtion. (The expression has been originally derived by Ikeguchi et. al. in a different way.)

$$\left\langle\Delta\mathbf{R}_\alpha\right\rangle_1=\left(k_BT_0\right)^{-1}\sum_\beta\left\langle\Delta\mathbf{R}_\alpha\Delta\mathbf{R}_\beta\right\rangle_0\cdot\mathbf{f}_\beta \qquad (7)$$

III. Structural response of protein to a thermodynamic perturbation [F. Hirata & K. Akasaka,*J. Chem. Phys.*, **142**, 044110 (2015)**]**

The (static) linear response theory, Eq. (7), can be applied to the structural change induced by a thermodynamic pertubation such as pressure, temperature, and denaturant. For examples, the structural change due to pressure can be described by the following equation.

$$\left\langle\Delta\mathbf{R}_\alpha\right\rangle=\left(k_BT_0\right)^{-1}\sum_\beta\left\langle\Delta\mathbf{R}_\alpha\Delta\mathbf{R}_\beta\right\rangle_0\left(\frac{\partial\Delta\bar{V}}{\partial\mathbf{R}_\beta}\right)_{P,T}P \qquad (8)$$

In the equation, P, $\Delta\bar{V}$, $\left\langle\Delta\mathbf{R}_\alpha\Delta\mathbf{R}_\beta\right\rangle_0$, and $\Delta\mathbf{R}_\alpha$ denote pressure (change), the partial molar volume of protein in water, the variance-covariance matrix of the structural fluctuation of protein concerning the unperturbed system, and the displacement or fluctuation of protein atom α from its equilibrium position. Eq. (8) implies the following physics for the pressure induced structural change of protein. The pressure (P) first induces the displacement in the position (or structure) of an atom β of protein so as to reduce the partial molar volume (Le Chatelier's law). The displacement is propagated to an atom α through the variance-covariance matrix $\left\langle\Delta\mathbf{R}_\alpha\Delta\mathbf{R}_\beta\right\rangle_0$ to

induce the displacment of α. Integrating the displacement over the contributions from all β atoms gives rise to $\langle \Delta \mathbf{R}_\alpha \rangle$.

The theory can be apllied to the computer-aided-drug-design (CADD) as a tool. A central issue in CADD is so called compound *screening*. It is a process to find a compound among as many as few hundred thousand compounds, that has the highest affinity to a target protein. A pin of the neck, that has been bothering scientists in the community, is the structural fluctuation of protein. As is emphasized in this report, conformation of a protein is fluctuating in our body to make a statistical ensemble around an equilibrium structure, the distribution of which is most likely a Gaussian. It is not nccessarily the equilibrium structure that has the highest affinity to a drug compound. It is very likely that one of *fluctuated structures* has higher affinity to a drug compound. Therefore, such a fluctuated structure should be regarded as a target for compound screening. However, it is not an easy task for conventional method. It is becuase a fluctuated structure is a "rare event" by definition, greater the fluctuation, harder for a coventinal method such as MD to sample the conformational space. The method proposed here can produce a structural fluctuation with any magnitude by controling an applied perturbation.

IV. **Dynamic linear response theory [**F. Hirata, *J. Chem. Phys.***, 145,** 234106 (2016)**]**

In the previous section, conformational changes of protein induced by a *static* pertubation, such as changes of thermodynamic variables, were described as an application of the theory of structural fluctuation. Here, a theory to describe the structural response of protein to a time-dependent perturbation is reveiewed briefly.

The time-dependent perturbation is defined by

$$H^{(1)}(t) = \theta(t)\left\{\delta F^S\left(\{\mathbf{R}\}\right) - \delta F^P\left(\{\mathbf{R}\}\right)\right\} \tag{9}$$

In the equation, $\delta F^P(\{\mathbf{R}\})$ and $\delta F^S(\{\mathbf{R}\})$ are the free energy of protein before and after the perturbation being applied, respectively. ($\delta F^S(\{\mathbf{R}\})$ is a non-equilibrium free energy surface) $\theta(t)$ is a step function defined by

$$\theta(t) = \begin{cases} 0 & t < 0 \\ 1 & t \geq 0 \end{cases} \tag{10}$$

That is, a perturbation corresponding to the non-equilibrium free energy is applied to the aqueous solution including protein at time t=0. Based on the linear response theory, the author has derived an equation to describe the structural response to the perturbation as follows.

$$\mathbf{R}_\alpha(t) = \mathbf{R}_\alpha(t=\infty) - \frac{1}{k_B T}\sum_\gamma \left\langle \Delta\mathbf{R}_\alpha(t)\Delta\mathbf{R}_\gamma \right\rangle_{eq}^{(0)} \cdot \mathbf{f}_\gamma(0) \tag{11}$$

where $\mathbf{f}_\gamma(0)$ is a force exerted on the protein due to the perturbation, defined by

$$\mathbf{f}_\gamma(t) = -\frac{\partial}{\partial \mathbf{R}_\gamma} H^{(1)}\left(\{\mathbf{R}\};t\right) . \tag{12}$$

Eq. (11) describes the time evolution of the structural relaxation of protein toward $\mathbf{R}_\alpha(t=\infty)$, which was $\mathbf{R}_\alpha(0)$ before t=0.

