**Chem-Bio Informatics Journal, Vol. 8, No. 1, pp.14-24(2008)**

# Solvent Site-Dipole Field Accompanying Protein-Ligand Approach Process

Norikazu Takano[1], Koji Umezawa[1], Jinzen Ikebe[1], Yuki Sonobe[1], Ryosuke Yagisawa[1], Jun-ichi Ito[1], Nobuyuki Hamasaki[1], Daisuke Mitomo[1], Hiroh Miyagawa[2], Akihiko Yamagishi[1], Junichi Higo[1,3,*]

*[1]School of Life Sciences, Tokyo University of Pharmacy and Life Sciences, 1432-1 Horinouchi, Hachioji, Tokyo, 192-0392, Japan.*
*[2]Taisho Pharmaceutical Co. Ltd, 1-403 Yoshino, Omiya, Saitama, 330-8530, Japan.*
*[3]The Center for Advanced Medical Engineering and Informatics, Osaka University, Open Laboratories for Advanced Bioscience and Biotechnology, 6-2-3, Furuedai, Suita, Osaka 565-0874, Japan*

**E-mail: higo@protein.osaka-u.ac.jp*

(Received January 16, 2008; accepted February 9, 2008; published online March 12, 2008)

## Abstract

**We did a molecular dynamics simulation of a system consisting of a peptide and a protein in explicit solvent to study biomolecular approach process. In the initial structure of simulation, the minimum inter-biomolecular distance was 30 Å. During the simulation, the biomolecules approached and contacted to each other. In spite of diffusive motions of water molecules, the orientations of water molecules tended to order in the inter-biomolecular zone showing coherent spatial patterns (solvent site-dipole field) of the ordering. The degree of ordering was synchronized well with the inter-biomolecular distance. This result strongly suggests that the biomolecules distant to each other can interact via the solvent site-dipole field. The effective range for the coherent ordering (i.e., the interaction range via the solvent site-dipole field) was larger than 20 Å. A bridge-like structure of the solvent orientational ordering connected the two biomolecules. Biological and physicochemical significance of the ordering is discussed.**

**Key Words:** Hydration, Molecular binding, Molecular recognition, Molecular interaction, Molecular dynamics simulation.

**Area of Interest:** Molecular Computing

# 1. Introduction

Function of biological molecules is exerted in solvent. Biomolecular binding consists of two processes: approach and fitting. At the beginning of the protein-ligand binding, two biomolecules distant to each other in space should approach to each other. After the approach, the ligand binds to the protein so as that the ligand fit to the binding site of protein. Recent trend for the protein-ligand binding study is to search a complex structure where the ligand fit well to the binding site of protein because the search for complex structure may lead to drug design. On the other hand, the molecular approach has not been studied well because it has been thought that this process happens through diffusive motions of biomolecules. If this conventional scheme is correct, the molecular approach has no importance for understanding the molecular binding. However, our recent molecular dynamics (MD) simulation of biomolecules has shown that the orientations of water molecules are ordered around the biomolecules [1][2][3][4][5][6]. The ordered orientations of water molecules constructed coherent spatial patterns around the biomolecules. The patterns were contributed not by particular water molecules but by a number of water molecules diffusing. We named the coherent patterns as "solvent site-dipole field" [3]. These studies suggest a mechanism that the solvent site-dipole field may mediate biomolecular attractive interactions. To make firmer this mechanism, it is essential to investigate a system consisting of a protein and a ligand, which are known experimentally to form a complex structure. The solvent site-dipoles field was also observed computationally around spherical hydrophobic solutes, although the effective range of the coherent patterns was short (a few angstroms) [7].

MD simulations have been used to computationally study static and dynamic properties of solutes in solvent. The simulation provides an MD trajectory consisting of a number of snapshots. The MD trajectory is used to detect solute surface sites where water molecules are observed with a high density (probability). A simulation study [4] has shown that the crystal-water sites, which are experimentally detected on the protein surface, coincide with the high solvent-density sites, which are computed from an MD simulation. Recent MD study has shown that the crystal-water and crystral-ion sites around RNA are reproduced well [8][9]. These works indicate that the MD simulation is reliable enough to study the solvent density around biomolecules at an atomic resolution and that the crystal-water sites are also trustable when the crystal-water sites are carefully determined as done in an X-ray crystallography [10].

The MD simulation can detect a quantity experimentally undetectable. Since solvent molecules are moving around solute [1][11][12], the hydration structure around the solute is too fragile to be detected experimentally. We have shown that the patterns of solvent-solvent and solvent-solute hydrogen bonds (H-bonds) in the vicinity of protein strongly depend on the surface shape of protein, and that the patterns are contributed not by particular water molecules but by a number of water molecules visiting the protein surface [13][14]. The computed H-bond patterns had a correlation with experimentally presumed H-bond patterns [14], where the crystal-water sites were connected by H-bonds when the relative positioning of the crystal-water sites were hydrogen bondable [15]. We named the H-bond patterns "H-bond field" [13][14]. Motions of water molecules are diffusive (isotropic) even in the vicinity of protein. However, when we view the motions in a time period of 10 ps, the motions showed various spatial flow patterns such as vertex, drying/wetting, and fair current [16]. A simulation work suggests that the solute-surface drying may be related to solute-solute attractive interactions [7].

