PCCP

RSC Publishing

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2013, **15**, 3646

A theoretical study of the two binding modes between lysozyme and tri-NAG with an explicit solvent model based on the fragment molecular orbital method

Takeshi Ishikawa,[ab] Raghunadha Reddy Burri,[ac] Yuji O. Kamatari,[ad] Shun Sakuraba,[ae] Nobuyuki Matubayasi,[ae] Akio Kitao[ac] and Kazuo Kuwata*[ab]

Received 8th August 2012,
Accepted 17th January 2013

DOI: 10.1039/c3cp42761g

www.rsc.org/pccp

To examine the stabilities and binding characteristics, fragment molecular orbital (FMO) calculations were performed for the two binding modes of hen egg-white lysozyme with tri-*N*-acetyl-D-glucosamine (tri-NAG). Solvent effects were considered using an explicit solvent model. For comparison with the computational results, we experimentally determined the enthalpic contribution of the binding free-energy. Our calculations showed that the binding mode observed by X-ray analysis was more stable than the other binding mode by -6.2 kcal mol^{-1}, where it was found that the interaction of protein with solvent molecules was crucial for this stability. The amplitude of this energy difference was of the same order as the experimental enthalpic contribution. Our detailed analysis using the energies divided into each residue was also consistent with a previous mutant study. In addition, the electron density analysis showed that the formal charge of the lysozyme (+8.0 e) was reduced to +5.16 e by charge transfer with solvent molecules.

1. Introduction

It is essential to accurately calculate the binding affinity between a protein and a small molecule for computational or theoretical studies. Such a calculation can be used for drug discovery studies and investigations of the recognition mechanism in biomolecular systems. Several approaches based on classical methods with empirical force fields have been proposed, such as the Poisson–Boltzmann surface area method[1] and the free-energy perturbation method.[2] These approaches have been successfully used for the calculation of the binding energy or the interaction analysis in biomolecular systems. However, one of the issues of these methods is that the result depends on the force field used in the calculations.

Recently, several quantum chemical calculations of large molecules including biomolecules have been reported, in which the dependence on empirical force fields is not an issue. For example, the fragment molecular orbital (FMO) method[3–6] is one of the most efficient quantum chemical approaches for large molecules. In this method, computational cost is reduced by dividing a target molecule into small fragments and calculating only the fragments and pairs of the fragments, in which distant pairs are approximately treated (*i.e.*, only electrostatic interaction between the fragments is considered). Accuracy of the FMO method has been discussed in some previous publications,[6–8] one of which demonstrated that a typical error on the total energy is a few kcal per mol for a small protein.[8] The advantage of this method is that it not only reduces the computational cost but also gives interaction energies between the fragments.[7] These values are called the inter fragment interaction energy (IFIE) or the pair interaction energy (PIE), by which we can obtain the detailed information about intra- and inter-molecular interactions. Thus, many FMO studies including calculation of the binding energy or analysis of the interaction between a protein and a small molecule have been reported.[9–21]

Hen egg-white lysozyme, which is one of the most examined proteins, has six subsites in the active site. The study with X-ray analysis revealed that an inhibitor of the lysozyme, tri-*N*-acetyl-D-glucosamine (tri-NAG), bound to the A–B–C subsites.[22] Such a binding mode is referred to as "binding mode1" in this paper (see Fig. 1). Many mutation studies have also been performed.

[a] *CREST Project, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan*
[b] *Center for Emerging Infectious Diseases, Gifu University, 1-1, Yanagido, Gifu 501-1194, Japan. E-mail: kuwata@gifu-u.ac.jp*
[c] *Laboratory of Computational Protein Science, Institute of Molecular and Cellular Biosciences, University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan*
[d] *Life Science Research Center, Gifu University, 1-1, Yanagido, Gifu 501-1194, Japan*
[e] *Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan*

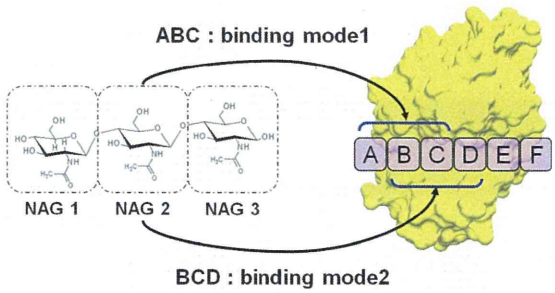

Fig. 1 Six subsites (A–F) in the active site of the lysozyme. In this study, binding mode1 (A–B–C) and binding mode2 (B–C–D) of tri-NAG were examined.

In the case of Trp62Tyr, Maenaka *et al.* reported that the complex with tri-NAG had another binding mode, *i.e.*, the B–C–D subsites[23] (this binding mode is referred to as "binding mode2" in this paper). In addition to experimental studies, several theoretical studies have been reported. Kamiya *et al.* performed classical simulations for the complex of lysozyme with tri-NAG using a multi-canonical sampling method.[24] They successfully demonstrated that the lowest free-energy conformation was binding mode1, *i.e.*, the X-ray structure of the complex. Binding mode2 was also observed as one of the other conformations in their simulation.

However, on first glance, the interaction of lysozyme with tri-NAG in binding mode2 seems to be stronger than that in binding mode1 because the contact area of binding mode2 is larger than that of binding mode1 (see Fig. 1). Actually, our preliminary docking simulations concluded that binding mode2 was the most stable conformation. Here, one question arises, *i.e.*, what types of interaction do cause the superiority of binding mode1? To answer this question, we performed FMO calculations for analyzing the protein–ligand interactions of the two binding modes. We consider that the FMO method is the best choice for such an investigation because it provides us with detailed information about intra- and inter-molecular interactions. In addition to the computational studies, we performed two experiments to determine the enthalpic contribution to the binding free-energy, *i.e.*, isothermal titration calorimetry (ITC) and surface plasmon resonance (SPR). In the following section, descriptions about computational and experimental methods are given, and after that, we discuss our results.

2. Materials and methods

2.1 ITC measurements and determination of thermodynamic parameters

Hen egg-white lysozyme and tri-NAG were purchased from Seikagaku Biobusiness Corp. (Tokyo, Japan).

