Biophysical Chemistry 156 (2011) 31–42

Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: http://www.elsevier.com/locate/biophyschem

Structural characteristics of hydration sites in lysozyme

Kunitsugu Soda [a,*], Yudai Shimbo [a,b], Yasutaka Seki [c], Makoto Taiji [a]

[a] Laboratory for Computational Molecular Design, Center for Computational Life Science, RIKEN, 7-1-26, Minatojima-Minami-machi, Chuo-ku, Kobe, Hyogo 650–0047, Japan
[b] Laboratory of Molecular Biophysics, Department of Bioengineering, Nagaoka University of Technology, 1603–1, Kamitomioka-machi, Nagaoka, Niigata 940–2188, Japan
[c] Laboratory of Structural Biology and Biophysics, School of Pharmacy, Iwate Medical University, Yahaba, Iwate 028–3694, Japan

ARTICLE INFO

Article history:
Received 17 December 2010
Received in revised form 17 February 2011
Accepted 17 February 2011
Available online 23 February 2011

Keywords:
Hydration water
Hydration matrix
Hydration site
Clustered hydration site
H-bond recombination
Crystal water

ABSTRACT

A new method is presented for determining the hydration site of proteins, where the effect of structural fluctuations in both protein and hydration water is explicitly considered by using molecular dynamics simulation (MDS). The whole hydration sites (HS) of lysozyme are composed of 195 single HSs and 38 clustered ones (CHS), and divided into 231 external HSs (EHS) and 2 internal ones (IHS). The largest CHSs, 'Hg' and 'Lβ', are the IHSs having 2.54 and 1.35 mean internal hydration waters respectively. The largest EHS, 'Clft', is located in the cleft region. The real hydration structure of a CHS is an ensemble of multiple structures. The transition between two structures occurs through recombinations of some H-bonds. The number of the experimental X-ray crystal waters is nearly the same as that of the estimated MDS hydration waters for 70% of the HSs, but significantly different for the rest of HSs.

1. Introduction

The 3D-structure formation, intermolecular interaction, ligand binding and so on are crucial processes through which a water soluble globular protein expresses its specific function. In these processes, some parts of the molecular surface of the protein are dehydrated to form intramolecular or intermolecular interactions, as a result of which the polypeptide chain is folded and various types of molecular recognition are fulfilled [1,2]. Many of the nonpolar groups are exposed to water in the early stage of protein folding. Subsequently they are collapsed by hydrophobic effects to form one or more hydrophobic clusters in native state, which contributes to structural stability. In the quaternary structure formation and ligand binding, similar processes to the above occur between nonpolar groups on the surface of subunits and ligand molecules respectively. Many studies have already been carried out on the hydration structure and thermodynamic properties of nonpolar groups and molecules [3–5].

In contrast with nonpolar groups, virtually all the polar groups (PGs) exposed to aqueous solvent are H-bonded to water independently of the conformational states of protein, i.e. in both native and nonnative states. As a result of it, they contribute not only to increase the solubility of proteins but also to minimize the structural stability of their folded or associated structure. It is because the neutral H-bonding energy in vacuum is about 5 kcal mol^{-1} [6,7], which is close to ten times the thermal energy at physiological temperatures. As summarized above, water plays essential roles in many molecular processes involved in the expression of protein functions. Hence, it is indispensable to elucidate the static and dynamic properties of hydration water (HW) both on the surface and in the interior of protein for understanding their molecular mechanisms. In the background, many studies have been made on the structure or spatial distribution [8–22], dynamics [8,18–29] and thermodynamics [30–32] of the hydration water of proteins.

Looked at hydration phenomena from the side of water, waters around a PG are H-bonded to it, while those around a nonpolar group are H-bonded to neighboring waters. This satisfies the thermodynamic condition that waters must almost always form hydrogen bonds to avoid large increase in the free energy of solution. As a result of it, some hydration structure characteristic of an H-bonded pair is formed around a PG of protein. On the other hand, a different hydration structure is formed around a nonpolar group depending on the surface structure of protein. Thus, while the hydration structure of a PG is formed "actively", that of a nonpolar one is formed "passively". In many of the conventional studies, the difference in spatial distribution between hydration water and bulk one has been used as an index for defining hydration sites [9–12,16–22,33–36].

Experimentally, HWs have been identified at the surface and in the interior of many proteins by using X-ray and/or neutron crystallographic analysis [9–12,33–36]. The method, however, has a weakness that HWs are lost at the region involved in crystal packing and X-ray cannot see those waters which are moving rapidly at the surface or whose position and numbers are fluctuating greatly in the interior cavity [37–39]. Several methods have been devised for

* Corresponding author. Tel.: +81 78 940 5645; fax: +81 78 304 4958.
E-mail address: soda@riken.jp (K. Soda).

0301-4622/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2011.02.006

identifying hydration sites by using molecular dynamics simulation (MDS): One method defines the hydration site (HS) of a protein molecule as the position of each peak of the solvent density distribution around it [12,16–19]. To minimize the effect of structural fluctuations in protein, Henchman et al. [20] defined a HS as the time averaged position (TAP) of HWs in the local reference frame (ARC) located on the main-chain of a neighboring surface residue. The other method defines a HS as a group of protein atoms mutually H-bonded by one or more waters for a time longer than a predetermined time [21,22], and so on.

It is established from experimental [40] and computational [41,42] studies that the structure of a protein as well as its hydration water is fluctuating continually under physiological conditions. Naturally the fluctuation of surface groups is larger than that of interior groups. In addition, owing to the recent advances in the performance of parallel computers and their availability, it has become possible to reproduce the molecular process of a protein in solution in the real-time range of μs [42,43]. It has enabled us to find large changes in side-chain conformation and fluctuations in main-chain structure close to local structural transitions near the molecular surface of protein. Though some pioneering studies have been carried out, [20] the common sense has prevailed until recent years that the hydration structure of a protein should be analyzed based on its fixed or average structure such as that obtained from X-ray crystal analysis. Under the present condition, however, the assumption does not hold true [40–42]. In other words, a new method for analyzing hydration structure needs to be developed where the effect of structural fluctuations in protein and hydration water is explicitly considered.

Recognizing those stated above, in the first half of this study, (1) we have worked out a new method for determining the HSs of a protein where its structural fluctuations are explicitly taken into account, and (2) applying the method to time-series data on the atomic coordinates of a lysozyme molecule and HWs, we have analyzed structural characteristics of the HSs of lysozyme. This paper describes results of these studies. Dynamical characteristics have been analyzed in the second half of this study, results of which will be published in a forthcoming paper.

2. Methods of hydration-site analysis

As described in Section 1, under physiological conditions, atomic groups on the surface of a protein molecule change their positions rapidly to be accompanied by some changes in the hydration structure. Therefore, it is inappropriate to describe the hydration structure of a solute protein as a function of spatial coordinates around a fixed protein molecule. Instead, we need to define a hydration structure by the spatial distribution of HWs seen from each atomic group. Then, we obtain its hydration structure as the average of those weighted with the occurrence frequencies of various surface structures. The hydration structure of an atomic group changes with not only its polarity but also its mobility and the shape of the space where it is located. Especially, the effect of mobility will be larger on nonpolar groups as their hydration structures are determined passively. In addition, as the interaction between a nonpolar group and water is weaker than thermal energy at physiological temperatures, there is no physically definite criterion for defining to which atomic group a hydration water should be assigned. On the other hand, one or more water molecules are almost always H-bonded to a PG exposed to water, independently of its location and mobility. Hence, HWs can be distinguished by a definite criterion of whether they are H-bonded or not. In either of the surface and interior of protein, there are many cases in which two atomic groups, A and B, are connected with each other by two hydrogen bonds (HB) through a water molecule. So, we will hereafter call it a bridging hydrogen bond (BHB) regarding it as a polar interaction via a bridging water, and say that "A and B are BH-bonded to each other".

Based on the above consideration, we define a hydration site (HS) as the one or more polar groups (PGs) that satisfy either of the following conditions:

(1) It is a single PG which is H-bonded to one or more water molecules with an average number of HBs larger than a predetermined threshold, n_{hb}.
(2) It is a group of two or more PGs which share one or more water molecules with an average number of HBs larger than a predetermined threshold, n_{sh}, where "two PGs share a water" means that "the PGs are H-bonded simultaneously to an identical water".

According to their definitions, the HW and HS in this study should be correctly called a polar HW and a polar HS respectively. Besides, while an atomic group such as a carboxyl group is in general treated chemically as a single PG each of the donors and/or acceptors in an atomic group is treated in this study as a PG which can be a HS. For example, a protonated carboxyl group is considered to consist of two acceptors, CO, and one donor, OH. As a PG is defined in this way, one water molecule can form one HB at most to any polar group. So, the number of HWs H-bonded to each PG is equal to the number of HBs it forms.

Choosing hen egg white (h.e.w.) lysozyme (PDB ID: 1rfp) as a sample protein and carrying out MDS on its aqueous solution, we have generated a statistical ensemble of diverse conformations for lysozyme and spatial distributions for HWs at thermodynamic equilibrium. These HSs are determined as one of the structural characteristics of this ensemble. The outline of MDS conditions in this study is summarized below:

(1) [Sample solution] one lysozyme molecule, 8351 waters, 25 Na^+ and 33 Cl^- in a dodecahedron cell with a periodic boundary condition;
(2) [MDS package] GROMACS 4.0.5 [44];
(3) [Ensemble] NPT (Nose-Hoover [45,46], Andersen [47,48]) with $P = 0.1$ MPa, $T = 298.15$ K, and the mean cell volume of $\overline{V} = 268.465$ nm^3;
(4) [Force fields] Amber-ff99SB (protein) [49,50], TIP4P-Ew (water) [51] and Ref. [47] (Na^+, Cl^-) [52], and the modified-Wolf method (long-range Coulombic forces) [53];
(5) [Simulation time] $T_S = 300$ ns (time step, $\Delta t_s = 2$ fs).

A combined mass of a lysozyme molecule and its crystal waters are energy-minimized in vacuum and immersed in an aqueous solution of NaCl with a final concentration of 0.15 M. All the non-hydrogen atoms of lysozyme and solvent molecules are given their respective initial velocities corresponding to the Maxwell distribution at $T = 10$ K. Then, the temperature is increased gradually to 298.15 K in 300 ps and the solution is thermally equilibrated for 1 ns. We decide the solution has reached thermal equilibrium when the rmsd between the MDS structure and the X-ray crystal structure has come to a stationary value.

After the equilibration steps, MDS is performed for a real time of 300 ns. Time-series data of the atomic coordinates for a lysozyme molecule, waters and salt ions are sampled at an interval of $\Delta t = 0.2$ ps. The formation of HBs between a PG on the surface and in the interior of protein and a HW is decided at every sample time by using Thornton's criterion [54]. All the HSs of lysozyme are determined by applying the definition of the HS to the results of analysis on HB formation to examine their structural characteristics, details of which are described in the following sections. In this paper, the one-letter code is used for denoting an amino-acid residue. For example, the main-chain CO group of the 35-th glutamic acid residue is denoted as E35_O.