In this work, we studied the solvent site-dipole field between a protein and a peptide, which are known to form a complex. The MD simulation demonstrates the orientational ordering of water molecules in the inter-biomolecular zone, and the degree of the ordering correlates with the inter-biomolecular distance. The effective range for the coherent patterns was larger than 20 Å.

Biological and physical significance of the solvent site-dipole field is discussed.

## 2. Materials and Methods

We studied a system consisting of a toxic protein α-Bungarotoxin and a short peptide (pdb entry = 1HAA), for which the complex structure was solved by an NMR experiment [17]. The peptide binds to the toxin with a highly affinity [17]. The protein and peptide are comprised of 76 and 13 amino-acid residues with net charges of $3e^+$ and $2e^-$, respectively.

We prepared the simulation system as follows: First, we calculated the mass centers of the two biomolecules in the complex and drew a straight line passing the two mass centers. Then, we shifted the peptide away from the protein by 30 Å along the straight line to generate isolated forms of the biomolecules. Next, we capped the N and C termini of the two biomolecules by acetyl and N-methyl groups, respectively, and immersed them in an explicit-solvent sphere (diameter = 100 Å), where the sphere center was put at the mid position of the mass centers of the two biomolecules. Last, we added 9 $Na^+$ and 10 $Cl^-$ ions to neutralize the net charge of the system. These procedures were done with using a LEAP option of AMBER 6.0 [18], which automatically sets the water density at 1 g/cc and put ions at energetically appropriate positions. The number of water molecules in the finally generated system was 15781, and the total number of atoms (i.e., atoms in protein, peptide, water molecules, and ions) was 48688. We call the molecular structure obtained here "reference structure".

After energy-minimizing the reference structure, and the energy-minimized structure was used for the initial structure of the MD simulation. First we did a preparative MD run to equilibrate the configuration of water molecules before a productive MD run (i.e., sampling run). In this preparative run, a harmonic potential was applied on main-chain heavy atoms of the two biomolecules to restraint them around the energy-minimized positions. The preparative run was done with the following stages: a 50ps constant-heating-rate run from 0 to 50 K, a 50ps constant-temperature run at 50 K, a 50ps constant-heating-rate run from 50 to 150 K, a 50ps constant-temperature run at 150 K, a 50ps constant-heating-rate run from 150 to 300 K, and a 50ps constant-temperature run at 300 K.

After the preparative run, we did a 7ns productive run at 300 K with saving a snapshot every 50 fs (140,000 snapshots saved). In the first 1.0 ns of the simulation, we restrained the main-chain heavy atoms of the two biomolecules to those in the reference structures. From 1 to 7 ns, the restraint was applied only to the main-chain heavy atoms of the protein (the peptide moved freely). This switch off of the restraint on the peptide is to analyze changes of the solvent site-dipole field before and after the switch off.

A computer program, Sander Classics of AMBER 6.0 was used for the MD simulation with the following condition: time step of 1 fs; AMBER Parm 96 force field [19]; the SHAKE method [20] to constrain covalent bonds between heavy atoms and hydrogen atoms; TIP3P water model [21] to express water molecules; and Berendsen's method [22] to control the temperature where the relaxation time = 0.2 ps. A harmonic potential was applied to water-oxygen atoms to avoid evaporation only when they were outside the water sphere. All of the electrostatic interactions in the system were exactly computed at every step of simulation without truncation. We used a special purpose computer, MD Engine II [23][24], developed to accelerate the computation of non-bonded interactions.

To analyze the MD trajectory, we computed the solvent site-dipole vector, which was introduced in our previous study [1][2][3][4][5][6]. Here, we explain the computation method of the solvent site-dipole vector: First, we defined a large box involving the whole system, where three

sides of the large box are parallel to the x-, y-, and z-axes of a coordinate system. Figure 1 schematically shows the axes, although the z-axis is not shown in this two-dimensional scheme. Next, the large box was divided into small cubes of 2.0Å × 2.0Å × 2.0Å. The representative position of a cube was assigned to the body center of the cube (see the cross mark at the center of the shaded cube in Figure 1). We specified each cube by three digits, *i*, *j* and *k*, which are the ordinal numbers of the cube along the x-, y-, and z-axes, respectively, in the large box (the shaded cube is specified by $(i,j,k)$ in Figure 1). The orientation of each water molecule, for which the position is $\vec{r}_{\mathrm{w}}$ at time *t*, was defined by a unit vector $\vec{e}_{\mathrm{w}}(\vec{r}_{\mathrm{w}};t)$ pointing from the oxygen atom toward the mid position of the two hydrogen atoms of the water molecule (see shaded arrows in Figure 1). Then, we re-assigned $\vec{e}_{\mathrm{w}}(\vec{r}_{\mathrm{w}};t)$ of each water molecule to a cube (*i*,*j*,*k*) where the oxygen atom of the water molecule was detected. This cube-assigned vector was designated as $\vec{e}_{\mathrm{c}}(i,j,k;t)$ (see black arrow at the center of the shaded cube in Figure 1). This vector is called "solvent site-dipole vector". If a cube does not involve a water molecule, no vector is assigned to the cube.

