A dry ligand-binding cavity in a solvated protein

Johan Qvist*, Monika Davidovic*, Donald Hamelberg†, and Bertil Halle‡

*Center for Molecular Protein Science, Department of Biophysical Chemistry, Lund University, SE-22100 Lund, Sweden; and †Howard Hughes Medical Institute, Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, CA 92039-0365

Edited by Brian W. Matthews, University of Oregon, Eugene, OR, and approved January 31, 2008 (received for review October 16, 2007)

Ligands usually bind to proteins by displacing water from the binding site. The affinity and kinetics of binding therefore depend on the hydration characteristics of the site. Here, we show that the extreme case of a completely dehydrated free binding site is realized for the large nonpolar binding cavity in bovine β-lactoglobulin. Because spatially delocalized water molecules may escape detection by x-ray diffraction, we use water 17O and 2H magnetic relaxation dispersion (MRD), 13C NMR spectroscopy, molecular dynamics simulations, and free energy calculations to establish the absence of water from the binding cavity. Whereas carbon nanotubes of the same diameter are filled by a hydrogen-bonded water chain, the MRD data show that the binding pore in the apo protein is either empty or contains water molecules with subnanosecond residence times. However, the latter possibility is ruled out by the computed hydration free energies, so we conclude that the 315 Å3 binding pore is completely empty. The apo protein is thus poised for efficient binding of fatty acids and other nonpolar ligands. The qualitatively different hydration of the β-lactoglobulin pore and carbon nanotubes is caused by subtle differences in water-wall interactions and water entropy.

β-lactoglobulin | free energy simulation | hydrophobic hydration | magnetic relaxation dispersion

Globular proteins are stabilized by dense atomic packing and by water exclusion from the nonpolar core region. However, the packing density is not uniform and a typical protein contains approximately four cavities per 100 residues of sufficient size to accommodate at least one water molecule (1). Small nonpolar cavities are usually empty and can be regarded as packing defects, whereas small polar cavities tend to be occupied by structural water molecules that stabilize the folded protein by H-bonding to otherwise unsatisfied peptide partners (2). Large cavities are frequently linked to protein functions, such as ligand binding and transport, membrane translocation, and enzyme catalysis. Some of these large cavities are lined exclusively or predominantly by nonpolar side-chains. The question whether such large nonpolar cavities are empty or contain water molecules is of fundamental biological importance (3–8).

The principal tool of structural biology, x-ray crystallography, cannot easily resolve this issue. There are three potential problems. (i) Because water molecules interact only weakly with the nonpolar cavity walls, they tend to be positionally disordered. As a result of such delocalization, the electron density of the water molecules may fall below the detection limit. (ii) At a typical resolution limit of ~2 Å, a contaminating nonpolar ligand may be misinterpreted as a water cluster (9, 10). (iii) For structures determined at cryogenic temperatures, the hydration status of the cavity may be altered by re-equilibration during the flash cooling process (11).

In favorable cases, water molecules in nonpolar protein cavities can be inferred from intermolecular nuclear Overhauser effects (NOEs) measured by solution NMR spectroscopy (12, 13). However, weak water–protein NOEs, as expected from disordered water molecules, are difficult to distinguish from intramolecular NOEs relayed by exchangable protons (14) or even from long-range NOEs to bulk water (15). Molecular dynamics and free-energy simulations can in principle provide detailed insights about cavity hydration, but their predictive power is limited by the known shortcomings of current force fields (16, 17).

Several attempts have been made to establish the hydration state of a nonpolar 40 Å3 cavity in human interleukin-1β. The two x-ray diffraction studies that focused on this issue came to different conclusions: the first study assigned two water molecules with 70% occupancy to the cavity (18), whereas the more recent one indicated that the cavity is empty (19). Free energy calculations on a nonpolar spherical model cavity indicated stable hydration by clusters of at least three cyclically H-bonded water molecules (20), but the interleukin-1β x-ray data are not consistent with a cluster of three or more water molecules unless the occupancy is very low (so the cavity is empty most of the time). In contradiction with the x-ray results, water NOEs with protein protons near the interleukin cavity were taken as evidence for two to four water molecules in the cavity (12) and a recent MD simulation study indicated stable hydration by a four-water cluster (21). The hydration status of the interleukin-1β cavity thus remains controversial.

Free-energy calculations on a large, weakly polar, globular cavity in the surface layer protein tetrabrachion predicted stable hydration by a cluster of seven to nine water molecules (22), in line with calculations on model cavities (20) and fully consistent with the crystal structure. Agreement between free energy calculations and x-ray crystallography was also found for a 160 Å3 nonpolar cavity engineered into T4 lysozyme, which was found to be empty under ambient conditions but cooperatively filled by several water molecules at kbar pressures (23).

Here, we investigate the hydration of a functional cavity that is both larger and less polar than any of the cavities mentioned above. Like other members of the lipocalin family, bovine β-lactoglobulin (BLG) has a β-barrel fold enclosing a cavity that binds nonpolar ligands, including palmitate (9, 24, 25). This cavity, known as the calyx, has a volume of 315 Å3 (as gauged by a 1.4 Å probe) with approximately cylindrical (5 × 15 Å) shape. The cavity is lined by 12 aliphatic and one aromatic side chain, with only one polar atom (Met-107.S) within 5 Å of the cavity axis. Access to the calyx is restricted by a flexible loop that forms a lid on the calyx entrance. Deprotonation of Glu-89, with an anomalously high pKₐ value, triggers a conformational transition that opens the lid (9, 26). Depending on ionic strength, this so-called Tanford transition occurs in the pH range 6–8.

The available crystallographic evidence does not conclusively establish whether the calyx is hydrated in the apo form of BLG. In several structures with the loop in the open conformation (at pH > 7), no electron density was found in the calyx (9, 24).