Interactions between the lysozyme and tri-NAG were analyzed using the MicroCal iTC200 system (GE Healthcare, Buckinghamshire, UK). The samples were degassed under vacuum prior to titrations. The reference cell was filled with water, and the sample cell was filled with 50 μM lysozyme in 10 mM sodium acetate at pH 4.5 or in 10 mM PBS at pH 7.3. The syringe was loaded with 500 μM tri-NAG in the same buffer in the sample cell. Each binding isotherm was obtained from 20 injections of tri-NAG into the lysozyme at 25 °C. For the first titration, 0.4 μl of tri-NAG was injected; for subsequent titrations, 2 μl tri-NAG was injected in 120 s intervals. The stirring speed and reference power were 1000 rpm and 5 μcal s^{-1}, respectively. Raw data were processed and integrated with Origin 7.0 software (MicroCal) provided by the manufacturer, and the binding parameters, such as binding constant (K_D), enthalpy changes (ΔH), and binding stoichiometry (n), were obtained using a one set of sites model.[25] The first injection in each experiment was not considered for the analysis. Gibbs free-energy changes (ΔG) and entropy changes (ΔS) can be determined using the standard thermodynamic relationships:

$$\Delta G = -RT \ln K_D, \quad (1)$$

$$\Delta G = \Delta H - T\Delta S. \quad (2)$$

2.2 SPR measurements and determination of thermodynamic parameters

Interactions between the lysozyme and tri-NAG were also analyzed by the Biacore T100 system (GE Healthcare, Buckinghamshire, UK). The lysozyme was immobilized on a sensor chip (CM5) using an amine coupling kit (GE Healthcare, Buckinghamshire, UK). Various concentrations of tri-NAG were injected into a running buffer (0.01 M HEPES pH 7.4 containing 0.15 M NaCl) for 1 min at a flow rate of 30 μl min^{-1}. Then the running buffer without tri-NAG was injected for 10 min at the same flow rate. Data were corrected using a blank sensor chip as a control. The binding constant (K_D) was calculated from the tri-NAG concentration dependence of the sensorgram response. The K_D values were determined at five different temperatures (15, 20, 25, 30, and 35 °C). van't Hoff analysis of the binding constant at various temperatures was used to determine ΔH. The ΔG and $T\Delta S$ were calculated using standard thermodynamic relationships (eqn (1) and (2)).

2.3 FMO method

In the FMO method, a target molecule is divided into small N_f fragments, and total properties are evaluated with calculations of the fragment monomers and dimers.[3,4] These calculations are performed including the electrostatic potential from the other fragments, which is called the "environmental electrostatic potential (ESP)".[3,4] For example, the total energy is expressed as follows:

$$E^{\text{total}} = \sum_{I<J}^{N_f} E_{IJ} - (N_f - 2)\sum_{I}^{N_f} E_I, \quad (3)$$

where E_I and E_{IJ} are energies of the monomer and dimer, respectively. This equation can be rewritten as[7]

$$E^{\text{total}} = \sum_{I}^{N_f} E'_I + \sum_{I<J}^{N_f} \Delta E_{IJ}, \quad (4)$$

where E'_I is the monomer energy without the contribution of the ESP, and ΔE_{IJ} is the IFIE. A detailed description of this equation could be found in the previous paper.[7] Similarly, the total electron density is described using the following equation:[4,6]

$$\rho^{\text{total}} = \sum_{I}^{N_f} \rho_I + \sum_{I>J}^{N_f} \Delta\rho_{IJ}, \quad (5)$$

where ρ_I is the electron density of the monomer, and $\Delta\rho_{IJ}$ is the two-body correction on the total electron density. This is calculated as

$$\Delta\rho_{IJ} = \rho_{IJ} - \rho_I - \rho_J, \quad (6)$$

where ρ_{IJ} is the electron density of the dimer. We can also obtain the Mulliken atomic population from the electron density of eqn (5).

2.4 Computational details

In this study, we performed FMO calculations for two binding complexes of lysozyme and tri-NAG, binding mode1 and 2 (see Fig. 1). To consider the solvent effect, we employed the explicit water model instead of implicit water models like the polarizable continuum method,[26] because we are also interested in investigating the charge transfer between the complex and solvent molecules. The atomic coordinates used here were prepared as follows:

1. Initial structures of binding mode1 and 2 were obtained from docking simulations, in which the X-ray structure of the lysozyme (PDB code: 1HEW[22]) was used. The N- and C-terminal residues were set to $-NH_3^+$ and $-COO^-$, respectively.

2. Water molecules, sodium ions, and chloride ions were added, and energy minimization was performed under the truncated octahedron boundary conditions using the AMBER9 package,[27] where FF99SB,[28] GLYCAM06,[29] and TIP3P[30] force fields were employed for the lysozyme, tri-NAG, and water molecules, respectively.

3. After heating (100 ps) and equilibrating (900 ps) processes, a constant temperature and pressure (300 K and 1 atm) ensemble simulation with the classical method was performed for 30 ns.

4. 40 snap-shots of the structure were randomly selected from the trajectory of the last 10 ns, and energy minimizations were performed.

5. Solvent molecules that are more than 8.0 Å away from the complex were excluded. As a result, we finally obtained 40 structures including the complex, water molecules, sodium ions, and chloride ions (one of them is graphically shown in Fig. 2).

Effects of the thickness of explicit solvent molecules were previously investigated,[32] in which it was confirmed that atomic charges and the internal energy of protein reached almost steady values with 8.0 Å thickness.

In this study, average values over the 40 calculations were used for the binding energy calculation, by which we could partially introduce the influence of geometrical fluctuation of the protein and solvent molecules at physiological temperature (simple additive averaging was used here).[33,34] The protein and the ligand were already placed near the global minimum point by the classical simulation, where the potential surface is considered to be globally flat. In this case a reasonable treatment of the geometrical effect is important rather than taking Boltzmann surface correction. Thus we simply took an average over the 40 snap-shots and actually the results were not bad as shown in the following sections.

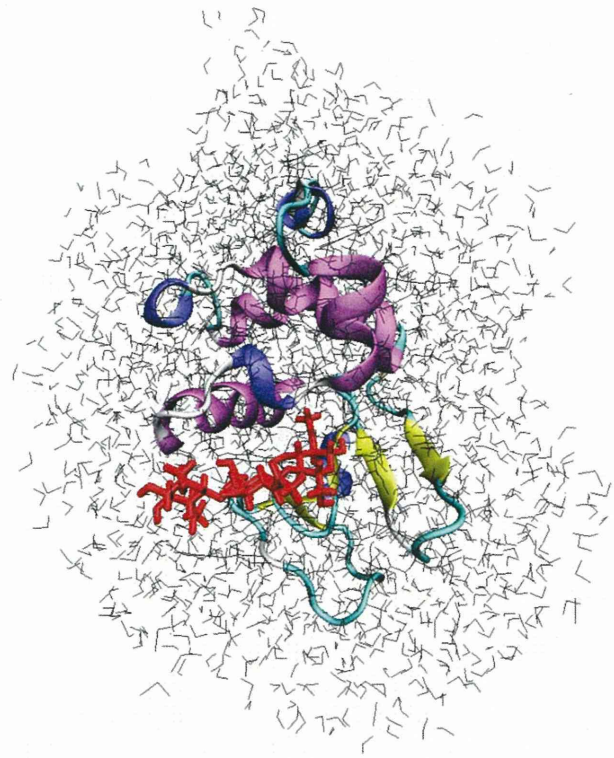

Fig. 2 Graphical representation of a structure of the complex (binding mode1) and solvent molecules used in our calculations. Tri-NAG is represented with red sticks and water molecules are represented with gray sticks. This image was created using VMD software.[31]