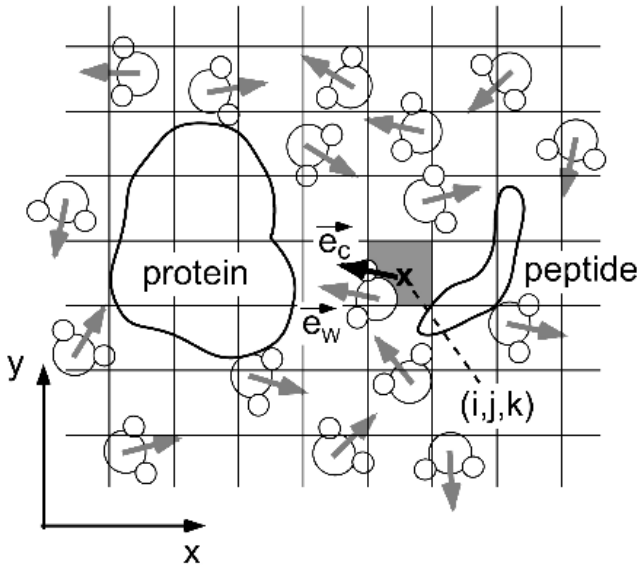


**Figure 1.** Scheme to explain the definition of solvent site-dipole vector $\vec{e}_c(i,j,k)$.

Since the solvent site-dipole vectors largely fluctuate with time, it is essentially important to take time-average of the solvent site-dipole vectors for analyzing the spatial patterns of the vectors. On the other hand, the relative positioning of the two biomolecules also changes with time. Thus, the time-average of the solvent site-dipole vectors should be taken in a short time window $[\bar{t}-\Delta t/2, \bar{t}+\Delta t/2]$, where $\bar{t}$ is the mean time (or mid time) of the window: i.e., the relative position of the peptide with respect to the protein should be small in the period of $\Delta t$ (remember that the protein is restrained in space). From some trials, we set as $\Delta t = 30$ ps. The change of the distance between the mass centers of the biomolecules was 1.02 Å on average in period of 30ps (the average was taken for the MD trajectory of 0 to 6 ns). This value is smaller than the cube size (2.0 Å). The 30ps time window involves 600 snapshots. We designate the time window of $[\bar{t}-15$ ps, $\bar{t}+15$ ps] as $\mathrm{TW}(\bar{t})$. The period of 30 ps is larger than the relaxation time for rotational motions of individual water molecules (2.2 ps from the current simulation). The time-averaged site-dipole vector for $\mathrm{TW}(\bar{t})$ is defined as:

$$\vec{d}(i,j,k;\bar{t}) = \frac{\sum_{t \in \mathrm{TW}(\bar{t})} \delta(i,j,k;t)\vec{e}_{\mathrm{c}}(i,j,k;t)}{\sum_{t \in \mathrm{TW}(\bar{t})} \delta(i,j,k;t)},$$

where $\delta(i,j,k;t) = 1$ when the cube $(i,j,k)$ involves a water molecule at time $t$ and otherwise 0. The summation is taken over 600 snapshots in $\mathrm{TW}(\bar{t})$. Although the norm of $\vec{e}_{\mathrm{c}}(i,j,k;t)$ is 1 ($|\vec{e}_{\mathrm{c}}(i,j,k;t)| = 1$), the norm of $\vec{d}(i,j,k;\bar{t})$ is smaller than 1 at most of cubes ($|\vec{d}(i,j,k;\bar{t})| < 1$). The larger the value of $|\vec{d}(i,j,k;\bar{t})|$, the more ordered the molecular orientations at site $(i,j,k)$ in the time window $\mathrm{TW}(\bar{t})$.

After calculating $\vec{d}(i,j,k;\bar{t})$ at cubes, we smoothed (i.e., coarse-grained) them as:

$$\vec{d}_{\mathrm{s}}(i,j,k;\bar{t}) = \frac{\sum_{\Delta i,\Delta j,\Delta k=-1}^{+1} \omega(i+\Delta i, j+\Delta j, k+\Delta k)\vec{d}(i+\Delta i, j+\Delta j, k+\Delta k;\bar{t})}{\sum_{\Delta i,\Delta j,\Delta k=-1}^{+1} \omega(i+\Delta i, j+\Delta j, k+\Delta k)},$$

where the summation is taken over 27 (= 3 × 3 × 3) cubes centered at the cube $(i,j,k)$. The weight factor ω determines the contribution of the 27 cubes to $\vec{d}_{\mathrm{s}}(i,j,k;\bar{t})$ as: $\omega(i,j,k) = 1$, $\omega(i \pm 1,j,k) = \omega(i,j \pm 1,k) = \omega(i,j,k \pm 1) = 0.5$, $\omega(i \pm 1,j \pm 1,k) = \omega(i,j \pm 1,k \pm 1) = \omega(i \pm 1,j,k \pm 1) = 0.3$, and $\omega(i \pm 1,j \pm 1,k \pm 1) = 0.1$. This weight factor is the same as that used in our previous study [6].