3. Results and discussion

3.1. Determining hydration sites

As described in the preceding section, we can obtain the information on which PG is H-bonded to how many HWs, and which PG shares a HW with which PG at each sample time of simulation. Using the data, we construct a hydration matrix determined definitely from the hydration state at each time of $t = t_k$ ($k = 1,\ldots,N_t$) as

$$\boldsymbol{h}(t_k) = \left\{h_{\mu\nu}(t_k)\right\}, \quad \left(\mu,\nu = 1,\ldots,N_p\right) \tag{1}$$

where N_t is the total number of sample times, N_p is the total number of PGs in protein and $h_{\mu\nu}(t_k)$ is the $\mu\nu$ element of the matrix $\boldsymbol{h}(t_k)$ given as follows:

(a) If the polar group μ is H-bonded to i waters,

$$h_{\mu\mu}(t_k) = i. \tag{2}$$

(b) If the two polar groups, μ and ν ($\mu \neq \nu$) , share j waters with each other,

$$h_{\mu\nu}(t_k) = h_{\nu\mu}(t_k) = j. \tag{3}$$

If a water λ is H-bonded to $w(\mu,\ \lambda;\ t_k)$ polar groups including the polar group μ, $1/w(\mu,\ \lambda;\ t_k)$ the water is assumed to be possessed by each polar group. Next, we obtain the mean hydration matrix $\boldsymbol{h}$ from the whole set of hydration matrices $\{\boldsymbol{h}(t_k)\}$ as

$$\boldsymbol{h} \equiv \frac{1}{N_t}\cdot\sum_{k=1}^{N_t}\boldsymbol{h}(t_k), \ \text{i.e.}\ h_{\mu\nu} \equiv \frac{1}{N_t}\cdot\sum_{k=1}^{N_t}h_{\mu\nu}(t_k). \tag{4}$$

As evident from the definition, all the $\boldsymbol{h}(t_k)$ are symmetric matrices, and so is the $\boldsymbol{h}$. The diagonal element $h_{\mu\mu}$ measures the average number of waters H-bonded to the PG μ and the off-diagonal element $h_{\mu\nu}$ indicates the average number of waters shared by or bridging the PGs μ and ν. Then, we repeat the operation in which, with its symmetry being kept, two appropriately chosen rows and columns of $\boldsymbol{h}$ are replaced with each other simultaneously so that larger $h_{\mu\nu}$'s may gather nearer to the diagonal. Some examples of matrix elements, the schematic view of hydration sites and the hydration matrix obtained after the above operation, which is again denoted as $\boldsymbol{h}$, are shown in Fig. 1.

The diagonal and off-diagonal elements are both virtually zero for the PGs that are completely buried in protein interior and are not H-bonded to any internal water. All pairs of row and column for these PGs are deleted from $\boldsymbol{h}$ in Fig. 1. The figure shows that $\boldsymbol{h}$ has many off-diagonal elements with nonzero but very small values. It indicates that there are many pairs of PGs which are slightly separated from each other and occasionally bridged by a HW. It means also that the smaller a matrix element is, the less stable the corresponding BHB is and the less frequently it occurs. Hence, if we set two thresholds of h^*_{hb} and h^*_{sh} to $h_{\mu\mu}$ and $h_{\mu\nu}$ respectively and neglect the small matrix elements for which the following relation holds,

$$h_{\mu\mu} < h^*_{hb}, h_{\mu\nu} < h^*_{sh}, (\mu \neq \nu), \tag{5}$$

the matrix $\boldsymbol{h}$ becomes a pseudo-diagonal one. It is approximately divided into an ensemble of symmetric matrices each of which corresponds to a small number of PGs combined with BHBs. In Fig. 1, the thresholds are set to

$$h^*_{hb} = h^*_{sh} = 0.20, \tag{6}$$

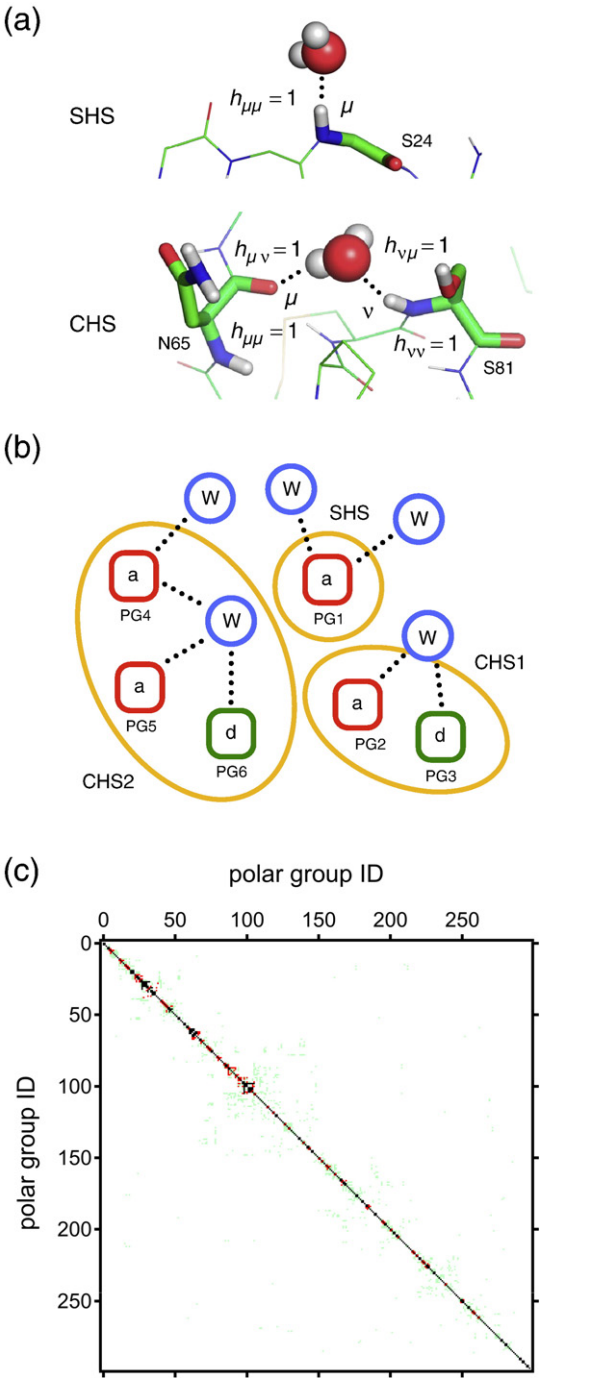

Fig. 1. Hydration matrix and hydration site. (a) A matrix element, $h_{\mu\mu}$, for a PG, μ = S24_NH, in a SHS and four elements, $h_{\mu\mu}$, $h_{\mu\nu}$, $h_{\nu\mu}$ and $h_{\nu\nu}$, for two PGs, μ = N65_O and ν = S81_NH in a CHS. (b) Schematic views of one SHS and two CHSs. The SHS has two EHWs. The two PGs in CHS1 are connected by a BHB via an EHW. CHS2 has an IHW and an EHW, and the three PGs in CHS2 are connected by three BHBs via an IHW. (c) Mean hydration matrix, $\boldsymbol{h}$, for lysozyme in aqueous solution. Both of the row and column numbers are the respective IDs of all polar groups with $h_{\mu\mu}$ larger than 0.20, which amounts to 298. Each colored dot indicates the magnitude of the matrix element, $h_{\mu\nu}$. Black, $h_{\mu\nu} \geq 0.20$; red, $0.10 \leq h_{\mu\nu} < 0.20$; green, $0.01 \leq h_{\mu\nu} < 0.10$. See text for the definition of $h_{\mu\mu}$ and $h_{\mu\nu}$.

and the off-diagonal elements for which Eq. (5) holds are colored red and green in Fig. 1 and will be neglected in the following analysis. The validity of choosing the threshold as above will be discussed later. Denoting again the matrix consisting of the elements colored black in Fig. 1 as $\boldsymbol{h}$, we assume that each of the approximately independent

groups of PGs forms a HS. In the following, the HSs consisting of one PG and two or more PGs are called the single hydration site (SHS) and clustered hydration site (CHS) respectively. Vishveshwara and her colleagues [20,21] define a HS as a group of protein atoms mutually H-bonded through one or more waters for a time longer than a predetermined time to apply it to lysozyme, RNase A and so on. Our definition of a CHS is similar to their definition in the respect that the formation of H-bonds by a bridging water is used in both definitions. However, their definition requires it as a necessary condition for a group of atoms to form a cluster that they are connected with one another by one or more waters for a time longer than a predetermined time. Owing to the difference, the set of HSs of lysozyme identified by us does not agree with that by them [19].

In connection with the definition of a HS, the following must be noted here. As the HS is defined as above, most HWs analyzed in this study are those of PGs composing a HS of lysozyme. There are some HWs of such a PG as not to be recognized as a HS because they are almost buried in protein interior. However, the amount of these waters is estimated to be only 0.9% of the whole HWs. In addition, binding of some other components such as salt ions to a HS decreases in fact the amount of its HWs. This effect, however, was also confirmed negligibly small under the present solution condition. Accordingly, we will consider the structural characteristics of HWs only at HSs in the following.

We will define here the internal and external hydration sites for later discussion. With the united-atom model where an H-atom is united with the atom to which it is bonded, effective van der Waals radii of the non-hydrogen atoms in protein were reported by Tsai et al. [55]. It is well-known that a water molecule can be modeled by a hard sphere with a radius of 0.14 nm [56]. If all the protein atoms and water molecules are approximated by hard spheres with their respective radii described above, the HWs of a protein can be classified into two groups: If a water molecule cannot go out of protein interior geometrically, we call it an internal water and otherwise an external water. In addition, we call an internal water H-bonded to any of the PGs of protein an internal hydration water (IHW). We can estimate the average number of IHWs, $\overline{n}_{\mathrm{hw\,i}}$, at each HS from analysis of the protein structures at N_t times. If the relation of $\overline{n}_{\mathrm{hwi}} \geq h^*_{\mathrm{in}}$ holds for a HS with a predetermined threshold, h^*_{in}, we call the HS an internal hydration site (IHS), and if the reverse relation of $\overline{n}_{\mathrm{hwi}} < h^*_{\mathrm{in}}$ holds, we call it an external hydration site (EHS). We adopted the following value for the threshold, h^*_{in}:

$$h^*_{\mathrm{in}} = 0.10. \tag{7}$$

This threshold is half of those for the HS in Eq. (6). The reason why the lower threshold is taken for h^*_{in} is as follows: It is expected that the HW of such a HS would show static and dynamic characteristics different from that of a site fully exposed to water. It is because the fact that one or more IHWs are shut up in protein interior even with the low probability of 0.1 indicates that the site should have some structural characteristics essential for producing an IHW.

3.2. Characteristic quantities of the hydration site in lysozyme

At the beginning of this section, we will list the notation of the quantities related to the hydration site (HS), hydration water (HW) and hydrogen bond (HB) of a protein molecule in the following:

(1) Numbers: 'N_α' and 'n_α'.
 (a) N_α, the total number of a quantity 'α' for the whole hydration sites;
 (b) n_α, the number of a quantity 'α' in a specified hydration site or polar group.
(2) Suffices: 'α'.
 (a) 'hs', hydration site; 'hs i', internal hydration site; 'hs e', external hydration site;
 (b) 'hw', hydration water; 'hw i', internal hydration water; 'hw e', external hydration water;
 (c) 'hb', hydrogen bond; 'hb i', internal hydrogen bond; 'hb e', external hydrogen bond;
 (d) 'pg', polar group.

The mean value of N_α or n_α is denoted by each variable with an overbar. For example, '$\overline{N}_{\mathrm{hw\,i}}$' denotes the average of the total number of IHWs for the whole HSs.

Denoting the mean numbers of HBs and HWs for a PG , μ, by $\overline{n}_{\mathrm{hb}}(\mu)$ and $\overline{n}_{\mathrm{hw}}(\mu)$, they are given by the following expressions,

$$\overline{n}_{\mathrm{hb}}(\mu) = h_{\mu\mu}, \tag{8}$$

$$\overline{n}_{\mathrm{hw}}(\mu) = \frac{1}{N_t} \cdot \sum_{k=1}^{N_t} \sum_{\lambda=1}^{N_w} 1 / w(\mu, \lambda; t_k), \tag{9}$$

where N_w is the total number of waters in solution. These are the most basic quantities among those characterizing the hydration properties of a PG in a HS. For later discussion, we will denote the mean number of HBs between a PG , μ, and the other PGs in protein by $\overline{n}_{\mathrm{hp}}(\mu)$, which is a quantity complementary to $\overline{n}_{\mathrm{hb}}(\mu)$.