In our FMO calculations, each amino acid residue was treated as a single fragment, except for pairs of cysteines connected with each other by disulfide bonds, which were treated as a single fragment. The tri-NAG molecule was divided into three fragments (Fig. 3). Although water molecules were basically treated as a single fragment, ions and their hydration water molecules (within 2.5 Å from the ion) were collected into one fragment. MP2 with the resolution of the identity approximation (RI-MP2)[35,36] was used for energy calculations, by which the electron correlation energy or the van der Waals interaction energy is reasonably evaluated. Here, the cc-pVDZ basis set[37] and its auxiliary basis set[38] were employed. Accuracy of the RI-MP2 method in the context of the FMO scheme has been discussed in the original paper,[36] which showed that the error of the RI approximation was lower than 0.01% of the electron correlation energy. All the FMO calculations were

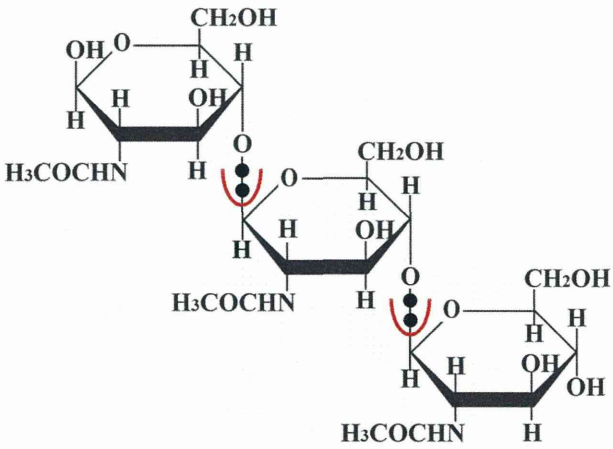

Fig. 3 Fragmentation pattern of tri-NAG used in the FMO calculations.

performed with our original quantum chemical program package PAICS.[33]

2.5 Definition of binding energy

One of characteristics of the FMO scheme is that the total energy can be easily divided into several components by restricting the summations of the index of fragment or fragment pair in eqn (4). Therefore, the internal energies of the lysozyme (E^{lys}), tri-NAG (E^{nag}), and solvent molecules (E^{solv}) can be expressed as

$$E^{\mathrm{lys}} = \sum_{I \in \text{lysozyme}} E'_I + \sum_{I<J \in \text{lysozyme}} \Delta E_{IJ}, \tag{7}$$

$$E^{\mathrm{nag}} = \sum_{I \in \text{tri-NAG}} E'_I + \sum_{I<J \in \text{tri-NAG}} \Delta E_{IJ}, \tag{8}$$

$$E^{\mathrm{solv}} = \sum_{I \in \text{solvent}} E'_I + \sum_{I<J \in \text{solvent}} \Delta E_{IJ}, \tag{9}$$

respectively. Additionally, the interaction energies between the solvent and lysozyme ($\Delta E^{\mathrm{solv-lys}}$), the solvent and tri-NAG ($\Delta E^{\mathrm{solv-nag}}$), and the lysozyme and tri-NAG ($\Delta E^{\mathrm{lys-nag}}$) can be expressed as

$$\Delta E^{\mathrm{solv-lys}} = \sum_{I \in \text{solvent}} \sum_{J \in \text{lysozyme}} \Delta E_{IJ}, \tag{10}$$

$$\Delta E^{\mathrm{solv-nag}} = \sum_{I \in \text{solvent}} \sum_{J \in \text{tri-NAG}} \Delta E_{IJ}, \tag{11}$$

$$\Delta E^{\mathrm{lys-nag}} = \sum_{I \in \text{lyzozyme}} \sum_{J \in \text{tri-NAG}} \Delta E_{IJ}, \tag{12}$$

respectively. One should note that the sum of these energies is exactly equal to the total energy; that is,

$$E^{\mathrm{total}} = E^{\mathrm{lys}} + E^{\mathrm{nag}} + E^{\mathrm{solv}} + \Delta E^{\mathrm{solv-lys}} + \Delta E^{\mathrm{solv-nag}} + \Delta E^{\mathrm{lys-nag}}. \tag{13}$$

Such a decomposition of the total energy is available for the isolated lysozyme and tri-NAG. Thus, we can define the binding energy using the following equation:

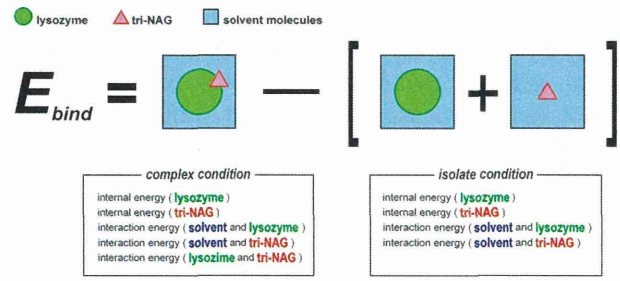

Fig. 4 Definition of binding energy using the results of FMO calculations. The green circle and the red triangle represent the lysozyme and tri-NAG, respectively, and the blue region represents the solvent molecules.

$$E_{\mathrm{bind}} = E^{\mathrm{lys}}_{\mathrm{comp}} + E^{\mathrm{nag}}_{\mathrm{comp}} + \Delta E^{\mathrm{solv-lys}}_{\mathrm{comp}} + \Delta E^{\mathrm{solv-nag}}_{\mathrm{comp}} + \Delta E^{\mathrm{lys-nag}}_{\mathrm{comp}} - (E^{\mathrm{lys}}_{\mathrm{iso}} + E^{\mathrm{nag}}_{\mathrm{iso}} + \Delta E^{\mathrm{solv-lys}}_{\mathrm{iso}} + \Delta E^{\mathrm{solv-nag}}_{\mathrm{iso}}), \tag{14}$$

where the subscripts "comp" and "iso" indicate the values calculated under complex and isolated conditions, respectively. Here, it should be noted that contribution of the internal energy of the solvent molecules is excluded. The definition of this binding energy is also illustrated in Fig. 4.

The primary purpose of our calculations is to compare binding mode1 and 2. The difference of the two binding energies can be evaluated using the following equation:

$$\Delta E^{1-2}_{\mathrm{bind}} = E^{\mathrm{mode1}}_{\mathrm{bind}} - E^{\mathrm{mode2}}_{\mathrm{bind}}, \tag{15}$$

where $E^{\mathrm{mode1}}_{\mathrm{bind}}$ and $E^{\mathrm{mode2}}_{\mathrm{bind}}$ are calculated using eqn (14) for mode1 and 2, respectively. If we need only $\Delta E^{1-2}_{\mathrm{bind}}$, the calculations for the isolated conditions are not necessary because the results of these calculations are the same for the two binding modes. Thus, in this study, only the calculations of the complexes were performed.