Next, we computed another solvent site-dipole vector focusing on the protein-peptide inter-molecular zone, as follows: First, we calculated the average atomic positions of the protein and peptide in each time window $\mathrm{TW}(\bar{t})$ (in the averaging, we did not use the structure superposition between biomolecular structures in snapshots). Next, we calculated the minimum heavy-atomic distance $L_{\min}(\bar{t})$ between the window-averaged biomolecular structures: $L_{\min}(\bar{t}) = |\vec{r}_{\mathrm{pep}}(\bar{t}) - \vec{r}_{\mathrm{pro}}(\bar{t})|$, where $\vec{r}_{\mathrm{pro}}(\bar{t})$ and $\vec{r}_{\mathrm{pep}}(\bar{t})$ are the heavy-atomic positions in the protein and the peptide, respectively, to provide $L_{\min}(\bar{t})$. The mid position is simply defined as:

$$\vec{r}_{\mathrm{mid}}(\bar{t}) = \frac{\vec{r}_{\mathrm{pro}}(\bar{t}) + \vec{r}_{\mathrm{pep}}(\bar{t})}{2} = \frac{L_{\min}(\bar{t})}{2}\vec{e}_{\mathrm{pp}}(\bar{t}) + \vec{r}_{\mathrm{pro}}(\bar{t}),$$

where $\vec{e}_{\mathrm{pp}}(\bar{t})$ is the unit vector parallel to the protein-peptide minimum-distance vector: $\vec{e}_{\mathrm{pp}}(\bar{t}) // \vec{r}_{\mathrm{pro}}(\bar{t}) - \vec{r}_{\mathrm{pep}}(\bar{t})$. Then, we defined a small sphere (SP) around the mid position (radius = 1.5 Å), and computed a weighted average over vectors $\vec{e}_{\mathrm{w}}(\vec{r}_{\mathrm{w}};\bar{t})$ that were detected in SP at time $\bar{t}$:

$$\vec{v}(\bar{t}) = \frac{\sum_{\vec{r} \in \mathrm{SP}} r_{\mathrm{c}}(\vec{r}_{\mathrm{w}})^{-1}\vec{e}_{\mathrm{w}}(\vec{r}_{\mathrm{w}};\bar{t})}{\sum_{\vec{r} \in \mathrm{SP}} r_{\mathrm{c}}(\vec{r}_{\mathrm{w}})^{-1}},$$

where $r_{\mathrm{c}}(\vec{r})$ is the distance from the mid position $\vec{r}_{\mathrm{mid}}$ to $\vec{r}_{\mathrm{w}}$ (the position at that a water molecule is detected at $t$): $r_{\mathrm{c}}(\vec{r}_{\mathrm{w}}) = |\vec{r}_{\mathrm{w}} - \vec{r}_{\mathrm{mid}}|$. Since the weight factor is the inversed distance $r_{\mathrm{c}}(\vec{r})^{-1}$, a vector $\vec{e}_{\mathrm{w}}(\vec{r};\bar{t})$ closer to the mid position $\vec{r}_{\mathrm{mid}}$ contributes more to the vector $\vec{v}(\bar{t})$.

Note that $\vec{v}(\bar{t})$ is not a time averaged quantity but a quantity defined at an instant $\bar{t}$, although $\vec{r}_{\mathrm{mid}}$ is the time-averaged quantity defined in time window $\mathrm{TW}(\bar{t})$. Finally, we took the time-window-average of $\vec{v}(t)$ in $\mathrm{TW}(\bar{t})$ as:

$$\vec{V}(\bar{t}) = \sum_{t \in \mathrm{TW}(\bar{t})} \vec{v}(t) / N_{\mathrm{detect}},$$

where $N_{\mathrm{detect}}$ is the number of snapshots where a water oxygen is detected in the sphere (SP) during the interval $\mathrm{TW}(\bar{t})$. The $\vec{V}(\bar{t})$ is simply recognized as: A large value of $|\vec{V}(\bar{t})|$ indicates that the orientations of water molecules visiting the mid-position sphere SP are well ordered in the time window. Contrarily, a small value of $\vec{V}(\bar{t})$ indicates less ordering. The vector $\vec{V}(\bar{t})$ largely fluctuates along the simulation trajectory, as shown later. We are interested in the component of $\vec{V}(\bar{t})$ parallel to the protein-peptide minimum-distance vector $\vec{r}_{\mathrm{pro}}(\bar{t}) - \vec{r}_{\mathrm{pep}}(\bar{t})$. Then we defined a quantity $P(\bar{t})$ by a scalar product of $\vec{V}(\bar{t})$ and $\vec{e}_{\mathrm{pp}}(\bar{t})$ as:

$$P(\bar{t}) = \vec{V}(\bar{t}) \cdot \vec{e}_{\mathrm{pp}}(\bar{t}).$$

If $P(\bar{t}) < 0$, the solvent site-dipole vectors tend to be oriented from the peptide to the protein at the protein-peptide mid position.

## 3. Results and Discussion

Figure 2 shows the time development of $L_{\mathrm{min}}(\bar{t})$ and $L_{\mathrm{cc}}(\bar{t})$, where $L_{\mathrm{cc}}(\bar{t})$ is the inter-mass center distance for the two biomolecules in $\mathrm{TW}(\bar{t})$. Time variations of $L_{\mathrm{min}}(\bar{t})$ and $L_{\mathrm{cc}}(\bar{t})$ correlate well to each other. For the first 1 ns, $L_{\mathrm{min}}(\bar{t})$ and $L_{\mathrm{cc}}(\bar{t})$ were almost constant because the biomolecules were under the restraint. After the removal of the restraint, $L_{\mathrm{min}}(\bar{t})$ and $L_{\mathrm{cc}}(\bar{t})$ fluctuated, and the two biomolecules began approaching after 1.5 ns (i.e., after 0.5 ns from the removal of the restraint). At $t \approx 5.75$ ns the two biomolecules contacted each other. After the first contact, the peptide moved around the protein surface like a rolling stone, but did not "correctly" bind to the binding site of the protein. In this study, we focus on the biomolecular approach process to elucidate the solvent site-dipole field in the inter-biomolecular zone. One may notice that $L_{\mathrm{cc}}(\bar{t})$ slightly decreased at the very beginning of simulation ($0 < t < 60$ ps). This decreasing is due to the difference of molecular structures used for the restraint: the energy-minimized structure during the preparative run and the reference structure during the productive run. Although this inconsistency for the restraint is drawback of the current study, no substantial error was seen on the solvent site-dipole field.