We will here consider the case where the two conditions below are satisfied by a pair of PGs, μ and ν:

(a) They can share only one water molecule 'Wc' at most, and
(b) they can exist in their respective configurational states where they are H-bonded to Wc independently of each other.

Then, the probability, $h_{\mu\nu}$, that the pair shares Wc is given by the following equation:

$$h_{\mu\nu} = p_{\mu c} \times p_{\nu c}, \tag{10}$$

where $p_{\mu c}$ and $p_{\nu c}$ are the probabilities that the PGs, μ and ν, exist in their respective H-bonded states. Though this is the simplest case, Eq. (10) is useful because the above conditions are satisfied at least approximately in many cases.

Based on the above consideration, we have obtained characteristic quantities for the whole HSs by analyzing MDS data of an aqueous solution of lysozyme, which are listed in Table 1.

A lysozyme molecule has 448 PGs. A third or 150 PGs are buried either completely or mostly in protein interior with $h_{\mu\mu}$<0.20 not to

Table 1
List of hydration-related quantities for h.e.w. lysozyme determined by the present analysis. See the note below for their definitions.

Hydration water [a]			Hydrogen bond [a]		
$\overline{N}_{\mathrm{hw}}$	$\overline{N}_{\mathrm{hw\,i}}$	$\overline{N}_{\mathrm{hw\,e}}$	$\overline{N}_{\mathrm{hb}}$	$\overline{N}_{\mathrm{hb\,i}}$	$\overline{N}_{\mathrm{hb\,e}}$
226.82	4.81	222.01	283.77	13.93	269.84
Hydration site [a]			**Polar group**		
N_{hs}	$N_{\mathrm{hs\,i}}$	$N_{\mathrm{hs\,e}}$	N_{p} [b]	$N_{\mathrm{p\,b}}$ [b]	N_{pg} [a]
233	2	231	448	150	298

Frequency of the cluster size of hydration sites [c]

n_{pg} [c]	1	2	3	4	6	9	11
$N_{\mathrm{hs}}(n_{\mathrm{pg}})$ [c]	195	29	5	1	1	1	1

[a] N_α, total number of α for the whole hydration sites. $\overline{N}_\alpha$, mean of N_α. Suffix: hw, hydration water; hb, hydrogen bond; hs, hydration site; pg, polar group; i, internal; e, external.
[b] N_p, total number of polar groups in the whole protein; N_{pb}, total number of polar groups buried in protein interior; $N_p = N_{pb} + N_{pg}$.
[c] n_{pg}, number of polar groups; $N_{hs}(n_{pg})$, total number of hydration sites with cluster size n_{pg}

be recognized as an HS. Either being exposed to solvent or H-bonded to internal water, the rest of 298 PGs are involved in HS formation. As shown in Table 1, 195 (84%) out of the whole 233 HSs are SHSs, which means that they account for about 2/3 (65%) the 298 substantially hydrated PGs. There are 38 CHSs, about 3/4 (76%) of which are the sites with a cluster size of 2, i.e. $n_{pg}=2$. There are only nine CHSs with $n_{pg}>2$.

Priya et al. [19] determined 150 HSs from the position of peaks with $g(\boldsymbol{r})>2$ for the density distribution function, $g(\boldsymbol{r})$, of solvent water obtained from MDS on an aqueous solution of h.e.w. lysozyme. The total number of HSs in the present study, 233, is as much as 55% larger than that of Priya et al. This difference results from the situation that, in our analysis, (a) a PG with $h_{\mu\mu}$ significantly smaller than one is recognized as a HS and (b) a PG around which the peak height of the density distribution of solvent water is apparently low due to its large-amplitude motion is correctly identified as a HS.

We can see from Table 1 that the whole 233 HSs form 283.77 ($\overline{N}_{hb}$) HBs to 226.82 ($\overline{N}_{hw}$) HWs on average. The number of the HWs corresponds to 0.97 HW per site, but to only 0.76 HW per PG. The result that the number of $\overline{N}_{hw}/N_{pg}$ is significantly smaller than 1.0 will be caused by the situation that (a) an identical water molecule is shared by more than one PG in many CHSs, (b) according to Thornton's criterion, the PGs exposed to water are not always H-bonded to surrounding waters, and (c) there are some PGs of which probability that they are H-bonded to water is significantly lower than 1.0, though their $h_{\mu\mu}$'s exceed the threshold h^{*}_{hb}. On the other hand, the mean numbers of $\overline{n}_{hb}/n_{pg}$ for carbonyl and carboxyl groups on EHS are both larger than 1.0 as 1.15 and 2.03 respectively. As an accumulated sum of various contributions, the mean number of $\overline{N}_{hb}/N_{pg}$ is apparently close to 1.0 as $283.77/298=0.95$.

Naturally, the mean number of IHWs is exceedingly smaller than that of EHWs: We have estimated at $\overline{N}_{hw\,i}=4.81$ and $\overline{N}_{hw\,e}=222.01$, which are 2.1% and 97.9% of the total HWs respectively. Compared with other proteins of similar sizes, however, this estimate of $\overline{N}_{hw\,i}$ for lysozyme corresponds to the class of large $\overline{N}_{hw\,i}$. The mean numbers of HBs to PGs per HW are 1.22 and 2.90 for EHWs and IHWs respectively, demonstrating that the latter is about 2.4 times larger than the former. It will show that most IHWs are shared among the PGs in protein to meet requirements of both IHWs and PGs for H-bonding.

The constitutive PGs and hydration-characteristic quantities are listed in Table 2 for the whole CHSs and IHSs of lysozyme. The three largest CHSs are Hg, Lβ and Clft with cluster sizes (n_{pg}) of 11, 9 and 6 respectively.

The largest two, i.e. Hg and Lβ, are IHSs and their mean numbers of IHWs, $\overline{n}_{hw\,i}$, are 2.54 and 1.38 respectively. The sum of both $\overline{n}_{hw\,i}$'s is 3.92, which accounts for 81.5% of the mean total number of IHWs in the whole HSs, $\overline{N}_{hw\,i}=4.81$, shown in Table 1. The rest of 0.89 IHWs is the sum of small contributions from the sites with $\overline{n}_{hw\,i}<0.10$. The five sites with $\overline{n}_{hw\,i}=0.05$–$0.1$ and the others have their contributions of 0.31 and 0.58 respectively. Among the former five sites, the two sites of J_{CD} and LSβ with $\overline{n}_{hw}\geq 1.0$ and the three sites of T69_$O_{\gamma 1}$, R68_NH and T40_O with $\overline{n}_{hw}<0.4$ have molecular mechanisms of producing an IHW distinct from each other: Though the PGs of J_{CD} and LSβ are mostly hydrated, their HWs cannot go out of protein for 5 to 10% of time owing to the fluctuation of the protein structure neighboring these HSs and these waters are recognized as IHWs. On the other hand, T69_$O_{\gamma 1}$, R68_NH and T40_O are located near LSβ, Lβ and Hg respectively. The former two are mostly H-bonded to the other PGs in protein, but T40_O has a low proportion of forming an H-bond to a PG in protein as 0.08. Because of the structural fluctuation of protein, these PGs are occasionally H-bonded to the respective IHWs belonging to the neighboring LSβ, Lβ and Hg to share them. As a result of it, their $\overline{n}_{hw\,i}$'s are estimated to be 0.05–0.09. However, even with this contribution added, the rate of H-bond formation for T40_O is as low as 0.31. Different from J_{CD} and LSβ, the $\overline{n}_{hw}$'s for these three sites are as small as 0.37, 0.11 and 0.09 respectively, because they are all hydrated only for a short time. While Clft is the largest CHS with an n_{pg} of 6 and an $\overline{n}_{hw}$ of 5.98 largest among all the EHSs, it has virtually no IHWs as its $\overline{n}_{hw\,i}$ is 0.01.

Table 2
List of major clustered hydration sites and single hydration sites with 0.05 or more internal hydration waters. To simplify notations, each of the CHSs with $n_{pg}\geq 3$ including IHSs is given an abbreviation (location) as follows: Hg (Hinge), Lβ (Loop in β-domain), LSβ (Loop and Sheet in β-domain), J_{CD} (Junction of α-helices, C and D), Clft (Cleft), J_{AB} (Junction of α-helices, A and B), HgS (Side of Hinge), Nt (N-terminal), H_A (A-Helix), J_{DE} (Junction of α-helices, D and E).

Site	Type[a]	Constitutive polar groups	$\overline{n}_{hw}$[b]	$\sigma(n_{hw})$[c]	$\overline{n}_{hw\,i}$[d]
Hg	C, I	Y53_O; I55_NH; L56_NH; A82_O; L83_O; S85_O; D87_O; A90_NH; S91_NH, O_γ, O_γ H_γ	4.77	0.60	2.54 (0.94)[e]
Lβ	C, I	S60_O; C64_O; D66_NH, $O_{\gamma 1}$; G67_NH; S72_ O_γ, O_γ H_γ; R73_NH; N74_NH	3.23	0.64	1.38 (0.53)[e]
Clft	C, E	F34_O; E35_$O_{\varepsilon 1}$, $O_{\varepsilon 2}$; L56_O; V109_NH; A110_NH	5.98	0.89	0.01
Nt	C, E	K1_NH_1, NH_2, NH_3; N39_$O_{\gamma 1}$	1.99	0.61	0.01
H_A	C, E	E7_O; R14_$N_{\eta 1}H_{\eta 1}$, $N_{\eta 2}H_{\eta 2}$	1.86	0.58	0.01
HgS	C, E	H15_ $N_{\gamma 1}H_{\gamma 1}$; D87_$O_{\delta 1}$, $O_{\delta 2}$	3.71	0.72	0.00
J_{AB}	C, E	D18_$O_{\delta 1}$, $O_{\delta 2}$; L25_NH	4.11	0.82	0.00
J_{CD}	C, ei	I98_O; N103_NH; G104_O	1.14	0.41	0.05
J_{DE}	C, E	R114_O; T118_ $O_{\gamma 1}$; W123_$N_{\varepsilon 1}H_{\varepsilon 1}$	2.15	0.57	0.00
LSβ	C, ei	T51_$O_{\gamma 1}H_{\gamma 1}$; D66_$O_{\gamma 2}$	1.33	0.45	0.07
	S, ei	T69_$O_{\gamma 1}$	0.37	0.25	0.09
	S, ei	R68_NH	0.11	0.11	0.05
	S, ei	T40_O	0.09	0.13	0.05

[a] Type of the hydration site: C, clustered; S, single; I, internal; E, external; ei, external site with internal hydration waters of $0.05\leq\overline{n}_{hwi}<0.10$.
[b] $\overline{n}_{hw}$, mean number of hydration waters in the hydration site.
[c] $\sigma(n_{hw})$, standard deviation of n_{hw}.
[d] $\overline{n}_{hw\,i}$, mean number of internal hydration waters in the hydration site.
[e] The value in parenthesis indicates the standard deviation of $n_{hw\,i}$, $\sigma(n_{hw\,i})$.

We can see from Tables 1 and 2 that, among the 298 substantially hydratable PGs, 20 are involved in IHSs and 278 in EHSs, and among the whole 226.82 HWs, 8.00 hydrate to IHSs and 218.82 to EHSs on average. Accordingly, 0.40 waters hydrate to a PG in IHSs and 0.79 waters hydrate to a PG in EHSs on average. It means that, on average, IHSs are only half hydrated compared with EHSs. It is because the proportion in which HWs are shared among the PGs is higher for an IHS than for an EHS. In fact, the mean numbers of HBs to HWs per PG, $\overline{n}_{hb}$, in an IHS and an EHS are nearly the same as 0.96 and 0.95 respectively. Besides, the mean numbers of HBs between a PG and the other PGs in protein, i.e. $\overline{n}_{hp}$, of 0.29 and 0.30 being added, the above numbers turn out to be just the same as 1.25 and 1.25 respectively.

