Coarse-grained Simulation of Azurin Crystal Complex System: Protein–protein Interactions

Micke Rusmerryani[a], Masako Takasu[b], Kazutomo Kawaguchi[a], Hiroaki Saito[a], Hidemi Nagao[a]

[a]Institute of Science and Engineering, Kanazawa University, Kakuma, Kanazawa 920-1192 Japan, [b]School of Life Sciences, Tokyo University of Pharmacy and Life Sciences, Hachioji, Tokyo 192-0392, Japan,
E-mail: micke@wriron1.s.kanazawa-u.ac.jp, nagao@wriron1.s.kanazawa-u.ac.jp

Abstract. *Most of protein function analyses focus mainly on the physical properties of a single protein. Nevertheless, the environments where proteins perform their biological functions are crowded with macromolecules, such as lipid, nucleic acids, and other proteins. The interactions between macromolecules may be affected by molecular crowding. Therefore, as an initial step we here investigate the protein–protein interactions for gaining insights into molecular crowding effects on protein conformational changes. Computational molecular simulation is one of the useful and important tools to study the protein interactions. Here we develop a coarse-grained model and a topology-based potential interactions to simulate dynamical properties of multiprotein complex crystal structure. We apply them to simulate complex crystal structure of* Pseudomonas Aeruginosa *azurin, a small cupredoxin, which functions as an electron carrier in bacterial respiration. Since electron transfer on azurin plays an important role in the biological system, it is important to characterize the protein interactions in azurin. In our simulation, the interactions between intra- and inter- domains are treated at the residue level with the implementation of the off lattice Gō-like model. In each domain, bonded interactions between residues are described by bond stretching, bond angle bending, and torsional angle potentials. The non-bonded interactions, which are represented by short range and long range potentials, describe the interactions both among residues and between proteins. We probe the protein–protein interactions by analyzing the protein binding. A simple clustering algorithm is applied to group the bound structures of protein complex. Moreover, we can investigate the importance of the long range interaction on the multiprotein complex system. These studies will serve as valuable insights for further investigation on molecular crowding effects.*

Keywords: azurin, protein–protein interaction, coarse-grained, Gō model, crowding

1 Introduction

Protein plays important roles in regulation of biological function in cells. Most biophysical studies of protein dynamics have been carried out in dilute solution in the test tubes. They assume that the proteins in the test tubes behave similarly as in a living cell. In fact, cells exist in a crowded environments where macromolecules such as nucleic acids, lipids, and proteins can take up to 40% of the total cellular volume[1].

However, it is very difficult to study the dynamics of macromolecules in crowded system experimentally. Computational studies may overcome this problem since it can provide the mechanistic explanation about the molecular dynamics up to the atomistic level. Recently, many computational studies have observed the behavior of protein by the presence of the crowding agents. Previous views of molecular crowding focused mostly on the volume exclusion by the presence of an inert spherical crowding agent, such as Ficoll 70 or Dextran 70[2, 3].

Nevertheless, the change in intermolecular interactions may be caused also by the presence of other proteins[4, 5]. Previous studies have found that the volume occupied by proteins affects the activity of the other proteins in solution. Thus the presence of 10% or more protein in a solution will tend to force proteins into compact configurations [6]. Motivated by their studies, it will be interesting to investigate the effects of crowding by the proteins as the crowding agent. We predict that the protein in the crowded environments will be more native-like along with the less space for the movement of the protein.

In this study we develop an approach to model the protein complex as a crowded system. In this approach, we implement a coarse-grained model and a topology-based potential interactions to simulate dynamical properties of protein complex. We focused on crystal structure of azurin, a small 128-residue cupredoxin which is composed of eight β-strands and one helix. Here, in order to further investigate the protein–protein interactions, the protein dynamics in an independent system is compared with the dynamics in the presence of the identical protein crowders. Steric repulsion, which is one of the intermolecular interactions, is also present in the multi-protein complex system. It will be helpful for gaining insights into molecular crowding effects on protein conformational changes.