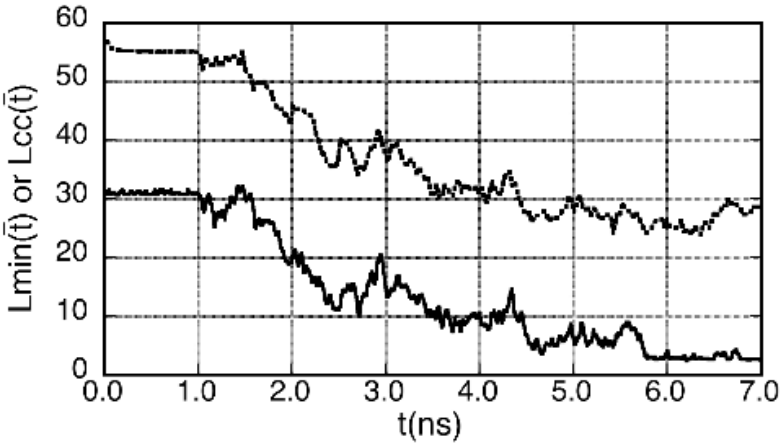


**Figure 2.** Time development of $L_{\min}(\bar{t})$ (solid line) and $L_{cc}(\bar{t})$ (dotted line).

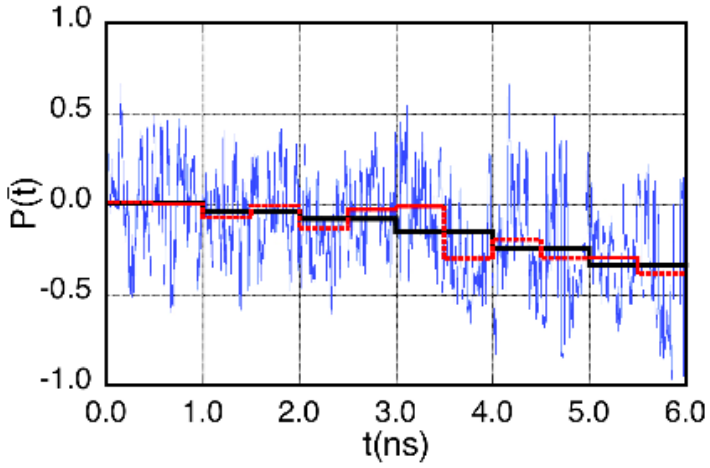


**Figure 3.** Time development of $P(\bar{t})$ (blue line), 1.0ns time-block average $< P(\bar{t}) >_{1ns}$ (black line), and 0.5ns time-block average $< P(\bar{t}) >_{0.5ns}$ (red line).

Figure 3 plots $P(\bar{t})$, $< P(\bar{t}) >_{0.5ns}$, and $< P(\bar{t}) >_{1ns}$ along the MD trajectory, where $< P(\bar{t}) >_{0.5ns}$ and $< P(\bar{t}) >_{1ns}$ are 0.5ns and 1.0ns time-block averages of $P(\bar{t})$, respectively. Fluctuations of $P(\bar{t})$ were large. However, $< P(\bar{t}) >_{1ns}$ monotonically decreased along the MD trajectory (see black line of Figure 3). The physical meaning of the negative $< P(\bar{t}) >_{1ns}$ is simple: Remember that the net charges of protein and peptide are $3e^{+}$ and $2e^{-}$, respectively. Then, it is expected the solvent site-dipole vectors aligned from peptide to protein, resulting in an inequality $P(\bar{t}) < 0$. In the first 1 ns, during which the restraint was applied on the two biomolecules, both block averages $< P(\bar{t}) >_{0.5ns}$ and $< P(\bar{t}) >_{1ns}$ were nearly zero. In the period [1-2ns], the value of $L_{\min}(\bar{t})$ was larger than 20 Å except for the very end of the period, and the value of $< P(\bar{t}) >_{1ns}$ was negative. This indicates that the water molecules are ordered at the protein-peptide mid position even when the two biomolecules are apart more than 20 Å.

In two periods $2.5 < t \leq 3.0$ ns and $4.0 < t \leq 4.5$ ns, the absolute value ($|< P(\bar{t}) >_{0.5ns}|$) of the 0.5ns block average decreased (see red line of Figure 3). Interestingly, $L_{\min}(\bar{t})$ increased in this

two periods (see Figure 2). This indicates that $P(\bar{t})$ correlates well with $L_{\min}(\bar{t})$ in spite of the large fluctuations of $P(\bar{t})$. In our previous simulation of protein-DNA approach [6], strong ordering of the solvent site-dipole vectors was seen in the inter-protein-DNA zone from the beginning to the end of simulation. However, we did not observe synchronized fluctuations between the protein-DNA distance and the ordering of solvent site-dipole vectors. Since DNA is a highly charged molecule, strong ordering was always induced. We presume that the synchronized fluctuations were hidden in the strong background. Thus, the present study showed the correlation for the first time. The current study supports our previous speculation that the solvent site-dipole filed may mediate the biomolecular approach [3][6].