The values of HW-related quantities given above are their respective mean values. Under physiological conditions, the values of these quantities are continually fluctuating owing to the structural fluctuations of HWs caused by the structural fluctuation of protein as well as the thermal motion of HWs themselves. For example, as shown in Table 2, the number of HWs for each HS is distributed around a mean, $\overline{n}_{hw}$, with a standard deviation, σ_n. Dynamic characteristics of these HWs will be described in detail in a forthcoming paper. As an example, the probability distribution of the total number of IHWs for the whole HSs, N_{hwi}, (Fig. 2(a)) and their spatial distribution (Fig. 2(b)) are shown in Fig. 2.

As shown in Fig. 2(a), the distribution of $N_{hw\,i}$ has a mean of $\overline{N}_{hw\,i}=4.81$ and a standard deviation of $\sigma_{Nhw\,i}=1.61$. It is fairly well approximated by the normal distribution function having the same parameters as above. It confirms that the value of $N_{hw\,i}$ fluctuates randomly around $\overline{N}_{hw\,i}=4.81$. The fluctuation will be brought about by both of the structural fluctuation of protein and thermal motion of HWs. As is naturally expected from the above result, Fig. 2(b) shows that the position of each IHW fluctuates in its HS. According to our database analysis of protein structures, the mean number of IHWs is estimated to be 3.3 for a protein with a size similar to that of lysozyme. Actually, our study estimates 4.81 IHWs

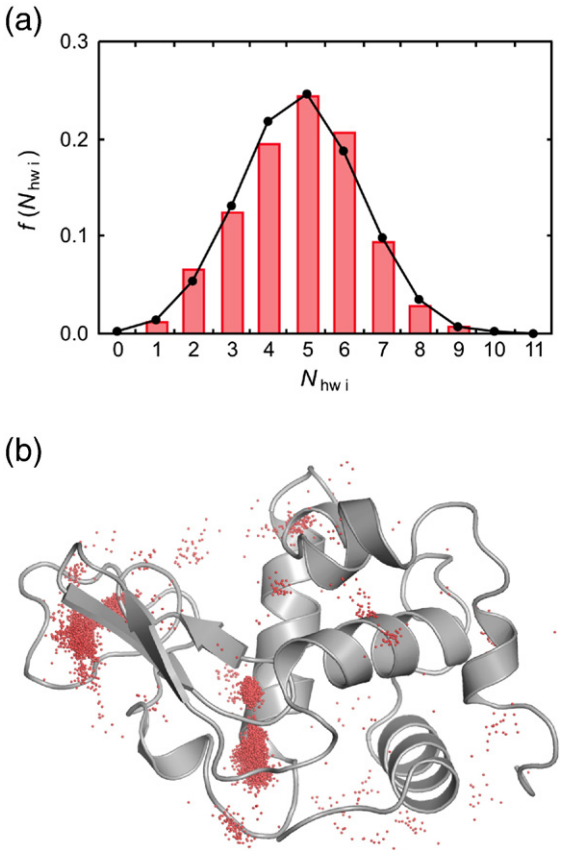

Fig. 2. (a) Probability distribution of the total number of internal hydration waters (IHW) for the whole hydration sites, $N_{\mathrm{hw\ i}}$, with its mean, $\overline{N}_{\mathrm{hw\ i}} = 4.81$, and standard deviation, $\sigma_{N\mathrm{hw\ i}} = 1.61$. (b) Distribution of IHWs in the interior of lysozyme. We can see that IHWs distribute fairly widely over the whole protein as well as in the major internal hydration sites of the hinge region (center) and the loop region in the β-domain (left side).

for a lysozyme molecule, which indicates the number of IHWs for lysozyme is significantly larger than that of an average protein.

We will discuss here the definition of a HS and the validity of the hydration analysis in this study. As described at the beginning of this paper, one of the important aims of this work is to obtain the definition of a HS capable of taking into account the dynamic realities of hydration phenomena as accurately as possible. The fact that all the PGs are not independent of each other but form fluctuating clusters by being bridged with HWs is one of the most important hydration characteristics of proteins. This is an attempt to develop a method for quantifying them. Naturally, the classification of the PGs into HSs and the prediction of hydration-related quantities derived from MDS analysis should depend on the values of three thresholds, h^*_{hb}, h^*_{sh} and h^*_{in}. The above-described results are those for their respective values of 0.20, 0.20 and 0.10. In this case, an off-diagonal element $h_{\mu\nu}$ for the two PGs, μ and ν, on two different HSs is smaller than 0.2. If another threshold h^*_{sh} is chosen for which the relation of $h_{\mu\nu} > h^*_{\mathrm{sh}}$ holds, then the two HSs merge into one larger HS. So, there is some ambiguity in determining the HSs of protein like that in whether or not to define the boundary of clouds close to each other. Choosing a threshold that minimizes the ambiguity will be the point of using this method. To see it, we have extracted all pairs of sites having PGs with $0.10 \le h_{\mu\nu} < 0.20$.

The total number of BHBs in these pairs of sites is 12.16, of which details are 6.06, 4.68, and 1.42 HBs between the SHSs, SHS and CHS, and CHSs respectively. This result indicates that, if the threshold h^*_{sh} is lowered from 0.20 to 0.10 with h^*_{hb} kept at 0.20, the total number of inter-HS BHBs is estimated larger by the amount. Especially, the increases in the first two account for 88% of the total increase, which bring about the merger of neighboring SHSs and the incorporation of SHSs into a CHS. The increase of 1.42 inter-CHS BHBs results in the merger of neighboring pairs of clusters, Hg and HgS, Lβ and LSβ, Clft and a CHS with $n_{\mathrm{pg}} = 2$, and J_{AB} and a CHS with $n_{\mathrm{pg}} = 2$. 11 HBs are involved in the total increase of 1.42 HBs. Six out of the 11 HBs are those bridging the two PGs in the same amino-acid residues. This fact along with the low bridging rate of these HBs confirms that, even if the name of these HBs changes from an inter-site HB to an intra-site HB, the images of these HSs are virtually unchanged.

3.3. Spatial distribution of the hydration sites

We can see from Fig. 3 that the HSs are distributed all over the surface of lysozyme. The large CHS with $\overline{n}_{\mathrm{hw}} = 5.98$ is located on the substrate binding cleft. As for the CHSs with $\overline{n}_{\mathrm{hw}} > 2$, we have the following sites: J_{AB} and J_{DE} at the respective connections of the two pairs of α-helices, (A and B) and (D and E), HgS on the side of the hinge region, Nt at the N-terminus, and H_A on the α-helix A. The two large IHSs, Hg and Lβ, are located in the hinge region and in the loop region of the β-domain respectively. The five EHSs having IHWs of $\overline{n}_{\mathrm{hw\ i}} = 0.05$–$0.09$, i.e. J_{CD}, LSβ, T69_$O_{\gamma 1}$, R68_NH, and T40_O are also shown.

3.4. Characteristics of individual hydration sites

In this section, choosing some typical HSs from among the SHSs, CHSs and IHSs, we will show the snapshots of these HSs and discuss their hydration characteristics.

Hydration structures for seven kinds of SHSs are drawn in Fig. 4. From their definition, SHSs show the simplest hydration states. While, in Figs. 4(a) and (b), each of G126_NH and S86_OH is H-bonded to one HW as a donor, in Fig. 4(c), G126_O is H-bonded to two HWs as an acceptor. In Figs. 4(d) and 4(e), the main-chain G67_O and side-chain Q41_O_ε are H-bonded to one and two donor HWs respectively. Though each main-chain CO group is H-bonded to one main-chain NH group in the regular secondary structure such as the α-helix and β-sheet, double-donor type HBs are frequently observed in the irregular structure of proteins. So, the double-donor type HB will not be an unusual form of HB for the HB to HWs. While, in Fig. 4(f), one HW is H-bonded to the NH^+ group of K116, five HWs are H-bonded to the COO^-

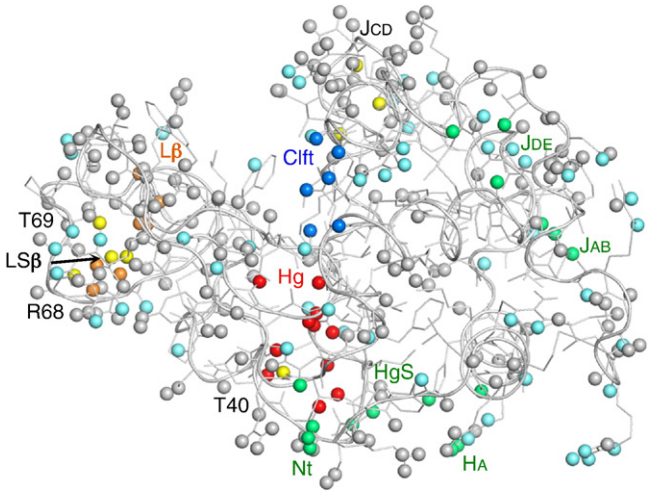

Fig. 3. Distribution of the polar groups in hydration sites (HS) over the whole surface of lysozyme. Abbreviations in the figure indicate the major HSs of lysozyme in Table 2. Small balls are located at either the accepter or the donor site of each HS. Red and orange, internal hydration sites (IHS); yellow, external hydration sites (EHS) having IHWs of $0.05 \le \overline{n}_{\mathrm{hwi}} < 0.10$. Clustered hydration sites (CHS): blue, $n_{\mathrm{pg}} = 6$; green, $n_{\mathrm{pg}} = 4$ and 3; light blue, $n_{\mathrm{pg}} = 2$. Single hydration sites (SHS): gray, $n_{\mathrm{pg}} = 1$. n_{pg} is the number of polar groups (PG) in HS.

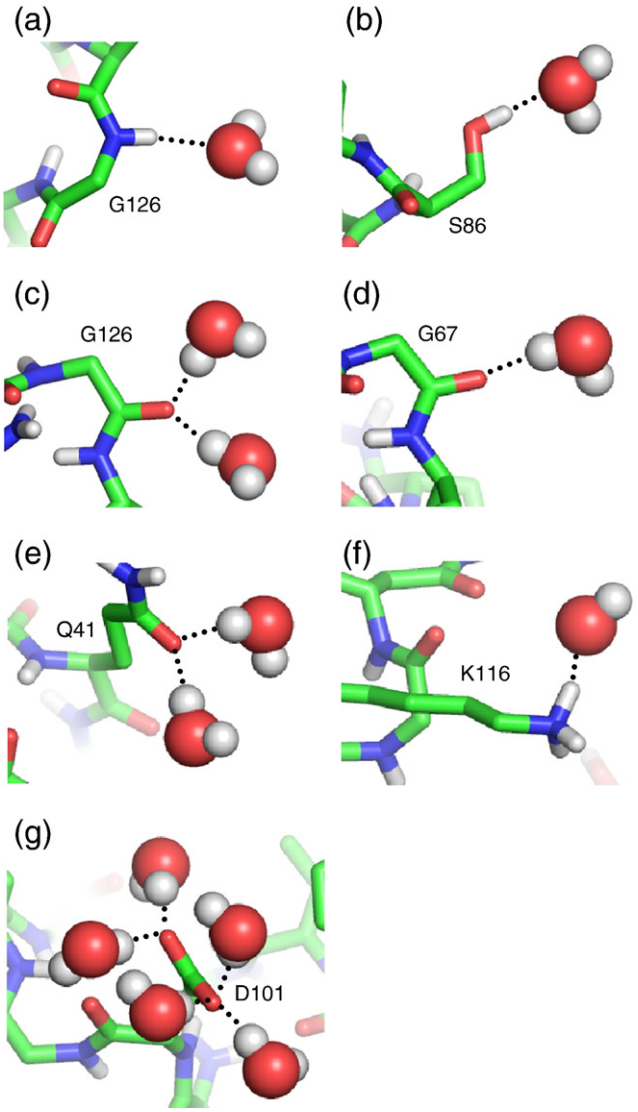

Fig. 4. (a) to (g) Snapshots of the hydration structures for seven kinds of single hydration sites (SHS). Each residue is indicated by its one-letter code.

group of D101 in Fig. 4(g). Thus, the hydration mechanism of SHS is simple but their forms are diverse. In the following, the HWs in a HS are distinguished by their symbols W*n* ($n = 1, 2, 3, \ldots$) and the acceptor and donor of W*n* are denoted as W*n* (a) and W*n* (d) respectively.