2 Methods

2.1 Model

The dynamical properties of azurin complex system was studied here to gain an overview of the effects of protein crowders. Azurin, one of blue copper protein, is a small 128-residues cupredoxin which is composed of eight β-strands and one helix. Its conformation forms a rigid β-barrel so that the azurin belongs to sandwich-like protein family [7]. In this study, the native structure of azurin complex system was obtained from X-ray crystal structure[1] of *Pseudomonas Aeruginosa* azurin (PDB ID: 4AZU) [8]. In this crystal structure, the unit cell[2] consists of one asymmetric unit[3], where its asymmetric unit is composed of a tetramer of azurin molecules.

From the native structure consisting of four identical chains, we build four systems consisting of different number of crowding agents. Our model systems are shown in Figure 1. The black chain represents a single protein chain that we choose as the main protein. Then the other proteins (shown in grey chains) act as the crowding agents to the single protein chain. First system represents the independent system where the distance between those two chains is more than the cutoff. The other systems have the interacted chains which contain two, three, or four chains, respectively.

2.2 Simulations

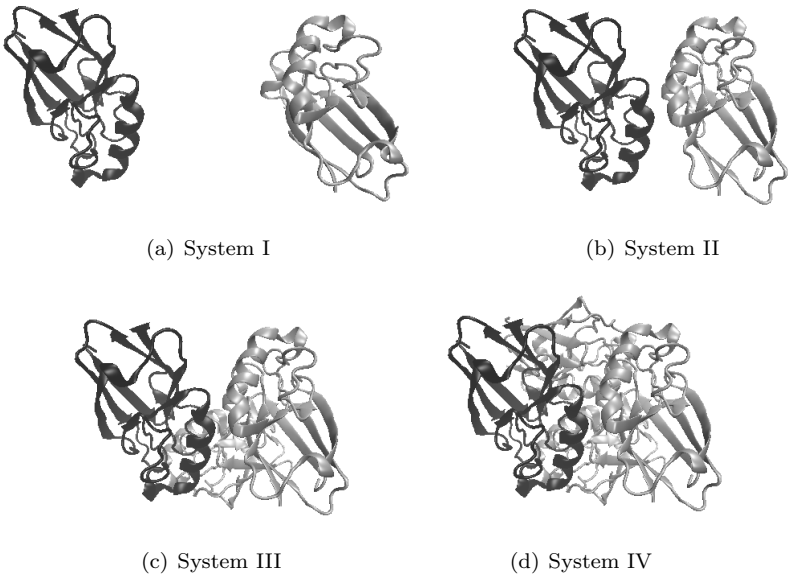
We carried out coarse-grained simulation and implement a topology-based potential interactions to simulate each configuration system. The potential energy for the entire system is:

$$U = E_p + E_{pp},$$

[1] **X-ray crystal structure** is a solid composed of repeated structural motifs in a three-dimensional lattice that generated by the scattering of X-rays, usually with some sort of internal rotational symmetry.

[2] **Unit cell** is the smallest unit that can generate the whole crystal by only translational symmetry operations, which enables compact packing of molecules in the crystal.

[3] **Asymmetric unit** is the smallest portion of the crystal structure to which rotation and translation using only the symmetry operators allowed by the crystallographic symmetry can be applied to generate one unit cell. It might contain one or more biological molecules.



(a) System I (b) System II

(c) System III (d) System IV

Figure 1: Initial configurations; system II, III, and IV are obtained from the crystal structure of tetramer azurin, while system I is obtained by separate two chains in system II so that those two chains have no intermolecular interaction.