The inter-biomolecular mid position $\vec{r}_{\text{mid}}$ moved when the atom pair $\vec{r}_{\text{pep}}$ and $\vec{r}_{\text{pro}}$ altered during the simulation, and the shift of $\vec{r}_{\text{mid}}$ may introduce inaccuracy in estimation of $P(\bar{t})$, although the shift was small: remember that the average of the shift was 1.02 Å during periods of 30 ps. However, $P(\bar{t})$ is the component of $\vec{V}(\bar{t})$ parallel to $\vec{e}_{\text{pp}}(\bar{t})$ at each time, whether the shift is large or not. Therefore, $< P(\bar{t}) >_{0.5\text{ns}}$ and $< P(\bar{t}) >_{1\text{ns}}$ are meaningful to assess the orientational ordering of water molecules at the inter-biomolecular zone.

For $t > 5.75$ ns the two biomolecules touched to each other and only a few water molecules visited the inter-biomolecular mid position during 30ps periods. Thus, the estimation of $P(\bar{t})$ has lack of statistics for $t > 5.75$ ns. However, the purpose of the present study is to investigate the solvent site-dipole field in the approach process of the two biomolecules. The behavior of water molecules in the biomolecular interface in the protein-ligand complex is out of the scope for the present study.

Next, we view the spatial patterns of the solvent site-dipole vectors in the whole system. Figures 4 and 5 show the patterns of $\vec{d}_{\text{s}}(i,j,k;\bar{t})$ in time windows TW(2.415 ns) and TW(3.730 ns), respectively. In Figure 4, where $L_{\min}(\bar{t})$ was about 12 Å, characteristic patterns of the solvent site-dipole field were observed (see region in between two purple lines), where the bridge-like field connected a negatively charged amino-acid residue (ASP87) of the peptide and a positively charged one (ARG37) of the protein. The residue numbering follows that in the original PDB. In Figure 5, $L_{\min}(\bar{t})$ was about 8 Å, smaller than that in Figure 4. A remarkably large solvent site-dipole bridge (circled by purple line) is seen between a negatively charged residue (GLU83) of the peptide and a positively charged one (LYS70) in the protein. Figures 4 and 5 clearly demonstrate a bridge-like structure of the solvent site-dipole vectors. However, the bridge was not always seen because the solvent site-dipole field fluctuated largely. The red circle in Figures 4 and 5 indicates characteristic patterns in the binding site. Similar patterns were often seen in the binding site.

The solvent site-dipole bridge was also seen in our previous MD simulation of DNA-protein approach [6]. We expected the appearance of the bridge before doing the simulation because DNA is a highly charged molecule. In contrast, the protein is a less charged molecule. We did another MD simulation of a system consisting of acidic and basic amino acids that were fixed taking an inter-molecular distance of 14 Å [3]. In the work, the solvent site-dipole bridge was visible from a 1ns time average. The current study has shown that 30ps time averaging without fixing the ligand position can provide the solvent site-dipole bridge. The observation of the bridge, therefore, suggests that the solvent site-dipole field generally mediates protein-ligand approach.

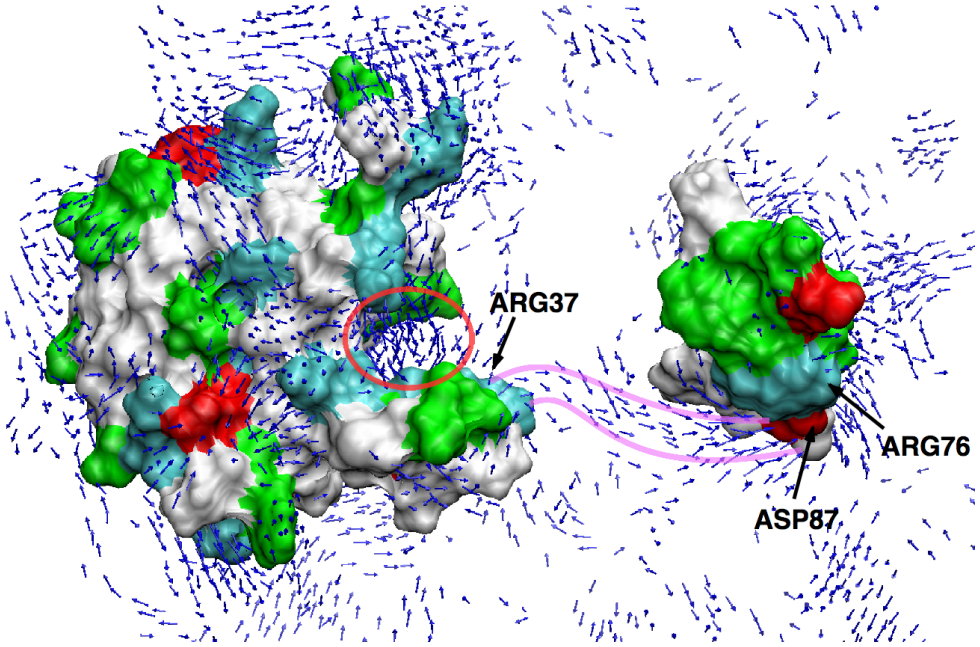


**Figure 4.** Solvent site-dipole vectors $\vec{d}_s(i,j,k;\bar{t})$ in time window TW(2.415 ns) (i.e., 2.400 ns-2.430 ns). Acidic and basic amino-acid residues are displayed by red and cyan, respectively. Non-polar ones are displayed by white, and polar ones by green. Arrows show solvent site-dipole vectors at cubes. Only arrows for that the norms are larger than 0.35 are displayed.

