

# The Low Barrier Hydrogen Bond in the Photoactive Yellow Protein: A Vacuum Artifact Absent in the Crystal and Solution

Timo Graen,<sup>†</sup> Ludger Inhester,<sup>‡</sup> Maike Clemens,<sup>†</sup> Helmut Grubmüller,<sup>\*,†</sup> and Gerrit Groenhof<sup>\*,§</sup>

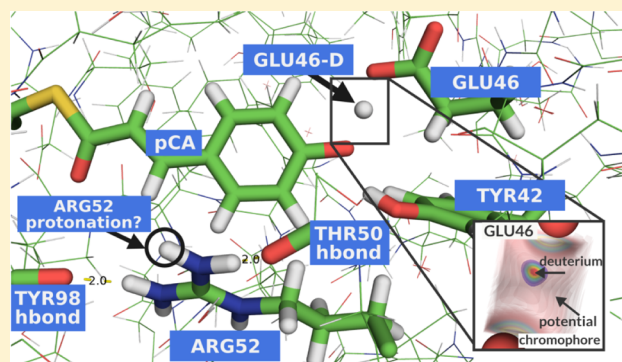
<sup>†</sup>Max Planck Institute for Biophysical Chemistry, Am Faßberg 11, 37077 Göttingen, Germany

<sup>‡</sup>Center for Free-Electron Laser Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany

<sup>§</sup>Department of Chemistry and Nanoscience Center, University of Jyväskylä, P. O. Box 35, 40014 Jyväskylän Yliopisto, Finland

## Supporting Information

**ABSTRACT:** There has been considerable debate on the existence of a low-barrier hydrogen bond (LBHB) in the photoactive yellow protein (PYP). The debate was initially triggered by the neutron diffraction study of Yamaguchi et al. (*Proc. Natl. Acad. Sci., U. S. A.*, **2009**, *106*, 440–444) who suggested a model in which a neutral Arg52 residue triggers the formation of the LBHB in PYP. Here, we present an alternative model that is consistent within the error margins of the Yamaguchi structure factors. The model explains an increased hydrogen bond length without nuclear quantum effects and for a protonated Arg52. We tested both models by calculations under crystal, solution, and vacuum conditions. Contrary to the common assumption in the field, we found that a single PYP in vacuum does not provide an accurate description of the crystal conditions but instead introduces strong artifacts, which favor a LBHB and a large <sup>1</sup>H NMR chemical shift. Our model of the crystal environment was found to stabilize the two Arg52 hydrogen bonds and crystal water positions for the protonated Arg52 residue in free MD simulations and predicted an Arg52 pK<sub>a</sub> upshift with respect to PYP in solution. The crystal and solution environments resulted in almost identical <sup>1</sup>H chemical shifts that agree with NMR solution data. We also calculated the effect of the Arg52 protonation state on the LBHB in 3D nuclear equilibrium density calculations. Only the charged crystal structure in vacuum supports a LBHB if Arg52 is neutral in PYP at the previously reported level of theory (*J. Am. Chem. Soc.*, **2014**, *136*, 3542–3552). We attribute the anomalies in the interpretation of the neutron data to a shift of the potential minimum, which does not involve nuclear quantum effects and is transferable beyond the Yamaguchi structure.



## INTRODUCTION

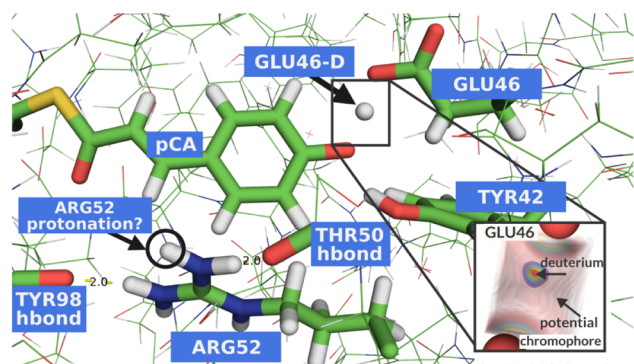
The photoactive yellow protein (PYP) is a small 125 amino acid protein that was purified from the phototropic bacterium *Haloerhodospira halophila*.<sup>1</sup> PYP was the first protein for which the X-ray structure provided evidence for a short strong hydrogen bond<sup>2</sup> with direct relation to function. Later, Yamaguchi et al.<sup>3</sup> classified this H-bond as a low barrier hydrogen bond (LBHB) with a delocalized proton. The classification was based on a combined X-ray and neutron diffraction study, which revealed an unusually large covalent deuterium bond distance of 1.21 Å for the deuterium (Glu46-D) between residue Glu46 and the oxygen of the PYP chromophore (Figure 1), *para*-coumaric acid (pCA). This bond length is significantly longer than the 1.0 Å expected for a regular covalent hydrogen bond, a feature indicative of a delocalized LBHB. Yamaguchi et al.<sup>3</sup> placed the Glu46-D deuterium at the minimum of an  $F_O - F_C$  omit map, in which the Glu46-D was removed. An experimental resolution of 1.5 Å was reported<sup>3</sup> for the light deuterium atoms. At this resolution, also other Glu46-D bond lengths agree with the experiment,