We can also obtain detailed information that was divided into each fragment. For example, in eqn (7), the internal energy of the lysozyme is rewritten as

$$E^{\mathrm{lys}} = \sum_{I \in \text{lysozyme}} \left(E'_I + \frac{1}{2} \sum_{J \neq I \in \text{lysozyme}} \Delta E_{IJ} \right). \tag{16}$$

Thus, the contribution of the fragment I to the internal energy is defined as

$$E^{\mathrm{lys}}_I = E'_I + \frac{1}{2} \sum_{J \neq I \in \text{lysozyme}} \Delta E_{IJ}. \tag{17}$$

Since a residue was basically treated as a single fragment in this study, the E^{lys}_I was considered as the contribution of each residue to the internal energy of the lysozyme. Here, a contribution of two cysteines connected with each other could not be divided into each residue because they were united into one fragment (only the sum of the two residues was obtained). Similarly, the interaction energies of the lysozyme with the

solvent molecules and tri-NAG can be rewritten as

$$\Delta E^{\text{solv-lys}} = \sum_{I \in \text{lysozyme}} \left(\sum_{J \in \text{solvent}} \Delta E_{IJ} \right), \quad (18)$$

$$\Delta E^{\text{lys-nag}} = \sum_{I \in \text{lysozyme}} \left(\sum_{J \in \text{tri-NAG}} \Delta E_{IJ} \right), \quad (19)$$

respectively. Thus, the contribution of each residue to the interaction energies of the lysozyme with the solvent molecules and tri-NAG is defined as

$$\Delta E_I^{\text{solv-lys}} = \sum_{J \in \text{solvent}} \Delta E_{IJ}, \quad (20)$$

$$\Delta E_I^{\text{lys-nag}} = \sum_{J \in \text{tri-NAG}} \Delta E_{IJ}, \quad (21)$$

respectively. In the next section, these values are used for the detailed analysis of the stability of the two complexes.

3. Results and discussion

3.1 Experimental thermodynamic parameters

To obtain experimental thermodynamic parameters, we performed ITC and SPR measurements. Fig. 5A and B show the isotherm and the plotted titration curve for the binding of tri-NAG to lysozyme at 25 °C, respectively. The fitting curves were drawn assuming a one set of sites model to obtain the thermodynamic parameters (binding constant (K_D), enthalpy change (ΔH), and binding stoichiometry (n), Gibbs free-energy change (ΔG), and entropy change (ΔS)) using standard thermodynamic relationships. The results are summarized in Table 1. The experimental thermodynamic parameters were also obtained using SPR. Fig. 6A shows the Biacore response during the binding experiments at 25 °C. The responses at the equilibrium state during the sample injection phase were plotted against the tri-NAG concentration (Fig. 6B) and fitted using two-state binding models to obtain K_D values. The K_D values were determined at five different temperatures (15, 20, 25, 30, and 35 °C), as shown in Fig. 6C. Thermodynamic parameters (ΔG, ΔH, and $T\Delta S$) were determined from van't Hoff analysis of the K_D values (Fig. 6D and Table 1).

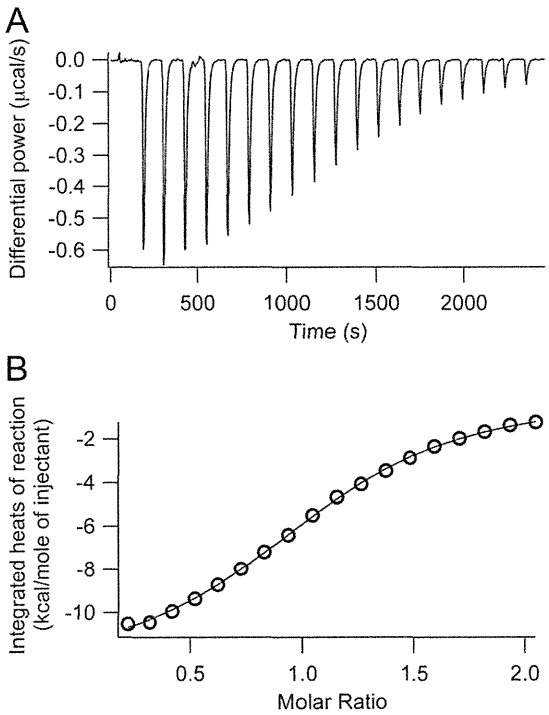

Fig. 5 ITC measurements for experimental thermodynamic parameters. The isotherm (A) and the plotted titration curve (B) for tri-NAG binding to lysozyme at pH 4.6 and 25 °C. The fitting curves were drawn assuming a one set of sites model to obtain the binding constant (K_D), enthalpy changes (ΔH), and binding stoichiometry (n).

Table 1 Experimental thermodynamic parameters for the binding of hen egg-white lysozyme and tri-NAG obtained from ITC and SPR experiments at 25 °C

	ITC at pH 4.6	ITC at pH 7.3	SPR at pH 7.4
K_D	7.3 μM	9.8 μM	19.3 μM
N	1.07	1.12	n.d.
ΔG	−6.98 kcal mol^{-1}	−6.81 kcal mol^{-1}	−6.45 kcal mol^{-1}
ΔH	−12.7 kcal mol^{-1}	−12.3 kcal mol^{-1}	−9.10 kcal mol^{-1}
$T\Delta S$	−5.69 kcal mol^{-1}	−5.48 kcal mol^{-1}	−2.6 kcal mol^{-1}

ITC is a standard quantitative technique for the direct measurement of thermodynamic binding parameters. The parameters obtained from SPR measurements were consistent with those obtained from ITC measurements, indicating that SPR is also useful for obtaining those thermodynamic parameters. An advantage of using SPR is that the amount of protein used for SPR experiments is much smaller in comparison with that used for ITC experiments. Therefore, it is cost-effective owing to its minimal sample amount and preparation requirements.

3.2 Calculation of the binding energy difference

In Table 2, the calculated energies are summarized for binding mode1 and 2. These are average values over the 40 calculations. Their standard deviations are also given. The contribution of the internal energies of the lysozyme and tri-NAG was favorable for binding mode2 because their internal energy differences were positive (+69.48 and +13.47 kcal mol^{-1} for $\Delta E_{\text{comp}}^{\text{lys}}$ and $E_{\text{comp}}^{\text{nag}}$, respectively). On the other hand, the three interaction energies were favorable for binding mode1 because their differences were negative (−53.63, −21.29, and −14.20 kcal mol^{-1} for $\Delta E_{\text{comp}}^{\text{solv-lys}}$, $\Delta E_{\text{comp}}^{\text{solv-nag}}$, and $\Delta E_{\text{comp}}^{\text{lys-nag}}$, respectively). As a result, $\Delta E_{\text{bind}}^{1-2}$ was determined to be −6.17 kcal mol^{-1} by summing up these five values. Thus, we can conclude that binding mode1 is more stable than binding mode2. This result is consistent with the structural information gathered from X-ray analysis. Additionally, these results can provide answer to the question given in the previous section, that is, the main reason why binding mode1 is more stable than binding mode2 is that the solvent interaction energy is very favorable for binding mode1.