However, in the structure 3BLG, in the closed conformation at pH 6.2, the electron density observed in the calyx was modeled as a string of five water molecules (Fig. 1) within H-bonding interactions. The available crystallographic evidence does not conclusively establish whether the calyx is hydrated in the apo form of BLG. In several structures with the loop in the open conformation (at pH > 7), no electron density was found in the calyx (9, 24). However, in the structure 3BLG, in the closed conformation at pH 6.2, the electron density observed in the calyx was modeled as a string of five water molecules (Fig. 1) within H-bonding interactions.
distance (2.7–3.0 Å) of each other (9, 27). The C–C separations across the calyx at the positions of these five water molecules are closely matching the corresponding dimension (8.1 Å) of the smallest carbon nanotube (CNT). Experimental and computational studies show that such CNTs are filled by a chain of water molecules under ambient conditions (28).

To clarify the hydration status of the calyx, we have performed water 17O and 2H magnetic relaxation dispersion (MRD) measurements (29) on BLG in aqueous solution. Because palmitate displaces any water molecules in the calyx, we can directly monitor calyx hydration via the difference in the MRD data recorded on the apo protein and the palmitate–BLG complex. This difference-MRD approach has been used to characterize internal water molecules in polar cavities (30–32). The MRD data indicate that the calyx is devoid of long-lived (ns) water, and the presence of more mobile water is ruled out by the free energy calculations reported here. We also obtain information about the smaller polar cavities, which are found to be occupied by ordered water molecules with residence times of ~20 ns, consistent with our MD simulations.

Results
Identification of Internal Water Molecules. Water 17O and 2H MRD profiles were recorded from aqueous solutions of BLG at neutral pH (Figs. 2 and 3). The apo form of BLG, without any hydrophobic ligand in the calyx, was examined at pH 7.4 (sample A1, calyx lid open) and at pH 6.2 (sample A2, calyx lid closed). Two independent samples (H1 and H2) of the holo form with a palmitate molecule bound in the calyx were studied, both at pH 7.4. At neutral pH, the protein exists predominantly as a homodimer. The dimer interface buries only 6% of the solvent-accessible surface area and does not interfere with ligand binding because the entrance to the calyx is 15 Å from the nearest dimer contact (9, 24). The structural changes in the protein induced by fatty acid binding are small and localized (25). Any observed difference between the concentration-normalized MRD profiles from the holo and apo proteins can therefore be attributed to water molecules in the calyx of the apo form, which are displaced by the bound palmitate molecule in the holo form. This argument assumes that the holo sample really has palmitate

Fig. 1. View of the crystal structure 3BLG (9) of β-lactoglobulin, showing two water molecules (red) buried in small polar cavities (labeled 175 and 177 in 3BLG, with 177 closer to the calyx) and five water molecules (blue) modeled in the nonpolar calyx. We label the calyx waters 1–5 starting from the bottom; the corresponding 3BLG labels are 178, 190, 179, 180, and 233. The cavity surface in the calyx is color-coded according to whether the contributing protein atoms are polar (red) or nonpolar (brown). The light green cartoon representation shows one of the two β sheets enclosing the nonpolar calyx. The figure was rendered with PyMOL (www.pymol.org) using a 1.4 Å probe to define the (external and internal) molecular surfaces of the protein and the same protein vdW radii as used for the cavity volume calculations.

Fig. 2. Water 17O MRD profiles from aqueous solutions of β-lactoglobulin with (holo) or without (apo) bound palmitate at 27°C. The data symbols refer to apo BLG at pH 7.4 (open squares, sample A1) and pH 6.2 (open diamonds, sample A2); holo BLG at pH 7.4 (filled triangles, sample H1; inverted filled triangles, sample H2); and protein-free buffer (open circles). Data from different BLG solutions have been normalized to the protein concentration of sample A1 (0.79 mM, Nw = 69300), using the scaling R1-R0/Nw implicit in Eq. 1. The curves were obtained by fitting the model to the two sets of apo or holo data as described in the text. The parameter values resulting from the fits are given in Table 1.

Fig. 3. Water 2H MRD profiles from aqueous solutions of β-lactoglobulin with (holo) or without (apo) bound palmitate at 27°C. Symbols and curves as in Fig. 2. The 2H and 17O data were measured on the same samples.
bound in the calyx. This was confirmed by \(^{13}\)C NMR spectroscopy, yielding 0.94 ± 0.03 (sample H1) and 0.91 ± 0.02 (sample H2) mole bound palmitate per mole BLG. If the calyx were determined by the structure 3BLG (9), water 175 makes three strong H-bonds in small polar cavities in nearly all crystal structures of BLG. In the aid of Eq. 6, the observed correlation time, \(\tau_{\text{obs}}\), can then only provide bounds on the water molecules in the calyx and on their order parameters and residence times (see Discussion).

The quantitative analysis of the MRD data is based on a model with two classes (a, b) of hydration water in fast exchange with bulk water. The \(\beta\) class contains the most long-lived internal water molecules, responsible for the dominant low-frequency component in the MRD profile. There are \(N_{\beta}\) such water molecules, with mean-square orientational order parameter \(S_2^\beta\). The \(\alpha\) class comprises the remaining hydration waters, interacting with the protein surface. They have a wide distribution of correlation times, that may range from the bulk water correlation time, \(\tau_0 = 1.9\) ps, up to the \(\beta\)-class correlation time, \(\tau_\beta = 11\) ± 2 ns. We model this distribution with a power law, but this choice hardly affects the \(\beta\)-class parameters of primary interest here.

The curves in Figs. 2 and 3 were obtained by fitting three model parameters to the MRD data: the order-parameter-weighted occupancy \((N_{\beta}S_2^\beta)\) and mean residence time \(\tau_{\text{w}}(N_{\beta}S_2^\beta)\) of long-lived (>10^{-8}s) internal water molecules, and the power-law exponent \((v)\) in the correlation time distribution for the external hydration layer. Because there is no significant difference between the two sets of data for the apo form (at pH 6.2 and 7.4), a single fit was performed on the combined apo data. Similarly, a single fit was performed on the combined data for the two equivalent holo samples. The parameter values resulting from these fits are collected in Table 1. The key finding is that there is no significant difference between the apo and holo samples in any of the three parameters. This is true for the \(^{17}\)O data as well as for the \(^{2}\)H data. Specifically, the \(^{17}\)O data yield \(N_{\beta}S_2^\beta(\text{apo}) - N_{\beta}S_2^\beta(\text{holo}) = -0.05 \pm 0.13\) with 95% confidence limits. This result rules out the presence of even a single moderately ordered water molecule in the calyx, provided that it has a residence time of order 10^{-8}s or longer.