The BHB diagram and a snapshot of Clft, a typical EHS-type CHS, are shown in Figs. 5(a) and 5(b) respectively.

Clft is located on the cleft between the α- and β-domains, where the two acidic residues, E35 and D52, express their catalytic activity of cleaving polysaccharides. It consists of the side-chain carboxyl group of E35 and the four main-chain PGs of F34_O, L56_O, V109_NH and A110_NH. It appears from the BHB diagram that the nucleus of E35 and the four PGs surrounding it are connected by the network of BHBs. However, it is obvious that the network of these HBs is not as structurally stable as it looks, because, among the values of nine $h_{\mu\nu}$'s, the maximum is 0.355 and the seven of them are smaller than 0.3. To see molecular details of the hydration structure, we examined the structure of Clft by using MDS data. Both E35_$O_{\varepsilon 1}$ and E35_$O_{\varepsilon 2}$ in the BHB diagram are two charged O atoms of the E35 side chain. It was found that, though the rotation of a carboxyl group around its $C_\gamma C_\delta$ bond is potentially free, actually it behaves as if it has only two stable positions: When $O_{\varepsilon 1}$ is BH-bonded to L56_O and either V109_NH or A110_NH, $O_{\varepsilon 2}$ is BH-bonded to F34_CO, the same hydration pattern is formed even when the two O_ε atoms are interchanged. Accordingly, the left figure is reduced to the right one where all the $h_{\mu\nu}$'s except those for V109 and A110 are virtually doubled. Nevertheless, the values of $h_{\mu\nu}$ are significantly smaller than 1.0. It is not because the BHBs between E35_$O_{\varepsilon 1,\ \varepsilon 2}$ and the four surrounding PGs are broken with a significant probability but because they are realized very diversely, for example, via more than one water molecule. From the above result, a dynamic view of HWs is derived that, in Clft, various networks with different BHB patterns appear and disappear repeatedly through rapid recombination of BHBs. It is also suggested that the reason why many of the PGs on the surface of protein are SHSs will be as follows: Even between two PGs close to each other, multiple BH-bonded structures are formed and a single BHB structure cannot exists stably. From analysis of hydration structures generated in MDS, it was also found that Clft can accommodate eleven HWs at its maximum, one of which distribution is shown in Fig. 5(b). We can see that six PGs and eleven HWs are connected with each other to form a close HB network.

The BHB diagram and two snapshots of an IHS, Hg, are shown in Fig. 6.

The site Hg is the largest IHS consisting of eleven PGs (9 amino-acid residues). Many studies have been devoted to analyze the hydration structure of this site [57–60]. As shown in Table 2, Hg has 2.54 IHWs and 2.23 EHWs on average. From the BHB diagram shown in Fig. 6(a), we can see that there are several BHBs with large $h_{\mu\nu}$'s in Hg not found in EHSs as Clft: The three pairs of PGs, i.e. L56_NH and Y53_O, I55_NH and L83_O, and D87_O and A90_NH, are BH-bonded to each other with large $h_{\mu\nu}$'s as 0.876, 0.772 and 0.636 respectively. It means that these PGs have their respective HWs with high probabilities. It was found from the result of MDS that, while the $n_{\mathrm{hw\ i}}$ of Hg changes between 1 and 4, the corresponding hydration structure has one part virtually unchanged and the other part variously changing with different hydration states. As the diagram in Fig. 6(a) is obtained as the average over these different hydration states, its interpretation cannot be made straightforwardly. Hence, we will consider here only the case of $n_{\mathrm{hw\ i}} = 3$. In this case, we could find that there are two major spatial configurations of three IHWs, which are shown in Figs. 6(b) and 6(c). In Fig. 6(b), the configuration is drawn where there exist three IHWs (W1, W2 and W3 in orange) and two EHWs (W4 and W5 in red), which will be called Conf_A in the following. The PGs in red ellipses are those H-bonded to IHWs.

The water W1 is located at the innermost of Hg and BH-bonded to the three PGs of L56_NH, Y53_O and S91_O_γ triangularly. All the three $h_{\mu\nu}$'s corresponding to these BHBs are fairly large, and so, W1 is the most stable IHW. Similar to W1, W2 can also be BH-bonded to the three PGs of I55_NH, L83_O and S91_O_γ triangularly. Compared with the $h_{\mu\nu}$ for the pair of I55_NH and L83_O, however, those for the other pairs are fairly smaller as 0.369 and 0.351. Equation (10) being applied to the three PGs, the probability that W2 is H-bonded to S91_O_γ is estimated to be about 0.41. As the side chain of S91 is structurally flexible, it can recombine its HB by changing the χ_2 angle of the terminal OH group. In fact, from analysis similar to the above, the probability that W2 is H-bonded to S91_$O_\gamma H_\gamma$ is estimated to be about 0.39. So, it is concluded that W2 is H-bonded to S91 with nearly the same probability as an acceptor and a donor, as a result of which they are H-bonded to each other with a probability as high as about 0.80. In this way, the side-chain OH group of S91 goes back and forth continually from one stable position to another, which involves a synchronous, cooperative change in hydration structure of the HS. Cooperating with each other, S91 and its hydration water repeat multi-basin-like structural fluctuations in Hg.

W3 is a unique IHW that has no pair of PGs to which it can be stably H-bonded. It is singly H-bonded to either S85_O or D87_O, and occasionally BH-bonded to both of them. In addition, it is H-bonded also to W2 and W4. Though W4 is an EHW, it is BH-bonded to two PGs in the same amino-acid residue, S91_$O_\gamma H_\gamma$ and S91_NH, with a fairly large $h_{\mu\nu}$ of 0.538. Besides, it can be H-bonded also to A82_O and W3. Though W5 is an EHW, it can be BH-bonded to a pair of D87_O and

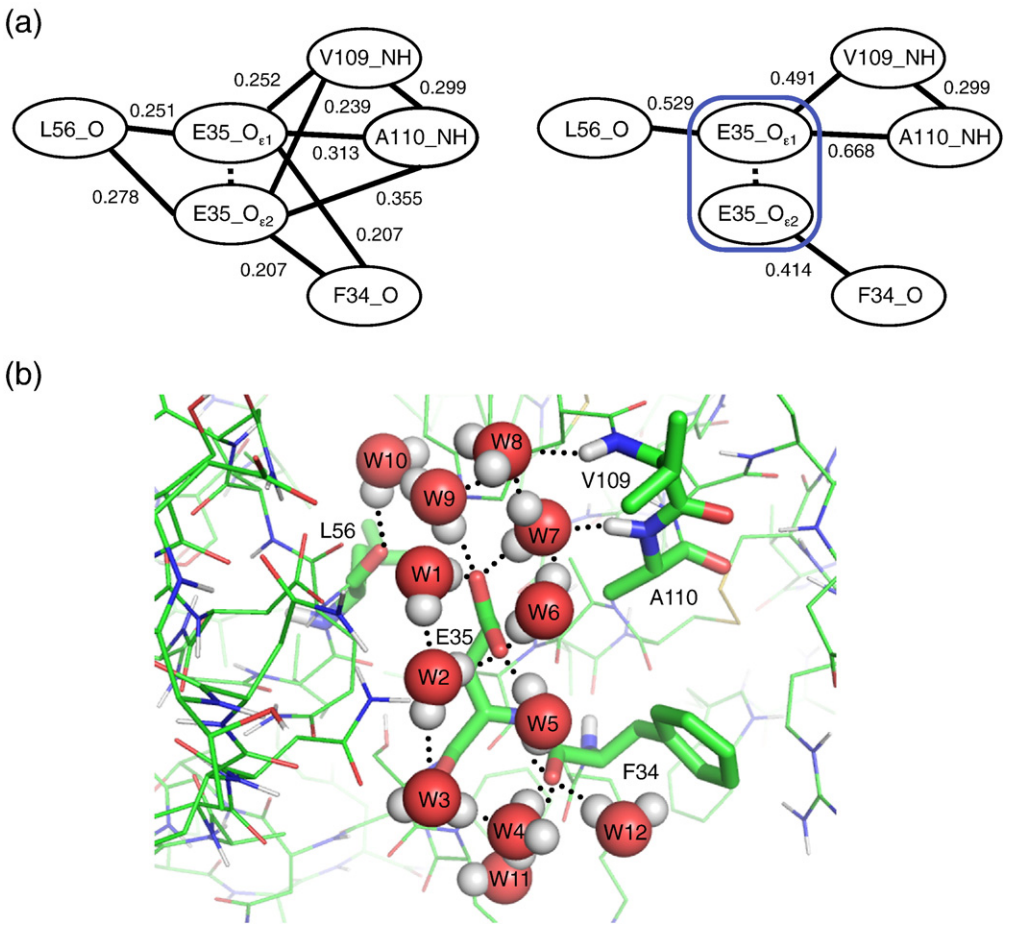

Fig. 5. (a) Bridging hydrogen bond (BHB) diagram (left) and its reduced one (right) for the clustered hydration site (CHS), Clft. Each polar group (PG) is connected with its neighboring PG by a bridging hydrogen bond (BHB) indicated by a solid line with the value of $h_{\mu\nu}$. See text for the definition of $h_{\mu\nu}$. (b) A snapshot of the Clft, in which eleven hydration waters (HW) are accommodated.

A90_NH with a large $h_{\mu\nu}$ of 0.636. Its detailed mechanism will be a very interesting problem to be solved in future.

In Fig. 6(c), another configuration called Conf_B is shown, in which W3 has vanished and W4 has become an IHW. Similar to those in Conf_A, the BHBs in which W1 is involved are completely kept and those in which W2 is involved are mostly kept also in Conf_B. BH-bonded to the four PGs, D87_O, S91_NH, S91_$O_\gamma H_\gamma$ and L83_O, W4 covers the deficit of W3. Besides, similar to the case of Conf_A, W5 is BH-bonded to D87_O and A90_NH. It was confirmed that the configuration with $n_{\mathrm{hw\ i}}=4$ is realized as a result of the two as follows: W4 in Conf_A becomes an IHW, or W4 enters the site of W3 in Conf_B and an external water enters the site of W4. In Conf_A, S85_O is H-bonded to both W3 and T40_$O_\gamma H_\gamma$. Even in the configuration of Conf_B lacking W3, S85_O escapes from being an unpaired PG as it can be H-bonded to T40_$O_\gamma H_\gamma$. This will serve to minimize the possible large effect of the structural destabilization of protein due to lacking W3.

Lβ is another IHS consisting of nine PGs (seven amino-acid residues) in the loop region of the β-domain. The BHB diagram and a snapshot of its hydration state are shown in Figs. 7(a) and 7(b) respectively.

From analysis of MDS data and Fig. 7, it was found that W1 is an IHW and BH-bonded to the four PGs of C64_O, N74_NH, R73_NH and S60_O or S72_O_γ. Especially, it is remarkable that the $h_{\mu\nu}$ for the BH-bonded pair of N74_NH and C64_O is close to 1.0 as 0.96. It means that W1 is almost surely captured at this position in the HS. In Fig. 7(b), the case is shown where W1 is H-bonded to R73_NH and S60_O , and W2 is BH-bonded to the four PGs of S60_O, S72_O_γ , G67_NH and D66_NH. Besides, S72_$O_\gamma H_\gamma$ is H-bonded to an EHW, W3. However, as W1(a) is H-bonded to R73_NH at the rate of about 0.4 and W1(d) is estimated to be H-bonded to S60_O or S72_O_γ at the ratio of about 6:4, this hydration structure is not always realized. The side chain of S72 is flexible and W1(d) recombines its HB partner between S60_O and S72_O_γ at the above ratio. When W1(d) is H-bonded to S72_O_γ, the HB between W2(d) and S72_O_γ is broken but W2(a) can be H-bonded to S72_$O_\gamma H_\gamma$ instead. Then, W2(d) is H-bonded to either C64_O or an external water. Besides, in Fig. 7(b), the side chain of D66 has a conformation different from that in X-ray crystal structure and its terminal carboxyl group is H-bonded to the three imino groups in the main- and side-chains of R68.