The binding energy evaluated with these FMO calculations is an enthalpic contribution of the binding free-energy. In this

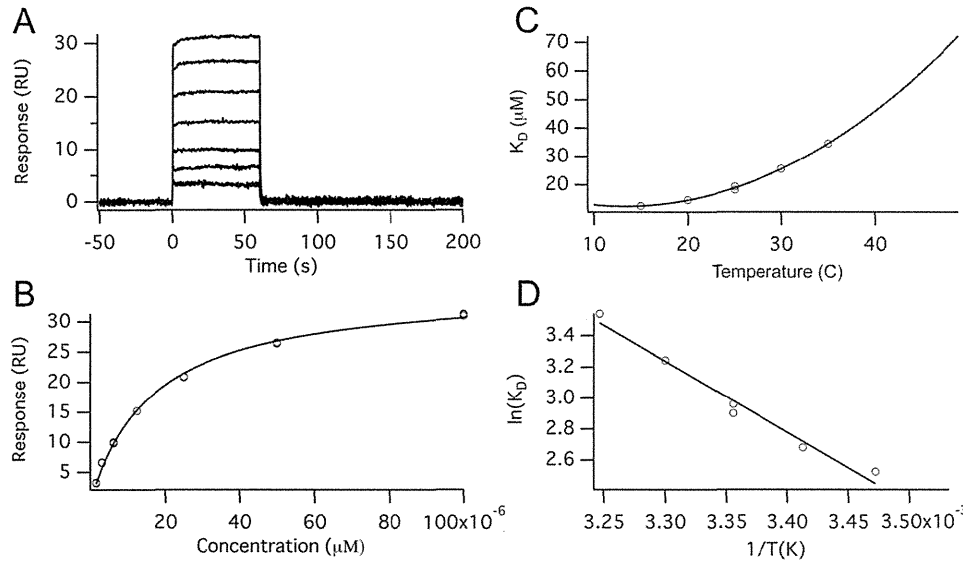

Fig. 6 SPR measurements for experimental thermodynamic parameters. Biacore response for the binding of tri-NAG to lysozyme at pH 7.4 and 25 °C (A). Tri-NAG concentrations were 1.56, 3.13, 6.25, 12.5, 25.0, 50.0, and 100 μM (from bottom to top). The responses at the equilibrium state during sample injection phase against tri-NAG concentration (B). The fitting curves were drawn assuming a 1 : 1 binding model to obtain the binding constant (K_D). Temperature dependence of the K_D values (C). van't Hoff analysis of the K_D values to obtain thermodynamic parameters (D).

Table 2 Averaged results over the 40 calculations for binding mode1 and 2 (kcal mol^{-1}) using the RI-MP2 method with cc-pVDZ. The definition of each value is given in the text. The standard deviations of the values are also given. The difference was obtained by subtracting the value of mode2 from that of mode1 (*i.e.*, a negative value is favorable for binding mode1)

	Mode1		Mode2		Difference
	Energy	Deviation	Energy	Deviation	Energy
ΔE_{comp}^{lys}	−32794793.62	59.75	−32794863.10	60.47	+69.48
ΔE_{comp}^{nag}	−1443674.35	6.42	−1443687.82	6.98	+13.47
$\Delta E_{comp}^{solv-lys}$	−6441.41	261.03	−6387.78	277.81	−53.63
$\Delta E_{solv-nag\ comp}$	−230.87	22.84	−209.58	19.49	−21.29
$\Delta E_{comp}^{lys-nag}$	−192.55	13.76	−178.38	14.60	−14.20
ΔE_{bind}^{1-2}					−6.17

study, not only binding free-energy but also enthalpic contribution were determined experimentally. Thus, although the value evaluated here was not the binding energy but their difference, it should be noted that the order of magnitude of this value was the same as those of the experimental values, *i.e.*, −9.10 kcal mol^{-1} (SPR at pH 7.4) or −12.3 kcal mol^{-1} (ITC at pH 7.3). In a previous study,[21] Nakanishi *et al.* reported that protein–ligand interaction energies calculated under vacuum conditions (70–100 kcal mol^{-1}) could be largely reduced by inclusion of the solvent effects (10–13 kcal mol^{-1}), which were comparable to a typical experimental value (this result was also discussed in another paper[8]). In their study, the solvent effects were evaluated using classical methods with empirical parameters. On the other hand, in the present study we evaluated the solvent effects using quantum chemical calculations of the interaction energies between the solvent molecules and protein, which could also provide the binding energies in the same order of magnitude as the experimental values.

Next, we focus our attention on the standard deviations, which reflect the influence of geometrical fluctuation on the enthalpic contribution of binding free-energy. One can immediately note that these standard deviations are large in comparison with the final energy value (−6.17 kcal mol^{-1}). Especially, the standard deviation of the interaction energy between the solvent molecules and lysozyme ($E_{comp}^{solv-lys}$) is the largest, which reflects high mobility of the solvent molecules. These large standard deviations are due to the fact that the global minimum of the potential energy surface, where the structures obtained from the classical simulation were placed, is globally flat but not necessarily smooth. Another reason for the large standard deviations is some mismatch between classical treatment and quantum chemical treatment. These standard deviations indicate that large number of sampling is needed for evaluation of the binding energy. Here, we do not consider that 40 snap-shots are sufficient for a rigorous evaluation of the binding energy. However, we can expect that our treatment with 40 snap-shots enables us to introduce influence of geometrical fluctuation into the following analyses of the interaction between lysozyme and tri-NAG.

3.3 Detailed analysis of the interaction

In this subsection, we give a detailed analysis of the stability of the two binding modes using the values divided into each residue. Fig. 7 shows the direct interaction energies between the lysozyme and tri-NAG (*i.e.*, $\Delta E_I^{lys-nag}$ in eqn (21)) for the residues within 3.0 Å from the tri-NAG. The values of the charged residues are given separately from the other residues.

For the charged residues, Asp101 had a very large interaction energy in binding mode1, and Glu35 and Asp52 had large interaction energies in binding mode2. This is consistent with the positions of the residues (see Fig. 8). From these results,

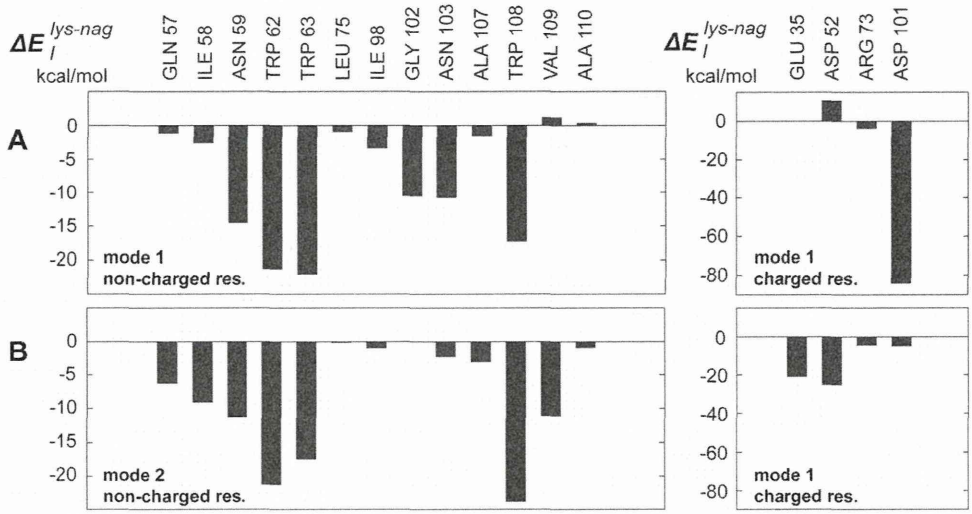

Fig. 7 Interaction energies between tri-NAG and the residues ($\Delta E_I^{lys-nag}$) within 3.0 Å from tri-NAG. These results were obtained using the averaged values over the 40 calculations.