The correlation time, \(\tau_{\text{w}}\), deduced from MRD profiles of protein solutions usually matches the protein rotational correlation time, \(\tau_{\text{RF}}\), and can then only provide bounds on the water residence time, \(\tau_{\text{w}} \leq \tau_{\text{RF}} \leq \tau_{\text{obs}} \leq 1/(6S_2^\beta\tau_{\text{RF}}) \approx 1\) μs (29). Here, however, the observed correlation time \(\tau_{\text{w}}\) for both sets of \(^{17}\)O data are a factor 2 shorter than the rotational correlation time of the BLG dimer, \(\tau_{\text{RF}} = 23\) ns, allowing us to determine \(\tau_{\text{w}}\) with the aid of Eq. 4. The long residence time, \(\tau_{\text{w}} \approx 20\) ns, indicates that these water molecules are at least partly buried, but they cannot be located in the calyx because they are present in the apo form as well as in the holo form. The value \(N_{\beta}S_2^\beta(\text{apo}) - N_{\beta}S_2^\beta(\text{holo}) = -0.05 \pm 0.13\) with 95% confidence limits. However, during the last 4.5 ns of the trajectory, both sites are simultaneously occupied, with order parameters \(S_2^{(17)O} = 0.61 \pm 0.05\) (W177a) and 0.40 \pm 0.05 (W177b). The sum of the \(S_2^{(17)O}\) values for the three water molecules, 1.7 \pm 0.1, is smaller than the experimental value, \(N_{\beta}S_2^\beta = 2.6 \pm 0.2\). If the simulation yields accurate order parameters, there must be at least one additional long-lived, fully or partly buried, water molecule. Whereas the simulations were performed on monomeric BLG, 95% of the protein in the MRD sample is present in dimeric form (supporting information SI Text). (The MRD parameter \(N_{\beta}S_2^\beta\) is quoted on a monomer basis.) The difference between the experimental and computed \(N_{\beta}S_2^\beta\) values might therefore be due to one or more long-lived water molecules at the dimer interface. In the BLG crystal dimer 1BSY (9), three water molecules are partly buried at the rim of the interface, each making two strong H-bonds with the protein.

Order parameters were also computed for the putative string of five water molecules in the calyx (Fig. 1). To prevent these water molecules from escaping from the calyx, they were harmonically restrained to their crystallographic positions with the same harmonic potential as used for the free energy calculations, but their orientations were not constrained. We thus obtained \(N_{\beta}S_2^\beta = 0.3 \pm 0.2\) for the three water molecules in the calyx. Because \(N_{\beta}S_2^\beta(\text{apo}) - N_{\beta}S_2^\beta(\text{holo}) = -0.05 \pm 0.13\) with 95% confidence, we can rule out this hydration motif provided that the water molecules are sufficiently long-lived to contribute to the \(\beta\) dispersion (see Discussion).

The \(^{2}\)H order parameter is generally smaller than the \(^{17}\)O order parameter, because only the former is affected by 180° flips about the water \(C_2\) axis (29). If the \(C_2\) flip is fast compared with \(\tau_{\text{w}} \approx 10\) ns and if other internal motions can be neglected, but makes only one strong H-bond to the backbone (Leu-39.O) in addition to two weaker polar interactions with side-chains (Met-24.S and Gln-120.O).

The \(^{2}\)H MRD profiles (Fig. 3) are similar to the \(^{17}\)O profiles (Fig. 2). As for \(^{17}\)O, there is no significant difference between the parameter values derived from the apo and holo profiles (Table 1). However, the \(^{2}\)H parameters \(N_{\beta}S_2^\beta\) and \(\tau_{\text{w}}\) are significantly larger than their \(^{17}\)O counterparts. These differences reflect a contribution from labile BLG hydrogens in hydroxyl, amino, and guanidino groups, exchanging with water deuterons on the millisecond time scale at pH 7.4. There are 82 such hydrogens in BLG, but they are in the slow-exchange regime and therefore add to \(N_{\beta}S_2^\beta\) only a small fraction of their maximum contribution (29). The water contribution to \(N_{\beta}S_2^\beta(\text{apo})\) is significantly less than \(N_{\beta}S_2^\beta(\text{holo})\), because \(S_2^\beta(\text{H}) < S_2^\beta(\text{O})\) (see below). Therefore, as much as half of \(N_{\beta}S_2^\beta(\text{H})\) may be produced by labile hydrogens. Because these labile hydrogens have residence times in the millisecond range, much longer than \(\tau_{\text{w}}\), the effective correlation time \(\tau_{\text{w}}\) should be longer for \(^{2}\)H than for \(^{17}\)O, as observed (Table 1). The \(^{2}\)H MRD data also rule out the (unlikely) possibility that the calyx contains water molecules with residence times in the range 1–100 μs, in which case they would contribute to the \(^{2}\)H MRD profile (fast exchange) but not to the \(^{17}\)O MRD profile (slow-exchange regime) (29).

**Table 1. Results of fits to water \(^{17}\)O and \(^{2}\)H MRD data**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Nuclide</th>
<th>(N_{\beta}S_2^\beta)</th>
<th>(\tau_{\text{w}}) ns</th>
<th>(v)</th>
<th>(\chi^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apo (A1 and A2)</td>
<td>(^{17})O</td>
<td>2.6 ± 0.2</td>
<td>20 ± 2</td>
<td>1.91 ± 0.02</td>
<td>1.1</td>
</tr>
<tr>
<td>Holo (H1 and H2)</td>
<td>(^{17})O</td>
<td>2.7 ± 0.2</td>
<td>18 ± 2</td>
<td>1.93 ± 0.02</td>
<td>0.7</td>
</tr>
<tr>
<td>Apo (A1 and A2)</td>
<td>(^{2})H</td>
<td>3.2 ± 0.2</td>
<td>27 ± 3</td>
<td>1.93 ± 0.02</td>
<td>0.7</td>
</tr>
<tr>
<td>Holo (H1 and H2)</td>
<td>(^{2})H</td>
<td>3.0 ± 0.2</td>
<td>28 ± 4</td>
<td>1.94 ± 0.02</td>
<td>0.5</td>
</tr>
</tbody>
</table>
we expect $S^2_{\text{H}_2O}/S^2_{\text{H}_2O} = 2.7$ (29). The simulations show that this limit is reached for the weakest polar hydration site (W177b), whereas the ratio is 1.3 for W175 and W177a. Internal water molecules without H-bonds to the protein are likely to undergo fast flip motions and are therefore more strongly manifested in the $^{17}$O MRD profile. Indeed, for the five calyx sites, the computed ratio $S^2_{\text{H}_2O}/S^2_{\text{H}_2O} = 2.0$ and 2.7.

