# Developing the Statistical Mechanics Theory of Liquids in Chemistry and Biophysics

**Department of Theoretical and Computational Molecular Science**
**Division of Theoretical Molecular Science II**

| | |
|---|---|
| HIRATA, Fumio | Professor |
| CHONG, Song-ho | Assistant Professor |
| YOSHIDA, Norio | Assistant Professor |
| MARUYAMA, Yutaka | Post-Doctoral Fellow |
| MIYATA, Tatsuhiko | Post-Doctoral Fellow |
| ISHIZUKA, Ryosuke | Post-Doctoral Fellow |
| PHONGPHANPHANEE, Saree | Post-Doctoral Fellow |
| KIYOTA, Yasuomi | Graduate Student |
| SUETAKE, Yasumi | Secretary |
| KONDO, Naoko | Secretary |
| YAMADA, Mariko | Secretary |

We have been developing a new theory for the molecular recognition by protein based on the statistical mechanics of liquids, or the 3D-RISM/RISM theory. The theory has demonstrated its amazing capability of "predicting" the process from the first principle.[1–3] However, what we have investigated so far is an entirely equilibrium process both in protein conformation and solvation.

Recently, we have started to incorporate the conformational fluctuation of protein into the molecular recognition process in two ways. The first of those is a "static" one in which we just shake the protein conformation to find the local minimum of the free energy surface by the combined 3D-RISM/RISM with conformational sampling algorithms, and to see if one can find the distribution of a guest molecule in the recognition site.[4] The other method is to take the "dynamic" fluctuation of protein conformation into account. The process can be described by hybridized 3D-RISM/RISM with the generalized Langevin theories.[5] The methodology is currently under construction, and some prospective view of the theory will be presented in the lecture.

## 1. 3D-RISM/RISM Studies on the Dissociation Pathway of CO in Myoglobin[6]

Myoglobin (Mb) is a globular protein which has important biological functions, or oxygen storage. Due to its biochemical function, many researchers have made intensive efforts to identify the escaping pathway of the ligand, experimentally and theoretically.[7]

Among many unresolved questions with respect to the CO escaping pathway, the dependence of the pathway on Xe concentration has recently been highlighted by Terazima based on the time-resolved partial molar volume (PMV) measured with the transient grating spectroscopy. The experimental results indicate that CO is trapped at the Xe site before escaping to solvent at room temperature. The difference of Xe and CO affinity of each Xe trapping site makes the CO escaping pathway different depending on Xe concentration. It is believed that the dissociated CO escapes to the solvent through the Xe1 trapping site under the Xe-free condition, dominantly, while CO escapes through the Xe4 site in a Xe-rich solution.

Shown in Figure 1 is the 3D-distribution of CO in the four Xe-sites calculated from 3D-RISM in the ternary mixture composed of water, Xe, and CO. It is apparent that each Xe-site has some affinity to a CO molecule. The comaprison of the population of CO and Xe in each Xe-site at high concentration revealed that Xe has greater affinity to the Xe1 site than CO, while CO is dominant in the Xe4-site.
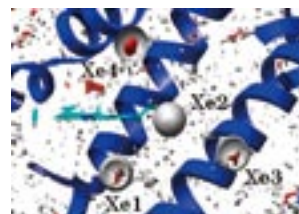


**Figure 1.** CO distribution in the Xe-sites: red, O; gray, C.

**Table 1.** The partial molar volume of the different association state of the CO-Mb complex. ($\Delta V_1$ = (Mb:CO) − (MbCO), $\Delta V_2$ = (Mb+CO) − (Mb:CO), and $\Delta V_{total}$ = (Mb+CO) − (MbCO)).

| models | PMV[$cm^3$/mol] |
|---|---|
| MbCO | 9029.0 |
| Mb:CO(Xe1) | 9030.4 |
| Mb:CO(Xe4) | 9032.8 |
| MB+CO | 9019.6 |

| | Xe1 | Xe4 | (exptl.[7]) |
|---|---|---|---|
| $\Delta V_1 =$ | 1.4 | 3.8 | (3±1) |
| $\Delta V_2 =$ | −10.8 | −13.2 | (−12.6±1.0) |
| $\Delta V_{total} =$ | −9.4 | −9.4 | (−10.7±0.5) |

(MbCO, CO bound at the Heme; Mb:CO(Xe1), CO is bound at the Xe1 site: Mb:CO(Xe4), CO is bound at the Xe4-site; MB+CO, CO is dissociated from Mb.)

In order to examine the hypothesis made by Terazima with respect to the escaping pathway, we have calculated PMV of the CO-Mb complexes in each of which a CO molecule is explicitly bound in one of the Xe-sites. The results are shown

in Table 1. As is obvious from the table, the theoretical results for Mb:CO(Xe4) show much better agreement with the experiment, indicating that CO escapes from the Xe4-site. The hypothesis made by Terazima is thus verified by the theory.

## 2. An Attempt toward the Generalized Langevin Dynamics Simulation[5]

It has been five decades since the molecular simulation scored its first step in the study of liquids and solutions. Accelerated by the increasing power of computer, the method has been enjoying the status of a standard tool to explore the molecular aspects of physical, chemical, and biological processes in liquids and solutions. However, the method is facing high barriers which may not be overcome by the improvement of computing power alone. One of those is the large and slow fluctuations taking place in a protein in solutions, which touches the zero wavelength and frequency limits. Straightforward applications of the molecular simulation to the limits are in danger to end up with an "animation" or a "science fiction."