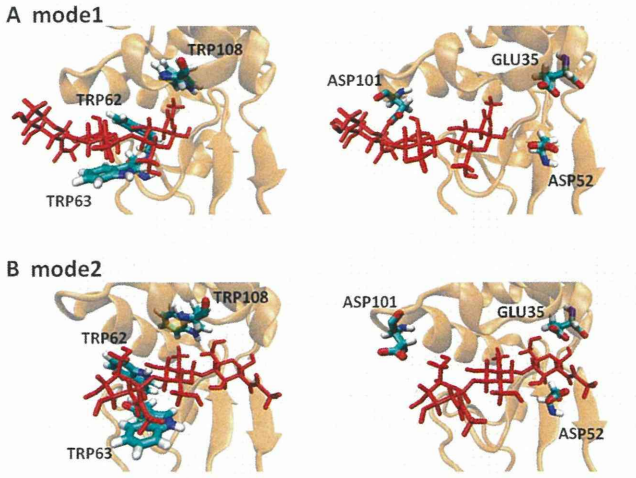

Fig. 8 Positions of tri-NAG and several residues. Tri-NAG is represented with red sticks. This image was created using VMD software.[31]

Table 3 Direct interaction energy with the lysozyme ($\Delta E_I^{lys-nag}$), the interaction energy with the solvent molecules ($\Delta E_I^{solv-lys}$), and the internal energy of the residues (ΔE_I^{lys}) for the two binding modes (kcal mol^{-1}) using the RI-MP2 method with cc-pVDZ. These values are averaged results over the 40 calculations. The difference is calculated by subtracting the value of mode2 from that of mode1 (*i.e.*, a negative value is favorable for binding mode1)

		$\Delta E_I^{lys-nag}$	$\Delta E_I^{solv-lys}$	ΔE_I^{lys}	Total
Glu35	Mode1	+0.16	−93.05	−297099.52	
	Mode2	−21.08	−11.81	−297113.31	
	Difference	+21.24	−81.24	+12.79	−47.21
Asp52	Mode1	+10.72	−95.30	−272515.45	
	Mode2	−25.33	−17.77	−272523.94	
	Difference	+36.05	−77.53	+8.49	−32.99
Asp101	Mode1	−84.09	−12.04	−272499.34	
	Mode2	−4.88	−49.17	−272518.67	
	Difference	−79.21	+37.13	+19.33	−22.75
Trp62	Mode1	−21.32	−19.97	−381645.03	
	Mode2	−21.27	−24.47	−381631.97	
	Difference	−0.05	+4.50	−13.06	−8.61

we can consider that Asp101 is important for binding mode1, and Glu35 and Asp52 are important for binding mode2. In other words, we can realize that Asp101 is the residue that is favorable for binding mode1, and Glu35 and Asp52 are the residues that are favorable for binding mode2. However, the contribution of each residue to the stability of the complex is caused not only by the direct interaction energy with the lysozyme ($\Delta E_I^{lys-nag}$ in eqn (21)) but also by the interaction energy with the solvent molecules ($\Delta E_I^{solv-lys}$ in eqn (20)) and the internal energy of the residues (E_I^{lys} in eqn (17)). In Table 3, we summarize the total contribution of these energies for Asp101, Glu35, and Asp52. As shown in this table, the differences in the total contribution between the two binding modes were −47.21 and −32.99 kcal mol^{-1} for Glu35 and Asp52, respectively, indicating that these residues were favorable for binding mode1. This was the opposite outcome compared to the case where only the direct interaction energy was considered. Table 3 shows that the Glu35 and Asp52 in binding mode1 do not contribute to the stability of the complex through the direct interaction (+0.16 kcal mol^{-1} and +10.72 kcal mol^{-1}, respectively), but they considerably contribute to the stability of the complex through the interaction with the solvent molecules (−93.05 kcal mol^{-1} and −95.30 kcal mol^{-1}, respectively), resulting in the opposite outcome.

The three tryptophans—Trp62, Trp63, and Trp108—had large interaction energies with tri-NAG in both binding modes. The positions of these tryptophans are shown in Fig. 8. As mentioned in the previous section, Maenaka *et al.* reported that binding mode2 is observed in addition to binding mode1 for a mutant of Trp62Tyr, while only binding mode1 is observed in the wild type.[23] This fact indicates that the relative stability of

Table 4 Sum of the Mulliken atomic charges over all atoms in the lysozyme and tri-NAG. The formal charges of the lysozyme and tri-NAG are +8.0 and 0.0, respectively. These results were obtained by the averaged values over the 40 calculations

	Mode1	Mode2
Lysozyme	+5.16 e	+4.41 e
Tri-NAG	−0.17 e	−0.15 e

binding mode1 is lost by the replacement of Trp62. The total contribution of this residue to the stability of the complex obtained from our calculations is shown in Table 3. If we consider only the direct interaction energy with tri-NAG, the difference between the two binding modes cannot be found because they were almost the same (−21.32 and −21.27 kcal mol^{-1} for binding mode1 and 2, respectively). But, the difference of the total contribution to the stability was negative (−8.61 kcal mol^{-1}), showing that Trp62 was favorable for binding mode1 in our studies as well. Thus, our results indicate that the relative stability of binding mode1 is reduced by the replacement of Trp62. This is consistent with the experimental study.

3.4 Analysis of the electron density

Herein, we discuss the electron density evaluated using eqn (5), where Mulliken population analysis with the HF method was employed. All the values used here are averaged values over the 40 calculations.

In the lysozyme, there are 17 positively charged residues (6 lysines and 11 arginines) and 9 negatively charged residues (7 asparagine acids and 2 glutamine acids), resulting in a formal charge of +8.0 e. This formal charge may be reduced by electron transfer from the solvent molecules. In Table 4, the sum of the atomic charges of the lysozyme and tri-NAG obtained using Mulliken population analysis is summarized for binding mode1 and 2. For binding mode1, the total atomic charge of the lysozyme was +5.16 e. Thus, we can consider that a negative charge of −2.84 e was transferred from the solvent molecules. Similarly, for binding mode2, a negative charge of −3.59 e could also be considered as being transferred from solvent molecules. This result indicates that an explicit solvent model should be used for calculations of highly charged proteins like lysozyme.