where E_p and E_{pp} represent the potential energy for protein and the interactions for protein-protein, respectively. We applied the off-lattice Gō model for the potential energy of protein to mimic the perfect funnel aspect of folding energy landscape [9, 10]. The following equation represents the Gō model interaction energy at a configuration Γ of a protein and Γ_0 as its native state that we use in this study[11, 12, 13, 15]:

$$\begin{aligned} E_p(\Gamma,\Gamma_0) = & \sum_{\text{bonds}} \frac{1}{2}K_r(r-r_0)^2 \\ & + \sum_{\text{angles}} \frac{1}{2}K_\theta(\cos(\theta)-\cos(\theta_0))^2 \\ & + \sum_{\text{dihedral}} K_\phi^{(n)}[1-\cos\left(n\times(\phi-\phi_0)\right)] \\ & + \sum_{i<j-3}^{\text{native}} \varepsilon_1(i,j)\left[5\left(\frac{r_{0ij}}{r_{ij}}\right)^{12}-6\left(\frac{r_{0ij}}{r_{ij}}\right)^{10}\right] \\ & + \sum_{i<j-3}^{\text{non-native}} \varepsilon_2(i,j)\left(\frac{C}{r_{ij}}\right)^{12}. \end{aligned} \tag{1}$$

Meanwhile, the interactions between proteins are non-bonded interactions. The potential is given by:

$$E_{pp}(\Gamma,\Gamma_0) = \sum_{p=1}^{p=M-1}\sum_{q=p+1}^{q=M}\sum_{i,j}^{\text{native}} \varepsilon_3(i,j)\left[5\left(\frac{r_{0i_pj_q}}{r_{i_pj_q}}\right)^{12} - 6\left(\frac{r_{0i_pj_q}}{r_{i_pj_q}}\right)^{10}\right] + \sum_{p=1}^{p=M-1}\sum_{q=p+1}^{q=M}\sum_{\text{non-native}} \varepsilon_4(i,j)\left(\frac{C}{r_{i_pj_q}}\right)^{12}, \tag{2}$$

where M is number of chains.

In equation 1, $r(r_0)$, $\theta(\theta_0)$, and $\phi(\phi_0)$ represent the distance between two adjacent residues, the angles formed by the subsequent residues, and the dihedral angle defined by four subsequent residues along the chain at the configuration $\Gamma(\Gamma_0)$. The last two terms in equation 1 correspond to non-local attractive native interactions and a short-range repulsive between non-native pairs, where $r_{ij}(r_{0ij})$ represents the distance between ith and jth amino acid residues at the configuration $\Gamma(\Gamma_0)$. The initial distances are obtained from the tetramer crystal structure of azurin. Since modern crystal structure of protein mostly add water molecules to the structures[14], we should consider the hydrogen bonding that exist between protein and water molecules. Then we use the 10-12 potential in our model as the attractive potential rather than Lennard-Jones potential because this potential is believed to correspond to the hydrogen bonds better.

In our simulation, we use $K_r = 100.0$ (kcal/mol/Å^2), $K_\theta = 20.0$ (kcal/mol), $K_\phi^{(1)} = 1.0$ (kcal/mol), $K_\phi^{(3)} = 0.5$ (kcal/mol), $\varepsilon_1 = 0.3$ (kcal/mol), $\varepsilon_2 = 0.2$ (kcal/mol), $C = 4.0$ Å as the parameters for all the systems studied [15, 16, 17]. The same parameters for the non-bonded interactions are used in the equation 2. Using these model, we simulate the dynamics of azurin complexes starting from its initial configuration as the native structure. The temperature during simulation time is controlled by Langevin dynamics with the friction coefficient, $\gamma = 0.25$ (τ^{-1}) where τ, a time unit, is 0.2 ps. We also removed the center of mass velocity during the simulation while we let each chain to move freely[18, 19].

2.3 Analysis

In order to investigate the importance of the intermolecular interactions among the chains, we compare the dynamical properties between system I which represents the independent system and system II which represents the interacted system. Furthermore, the crowding effects also can be observed by comparing the dynamical properties of all systems. As the dynamical properties, we monitor the fluctuation of the distance between the center of mass of two representative chains. Then, we also consider the autocorrelation of the distance to observe the tendency for the systems whether the system remains in the same state from one observation to the next. We calculate the autocorrelation by the sufficient statistical average of the time series of the distance ($D(t)$) as following equation:

$$R(\tau) = \frac{\langle (D(t) - \langle D\rangle)\cdot(D(t+\tau) - \langle D\rangle)\rangle}{\langle (D(t) - \langle D\rangle)\rangle^2}, \tag{3}$$

where τ is time lag.