From the power-law exponent, $v = 1.92 \pm 0.02$ (Table 1), deduced from the $^{17}$O MRD profiles, we can obtain the product of the number, $\Delta N_o$, of surface hydration water molecules in a given correlation time range and their mean-square order parameter, $S^2_{\text{H}_2O}$. We thus find $\Delta N_o S^2_{\text{H}_2O} = 12 \pm 1$ for $\tau_N > 100$ ps. From the MD trajectory, we identify $\Delta N_o = 17 \pm 1$ surface hydration sites that are occupied by water molecules with residence times longer than 100 ps. This is consistent with the experimental result: correlation times as long as 100 ps and above should reflect exchange out of the hydration site, rather than the much faster restricted rotation within the site which makes $S^2_{\text{H}_2O} < 1$.

### Discussion

An observed difference between the $^{17}$O MRD profiles recorded before and after a well-defined molecular perturbation, such as palmitate binding to the calyx of BLG, would have provided unambiguous evidence for the presence of water in the calyx of the apo protein. Our finding that there is no significant difference does not categorically rule out the presence of water in the calyx, but it does impose bounds on the possible values of the order-parameter-weighted occupancy, $N_o S^2_{\text{H}_2O}$, and on the mean residence time, $\tau_W$, for any water molecules in the calyx. These bounds can be established by computing the difference MRD profile, $\Delta R_1 = R_1(\text{apo}) - R_1(\text{holo})$, for a range of $N_o S^2_{\text{H}_2O}$ and $\tau_W$ values and comparing these with experimental $\Delta R_1$ data (Fig. S1).

Given the experimental uncertainty, $\Delta R_1 = 2 \text{ s}^{-1}$, we can then divide the $N_o S^2_{\text{H}_2O}$ plane into allowed and excluded regions (Fig. S2). Using the MD estimate, $N_o S^2_{\text{H}_2O} = 1.3$, we find that the MRD data rule out a five-water chain in the calyx if $\tau_W > 2$ ns.

On the other hand, water molecules in the calyx would escape detection by MRD if they were sufficiently short-lived ($\tau_W < 2$ ns). MD simulations of single-file water chains in CNTs yield residence times of order 0.1–1 ns (33). However, the calyx is closed at one end and is not as straight and smooth as a CNT, and these factors would slow down water exchange from the calyx. In particular, the flexible loop at the calyx mouth would partly block water exchange in the “closed” conformation at pH 6.2 (9, 26). Nevertheless, it is not certain that these factors would make $\tau_W > 2$ ns. Therefore, we rely on the free energy calculations (Table 2) to rule out the case of short-lived calyx hydration. The methodological limitations in these calculations, notably the incomplete representation of polarization effects, are not likely to overturn the decisive outcome.

In conclusion, the experimental and theoretical results presented here are not consistent with a hydrated calyx in BLG. The electron density observed in the calyx in the crystal structure of 3BLG must therefore be attributed to binding of a nonpolar impurity, a possibility that was entertained in the original publication (9) but that was later dismissed (27). Similar problems have been encountered with other lipocalins (10).

In the absence of other interactions than a hard confining potential, the mean water occupancy of a cavity (of arbitrary size and shape) is simply $V_{\text{cav}}/V_{\text{cav}}$, where $V_{\text{cav}}$ is the (mean) volume of the (fluctuating) cavity and $V_{\text{cav}}$ is the molar volume of water vapor in equilibrium with bulk water, which is $1.16 \times 10^6 \text{Å}^3$ at 300 K. Identifying $V_{\text{cav}}$ with the volume, 27 Å$^3$, accessible to the center of a 1.4 Å radius probe in the calyx of the 3BLG structure, divided by 6 because the calyx contains approximately six potential hydration sites, we obtain $\xi(n) = 4 \times 10^{-6}$. As expected, the computed single-water occupancies are larger, $>10^{-4}$ (Table 2), indicating that water–protein interactions provide significant stabilization. Direct calculation of the mean water–protein interaction energy from the 10-ns MD simulation with “forced” hydration of the calyx yields $-16 \pm 3 \text{kJ mol}^{-1}$ for each of the five hydration sites (Table S1), essentially the same value for a spontaneously hydrated CNT of similar dimensions (28).

Polarization effects are likely to contribute to the different (observed) hydration of CNT and calyx. In contrast to the CNT, with polarizable $\pi$ electrons in the graphene sheet, the calyx is lined by 12 aliphatic (Leu-39, Val-41, Val-43, Val-45, Leu-54, Ile-56, Leu-58, Ile-71, Ile-84, Leu-87, Val-92, Leu-103) and only one aromatic (Phe-105) side chain. The incomplete treatment of polarization effects in the force field used in both of these simulations is likely to underestimate the stabilization energy in the CNT (because the O–H–$\pi$ interaction is not properly described by the Leonard–Jones parameters) and overestimate it in the calyx (because the TIP3P effective dipole moment is larger than expected for a single water molecule in a nonpolar environment).