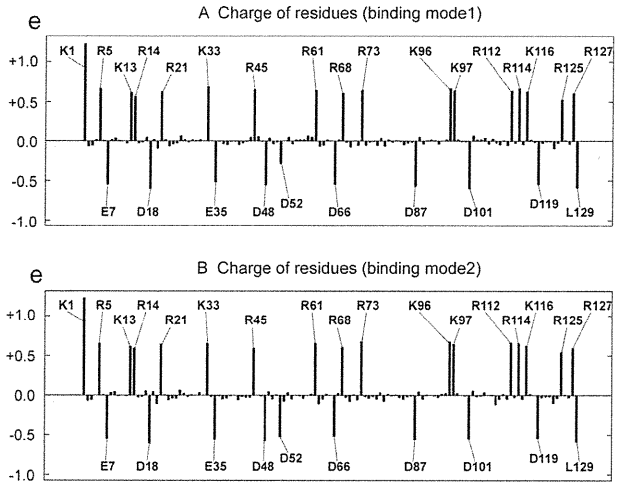

Fig. 9 Charge of each residue evaluated using the Mulliken atomic populations for the two binding modes. These results were obtained using the averaged values over the 40 calculations.

By summing up the atomic charges over the residue, we can obtain the total charge divided into each residue. These data are illustrated in Fig. 9. All charged residues have an amplitude of 0.5–0.6 e; on the other hand, all non-charge residues have a value around 0.0 e. Note that the charge of the N-terminal residue (Lys1) was approximately 1.20 e because the main chain of this residue was capped with $-NH_3^+$ (*i.e.*, its formal charge was +2.0 e). Similarly, the charge of the C-terminal residue (Leu129) was approximately −0.60 e (the main-chain was capped with $-COO^-$ and its formal charge was −1.0 e). This result indicates that the formal charge of charged residues is largely reduced by the electron transfer with the solvent molecules. We consider that atomic charges obtained by the FMO method can be used for determination of empirical potentials of amino acid residues. Actually, studies to determine the atomic charges in empirical force fields with the FMO method have been reported by Okiyama *et al.*[39,40]

4. Summary

In this study, we calculated the difference of the two binding modes for tri-NAG and lysozyme using the FMO method. Here, the RI-MP2 method was employed to evaluate the van der Waals interaction. The solvent effect was considered with the explicit solvent model, which enabled us to discuss charge transfer interactions. The effect of the geometrical fluctuation could be included using the averaged results over 40 calculations with various structures. In addition to theoretical investigations, we performed ITC and SPR measurements to determine the enthalpic contribution in the binding free-energy.

As the difference of the binding energies between the two binding modes is −6.17 kcal mol^{-1}, we concluded that binding mode1 was more stable than binding mode2. This is consistent with an experimental result obtained from the X-ray structure analysis. Our calculations also revealed the reason why binding mode1 is more stable than binding mode2. The amplitude of this binding energy difference was of the same order as the enthalpic contribution determined by ITC (−12.3 kcal mol^{-1}) or SPR (−9.10 kcal mol^{-1}). This result showed that reasonable enthalpy could be obtained by including the solvent effect with an explicit solvent model. By analysis of the energies divided into each residue, Trp62 was favorable for binding mode1, which was consistent with a previous mutation study. The electron density obtained in our calculation showed that the formal charge of the lysozyme (+8.0 e) was reduced to +5.16 and +4.41 e by charge transfer for binding mode1 and 2, respectively.

Acknowledgements

We thank Dr Yasushi Sakaguchi of DKSH Japan K.K. for performing ITC measurements. The calculations were performed using the computational resources offered at Gifu University.

References

1 J. Srinivasan, T. E. Ceatham III, P. Cieplak, P. A. Kollman and D. A. Case, *J. Am. Chem. Soc.*, 1998, **120**, 9401.
2 P. A. Bash, U. C. Singh, F. K. Brown, R. Langridge and P. A. Kollman, *Science*, 1987, **235**, 574.
3 K. Kitaura, T. Sawai, T. Asada, T. Nakano and M. Uebayasi, *Chem. Phys. Lett.*, 1999, **312**, 319.
4 K. Kitaura, E. Ikeo, T. Asada, T. Nakano and M. Uebayasi, *Chem. Phys. Lett.*, 1999, **313**, 701.
5 D. G. Fedorov and K. Kitaura, *J. Phys. Chem. A*, 2007, **111**, 6904.
6 D. G. Fedorov and K. Kitaura, *The Fragment Molecular Orbital method: Practical Applications to Large Molecular System*, CRC Press, Boca Raton, FL, 2009.
7 T. Nakano, T. Kaminuma, T. Sato, K. Fukuzawa, Y. Akiyama, M. Uebayasi and K. Kitaura, *Chem. Phys. Lett.*, 2002, **351**, 475.
8 D. G. Fedorov, T. Nagata and K. Kitaura, *Phys. Chem. Chem. Phys.*, 2012, **14**, 7562.
9 M. Ito, K. Fukuzawa, Y. Mochizuki, T. Nakano and S. Tanaka, *J. Phys. Chem. B*, 2007, **111**, 3525.
10 M. Ito, K. Fukuzawa, Y. Mochizuki, T. Nakano and S. Tanaka, *J. Phys. Chem. A*, 2008, **112**, 1986.
11 M. Ito, K. Fukuzawa, T. Ishikawa, Y. Mochizuki, T. Nakano and S. Tanaka, *J. Phys. Chem. B*, 2008, **112**, 12081.
12 T. Iwata, K. Fukuzawa, K. Nakajima, S. H. Aida, Y. Mochizuki, H. Watanabe and S. Tanaka, *Comput. Biol. Chem.*, 2008, **32**, 198.
13 K. Fukuzawa, Y. Mochizuki, S. Tanaka, K. Kitaura and T. Nakano, *J. Phys. Chem. B*, 2006, **110**, 16102.
14 K. Fukuzawa, Y. Komeiji, Y. Mochizuki, A. Kato, T. Nakano and S. Tanaka, *J. Comput. Chem.*, 2006, **27**, 948.
15 T. Ishikawa, Y. Mochizuki, S. Amari, T. Nakano, H. Tokiwa, S. Tanaka and K. Tanaka, *Theor. Chem. Acc.*, 2007, **118**, 937.
16 T. Ishikawa, Y. Mochizuki, S. Amari, T. Nakano, S. Tanaka and K. Tanaka, *Chem. Phys. Lett.*, 2008, **463**, 189.
17 S. Amari, M. Aizawa, J. Zhang, K. Fukuzawa, Y. Mochizuki, Y. Iwasawa, K. Nakata, H. Chuman and T. Nakano, *J. Chem. Inf. Comput. Sci.*, 2006, **46**, 221.
18 M. P. Mazanetz, O. Ichihara, R. J. Law and M. Whittaker, *J. Cheminf.*, 2011, **3**, 2.
19 T. Sawada, D. G. Fedorov and K. Kitaura, *J. Am. Chem. Soc.*, 2010, **132**, 16862.
20 K. Fukuzawa, K. Kitaura, M. Uebayasi, K. Nakata, T. Kaminuma and T. Nakano, *J. Comput. Chem.*, 2005, **26**, 1.
21 I. Nakanishi, D. G. Fedorov and K. Kitaura, *Proteins: Struct., Funct., Bioinf.*, 2007, **68**, 145.
22 J. C. Cheetham, D. J. Artymiuk and D. C. Phillips, *J. Mol. Biol.*, 1992, **224**, 613.
23 K. Maenaka, M. Matsushima, H. Song, F. Sunada, K. Watanabe and I. Kumagai, *J. Mol. Biol.*, 1995, **247**, 281.
24 N. Kamiya, Y. Yonezawa, H. Nakamura and J. Higo, *Proteins: Struct., Funct., Bioinf.*, 2008, **70**, 41.
25 T. Wiseman, S. Williston, J. F. Brandts and L.-N. Lin, *Anal. Biochem.*, 2012, **421**, 26.
26 D. G. Fedorov, K. Kitaura, H. Li, J. H. Jensen and M. S. Gordon, *J. Comput. Chem.*, 2006, **27**, 976.
27 D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling III, J. Wang, F. Paesani, R. E. Duke, R. Luo, K. M. Merz, D. A. Pearlman, M. Crowley, R. C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K. F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Peroza, D. H. Mathews, C. Schafmeister, W. S. Ross and P. A. Kollman, *AMBER 9*, University of California, San Francisco, 2006.
28 V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins: Struct., Funct., Bioinf.*, 2006, **65**, 712.
29 K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley and R. J. Woods, *J. Comput. Chem.*, 2008, **29**, 622.
30 W. L. Jorgensen, J. Chandrasekhar, J. Madura and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926.
31 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33.
32 Y. Komeiji, T. Ishida, D. G. Fedorov and K. Kitaura, *J. Comput. Chem.*, 2007, **28**, 1750.
33 T. Ishikawa, T. Ishikura and K. Kuwata, *J. Comput. Chem.*, 2009, **30**, 2594.
34 T. Ishikawa and K. Kuwata, *J. Chem. Theor. Comput.*, 2010, **6**, 538.
35 M. Feyereisen, G. Fitzgerald and A. Komornicki, *Chem. Phys. Lett.*, 1993, **208**, 359.
36 T. Ishikawa and K. Kuwata, *Chem. Phys. Lett.*, 2009, **474**, 195.
37 T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007.
38 F. Weigend, A. Köhn and C. J. Hätting, *Chem. Phys.*, 2002, **116**, 3175.
39 Y. Okiyama, H. Watanabe, K. Fukuzawa, T. Nakano, Y. Mochizuki, T. Ishikawa, S. Tanaka and K. Ebina, *Chem. Phys. Lett.*, 2007, **449**, 329.
40 Y. Okiyama, H. Watanabe, K. Fukuzawa, T. Nakano, Y. Mochizuki, T. Ishikawa, K. Ebina and S. Tanaka, *Chem. Phys. Lett.*, 2009, **467**, 417.