but no alternative structural interpretations have been reported in the literature.

In addition to the peculiar position of the Glu46-D deuterium, residue Arg52, which shields the hydrophobic chromophore pocket from the solvent, was suggested to be deprotonated<sup>3</sup> even at the experimental pD of 9. A deprotonated Arg52 was concluded based on a lack of nuclear density at two out of four deuterium positions, DH12 and DH22, with occupation values of 0.22 and 0.68, respectively. Prior to the Yamaguchi study, Arg52 was assumed to be protonated<sup>2</sup> because it is located solvent exposed at the PYP surface where it shields the chromophore pocket tightly in the ground state. The hydrogen bond network was believed to loosen only in the blue-shifted pB state further down in the photocycle.<sup>4</sup> Furthermore, theoretical results suggest that a protonated arginine is required to control the photoisomerization of the chromophore.<sup>5,6</sup>

Received: June 1, 2016

Published: November 30, 2016



**Figure 1.** PYP chromophore pocket. The chromophore pocket of PYP is shown, including the hydrogen bonding partners of the chromophore and Arg52.

A model connecting the two observations, which we here refer to as the Yamaguchi model,<sup>3</sup> was proposed in which a LBHB at the Glu46-D deuterium position is a consequence of the deprotonation of the Arg52 residue. This model has been investigated in several theoretical works. The earliest works by Saito et al.<sup>7–10</sup> focused on the energetics of the Glu46-D position, the neutral Arg52 protonation state, and the discrepancy between the LBHB and NMR data published by Sigala et al.<sup>11</sup> In their work, Sigala et al.<sup>11</sup> measured the <sup>1</sup>H NMR chemical shift of Glu46-H for PYP in solution but did not find the large chemical shifts expected for a LBHB (i.e., 17–19 ppm). Instead, Sigala et al. reported a solution chemical shift of  $\delta_{\text{H}} = 15.2$  ppm.<sup>11</sup> Furthermore, Oktaviani and Yoshimura et al.<sup>12,13</sup> did not observe changes associated with Arg52 over the pH range 3.4 to 11.2, suggesting that Arg52 remains protonated, unless its  $\text{pK}_{\text{a}}$  is below 3.4. The latter, however, would require a  $\text{pK}_{\text{a}}$  shift of 8 units, which can be considered highly unlikely for solvent exposed arginines.<sup>14</sup> Both observations seem to argue against the Yamaguchi model in solution.

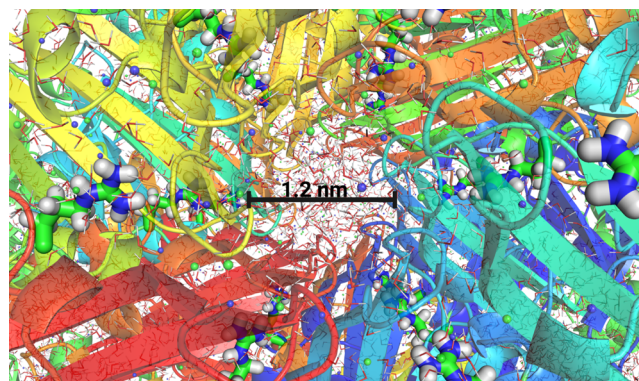
Further support that Arg52 is protonated for PYP in solution comes from calculations by Saito and Ishikita,<sup>7,8</sup> who estimated a  $\text{pK}_{\text{a}}$  above 12 based on a PB continuum model, in line with results from the NMR titration experiments of Oktaviani and Yoshimura et al.<sup>12,13</sup> They further computed Glu46 <sup>1</sup>H NMR chemical shifts of  $\delta_{\text{H}} = 14.5$ – $14.6$  ppm for QM/MM optimized geometries<sup>7</sup> in agreement with Sigala et al.<sup>11</sup> In contrast, the unoptimized crystal geometry in vacuum led to a much larger value of  $\delta_{\text{H}} = 19.7$  ppm. The authors concluded their work by stating that the hydrogen positions in the Yamaguchi structure were “essentially reasonable”<sup>7</sup> given the 1.5 Å resolution, thus questioning their own strong arguments against a LBHB in PYP.