We also compare the thermodynamical property by calculating the free energy profile. In this study, the free energy profile is obtained by the histogram method[20] as a function of reaction coordinate Q to compare the nativeness of all systems. The Q itself is a measure of nativeness

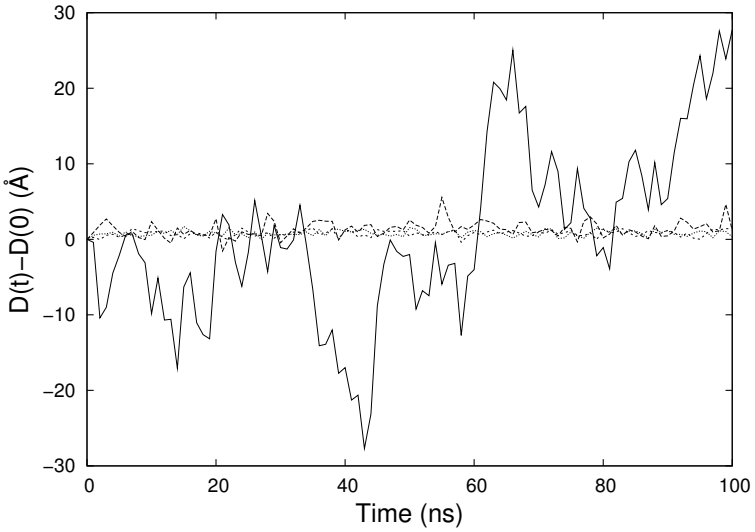


Figure 2: Distance between the centers of mass of two representative chains. The solid line represents the interchain distance for independent system. The long-dashed line, dashed line, and short-dashed line represent the interchain distance for interacted system: system II, system III, and system IV, respectively.

Table 1: Standard deviation of the distance between the centers of mass amongst the chains

System	chain A-B	chain A-C	chain A-D
System I	11.448	-	-
System II	0.944	-	-
System III	0.492	0.329	-
System IV	0.39	0.243	0.441

[10], defined as the fraction of the native contacts formed in a given conformation. The value of Q ranges between 0 to 1. A Q score close to unity means that the conformation is similar to the native structure. Otherwise, a Q-score close to zero means that the conformation is dissimilar to the native structure.

3 Results and Discussions

In this study, we applied a residue level C_α model to represent the structures of azurin. We applied Gō -like model and performed coarse-grained simulation of azurin complex crystal system. During the simulation, the dynamics of the protein have been monitored by calculating the interchain distance. Two chains, called chain A and B, are chosen as the representative chains which are being compared among all systems. From the viewpoint of these two representative chains, we calculated and compared the distance between the centers of mass as shown in Figure 2.

By comparing the dimer system, system I (solid line) and system II (long-dashed line), in Figure 2, we can see that the system which has the intermolecular interactions is less fluctuating than the system which has no interaction amongst the chains. It shows that the intermolecular interactions

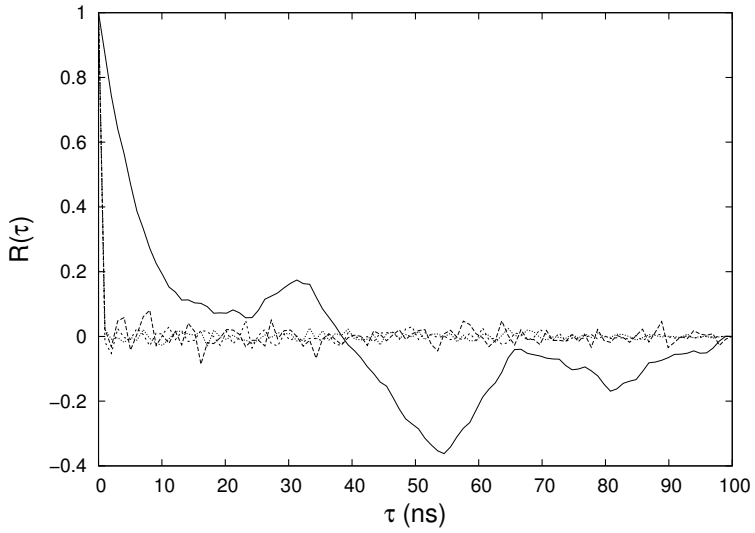


Figure 3: Autocorrelation of distance between two representative chains as a function of time lag (τ). The solid line, long-dashed line, dashed line, and short-dashed line represent system I, system II, system III, and system IV, respectively.