In this report, we have attempted a new step toward the molecular dynamics simulation which is not based on the Newton equation, but on the generalized Langevin theory, in which all the degrees of freedom concerning the solvent molecules are "coarse-grained" or "projected" in term of the pair correlation functions. Choosing the coordinates $R$ of protein atoms and the density field $\rho$ of solvent atoms, as well as their conjugated momentum, as dynamic variables in the phase space, we have derived the generalized Langevin equations for protein dynamics in solutions.

$$\frac{d\mathbf{R}(t)}{dt} = \frac{\mathbf{P}(t)}{M}$$

$$\frac{d\mathbf{P}(t)}{dt} = -k_B T \mathbf{L}^{-1} \cdot \mathbf{R}(t) - \frac{1}{Mk_B T} \int_0^t ds \left\langle \mathbf{W} e^{i(t-s)QL} \right\rangle \cdot \mathbf{P}(s) + e^{itQL} \mathbf{W}$$

$$\frac{d\rho_k(t)}{dt} = ik J_k^L(t)$$

$$\frac{dJ_k^L(t)}{dt} = \frac{ikk_B T}{mS_k} \delta\rho_k(t) - \frac{m}{Nk_B T} \int_0^t \left\langle R_{-k} e^{i(t-s)QL} R_k \right\rangle J_k^L(s) + e^{itQL} f_k$$

where $R$ and $P$ denote the coordinates and conjugated momenta of protein atoms, $\rho$ and $J$ the density and momentum field of solvent atoms, respectively. The first two equations describe the dynamics of protein, while the last two are concerned with the solvent dynamics. However, they should be solved simultaneously, since the equations are closely coupled each other. It is our plan to solve the equation in the way just as is done in the molecular dynamic simulation, or the numerical integration of the equation.

The equations for protein have a typical expression of the original Langevin equation, but each term in the right hand side has a microscopic description in contrast to the original one. The first term is related to the variance-covariance matrix of mean square displacement of protein, which signifies the conformational fluctuation of the molecule. The factor $k_B T L^{-1}$ is related to a frequency matrix of the fluctuation, or the variance-covariance matrix ($<\Delta R \Delta R>$)), diagonalization of which gives rise to an "effective normal mode" of the fluc-

tuation. The second term is the damping or drag term. In the original Langevin equation, this term is local in time, and is described by the phenomenological expression such as the Oseen tensor which models the hydrodynamic interactions. The third term stands for the "random" force acting on the solute, which of course is orthogonal to the dynamic variables at time zero.

It is a highly nontrivial problem to solve the equation by means of the numerical integration. Our working hypothesis to solve the equations is following.
(1) For each time step $\Delta t$, the protein structure is at "local equilibrium," and the fluctuation around the equilibrium follows the "central limiting theorem," or the "Gaussian fluctuation."
(2) By virtue of the Gaussian fluctuation, the variance-covariance matrix ($<\Delta R \Delta R>$) can be the second derivative of the free energy of protein (conformational energy + solvation free energy) with respect to protein atom positions.
(3) The solvation free energy of protein and its derivatives can be evaluated from the 3D-RISM theory. (The variance-covariance matrix ($<\Delta R \Delta R>$) is proportional to the second derivative.
(4) The time step $\Delta t$ of integration should be sufficiently large so that the central limiting theorem is valid, while it should be small enough so that the memory term can be approximated with a short time memory.
(5) The perturbation onto solution from protein can be treated by a renormalized potential or the direct correlation functions.

There is another concern in the actual implementation of the theory to the computational science. Each time step of the numerical integration requires the solution of the 3D-RISM equation, which has been notorious with respect to the computational time. For example, when the equation was first solved for protein about five years ago, it took "a month" by ordinary workstations to get the 3D-distribution around "one" conformation of the solute. It was essentially due to the 3D-FFT program which is notorious about the parallelization. If it is the case, the new approach described above is useless, because the simulation requires thousands of steps to explore a meaningful region of the conformational space. However, we could have made a dramatic progress during the year of 2009, thanks to the collaboration with the computer scientists in Tsukuba: The current achievement is "a minute per a conformation" of protein with T2K in Tsukuba.

## References

1) F. Hirata, *Molecular Theory of Solvation*, Kluwer; Dordrecht, Netherlands (2003).

2) A. Kovalenko and F. Hirata, *J. Chem. Phys.* **110**, 10095–10112 (1999).

3) T. Imai, R. Hiraoka, A. Kovalenko and F. Hirata, *J. Am. Chem. Soc. (Communication)* **127**, 15334–15335 (2005).

4) T. Miyata and F. Hirata, *J. Comput. Chem.* **29**, 872 (2007).

5) B. Kim, S-H. Chong, R. Ishizuka and F. Hirata, *Cond. Matt. Phys.* **11**, 179 (2008).

6) Y. Kiyota, R. Hiraoka, N. Yoshida, Y. Maruyama, T. Imai and F. Hirata, *J. Am. Chem. Soc. (Communication)* **131**, 3852 (2009).

7) M. Sakakura, S. Yamaguchi, N. Hirota and M. Terazima, *J. Am. Chem. Soc.* **123**, 4286 (2001).