### Table 2. Hydration free energies and occupancies for calyx sites

<table>
<thead>
<tr>
<th>$n^*$</th>
<th>Single</th>
<th>$1 \rightarrow 5$</th>
<th>$5 \rightarrow 1$</th>
<th>$10^6 \log \xi(n)$</th>
<th>$1 \rightarrow 5$</th>
<th>$5 \rightarrow 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$24 \pm 2$</td>
<td>$24 \pm 2$</td>
<td>$12 \pm 1$</td>
<td>$-4.1 \pm 0.4$</td>
<td>$-4.2 \pm 0.4$</td>
<td>$-2.1 \pm 0.2$</td>
</tr>
<tr>
<td>2</td>
<td>$23 \pm 1$</td>
<td>$17 \pm 2$</td>
<td>$14 \pm 1$</td>
<td>$-3.9 \pm 0.1$</td>
<td>$-2.9 \pm 0.4$</td>
<td>$-2.5 \pm 0.2$</td>
</tr>
<tr>
<td>3</td>
<td>$12 \pm 2$</td>
<td>$9 \pm 2$</td>
<td>$19 \pm 1$</td>
<td>$-2.1 \pm 0.4$</td>
<td>$-1.6 \pm 0.4$</td>
<td>$-3.3 \pm 0.2$</td>
</tr>
<tr>
<td>4</td>
<td>$14 \pm 3$</td>
<td>$11 \pm 4$</td>
<td>$7 \pm 1$</td>
<td>$-2.5 \pm 0.6$</td>
<td>$-1.8 \pm 0.8$</td>
<td>$-1.3 \pm 0.2$</td>
</tr>
<tr>
<td>5</td>
<td>$10 \pm 4$</td>
<td>$10 \pm 2$</td>
<td>$7 \pm 1$</td>
<td>$-1.8 \pm 0.6$</td>
<td>$-1.8 \pm 0.4$</td>
<td>$-1.3 \pm 0.2$</td>
</tr>
<tr>
<td>All</td>
<td>$71 \pm 6$</td>
<td>$60 \pm 3$</td>
<td>$-12.4 \pm 1.1$</td>
<td>$-10.5 \pm 0.4$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Hydration sites labeled as in Fig. 1.

$^1$Free energies in kJ mol$^{-1}$ refer to a standard concentration of 30.02 Å$^{-3} = 55.32$ M.
Hydration of nonpolar cavities can be promoted by electric fields produced by remote charges and dipoles (12, 34). A simple model calculation, taking into account the dipole moment (μ = 1.855 D) and polarizability (α = 1.662 × 10⁻¹¹ C m² V⁻¹) of the water molecule in a singly occupied cavity, shows that an electric field E enhances the occupancy ξ above the value ξ₀ for an ideal cavity according to ξ(1 - ξ) = ξ₀ exp(A sin(μ/α)M, where A = αE²(2kBT) and M = μE/((μkBT). It thus follows that a field of ~10 GV m⁻¹ is required to obtain roughly half occupancy of a small nonpolar cavity, whereas a field of ~5 GV m⁻¹ has negligible effect (ξ < 10⁻³ in our case). We computed the protein-generated electric field E at the oxygen of internal water molecules from the 10-ns MD trajectory, finding 12.5 GV m⁻¹ for W175, in the typical range for small polar cavities, but only 1.7–2.8 GV m⁻¹ for the three deepest calyx sites (Table S1). Calyx hydration should thus not be significantly enhanced by protein-induced water polarization. Entropic factors also make the calyx more hydrophobic than a CNT. By occupying space, water molecules in the calyx suppress conformational fluctuations of the lining nonpolar side-chains, thereby increasing the free energy.

The total interaction energy per water molecule in the five-water chain (Fig. 1) is about ~60 kJ mol⁻¹ (Table S1), ~15% less than in the CNT (28). Each water–water interaction within the calyx contributes about ~20 kJ mol⁻¹, as expected for a well formed H-bond with mean O–O separation of 2.84 Å (Table S1). Each water molecule in the chain switches H-bond donor and rotates about the calyx axis on a picosecond time scale, consistent with the relatively small order parameters and the large S₂(17O)/S₂(1H) ratios. Yet, the H-bonds are highly linear, with mean (OH,OO) angle of 15–25°. Higher water ordering (lower entropy) may thus contribute, along with the slightly weaker stabilization energy, to explain why the calyx is dehydrated while the CNT is hydrated.

BLG belongs to the calycin superfamily (35). The calycins share a common antiparallel β-barrel folding motif, and they all bind nonpolar molecules in the enclosed cavity. However, the mechanism and energetics of ligand binding differ widely among the calycins. For example, rat intestinal fatty acid binding protein (I-FABP), an abundant cytoplasmic calycin, binds palmitate (in an extended conformation, as in BLG) in a ~500 Å³ globular cavity of mixed polarity and occupied by ~20–30 water molecules (31, 36). This contrasts sharply with the completely nonpolar, empty binding cavity in BLG that seems to be poised for receiving an acyl chain or other nonpolar ligand. Yet, the palmitate affinity is lower for BLG (~10⁶ M⁻¹) than for I-FABP (~10⁸ M⁻¹) (37, 38). The different strategies for fatty acid binding used by these proteins may provide important clues to their still incompletely understood biological functions. All but one of the 13 nonpolar side-chains lining the calyx in BLG are conserved, or replaced by another nonpolar residue, in human glycodelin and in mouse major urinary protein (35). The binding cavity in the latter protein is not entirely nonpolar and appears to contain ~4 water molecules that are displaced by the phosphorenone ligand (6). A recent MD simulation indicates that the binding cavity of apo cyclooxygenase-2 is empty 80% of the time (8), but the crystal structure of the complex with arachidonic acid shows 26 polar atoms, including five water molecules, within 5 Å of the ligand (39). To the best of our knowledge, BLG is the only clearly documented case of a large dehydrated binding cavity.

What are the biological implications of the complete exclusion of water from the 315 Å³ calyx in BLG? It has been suggested that a hydrophobic pore in a membrane protein may act as a “vacuum lock,” where polarity-modulating conformational changes control the permeability to polar molecules by expelling water molecules in a microscopic version of capillary evaporation (or dewetting) (4, 40). In a hydrophobic pore with only one entrance, the effect of the void is instead to facilitate binding of nonpolar ligands. In I-FABP, the molecules that are displaced as the ligand enters the cavity can exit through another transient channel. Because such a “back door” does not exist in BLG, productive ligand binding can only occur when the calyx has been evacuated. Stable hydration of the calyx would therefore drastically reduce the rate of ligand association as well as the binding affinity.

Materials and Methods

Sample Preparation and Characterization. Bovine β-lactoglobulin isofrom A was purified and complexed with 13C-labeled palmitate as detailed in SI Text. The amount of palmitate bound to BLG in the holo samples was determined by ¹³C NMR spectroscopy as described in SI Text.