affect the dynamics of proteins. Since we let each chain to move freely, the chain may move toward or away other chains. When we have independent system, then the distance between two centers of mass becomes more fluctuating. Otherwise, in system II, two chains were placed close enough to have intermolecular interactions. The residue in chain A may have attractive interaction with the residue in chain B. Yet, the steric repulsion interactions between these chains prevent the chains from collapsing. Then the combination of the attractive and repulsive interactions gave us less movement of each chain.

Moreover, we can also compare the dimer, trimer, and tetramer system. Figure 2 shows that as the increasing number of chains, the fluctuation of distance between the centers of mass become smaller. It is also confirmed by the standard deviation on Table 1. The trimer system is less fluctuating than the dimer system and so the tetramer is less fluctuating compared with the trimer system. This similar tendency shows the importance of the long range interactions to the stability of the azurin complex system. According to this result, the autocorrelation of the distance (Figure 3) also show that the more crowded the system, the less fluctuation and the faster to reach the relaxation time.

Furthermore, we also monitored the thermodynamic configurations of the main protein as a function of the reaction coordinate (Q) which measures the nativeness. By using this reaction coordinate, we obtained the free energy profile ($F(Q)$) as shown in Figure 4. Here we compared the nativeness of chain A as the main protein. We neglected the nativeness of other chains since we treated them as the crowding agents. From Figure 4, we can see that system II is more native-like than system I. Furthermore, the more crowded the system, it becomes more native-like as our expectation. In the trimer and tetramer system, each residue will have less space for the movement. Therefore, the configuration becomes more compact as the number of the crowding agents increases. Since the same crystal structure of azurin are used in the systems, the results are very natural and show that the tetramer system has the most compact and most native-like

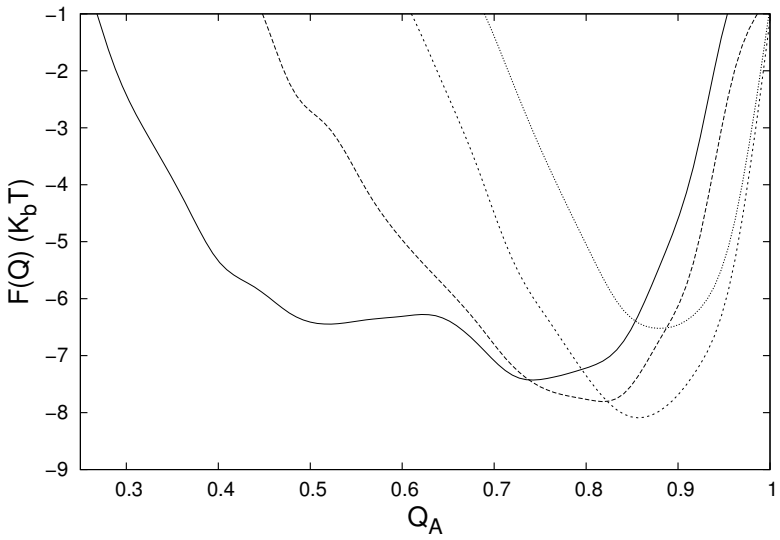


Figure 4: Free energy profile as a function of reaction coordinate Q. The solid line, long-dashed line, dashed line, and short-dashed line represent the free energy profile for simulation of system I, system II, system III, and system IV, respectively.

structure.