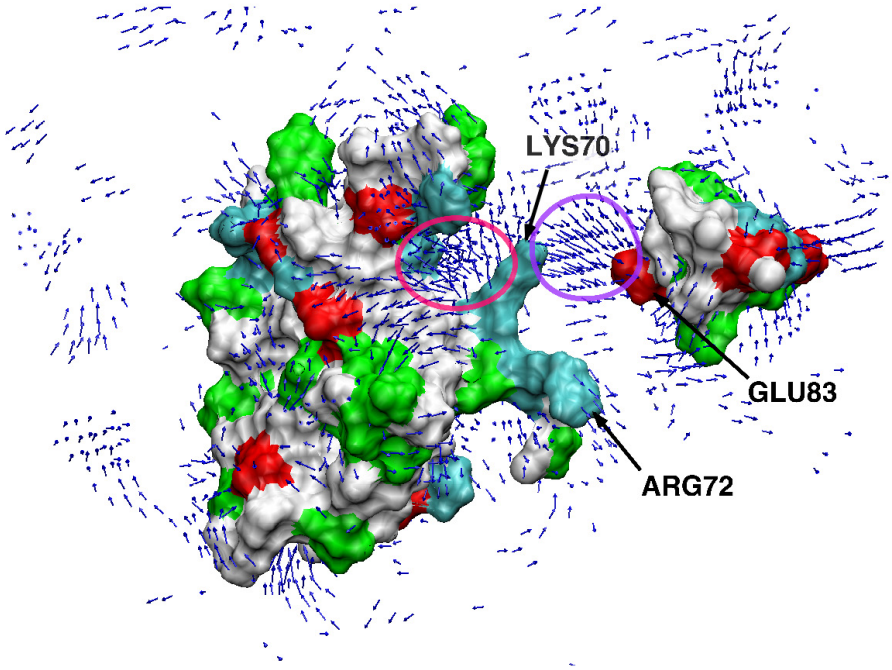


**Figure 5.** Solvent site-dipole vectors $\vec{d}_s(i,j,k;\bar{t})$ in time window TW(3.730 ns) (i.e., 3.715 ns-3.745 ns). See caption for Figure 4 for colors of amino-acid residues and for solvent site-dipole vectors.

The existence of the solvent site-dipole field suggests a novel mechanism for biomolecular interaction. The current study, on the other hand, proposes a deep question on the interaction mechanism: What is the physical type of the interactions resulted from the field? The solvent site-dipole ordering indicates that the orientational motions of water molecules are suppressed in the inter-biomolecular zone. This leads to a decrease of the dielectric constant in the zone, as previously discussed [3][4]. In the Poisson-Boltzmann scheme, the dielectric constant rapidly reaches the bulk-water value (≈ 80) at a probe point distant from the protein surface by a few angstroms. This means that the Poisson-Boltzmann field is effective only in a narrow region a few angstroms distant from the protein surface. Contrarily, our study suggests that biomolecules may interact even when the inter-biomolecular distance is larger than 20 Å. The driven force based on the dielectric-constant decrement is enthalpic. On the other hand, another mechanism is possible: The coherent patterns of the solvent site-dipole vectors indicate that the solvent molecules are structured in the inter-biomolecular zone, although the patterns are fragile. Note that the structured solvent decreases the entropy of the system. Therefore, approach of two biomolecules, which decreases volume of the inter-biomolecular zone, increases the entropy. We do not have definite answer which enthalpy or entropy is more effective in the biomolecular interactions.

As explained in Introduction section, the biomolecular binding should be separated in two processes: biomolecular approach and molecular fitting. We showed that the biomolecular approach is not a simple diffusive process. On the other hand, the "correct" complex structure determined by the experiment was not obtained in the current simulation. We have recently shown that an enhanced conformational sampling method (multicanonical MD method) is useful for studying the fitting process: the correct complex structure of a protein and a ligand was obtained from an ab initio calculation, where no empirical score function was not used, and the lowest free-energy complex was assigned to the correct complex structure [25]. A novel method for biomolecular recognition may be possible when the solvent site-dipole field is taken into account in molecular docking study.

## References

[1] Higo, J., Kono, H., Nakajima, N., Shirai, H., Nakamura, H., Sarai, A., Molecular dynamics study on mobility and dipole ordering of solvent around proteins: effects of periodic-box size and protein charge, *Chem. Phys. Lett.*, **306**, 395-401 (1999).

[2] Higo, J., Kono, H., Nakamura, H., Sarai, A., Solvent density and long-range dipole field around a DNA-binding protein studied by molecular dynamics, *Proteins*, **40**, 193-206 (2000).

[3] Higo, J., Sasai, M., Shirai, H., Nakamura, H., Kugimiya, T., Large vortex-like structure of dipole field in computer models of liquid water and dipole-bridge between biomolecules, *Proc. Natl. Acad. Sci. USA.*, **98**, 5961-6964 (2001).