Chemical Physics Letters 559 (2013) 94–98

Contents lists available at SciVerse ScienceDirect

Chemical Physics Letters

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

Free-energy analysis of lysozyme–triNAG binding modes with all-atom molecular dynamics simulation combined with the solution theory in the energy representation

Kazuhiro Takemura [a], Raghunadha Reddy Burri [a,b], Takeshi Ishikawa [b,c], Takakazu Ishikura [c], Shun Sakuraba [d,e], Nobuyuki Matubayasi [b,d], Kazuo Kuwata [b,c], Akio Kitao [a,b,*]

[a] Institute of Molecular and Cellular Biosciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan
[b] Japan Science and Technology Agency, Core Research for Evolutionary Science and Technology, Japan
[c] Center for Emerging Infectious Diseases, Gifu University, 1-1, Yanagido, Gifu 501-1194, Japan
[d] Institute for Chemical Research and Elements Strategy Initiative for Catalysts and Batteries, Kyoto University, Uji, Kyoto 611-0011, Japan
[e] Institute for Molecular Science, National Institute of Natural Sciences, Myodaiji, Okazaki 444-8585, Japan

ARTICLE INFO

Article history:
Received 21 September 2012
In final form 31 December 2012
Available online 7 January 2013

ABSTRACT

We propose a method for calculating the binding free energy of protein-ligand complexes using all-atom molecular dynamics simulation combined with the solution theory in the energy representation. Four distinct modes for the binding of tri-*N*-acetyl-D-glucosamine (triNAG) to hen egg-white lysozyme were investigated, one from the crystal structure and three generated by docking predictions. The proposed method was demonstrated to be used to distinguish the most plausible binding mode (crystal model) as the lowest binding energy mode.

1. Introduction

Proteins function by interacting with other molecules. Investigating the complex structures of proteins and their intermolecular interactions is thus a key to understanding protein function. Molecular dynamics (MD) simulation is widely employed to investigate molecular interactions. Free energy analysis methods [1–3] such as free energy perturbation (FEP) [1,4] and linear interaction energy (LIE) [5] are useful for evaluating binding free energies during MD simulations. FEP-type calculations often require simulating multiple intermediate states between the two states of interest [1,2]. The LIE method enables faster calculation of the binding free energy without the need to simulate intermediate states; however, three parameters (α, β, and χ) must be adjusted empirically [5].

One of the authors developed a theory of solution in the energy representation (ER) [6–8] in which the solvation fee energy can be obtained from relatively short molecular simulations performed at two end-points, the pure solvent and solution system, and that does not involve any adjustable parameters. This method was employed to calculate highly accurate solvation free energies of various amino acid side chain analogs [9], and the data were comparable to the experimental values [10] and the values calculated using the Bennett acceptance ratio method [11].

We utilized the ER method to analyze the protein-ligand binding mode. Considering thermodynamics cycle that connects complex formations *in vacuo* and in solution, we calculated the binding free energy from the conformational and solvation free energies and entropy of the solute molecules. In this Letter, we focus on the binding of tri-*N*-acetyl-D-glucosamine (triNAG) to hen egg-white lysozyme, which is an exceptionally well-characterized glycoside hydrolase protein. The main purposes of this Letter were to examine if the proposed energy calculation method can distinguish the crystal binding mode as the lowest binding energy mode from the complex model structures generated by docking predictions. Protein-ligand docking comprises generation of various complex structures and evaluation of the generated structures. We focus on the latter stage in this Letter.

The lysozyme molecule has six binding pockets for glucosamine, designated as A–F. TriNAG and tetraNAG bind to the A–B–C and A–B–C–D sites, respectively. Maenaka et al. found that the Trp62Try mutant of lysozyme has two binding modes, A–B–C and B–C–D [12]. Kamiya et al. performed flexible docking simulations of triNAG binding to lysozyme using an extensive multicanonical MD simulation with explicit water sphere around the binding sites and showed that the most populated binding mode is separated from the other binding modes by a free energy barrier [13]. Using three different docking programs, AutoDock3 [14], GOLD [15], and Molegro Virtual Docker (MVD) [16], we generated three binding modes which are all significantly different from the most stable A–B–C binding mode in crystal [17]. We performed short MD simulations on the aforementioned binding modes and

* Corresponding author at: Institute of Molecular and Cellular Biosciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan. Fax: +81 3 5841 2297.
E-mail address: kitao@iam.u-tokyo.ac.jp (A. Kitao).

0009-2614/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cplett.2012.12.063