We can conclude from the three examples described above that (a) the ordinary hydration structure of a HS is an average of diverse structures weighted with their respective occurrence frequencies, and (b) the transition between these structures occurs through various H-bond recombinations among HWs and protein PGs resulting from their thermal motion.

3.5. Comparison with X-ray crystal-structure analysis

Atomic-coordinate data for the 3D-structure of h.e.w. lysozyme have been reported for various samples such as are crystallized into different crystal forms under mutually different solvent conditions, though their global structures are almost the same [33–35,61]. Generally, atomic coordinates for the oxygen atoms of many HWs

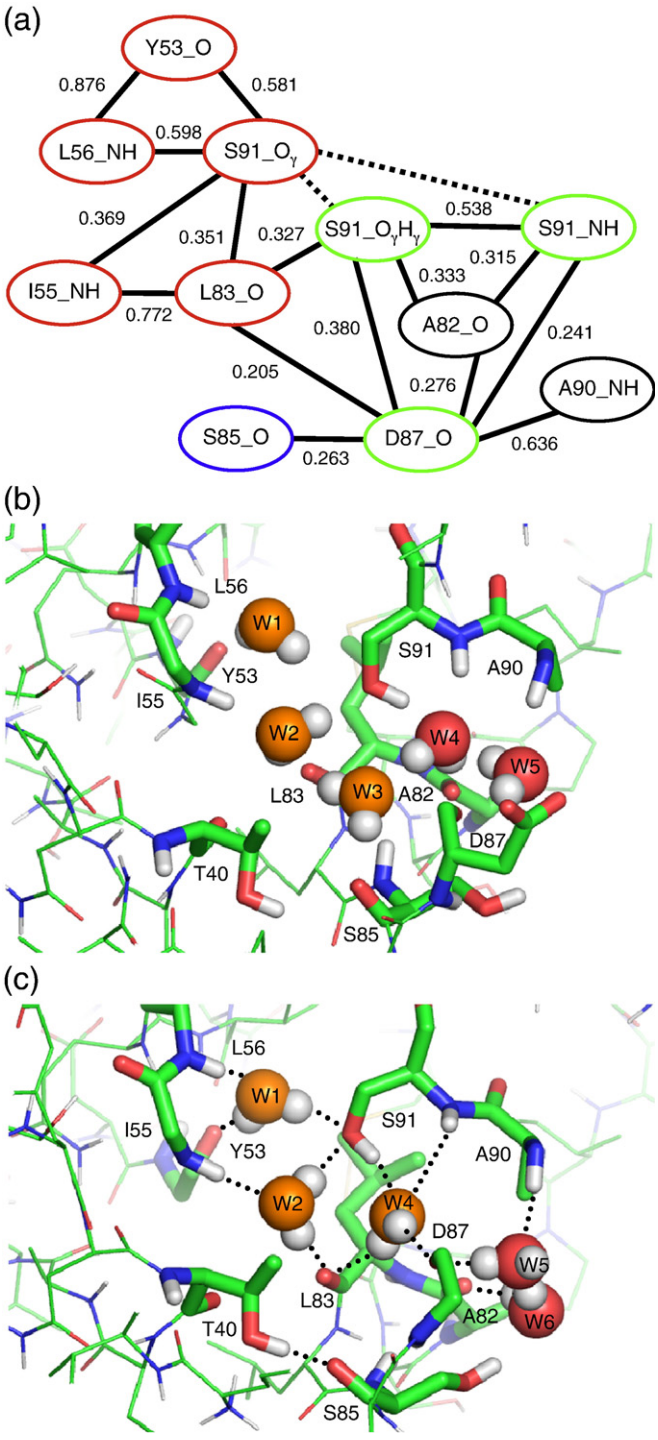

Fig. 6. (a) Bridging hydrogen bond (BHB) diagram for the largest internal hydration site (IHS), Hg, similar to Fig. 5(a). Red, polar groups (PG) H-bonded to internal hydration waters (IHW) common to (b) and (c); blue, a PG H-bonded to IHW only in (b); green, PGs H-bonded to IHW only in (c). The two PGs connected by a broken line are those in the same amino-acid residue, but the corresponding $h_{\mu\nu}$ is negligibly small. (b) and (c) Snapshots for the two major spatial configurations of Hg when it has three internal hydration waters (IHW). See text for details of these configurations.

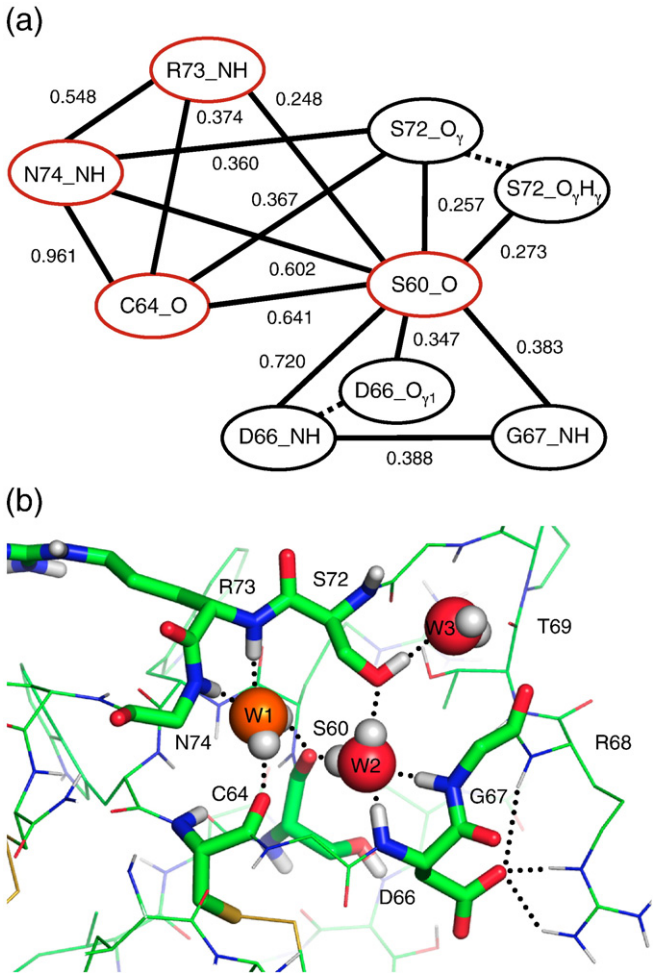

Fig. 7. (a) Bridging hydrogen bond (BHB) diagram for another internal hydration site (IHS), Lβ, similar to Fig. 6(a). Red, polar groups (PG) H-bonded to an internal hydration waters (IHW). (b) A snapshot for Lβ when it has one internal hydration water (IHW). See text for details of the configuration.

are included in these data. These waters have been identified by crystallographic analysis and are expediently called crystal waters (CWs) in the following. First, using information on the position of CWs, we estimate their H-bonding state. Then, comparing it with the H-bonding state derived from MDS, we examine the correlation between the two sets of information. Because no atomic coordinate for H-atoms is given in most X-ray crystal-structure data, and so, the orientation of all CWs is not known, we cannot determine whether a CW is H-bonded or not by applying Thornton's criterion to it. Therefore, if the distance, d, between a CW and the non-hydrogen atom of a PG is shorter than 0.35 nm, i.e.

$$d < 0.35\text{nm}, \tag{11}$$

we assume here that the CW is H-bonded to the PG independently of their mutual orientation. Because the maximum distance of 0.35 nm is equal to that of Thornton' criterion where additionally some orientational restrictions are imposed, the above criterion of Eq. (11) is an easier one than Thornton's. As a result of it, even if a CW and a PG is determined as H-bonded under the criterion of Eq. (11), they may not be H-bonded according to Thornton's criterion.

The purpose of this study is to analyze the hydration structure of a protein in solution. So, in order for us to be able to make a physically meaningful comparison between a HW obtained from MDS and a CW given by crystal-structure analysis, the hydration state in crystal must be similar to that in solution at least approximately. Concerning the

point, a protein molecule in crystal has a problem characteristic of it as follows:

(a) It is surrounded by the same kind of proteins in crystal contacting with it either directly or through a small number of solvent molecules. The hydration state of these PGs on the molecular surface that are involved in the so called crystal packing should differ greatly from that in solution.

(b) Salt ions and crystallizing reagent molecules are often bound to some surface PGs, for which information on their HWs cannot be obtained. To avoid these difficulties, we excluded from analysis such sites as are applicable to (a) and (b). It has been known [18,19] that there are many sites where a HW is found only on either the site predicted by MDS or the site identified in crystal. Effects of the above crystal packing and the structural fluctuation of both protein and water are expected for the observation. It is also reported that 30 to 40% of the accessible surface area of protein is excluded from solvent as a result of crystal packing [62], and, moreover, the protein packing density ρ_p by volume in crystal varies with its crystal form [63]. We estimated the ρ_p of lysozyme structure in crystal to be 0.651 for 4lzt (triclinic) [33], 0.642 for 1lj3 (monoclinic) [61] and 0.613 for 1iee (tetragonal) [34].

For the crystal structure of lysozyme, we have used mainly the data with a PDB code (resolution) of 1iee (94 pm). Its tetragonal crystal has a space group of $P4_32_12$ symmetry. Depending on the need, we use also the structure of 4lzt (95 pm) for a triclinic crystal with P1 symmetry. Both structures are determined with a high resolution of less than 100 pm at room temperature. It was found that, for the structure of 1iee, as much as 94 HSs out of the 233 HSs, i.e. 40.3%, are applicable to the above (a) and (b), and cannot be used for analysis. Nevertheless, there are 139 available HSs in the 1iee structure, a larger number of HSs than that for 4lzt, which is the chief reason that we chose the 1iee structure as the target for analysis. In addition, we could confirm following situations: (a) Crystals for the 1rfp and 1iee structures have the same tetragonal form, and (b) the rmsd between the coordinates of main-chain non-hydrogen atoms in the two structures is very small as 0.031 nm. (c) The 1rfp structure lacks an internal water which exists in the Lβ region of the 1iee structure. However, as a water molecule enters the site at the step of equilibration, the initial structure for MDS can be regarded virtually equivalent whichever structures are used.

The structural data of 1iee give atomic coordinates of the O atoms for 220 CWs. Among them, there are 51 CWs which are more than 0.35 nm away from all PGs in protein and, practically, cannot be H-bonded to any of the PGs. Most of them are the second HWs of PGs in lysozyme, among which there are six HWs of nonpolar groups. 108 CWs, which can be a HW of one of the 139 HSs available for analysis, were extracted from among the 169 (=220−51) CWs that can be H-bonded to any one of the PGs, i.e. can be a first HW. It reveals that 61 (=169−108) CWs, i.e. 36% of all, are involved in the HSs applicable to the above (a) and (b). These 108 CWs were found to be the HWs of 106 HSs out of the whole 139 HSs. It means that no CW is found for 33 (=139−106) sites.