In the equation, the time evoulution of the variance-covariance matrix of the fluctuation ($\left\langle \Delta\mathbf{R}_\alpha(t)\Delta\mathbf{R}_\gamma \right\rangle_{eq}^{(0)}$) is included as a response function. This is concerned with the reference system, and it can be evaluated by the theory described above. Multipling both sides Eq. (1) by $\Delta\mathbf{R}_\gamma(0)$ and taking a statistcial average, one gets the following equation.

$$\frac{d}{dt}\begin{pmatrix} \mathbf{C}(t) \\ \dot{\mathbf{C}}(t) \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{1} \\ -\mathbf{K} & -\zeta \end{pmatrix}\begin{pmatrix} \mathbf{C}(t) \\ \dot{\mathbf{C}}(t) \end{pmatrix} \tag{13}$$

where **K** and ζ correspond, respectively, to the "force constant" (Eq. (3)) and friction cefficient matrices, and $\mathbf{C}(t)$ and $\dot{\mathbf{C}}(t)$ are defined by the following equations.

$$\mathbf{C}_{\alpha\gamma}(t) \equiv \left\langle \Delta\mathbf{R}_\alpha(t)\Delta\mathbf{R}_\gamma(0) \right\rangle, \quad \dot{\mathbf{C}}_{\alpha\gamma}(t) \equiv \left\langle \Delta\mathbf{V}_\alpha(t)\Delta\mathbf{R}_\gamma(0) \right\rangle \tag{14}$$

The formal solution of Eq. (13) can be obtained readily to give

$$\mathbf{C}(t) = \exp\left(\mathbf{A}t\right)\mathbf{C}(\infty)$$

where A is defined by the following equation.

$$\mathbf{A} = \begin{pmatrix} \mathbf{0} & \mathbf{1} \\ -\mathbf{K} & -\zeta \end{pmatrix}$$

The problem is thus reduced to calculate the parameters **K** and ζ. **K** can be obtained from the free energy surface of protein as the second derivative with respect to the atomic coordinates, based on Eq. (3). On the other hand, the author devised the following model for the friction coefficient matrix.

$$\zeta_{\alpha\beta} = \begin{cases} 0 & (\alpha \neq \beta) \\ f_\alpha \zeta_{\alpha,bulk} & (\alpha = \beta) \end{cases}$$

in which $\zeta_{\alpha,bulk}$ is the friction exerted on atom α in bulk solvent, and f_α is the fraction of the atom contacting with solvent. We define f_α by the following equation based on the radial distribution function (RDF) of solvent around the atom,

$$f_\alpha = \frac{g_w(\sigma; protein)}{g_w(\sigma; bulk)} \tag{56}$$

where $g_w(\sigma;bulk)$ is the RDF of water in bulk at the contact separation between the atom and a solvent molecule, and $g_w(\sigma;protein)$ is that corresponding to the atom in protein. If the atom is at surface of protein, f_α is close to unity, because the atom is well exposed to solvent, and $g_w(\sigma;protein)$ will become close to $g_w(\sigma;bulk)$. On the other hand, $g_w(\sigma;protein)$ will become small or zero if the atom is buried inside the protein, since there are no or few water molecules around such an atom. Both $g_w(\sigma;bulk)$ and $g_w(\sigma;protein)$ can be readily calculated based on the 3D-RISM/KH theory by making an appropriate definition for the contact separation σ between the protein atom and solvent. A typical choice of σ can be the position of the first peak in RDF in the bulk solvent.

It is a non-trivial problem to determine $\zeta_{\alpha,bulk}$ by experimental means, because an atom in protein has a partial charge in general, which is not the case if an atom is isolated by itself in solution: a charged atom in solution exists just in the form of an "ion" which of course has full charges, such as monovalent and divalent ions. Such an atom with a partial charge in solution is just an "imaginary" atom. We have proposed a recipe to determine the friction of such an imaginary atom in solution based on the site-site mode coupling theory for the dynamics of ions [7]. (Alternatively, the quantity can be estimated readily by the standard MD simulation for an ion with hypothetical *partial-charges* in solution.)