4 Conclusions

Since the distance between the centers of mass and the measure of nativeness are in a good agreement, the protein–protein interactions play an important role in the protein complex system. From these studies we found that increasing the number of the crowding agents affects the protein movement. Combination of the attractive and repulsive interactions gave us less fluctuation of the interchain distance. In addition, the increasing the number of the protein crowders also impacts the nativeness of the protein. As we expected before, the more crowded the system, the more native protein will become. It is very clear that the tetramer system, the unit cell of crystal structure of azurin (PDB ID: 4AZU), has the most compact and most native-like structure. So, this model can be expanded by generating the unit cell to a larger crystal system. Further studies with higher resolution of the potential model may also give us broader insights into the molecular crowding effects.

References

[1] A. P. Minton (2006). How can biochemical reactions within cells differ from those in test tubes? *J. Cell. Sci.*, **119**, 2863 – 2869.

[2] D. Homouz, L. Stagg, P. W. Stafshede, and M. S. Cheng (2009). Macromolecular crowding modulates folding mechanism of α/β protein apoflavodoxin. *Biophys. J.*, **96**, 671 – 680.

[3] D. L. Pincus and D. Thirumalai (2009). Crowding effects on the mechanical stability and unfolding pathways of ubiquitin. *J. Phys. Chem. B*, **113**, 359 – 368.

[4] A. V. Predeus, et. al. (2012). Conformational sampling of peptides in the presence of protein crowders from AA/CG-multiscale simulations. *J. Phys. Chem. B*, **116**, 8610 – 8620.

[5] Y. C. Kim and G. Hummer (2008). Coarse-grained models for simulations of multiprotein complexes: application to ubiquitin binding. *J. Mol. Biol.*, **375**, 1416 – 1433.

[6] A. B. Fulton (1982). How crowded is the cytoplasm? *Cell.* **30**, 345 – 347.

[7] G. E. Norris (1982). *The three dimensional structure of azurin, a blue copper protein, at 3Å resolution*, Massey University, New Zealand.

[8] H. Nar, et. al. (1991). Crystal structure analysis of oxidized *Pseudomonas Aeruginosa* azurin at pH 5.5 and pH 9.0. *J. Mol. Biol.*, **221**, 765 – 772.

[9] N. Gō (1983). Theoretical studies of protein folding. *Ann. Rev. Biophys. Bioeng.*, **12**, 183 – 210.

[10] J. N. Onuchic, Z. L. Schulten, and P.G. Wolynes (1977). Theory of protein folding: the energy landscape perspective. *Ann. Rev. Phys. Chem.*, **48**, 545 – 600.

[11] C. Clementi, H. Nymeyer, and J.N. Onuchic (2000). Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.*, **298**, 937 – 953.

[12] H. Bekker (1996). *Molecular dynamics simulation methods revised*, Proefschrift, Groningen.

[13] M. Griebel, et al. (2007). *Numerical simulation in molecular dynamics*, Springer, Germany.

[14] K. R. Acharya and M. D. Lloyd. (2005). The advantages and limitations of protein crystal structures. *TRENDS in Phar. Sci.*, **26**, 10 – 14.

[15] M. Rusmerryani, et al. (2013). Transition state analysis of azurin *via* Gō-like model. *AIP Conf. Proc.*, **1518**, 641 – 644.

[16] H. Kenzaki, et al. (2011). CafeMol: a coarse-grained biomolecular simulator for simulating proteins at work. *J. Chem. Theory Comp.*, **7**, 1979 – 1989.

[17] W. Li, P. G. Wolynes, and S. Takada (2011). Frustration, specific sequence dependence, and nonlinearity in large-amplitude fluctuations of allosteric proteins. *PNAS* **108**, 3504 – 3509.

[18] G. A. M. Maldonado, et al. (2012). On the centre of mass velocity in molecular dynamics simulations. *Revista Mexicana de Fisica*, **58**, 55 – 60.

[19] M. Guenza (2002). Intermolecular effects in the center-of-mass dynamics of unentangled polymer fluids. *Macromolecules*, **35**, 2714 – 2722.

[20] S. Cheluvaraja and H. Meirovitch (2006). Calculation of the entropy and free energy of peptides by molecular dynamics simulations using the hypothetical scanning molecular dynamics method. *J. Chem. Phys.*, **125**, 024905.