As shown in Table 1, the total numbers of HSs and HWs predicted from MDS are 233 and 226.82 respectively. Similarly, 127.97 HWs are predicted from MDS for the 139 sites available for analysis. On the other hand, 108 CWs were found for the available sites. They could be assigned to 104 HSs for which MDS predicts 103.84 HWs on average. There are 76 HSs in which the numbers of CWs and HWs can be regarded as practically equal to each other as the difference is smaller than 0.5. For 21 HSs out of the rest of 28 HSs, we could conclude that the hydration structures in both solution and crystal are essentially the same, though the number of CWs is slightly larger than that of HWs. Summing up these HSs, there are 97 HSs with practically equal numbers of CWs and HSs, seven HSs with significantly different numbers, and 35 HSs with no CW detected. We can see from the above that, while nearly the same numbers of CWs and HWs are observed for about 70% of the HSs, either a part of or the whole CWs are not found for the rest of 42 or 30% of the HSs out of the 139 HSs taken for analysis. Several researchers have so far pointed out problems in determining the position of crystal waters of a protein in crystal [63–66]. The above disagreement between the numbers of CWs and HWs seems to indicate that the "crystal water" cannot reflect accurately the hydration state of a protein in crystal. According to our analysis on the structure of HSs, it seems basically impossible to represent the dynamic behavior of a HW which moves following various structural fluctuation of protein by a crystal water located at a fixed point. In the following, we will discuss some of the examples among the 42 HSs.

The seven HSs with significantly different numbers of CWs and HWs consist of one CHS with $n_{pg}=3$, three CHSs with $n_{pg}=2$, and three SHSs. The CHS with $n_{pg}=3$ is composed of the three PGs of D18_$O_{\delta1}$, D18_$O_{\delta2}$ and L25_NH and is estimated to have 4.1 HWs from MDS. However, only one CW bridging L25_NH and D18_$O_{\delta1}$ is identified and there is no CW hydrating $O_{\delta2}$ in the crystal structure. The number of CWs is significantly smaller than the mean number of HWs from MDS for the rest of six HSs. It will result from the fact that a part of HWs are not identified as CWs, as each HS can have multiple hydration states.

Details of the HSs of which HWs cannot be found among those available for analysis are one CHS with $n_{pg}=2$ and 34 SHSs. From the result of individual studies on these sites, we could expect that the reason why no CW is found in them will be mainly the following two:

(α) Though the PG in the site concerned is surely H-bonded to a HW, the electron density distribution necessary for identifying the position of the HW is not generated, because the mobility of the chain segment on which the site is located is so high that it takes diverse conformations.

(β) Because of a fairly high probability with which the request of a PG in the site for H-bonding is satisfied by an HB to the other PG, the rate that it is H-bonded to a HW is low and the occupancy of a HW in the site is low.

The quantities that measure the rates of H-bonding for a PG, μ, to be H-bonded to a HW and the other PG are $\overline{n}_{hb}(\mu)$ $(=h_{\mu\mu})$ and $\overline{n}_{hp}(\mu)$ respectively. To confirm the expectation of (α) and (β), we examined the correlation of $\overline{n}_{hb}(\mu)$ and $\overline{n}_{hp}(\mu)$ for the 35 PGs. As a result of it, we could confirm that there is a strong negative correlation for 83% of the PGs: It holds that $\overline{n}_{hb}(\mu)\geq0.50$ and $\overline{n}_{hp}(\mu)<0.50$ for 13 PGs, and $\overline{n}_{hb}(\mu)<0.50$ and $\overline{n}_{hp}(\mu)\geq0.50$ for 16 PGs. It is supposed that the former PGs correspond to the case (α) and the latter PGs to the case (β). All the four PGs with $\overline{n}_{hb}(\mu)\geq0.50$ and $\overline{n}_{hp}(\mu)\geq0.50$ were found to be carboxyl groups and are H-bonded to both of a HW and a PG in protein. As the mechanism of (α) is apparent, we will describe here two examples of the sites where, through the mechanism of (β), no CW is detected even at a low temperature of 120 K, which was experimentally confirmed.

The two PGs of N44_O and D52_O form a CHS with $n_{pg}=2$. These residues are located in the two β-strands forming an antiparallel β-sheet in the β-domain and are involved in the two HBs of N44_NH···O_D52 and N44_O···HN_D52. Though the one side of the β-sheet is exposed to solvent waters, they cannot be H-bonded to the PGs because a HB of the NH···O type generally has a nearly linear conformation. However, only when the β-sheet is distorted and the HB is bent, two water molecules are independently or, forming a BHB, one water molecule is H-bonded to the two vacant acceptor sites of the two carbonyl groups. As a result, the values of $h_{\mu\mu}=0.431$, $h_{\nu\nu}=0.497$ and $h_{\mu\nu}=h_{\nu\mu}=0.254$ (μ=N44_O, ν=D52_O) are obtained from MDS. This shows that the two types of HBs are formed with nearly the same probability.

All the three PGs of N106_O, R112_$N_\varepsilon H_\varepsilon$ and R112_$N_{\eta2}H_{\eta21}$ are determined as an SHS by MDS, but have the small numbers of mean HBs as $h_{\mu\mu}=0.299$, 0.235, and 0.322 respectively. This results from the situation that the PG of N106_O is H-bonded to the two PGs of R112

with a percent higher than 70%, and so, the probability that the HW can form a HB is lower than 0.3.

J_{CD} is a CHS with $n_{pg}=3$ consisting of the three main-chain PGs of λ=I 98_O, μ=G104_O and ν=N103_NH located at the junction of two α-helices, C and D. The hydration matrix has the following elements as $h_{\lambda\lambda}=0.856$, $h_{\mu\mu}=0.751$, $h_{\nu\nu}=0.531$, $h_{\lambda\mu}=0.592$, $h_{\lambda\nu}=0.214$ and $h_{\mu\nu}=0.142$. The two typical conformations of J_{CD} and the corresponding distribution of HWs, which were generated in MDS, are illustrated in Fig. 8.

In Fig. 8(a), the HW of W1 is BH-bonded to the two PGs of λ and μ and connected with the third PG of ν via the HW of W2. However, various hydration structures other than this can also be formed. For example, there occur such configurations as the W1 is H-bonded to ν either through two waters or directly. The structure of 1iee lysozyme has the latter configuration [34], but, in MDS, the frequency of its occurrence is much lower than that in the figure. The probability that the W1 is determined as an IHW is so low as $\overline{n}_{hw}=1.14$ and $\overline{n}_{hw\,i}=0.05$ as shown in Table 2. In Fig. 8(b), another conformation of the J_{CD} loop where the BHB in Fig. 8(a) is broken and both N103_NH and N103_$N_\gamma H_\gamma$ are directly H-bonded to I 98_O. As the three PGs are H-bonded to one another and surrounding PGs, there is no HW in this structure. This is the same conformation as that of the crystal structure of 4lzt [33], but its occurrence frequency in MDS is also low. This result proves that the structure of a protein and its hydration water shows diverse fluctuations at physiological temperatures. It also suggests that the crystal and solution structures of a protein can differ greatly from each other in the occurrence frequency of diverse conformations.

4. Conclusions

The hydration site (HS) of a protein molecule has so far been defined as each peak position of the number density of hydration waters (HWs) around it. In regard to the definition, however, a problem has been pointed out that estimated peak positions are liable to be blurred with thermal fluctuations of the protein structure. Changing our viewpoint to overcome the problem, we introduced a new method that a HS is defined not as a peak position of water density but as a polar group (PG) H-bonded to HWs with an average number exceeding a predefined threshold. A great advantage of adopting this definition is that the structural fluctuation of a protein brings about no error in identifying its HSs. In addition, we have newly introduced a hydration matrix obtained from applying Thornton's criterion for determining H-bond (HB) formation to time-series data on the atomic coordinates of both protein and HW. It enables us to identify automatically all the HSs of a protein and to analyze its hydration structure and characteristic properties.

Applying our method to a lysozyme molecule in solution, we could identify its 233 HSs in all. Our method for finding HB formation can detect not only the HB between a PG and a HW but also the bridging hydrogen bond (BHB) where a HW connects two PGs. Making use of it, we could find that there are two kinds of HSs in lysozyme: 195 single hydration sites (SHS) each consisting of a single PG and 38 clustered hydration sites (CHSs) each consisting of two or more PGs connected with each other by one or more BHBs. The structures of protein and HW are both thermally fluctuating rapidly, as a result of which a BHB in a CHS as well as a single HB in a SHS alternates connection with disconnection continually. Hence, the BHB diagram, where two PGs forming a BHB in a CHS are connected by a line, serves as a very useful tool for representing structural characteristics of the HS. Several examples of analysis on the structure of major CHSs have led to the conclusion as follows: (a) The mean structure of a HS in lysozyme is given by a weighted average over the multiple elementary structures formed by its PGs and HWs, and (b) each of the transition processes between different elementary structures is realized through various recombination steps of HBs among the PGs and HWs.

Referring to the facts and conclusions described above, we will be able to expect the following: Judging from the structure of CHSs in lysozyme such as Hg, Lβ, Clft and so on, the structural characteristic of a CHS that it consists of several PGs connected by BHBs will be common to most of the globular proteins. It is also readily expected that (a) both connection and disconnection of a BHB between two PGs will possibly be linked to various structural fluctuations of protein, and (b) the respective sets of HSs expected for a natural mobile protein and its immobilized counterpart will naturally be different from each other. Combining the prediction and preliminary results of analysis on the dynamic characteristics of protein hydration, which is now in progress, a possibility is expected that the structural fluctuation of CHSs plays important roles in the folding, unfolding and functioning of globular proteins.

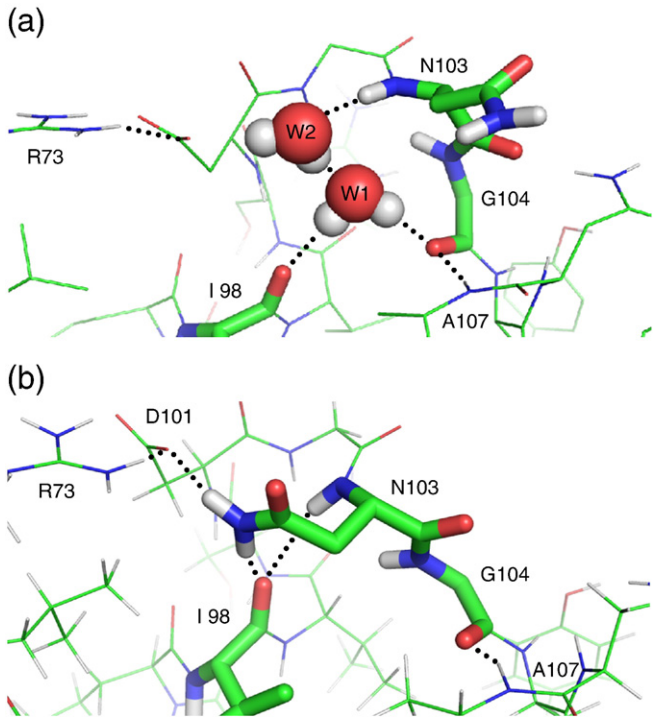

Fig. 8. Two snapshots for the clustered hydration site (CHS), JCD, with $n_{pg}=3$ demonstrating diverse configurations of protein and hydration water (HW). (a) A configuration having two HWs and (b) another configuration having no HW.

Acknowledgments

K.S. is very grateful to Dr. A. Wada, the Emeritus Professor of the University of Tokyo, for his continuous encouragement and Prof. K. Akasaka for his valuable discussion. MDSs in this work were performed by using the RIKEN Integrated Cluster of Clusters (RICC) facility.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.bpc.2011.02.006.

References

[1] T.E. Creighton, Protein Folding, W.H. Freedman, New York, 1992.
[2] Y. Levy, J.N. Onuchic, Water mediation in protein folding and molecular recognition, Annu. Rev. Biophys. Biomol. Struct. 35 (2006) 389–415.
[3] N.T. Southall, K.A. Dill, A.D.J. Haymet, A view of hydrophobic effect, J. Phys. Chem. B 106 (2002) 521–533.
[4] D. Chandler, Interfaces and the driving force of hydrophobic assembly, Nature 437 (2005) 640–647.