Magnetic Relaxation Dispersion. The longitudinal relaxation rate, R₁, of the water ²H and ¹⁷O magnetizations was measured with 0.5–1.0% accuracy at 27.0 ± 0.1°C using five different NMR spectrometers, as described in SI Text. Fits to the MRD data were performed with the Levenberg–Marquardt algorithm and the parameter errors quoted in Table 1 were obtained with the Monte Carlo method using 1,000 synthetic data sets.

The MRD data were interpreted in terms of a model with two classes of hydration water (ω₁, ω₂) in fast exchange (on the spin relaxation time scale) with bulk water. The observed spin relaxation rate can then be decomposed as (29)

\[
R₁(ω) = R₁^0 + Nω/Nw [(R₁^0(ω₁)) - R₁^0] + Nω/Nw [(R₁^0(ω₂)) - R₁^0],
\]

where R₁^0 is the relaxation rate measured on the protein-free buffer solution (open circles in Figs. 2 and 3) and the angular brackets signify an average over all water molecules in the class. Nω is the analytically determined water/BLG mole ratio; see SI Text. Class ω₁ comprises the N½ (~500) water molecules in contact with the external protein surface, whereas class ω₂ contains a small number of internal water molecules with mean residence time τω₂ comparable to, or longer than, the rotational correlation time τR (23 ns) of the dimeric protein. The average intrinsic relaxation rates in Eq. 1 are given by (29)

\[
(R₁^0(ω₁)) = 0.2 J(2.0ω₁) + 0.8 J(2.0ω₂),
\]

and similarly for R₁(ω₂). Here, J(ω) is the known nuclear quadrupole frequency (29) and the motional spectral density for class ω has the Lorentzian form

\[
J(ω) = S_p^2/1 + (ωτ_p)^2 ,
\]

where S_p is the root-mean-square orientational order parameter (29) and the correlation time τ_p is determined by protein tumbling and/or internal water exchange according to:

\[
1/τ_p = 1/τ_W + 1/τ_R,
\]

For further details, see SI Text.

Molecular Dynamics Simulations. MD simulations of monomeric BLG in water were performed with the NAMD code (41) and the Amber ff99 force field (42). The protein, with atomic coordinates from the crystal structure 3BLG (9) and with added H atoms and disulfide bonds, was solvated with 6,249 TIP3P water molecules (including the crystal waters) in a truncated octahedron. The net charge of the protein, with ionizable residues modeled in their dominant protonation states at neutral pH (also Glu-89 was deprotonated), was neutralized by nine sodium ions. After energy minimization and equilibration, the MD simulation was performed at 300 K and 1 atm with periodic boundary conditions, particle-mesh Ewald summation for long-range electrostatics, 10 Å cutoff for nonbonded interactions, constrained X-H bonds, and 2-fs time step. Atomic coordinates were saved every 2 ps of the 20-ns MD trajectory. Order parameters, electric fields and mean interaction energies for internal water molecules were calculated as described in SI Text.

Free Energy Calculations. The standard free energy of hydration, ΔG_H₂O(n), for the process where a water molecule is transferred from bulk water to hydration site n in the calyx (labeled as in Fig. 1) was computed by thermodynamic integration using the double-decoupling method (43, 44) as described in SI Text. Calculations were performed for hydration of the calyx by a single water molecule, ΔG_H₂O(n), and for sequential addition of water molecules to all of...
the five hydration sites shown in Fig. 1. The latter filling process was carried out in two ways: starting from the bottom, \(G_{\text{hyd}}(n;1 \rightarrow 5)\), or the mouth, \(G_{\text{hyd}}(1)\), of the calyx. With the standard state concentration corresponding to bulk water at 300 K (55.32 M), the site occupancy \(\xi(n)\) is related to the standard free energy of hydration as

\[
\frac{\xi(n)}{1 - \xi(n)} = \exp \left[ - \frac{\Delta G_{\text{hyd}}(n)}{k_BT} \right].
\]

[5]

Supporting Information

Qvist et al. 10.1073/pnas.0709844105

SI Text

Sample Preparation. Bovine β-lactoglobulin (BLG) isoform A (Sigma, L-7880) was purified by anion exchange chromatography at pH 6.0 (10 mM Mes buffer) on a DEAE-sephacel column, eluted with 25–300 mM NaCl gradient. The main fraction from the column was further purified by size-exclusion chromatography on a Sephadex G-50 column at pH 2.7 (10 mM citric acid, 50 mM NaCl). The protein was then exhaustively dialyzed against milli-Q water and lyophilized. The high purity was confirmed by agarose gel electrophoresis and SDS/PAGE. In addition, we recorded a 1H MRD profile (1 kHz to 10 MHz), which confirmed that any high molecular weight impurities/aggregates had been removed by the gel filtration (1). The low pH in this step also removes any endogenous lipids bound to the protein (2), so the purified protein is in the apo state.

NMR samples were made by dissolving the purified protein in 25 mM potassium phosphate buffer at pH 7.4, made from 52 wt% D2O (99.9%, low paramagnetic content, Cambridge Isotope Laboratories) and 48 wt% H217O (19 atom% 17O, Isotec) with addition of 0.05 wt% NaN3. Quoted pH values are uncorrected for H/D isotope effects. Final protein concentrations were determined with an accuracy of ±1% by complete amino acid analysis (performed at the Amino Acid Analysis Center, Department of Biochemistry and Organic Chemistry, Uppsala University, Uppsala, Sweden). Two apo samples with 0.79 mM BLG and pH 7.4 (sample A1) or pH 6.2 (sample A2) were investigated by MRD.

Samples with palmitate in complex with BLG were prepared with the wall-to-protein transfer protocol (3). Palmitic acid was deposited on the glass wall of the NMR tube as the chloroform solvent was evaporated by a stream of argon gas. The protein solution was then added to the NMR tube and incubated at 37°C for 24 h. The BLG concentrations in the two holo samples were 0.74 mM (sample H1) and 0.80 mM (sample H2).