Concluding Remarks

Although the author proposed a concept concerning the structural fluctuation of protein, the concept is not specific just to a biomolecule, but applies more universally to the elasiticity of materials. There have been two physical concepts well regarded for the elasiticity of materials. [8] One of those is the reponse (deformation or strain) of a solid material to a perturbation (stress), in which the restoring force is proportional to the deformation, and the proportional constant, or force constant, is given by the second derivative of the potential energy with respect to the deformation. Such an elasiticity is referred to as "energy elasiticity." The other elasiticity is the "rubber elasiticity." In this case, too, the restoring force is proportional to the magnitude of deformation. However, the origin of the force is not energy, but entropy. A rubber, which is a polymer, can take a large number of possible conformations with different dihedral angles among atoms, or a state of larger entropy, when it is a shrinked state. But, it can take only one conformation, or a state of the least entropy, when it is fully stretched. So, a rubber tends to restore spontaneously the state of larger entropy (a shrinked state) from that of less entropy (a stretched sate). Such an elasiticity is called "entropy elesiticity."

The concept described in this report is different from either of the conventional physics concerning elesiticity. It is something like an *elasiticity (or linearity) induced by solvent.* The author refers to the elasiticty as "solvent induced elasiticity" or "free energy elasiticity."

Recently, the author has found an interesting phenomenon, which may or may not be related to the new concept of elasiticity. The author wears contact-lenses everyday, that are supposed to be stocked in saline solution over night. One night, he just threw them away unconsciously into a garbage can, and found them next morning in miserable shape, which was entirely dried out. Unfortunately, he did not have extra lenses at that time, thereby he immersed the lenses in saline solution, hoping that it would recover the original shape and function. After few minutes, the lenses recovered their shape and function completely. If one folds a contact lense with fingers in water, the lense is largely distorted. However, the lense restores its original shape completely after the pressure is removed. On the other hand, when the lense is dried out, it will be clashed if one folds it with fingers, and the original shape will never be restored. The phenomenon is akin to the case of protein discussed in the review, suggesting that a

same natural law is govering the both phenomena.

References

[1] N. Go, T. Noguchi, and T. Nishikawa, Proc. Natl. Acad. Scie., USA, 80, 3696 (1983).

[2] B. Brooks and M. Karplus, *Proc. Natl. Acad. Sci*. USA, **80**, 6571 (1983).

[3] K. Akasaka, *Chem. Rev*., **106**, 1814 (2006).

[4] M. Ikeguchi, J. Ueno, M. Sato, and . Kidera, *Phys. Rev. Lett.*, 94, 078102 (2005).

[5] R. Kubo, J. Phys. Soc. Jpn, 12, 570 (1957).

[6] H. L. Friedman, "*A course in Statistical Mechanics*," (Prentice Hall, New Jearsey, 1985).

[7] S-H. Chong and F. Hirata, *J. Chem. Phys*., **108**, 7339 (1998); **111**, 3654 (1999).

[8] T. Nakagawa, "Rheology," Iwanamizensho. (Japanese).

［Pulication list in 2016］

1. Jiraphorn Phanich, Thanyada Rungrotmongkol, Daniel Sindhikara, Saree Phongphanphanee, Norio Yoshida, Fumio Hirata, Nawee Kungwan, and Spot Hannongbua, "A 3D-RISM/RISM Study of the Oseltamivir Binding Efficiency with the Wild-type and Resistance-Associated Mutant Forms of the Viral Influenza B Neuraminidase," Protein Science, 25, 147 – 158 (2016)

2. Fumio Hirata, Bogsoo Kim, "Multi-scale dynamics simulation of protein based on the generalized Langevin equation combined with 3D-RISM theory," *J. Mol. Liquids*, 217 23-38 (2016).

3. Myroslav Holovko, Andriy Kovalenko, and Fumio Hirata, "Partial molar volume of nonionic surfactants in aqueous solution studied by the KB/3D-RISM-KH theory," *J. Mol. Liquids*, 217, 103 – 111 (2016)

4. Masatake Sugita and Fumio Hirata, "Predicting the binding free energy of the inclusion process of 2-hydroxypropyl-b-cyclodextrin and small molecules by means of the MM/3D-RISM method," *J. Phys.:Condens. Matter*, **28**, 384002-384013 (2016)

5. Fumio Hirata, “A molecular theory of the structural dynamics of protein induced by a perturbation,” *J. Chem. Phys*., **145,** 234106 (2016); doi: 10.1063/1.4971799

［**Book Chapter**］

Fumio Hirata, Norio Yoshida, and Bongsoo Kim, “Theory of Molecular Recognition and Structural Fluctuation of Biomolecules,” in “Molecular Science of Fluctuations Toward Biological Functions,” eds., M. Terazima, M. Kataoka, R. Ueoka, and Y. Okamoto, Springer, 2016.