[4] Higo, J., Nakasako, M., Hydration structure of human lysozyme investigated by molecular dynamics simulation and cryogenic X-ray crystal structure analyses: on the correlation between crystal-water sites, solvent-density and solvent-dipole, *J. Comp. Chem.*, **23**, 1323-1336 (2002).

[5] Yokomizo, T., Yagihara, S., Higo, J., Rotational motions of solvent site-dipole field around a protein, *Chem. Phys. Lett.*, **374**, 453-458 (2003).

[6] Hamasaki, N., Miyagawa, H., Mitomo, D., Yamagishi, A., Higo, J., DNA-protein binding mediated by a solvent site-dipole field, *Chem. Phys. Lett.*, **431**, 160-163 (2006).

[7] Hotta, T., Kimura, A., Sasai, M., Fluctuating hydration structure around nanometer-size hydrophobic solutes. I. Caging and drying around $C_{60}$ and $C_{60}H_{60}$ spheres, *J. Phys. Chem. B*,

**109**, 18600-18608 (2005).
[8] Auffinger, P., Bielecki, L., Westhof, E., The $Mg^{2+}$ binding sites of the 5S rRNA loop E motif as investigated by molecular dynamics simulations, *Chem. Biol.*, **10**, 551-561 (2003).
[9] Vaiana, A. C., Westhof, E., Auffinger, P., A molecular dynamics simulation study of an aminoglycoside/A-site RNA complex: conformational and hydration patterns, *Biochimie*, **88**, 1061-1073 (2006).
[10] Nakasako, M., Large-scale networks of hydration water molecules around bovine β-trypsin revealed by cryogenic X-ray crystal structure analysis, *J. Mol. Biol.*, **289**, 547-564 (1999).
[11] Komeiji, Y., Uebayashi, M., Someya, J., Yamato I., A molecular dynamics study of solvent behavior around a protein, *Proteins*, **16**, 268-277 (1993).
[12] Makarov, V. A., Feig, M., Andrews, B. K., Pettitt, B. M., Diffusion of solvent around biomolecular solutes: a molecular dynamics simulation study, *Biophys. J.*, **75**, 150-158 (1998).
[13] Yokomizo, T., Nakasako, M., Yamazaki, T., Shindo, H., Higo, J., Hydrogen-bond patterns in the hydration structure of a protein, *Chem. Phys. Lett.*, **401**, 332-336 (2005).
[14] Yokomizo, T., Higo, J., Nakasako, M., Patterns and networks of hydrogen-bonds in the hydration structure of human lysozyme, *Chem. Phys. Lett.*, **410**, 31-35 (2005).
[15] Nakasako, M., Large-scale networks of hydration water molecules around proteins investigated by cryogenic X-ray crystallography, *Cell. Mol. Biol.*, **47**, 767-790 (2002).
[16] Umezawa, K., Higo, J., Shimotakahara, S., Shindo, H., Collective solvent flows around a protein investigated by molecular dynamics simulation, *J. Chem. Phys.*, **127**, 045101 (2007).
[17] Scherf, T., Kasher, R., Balass, M., Fridkin, M., Fuchs, S., Katchalski-Katzir, E., A β-hairpin structure in a 13-mer peptide that binds α-bungarotoxin with high affinity and neutralizes its toxicity, *Proc. Natl. Acad. Sci. USA*, **98**, 6629-6634 (2001).
[18] Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham III, T. E., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K., Kollman, P. A., AMBER 6; University of California San Francisco: San Francisco, CA (1996).
[19] Kollman, P. A., Dixon, R. W., Cornell, W. D., Chipot, C., Pohorille, A., The development/application of a "minimalist" organic/biochemical molecular mechanic force field using a combination of *ab initio* calculations and experimental data, In: van Gunsteren, W. F. (ed) Computer simulations of biological systems, ESCOM, Dordrecht, The Netherlands, 83-96 (1997).
[20] Ryckaert, J.-P., Ciccotti, G., Berendsen, H. J. C., Numerical integration of the cartesian equation of motion of a system with constraints: molecular dynamics of N-alkanes, *J. Comput. Phys.*, **23**, 327-341 (1977).
[21] Jorgensen, W.L., Chandrasekhar, J., Madura, J. D., Impey, R. W., Klein, M. L., Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.*, **79**, 926-935 (1983).
[22] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., Haak, J. R., Molecular dynamics with coupling to an external bath, *J. Chem. Phys.*, **81**, 3684-3690 (1984).
[23] Amisaki, T., Toyoda, S., Miyagawa, H., Kitamura, K., Development of hardware accelerator for molecular dynamics simulations: A computation board that calculates nonbonded interactions in cooperation with fast multipole method, *J. Comp. Chem.*, **24**, 582-592 (2003).
[24] Toyoda, S., Miyagawa, H., Kitamura, K., Amisaki, T., Hashimoto, E., Ikeda, H., Kusumi, A., Miyakawa, N., Development of MD Engine: High-speed accelerator with parallel processor design for molecular dynamics simulations, *J. Comp. Chem.*, **20**, 185-199 (1999).
[25] Kamiya, N., Yonezawa, Y., Nakamura, H., Higo, J., Protein-inhibitor flexible docking by a multicanonical sampling: native complex structure with the lowest free energy and a free-energy barrier distinguishing the native complex from the others, *Proteins*, **70**, 41-53 (2008).