13C NMR Spectroscopy. The holo solutions contained palmitate with 13C-labeled carbonyl carbon (99%, Cambridge Isotope Laboratories) and, as an internal reference, 4.76 mM (sample H1) or 2.50 mM (H2) sodium acetate with 13C-labeled methyl carbon (99%, Cambridge Isotope Laboratories). 13C spectra were recorded on a Varian Unity Plus 600 spectrometer using 6336 transients, the carrier frequency precisely halfway between the acetate and palmitate resonances, 7 μs 90° pulse length, no proton decoupling, 12 s repetition time, and 24 h total acquisition time. The concentration bound palmitate was obtained from the integrated intensities of the palmitate and acetate peaks. The 13C signal from excess palmitate (∼4 equivalents) in the holo samples, present as vesicles or microcrystallites, is broadened beyond detection. Neither excess palmitate nor acetate affects the water relaxation significantly.

1H and 17O MRD Experiments. The longitudinal relaxation rate, R1, of the water 1H and 17O magnetizations was measured at 11 magnetic fields, corresponding to the Larmor frequency range 0.70–92.1 MHz for 1H and 0.67–81.3 MHz for 17O. Frequencies below 10 MHz were accessed with an iron-core magnet (Drusch EAR-35N) interfaced to a Teagc Discovery console, while higher frequencies were sampled with four NMR spectrometers with conventional cryomagnets (Bruker Avance DMX 100 and 200, Varian Unity Plus 500 and 600). A sufficient number of transients was accumulated to obtain a signal-to-noise ratio of at least 100. R1 was determined from a three-parameter fit of single-exponential magnetization curves obtained with the inversion recovery pulse sequence. The frequency-independent relaxation rate of a protein-free buffer sample was measured at each field, allowing the experimental uncertainty in R1 to be estimated to 0.5–1.0%. The sample temperature was regulated to ±0.1°C and was measured with a thermocouple referenced to an ice-water bath.

Analysis of MRD Data. The definition of the α and β classes differs slightly from that used in most earlier MRD studies of protein hydration (4). Previously, β waters accounted for the entire observed frequency dependence in R1, while the α waters were responsible for the frequency-independent relaxation enhancement (above the bulk water value, R10). Because the highest accessed frequency is ~100 MHz, the two classes corresponded to water molecules with correlation times shorter than (α) or longer than (β) ~1 ns. The MRD profiles in Figs. 2 and 3 cannot be adequately described by a single correlation time τp. We therefore generalize the model by including in class α all hydration water molecules with correlation times shorter than τp (but longer than the bulk water correlation time τo). Furthermore, we adopt a power-law distribution, for the correlation time τp of α waters in the range [τo; τp] (5). This form is suggested by MD simulations (6) and low-temperature MRD studies (C. Mattea, J.Q., B.H., unpublished data). Class α comprises the Ne water molecules in contact with the external protein surface (7). This number (per BLG monomer) was estimated from the solvent-accessible surface area of the BLG dimer. With 1.4 Å probe radius, we obtain 15,250 Å2 as an average of three crystal dimers. Dividing this by 2 (to convert to monomer basis) and by 15 Å2, the mean surface area occupied by one water molecule (4, 8), we obtain Nα = 508.

With a monomer-dimer association constant of 2.03 × 105 M−1 at pH 6.5 (1), BLG in our samples (~0.8 mM) should be 95% dimeric. The experimentally determined (10) rotational correlation time, τRP, of monomeric BLG is reproduced by hydrodynamic calculations on the crystal structure, using an effective atomic radius of 2.97 Å (1), close to generally recommended value of 3.0 Å (11). For this radius, hydrodynamic calculations yield τRP(dimer)/τRP(monomer) = 2.26 (1) (using this ratio and scaling the experimental monomer value (τRP ∼ η/T) to our temperature (27.0°C) and solvent viscosity (0.959 cP), we obtain τRP(dimer) = 23.2 ns.

Cavity Volumes. Cavity volumes were calculated with the program VOIDOO (12), using the default set of united-atom vdW radii (from the original AMBER force field). To define the calyx volume in 3BLG unambiguously, water 277 was replaced by a dummy atom with 2 Å vdW radius that blocks the entrance to the hydrophobic part of the cavity. The cavity was refined in 10–20 steps with an initial grid size of 0.25 Å and with a grid shrinkage factor of 0.9 per step. The convergence criteria were either <0.1% change in cavity volume in consecutive steps or <1% standard deviation for the last four steps. The cavity volume is defined here as the probe-occupied volume, i.e., the complement to the molecular surface. This is the most appropriate volume measure for assessing the hydration capacity of a cavity. The cavity volume depends strongly on the probe radius. For the calyx volume in 3BLG we thus obtained 397 Å3 with 1.2 Å probe, 315 Å3 with 1.4 Å probe, and 212 Å3 with 1.6 Å probe. The volume reported in the main text refers to a probe radius of 1.4 Å. Using the same procedure, we computed the volume of the
nonpolar cavity in interleukin-1β (Protein Data Bank structure 11B) to 38 Å² (1.4 Å probe), in agreement with the value (39 Å²) quoted in a recent x-ray diffraction study (13). The hydration capacity of the BLG calyx is thus eight times that of the interleukin-1 β cavity.

Order Parameters. Orientational order parameters for the water 17O and 2H nuclear quadrupole tensors in internal water molecules were computed from equations given elsewhere (4) and newly developed code. The 20 ns MD trajectory was used for the two internal water molecules 175 and 177, whereas a separate 10 ns MD trajectory was used to compute order parameters for the five water molecules in the calyx. The 10 ns trajectory was generated in the same way, except that a harmonic restraining potential, \( V_{\text{r}} = (1/2) k_{\text{r}} r_0^2 - r_0^2 \), was applied to prevent the water molecules from escaping from the calyx (as they did at the equilibration stage of the 20-ns trajectory). Here, \( r_0 \) is the actual position of the water oxygen and \( r_0 \) is its position in the crystal structure 3BLG (14). The force constant \( k_{\text{r}} \) was set to 25.1 kJ mol\(^{-1}\) Å\(^{-2}\), corresponding to a root-mean-square displacement of 0.55 Å or an isotropic thermal B factor of 7.8 Å\(^2\). Protein rotation was suppressed by harmonic restraints on the C\(^\alpha\) atoms of Glu-55, Lys-91 and Phe-136. Both trajectories were subdivided in 500 ps blocks, the first of which was discarded. Quoted \( S^2 \) values refer to the mean and standard deviation of the block averages.