[5] D. Paschek, Heat capacity effects associated with the hydrophobic hydration and interaction of simple solutes: a detailed structural and energetical analysis based on molecular dynamics simulations, J. Chem. Phys. 120 (2004) 10605–10617.
[6] S.Y. Sheu, D.Y. Yang, H.L. Selzle, E.W. Schlag, Energetics of hydrogen bonds in peptides, Proc. Natl Acad. Sci. USA 100 (2003) 12683–12687.
[7] N. Ben-Tal, D. Sitkoff, I.A. Topol, A.S. Yang, S.K. Burt, B. Honig, Free energy of amide hydrogen bond formation in vacuum, in water, and in liquid alkane solution, J. Phys. Chem. B 101 (1997) 450–457.
[8] A.R. Bizzarri, S. Cannistraro, Molecular dynamics of water at protein-solvent interface, J. Phys. Chem. B 106 (2002) 6617–6633.
[9] W.J. Scanlon, D. Eisenberg, Solvation of crystalline proteins: theory and its application to available data, J. Mol. Biol. 98 (1975) 485–502.
[10] W.J. Scanlon, D. Eisenberg, Solvation of crystalline proteins: solvent bound in sperm whale metmyoglobin type A crystals at 6.1 and 23.5 °C, J. Phys. Chem. 85 (1981) 3251–3256.
[11] M. Nakasako, Water-protein interactions from high-resolution protein crystallography, Philos. Trans. R. Soc. Lond. B 359 (2004) 1191–1206.
[12] J. Higo, M. Nakasako, 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. Comput. Chem. 23 (2002) 1323–1336.
[13] T. Yokomizo, M. Nakasako, T. Yamazaki, H. Shindo, J. Higo, Hydrogen-bond patterns in the hydration structure of a protein, Chem. Phys. Lett. 401 (2005) 332–336.
[14] W. Gu, B.P. Schoenborn, Molecular dynamics simulation of hydration in myoglobin, Proteins 22 (1995) 20–26.
[15] F. Merzel, J.C. Smith, Is the first hydration shell of lysozyme of higher density than bulk water? Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 5378–5383.
[16] V. Lounnas, B.M. Pettitt, A connected-cluster of hydration around myoglobin: correlation between molecular dynamics simulations and experiment, Proteins 18 (1994) 133–147.
[17] G. Hummer, A.E. García, D.M. Soumpasis, Hydration of nucleic acid fragments: comparison of theory and experiment for high-resolution crystal structures of RNA, DNA, and DNA-drug complexes, Biophys. J. 68 (1995) 1639–1652.
[18] V.A. Makarov, B.K. Andrews, P.E. Smith, B.M. Pettitt, Residence times of water molecules in the hydration sites of myoglobin, Biophys. J. 79 (2000) 2966–2974.
[19] M.H. Priya, J.K. Shah, D. Asthagiri, M. Paulaitis, Distinguishing thermodynamic and kinetic views of the preferential hydration of protein surface, Biophys. J. 95 (2008) 2219–2225.
[20] R.H. Henchman, J.A. McCammon, Extracting hydration sites around proteins from explicit water simulations, J. Comput. Chem. 23 (2002) 861–869.
[21] M.S. Madhusudhan, S. Vishveshwara, Deducing hydration sites of a protein from molecular dynamics simulations, J. Biomol. Struct. Dyn. 19 (2001) 1–10.
[22] B.S. Sanjeev, S. Vishveshwara, Protein-water interactions in ribonuclease A and angiogenin: a molecular dynamics study, Proteins 55 (2004) 915–923.
[23] V. Lounnas, B.M. Pettitt, Distribution function implied dynamics versus residence times and correlations: solvation shells of myoglobin, Proteins 18 (1994) 148–160.
[24] F. Sterpone, M. Ceccarelli, M. Marchi, Dynamics of hydration in hen egg white lysozyme, J. Mol. Biol. 311 (2001) 409–419.
[25] N. Somolin, R. Winter, Molecular dynamics simulations of staphylococcal nuclease: properties of waters at the protein surface, J. Phys. Chem. B 108 (2004) 15928–15937.
[26] A. Lehoux, M. Krzystyniak, E. Baguet, A faster way to characterize by triple-quantum-filtered ^{17}O NMR water molecules strongly bound to macromolecules in solution, J. Magn. Reson. 148 (2001) 11–22.
[27] E. Persson, B. Halle, Nanosecond to microsecond protein dynamics probed by magnetic relaxation dispersion of buried water molecules, J. Am. Chem. Soc. 130 (2008) 1774–1787.
[28] G. Otting, E. Liepinsh, K. Wüthrich, Protein hydration in aqueous solution, Science 254 (1991) 974–980.
[29] A. Yokota, K. Hirai, H. Miyauchi, S. Iimura, Y. Noda, K. Inoue, K. Akasaka, H. Tachibana, S. Segawa, NMR characterization of three-disulfide variants of lysozyme, C64A/C80A, C76A/C94A, and C30A/C115A—a marginally stable state in folded protein, Biochemistry 43 (2004) 6663–6669.
[30] G.I. Makhatadze, P.L. Privalov, Energetics of protein structure, Adv. Protein Chem. 47 (1995) 307–425.
[31] T. Imai, Y. Harano, M. Kinoshita, A. Kovalenko, F. Hirata, A theoretical analysis on hydration thermodynamics of proteins, J. Chem. Phys. 125 (2006) 24911.
[32] T. Chalikian, On the molecular origins of volumetric data, J. Phys. Chem. B 112 (2008) 911–917.
[33] M.A. Walsh, T.R. Schneider, L.C. Sieker, Z. Dauter, V.S. Lamzin, K.S. Wilson, Refinement of triclinic hen egg-white lysozyme at atomic resolution, Acta Crystallogr. D 54 (2001) 522–546.
[34] C. Sauter, F. Otálora, J. Gavira, O. Vidal, R. Giegé, J.M. García-Ruiz, Structure of tetragonal hen egg-white lysozyme at 0.94 Å from crystals grown by the counter-diffusion method, Acta Crystallogr. D 51 (2001) 1119–1126.
[35] N. Niimura, Y. Minezaki, T. Nonaka, J.-C. Castagna, F. Cipriani, P. Høghøj, M.S. Lehmann, C. Wilkinson, Neutron Laue diffractometry with an imaging plate provides an effective data collection regime for neutron protein crystallography, Nature Struct. Mol. Biol. 4 (1997) 909–914.
[36] M.D. Collins, M.L. Quillin, G. Hummer, B.W. Matthews, S.M. Gruner, Structural rigidity of a large cavity-containing protein revealed by high-pressure crystallography, J. Mol. Biol. 367 (2007) 752–763.
[37] A. Damjanovic, J.L. Schlessman, C.A. Fitch, A.E. Garcia, B. Garcia-Moreno, Role of flexibility and polarity as determinants of the hydration of internal cavities and pockets in proteins, Biophys. J. 93 (2007) 2791–2804.
[38] G. Otting, E. Liepinsh, B. Halle, U. Frey, NMR identification of hydrophobic cavities with low water occupancies in protein structures using small gas molecules, Nature Struct. Biol. 4 (1997) 396–404.
[39] T. Imai, R. Hiraoka, A. Kovalenko, F. Hirata, Locating missing water molecules in protein cavities by the three-dimensional reference interaction site model theory of molecular solvation, Proteins 66 (2007) 804–813.
[40] R. Kitahara, S. Yokoyama, K. Akasaka, NMR snapshots of a fluctuating protein structure. Ubiquitin at 30 bar—3 kbar, J. Mol. Biol. 347 (2005) 277–285.
[41] M. Karplus, J.A. McCammon, Molecular dynamics simulations of biomolecules, Nature Struct. Biol. 9 (2002) 646–652.
[42] D.E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R.O. Dror, M.P. Eastwood, J.A. Bank, J.M. Jumper, J.K. Salmon, Y. Shan, W. Wriggers, Atomic-level characterization of the structural dynamics of proteins, Science 330 (2010) 341–346.
[43] A. Suenaga, T. Narumi, N. Futatsugi, R. Yanai, Y. Ohno, N. Okimoto, M. Taiji, Folding dynamics of 10-residue beta-hairpin peptide chignolin, Chem. Asian J. 2 (2007) 591–598.
[44] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation, J. Chem. Theory Comput. 4 (2008) 435–447.
[45] S. Nose, A unified formulation of the constant temperature molecular dynamics methods, Mol. Phys. 52 (1984) 255–268.
[46] W.G. Hoover, Canonical dynamics: equilibrium phase-space distribution, Phys. Rev. A 31 (1985) 1695–1697.
[47] H.C. Andersen, Molecular dynamics simulations at constant pressure and/or temperature, J. Chem. Phys. 72 (1980) 2384–2393.
[48] S. Nose, M.L. Klein, Constant pressure molecular dynamics for molecular systems, Mol. Phys. 50 (1983) 1055–1076.
[49] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple Amber force fields and development of improved protein backbone parameters, Proteins 65 (2006) 712–725.
[50] E.J. Sorin, V.S. Pande, Exploring the helix-coil transition via all-atom equilibrium ensemble simulations, Biophysical J. 88 (2005) 2472–2493.
[51] H.W. Horn, W.C. Swope, J.W. Pitera, J.D. Madura, T.J. Dick, Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew, J. Chem. Phys. 120 (2004) 9665–9678.
[52] I.S. Joung, T.E. Cheatham, Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations, J. Phys. Chem. B 112 (2008) 9020–9041.
[53] C.J. Fennell, J.D. Gezelter, Is the Ewald summation still necessary? Pairwise alternatives to the accepted standard for long-range electrostatics, J. Chem. Phys. 124 (2006) 234104.
[54] I.K. McDonald, J.M. Thornton, Satisfying hydrogen bonding potential in proteins, J. Mol. Biol. 238 (1994) 777–793.
[55] J. Tsai, R. Taylor, C. Chothia, M. Gerstein, Packing density in proteins: standard radii and volumes, J. Mol. Biol. 290 (1999) 253–266.
[56] A.K. Soper, The radial distribution functions of water and ice from 220 K to 673 K and at pressures up to 400 MPa, Chem. Phys. 258 (2000) 121–137.
[57] K.P. Wilson, B.A. Malcolm, B.W. Matthews, Structural and thermodynamic analysis of compensating mutations within the core of chicken egg white lysozyme, J. Biol. Chem. 267 (1992) 10842–10849.
[58] P. Shih, D.R. Holland, J.F. Kirsch, Thermal stability determinants of chicken egg-white lysozyme core mutants: hydrophobicity, packing volume, and conserved buried water molecules, Protein Sci. 4 (1995) 2050–2062.
[59] K. Takano, Y. Yamagata, K. Yutani, Buried water molecules contribute to the conformational stability of a protein, Protein Eng. 16 (2003) 5–9.
[60] M. Refaee, T. Tezuka, K. Akasaka, M.P. Williamson, Pressure-dependent changes in the solution structure of hen egg-white lysozyme, J. Mol. Biol. 327 (2003) 857–865.
[61] N.T. Saraswathi, R. Sankaranarayanan, M. Vijayan, Effect of stabilizing additives on the structure and hydration of proteins: a study involving monoclinic lysozyme, Acta Crystallogr. D 58 (2002) 1162–1167.
[62] S.A. Islam, D.L. Weaver, Molecular interactions in protein crystals: solvent accessible surface and stability, Proteins 8 (1990) 1–5.
[63] M. Frey, Water structure associated with proteins and its role in crystallization, Acta Crystallogr. D 50 (1994) 663–666.
[64] G.N. Phillips, B.M. Pettitt, Structure and dynamics of the water around myoglobin, Protein Sci. 4 (1995) 149–158.
[65] V. Lounnas, B.M. Pettitt, G.N. Phillips, A global model of the protein-solvent interface, Biophys. J. 66 (1994) 601–614.
[66] J.S. Jiang, A.T. Brunger, Protein hydration observed by X-ray diffraction, J. Mol. Biol. 243 (1994) 100–115.