Mean Interaction Energies. The mean interaction energy, \( U \), of a given water molecule with a neighboring (H-bonded) water molecule, with protein atoms, or with all atoms was obtained from the 10 ns MD trajectory used to compute order parameters. The results given in Table S3 are based on 300 time frames taken at 10 ps intervals. In the same way, we computed the mean electric field at the position of the oxygen atom of internal water molecules. The contribution to this field from the protein atoms is given in Table S3.

Hydration Free Energies. The free energy calculations were performed with the AMBER 8.0 package (amber.scripps.edu). The simulation details were the same as for the 10 ns MD trajectory, except that the truncated octahedron was smaller (2,902 water molecules).

The standard free energy of hydration, \( \Delta G_{\text{hydr}}^0(n) \), was obtained as the difference \( \Delta G_W = \Delta G_W^0(n) \), where \( \Delta G_W^0(n) \) refers to transfer of a water molecule from hydration site \( n \) in the protein cavity (P) to vacuum and \( \Delta G_W \) refers to the transfer of a water molecule from bulk water (W) to vacuum. \( \Delta G_W^0(n) \) was computed by thermodynamic integration along a decoupling path where the electrostatic part of the interaction of the water molecule with the rest of the system was gradually turned off before the first term was obtained by thermodynamic integration as described above. The second term corrects for the confining effect of the restraint potential and defines the standard state.

The restraint volume is \( V_{\text{res}} = (2\pi \kappa_0 T/k_{\text{res}})^{3/2} \), where \( \kappa_0 \) is the effective constraint constant (16). Our choice of standard state thus corresponds to a bulk water concentration of \( 1/V_W = 55.32 \) M. Because the water molecules in the calyx undergo symmetric C\(^2\) flips on a picosecond time scale, the configuration space is fully sampled and Eq. 6 should not contain any orientational constraint correction (16, 17).

Since the restraint potential restricts the sampled configuration space, the computed hydration free energy is overestimated. To demonstrate that this effect is insignificant, we computed \( \Delta G_{\text{hydr}}^0(3!5) \) with the force constant reduced by a factor 3 to \( k_{\text{res}} = 8.4 \) kJ mol\(^{-1}\) Å\(^{-2}\). The result with the weaker restraint, 8.5 ± 1.7 kJ mol\(^{-1}\), did not differ significantly from the result, 9.2 ± 2.5 kJ mol\(^{-1}\), with the stronger restraint. We also checked that the trajectory segments used in each step of the thermodynamic integration were sufficiently long to adequately sample configuration space. We thus used trajectories of either 40 ps or 400 ps total length (divided into four blocks) for each \( \lambda \) value when computing the hydration free energy for site 3 while the other four sites were occupied. The results did not differ significantly: 1.3 ± 1.4 kJ mol\(^{-1}\) for the short and 1.2 ± 2.2 kJ mol\(^{-1}\) for the long trajectory. Even though 10-ps blocks thus appear to be sufficient, we used 20-ps blocks for all reported results.

Fig. S1. Water $^{17}$O difference MRD profiles (0.79 mM BLG, pH 7.4, 27°C), showing the expected contribution from water molecules in the calyx that are displaced by palmitate in the holo protein. The red circles were obtained from the experimental $R_1$ data (for samples A1 and H1) and the blue curves were calculated from Eqs. 1–4 with $N_pS_p^2 = 1.3$ (from the MD simulation), $\tau_{WP} = 23$ ns and the indicated water residence time, $\tau_W$. The palmitate occupancy was set to 94%, as in sample H1. Note that only the last term in Eq. 1 contributes to $\Delta R_1 = R_1$(apo) – $R_1$(holo).
Fig. S2. Contour diagram showing curves of constant $^{17}$O $\Delta R_1$ at the lowest measurement frequency (0.67 MHz) for different combinations of the order-parameter-weighted occupancy, $N_0 S^2_0$, and residence time, $\tau_W$. The contours were calculated in the same way as the MRD profiles in SI Fig. S1. The dashed line corresponds to the $N_0 S^2_0$ value obtained from the MD simulation with 5 water molecules confined to the calyx. If the detection level is taken as $\Delta R_1 = 2$ s$^{-1}$, this $N_0 S^2_0$ value implies that the calyx waters would escape detection only if $\tau_W < 2$ ns.
Table S1. Geometry and energetics of internal water molecules in BLG

<table>
<thead>
<tr>
<th>Property</th>
<th>Calyx hydration site n</th>
<th>Small cavity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 1</td>
<td>n = 2</td>
</tr>
<tr>
<td>R(On – O_{n+1}), Å</td>
<td>2.83 (1)</td>
<td>2.85 (1)</td>
</tr>
<tr>
<td>β_{nb}, deg*</td>
<td>19.0 (7)</td>
<td>15.7 (5)</td>
</tr>
<tr>
<td>E_{b}, GV m^{-1}</td>
<td>1.7 (4)</td>
<td>2.8 (7)</td>
</tr>
<tr>
<td>U(W_{n} – W_{n+1}), kJ mol^{-1}</td>
<td>-20.8 (2)</td>
<td>-20.9 (2)</td>
</tr>
<tr>
<td>U(W_{n} – P), kJ mol^{-1}</td>
<td>-14.9 (2)</td>
<td>-18.8 (3)</td>
</tr>
<tr>
<td>U_{tot}(W_{n}), kJ mol^{-1}</td>
<td>-38.4 (2)</td>
<td>-64.2 (4)</td>
</tr>
</tbody>
</table>

Results obtained by averaging over 300 configurations, taken at intervals of 10 ps from a 10-ns MD trajectory where five water molecules were harmonically confined to the crystallographic hydration sites in the calyx (see Fig. 1). Digits within parentheses give the standard error of the mean.

*Mean H-bond angle, between O—H bond of Wn and O_{n+1} vector.
†Mean electric field at water oxygen produced by protein atoms.
‡Mean interaction energy of water molecule n with water molecule n + 1 [U(W_{n} – W_{n+1})], with all protein atoms [U(W_{n} – P)], or with all atoms [U_{tot}(W_{n})].