The Yamaguchi model remained appealing and triggered further interest in the theoretical community, in which the Yamaguchi model regained support in two recent theoretical works by Nadal-Ferret et al.<sup>15</sup> and Kanematsu et al.<sup>16</sup> The first study by Nadal-Ferret et al.<sup>15</sup> supported the Yamaguchi model based on calculations of the nuclear deuterium wave functions with a QM/MM description of the hydrogen bond potential. Their calculations showed that only for a deprotonated Arg52 does Glu46-D delocalize, whereas it reverts into a regular hydrogen bond for a protonated Arg52. Kanematsu et al.<sup>16</sup> reached the same conclusions independently with their multicomponent quantum mechanical ONIOM method. Both works relied on CAM-B3LYP/6-31+G(d,p)<sup>17</sup> density func-

tional theory [DFT] calculations of the crystal coordinates in vacuum.

Previous theoretical studies of the LBHB rested on the assumption that the influence of the environment surrounding the protein on the LBHB is small and that crystal conditions were therefore sufficiently well described by a single copy of the charged protein in vacuum. Based on this assumption, support for the hypothesis that the formation of the LBHB in PYP in the crystal is triggered by the deprotonation of Arg52 was reported,<sup>15,16,18</sup> but without an explanation for the rather unusual protonation state of this intrinsically high  $\text{pK}_{\text{a}}$  residue.<sup>14</sup>

A visual inspection of the crystal unit cell [Figure 2] reveals a wide solvent accessible pore through the crystal lattice, with



**Figure 2.** PYP crystal pore. Semi-opaque rendering of the solvated 2ZOI crystal model illustrating the environment previously approximated by vacuum. In the center, the solvent pore is visible, with which Arg52 (sticks) is contact.

which Arg52 is in direct contact. Since water and ions can move freely through that pore, Arg52 is in direct contact with solvent molecules. As the buffer solution in the experiment has a high ionic strength with 1.1 M sodium chloride and 2.5 M ammonium sulfate, the crystal is likely highly polar but charge neutral, unlike the crystal structure in vacuum, which has a net charge. In the crystal, the adjacent proteins and solvent provide, due to their polarizability, markedly more shielding for the electrostatic interaction between charged residues and the LBHB than a vacuum environment. This effect reduces the interaction of charged residues at the PYP surface with the LBHB candidate in the crystal compared to a vacuum environment.

Several open questions remain regarding Arg52 and the LBHB. If the Yamaguchi model is valid, how does the crystal environment shift the  $\text{pK}_{\text{a}}$  of Arg52 from above 12 in solution to below 9 in the crystal? Furthermore, it remains unclear how the large extent of Glu46-D delocalization, as observed in previous calculations on an isolated protein in vacuum,<sup>15</sup> changes when basic elements of the crystal unit cell are included into the calculations such as symmetry mates, water molecules, and ions. Finally, given the low signal-to-noise for the light atom positions in the Yamaguchi et al.<sup>3</sup> neutron data, can there be alternative structural models, which do not require a deprotonated Arg52 or nuclear quantum effects at room temperature, to account for the measured neutron structure factors?

To address these open questions, we have performed a series of MM and QM/MM computations, in which we calculated the

difference in the  $pK_a$  of Arg52 between PYP in solution and crystal, the nuclear equilibrium density of the Glu46-D, and  $^1\text{H}$  chemical shift. The previously suggested Arg52 trigger mechanism in the formation of a Glu46-D LBHB was only observed for the charged crystal structure in vacuum but not in a solvated or crystal environment. Our results suggest that Arg52 is protonated and the LBHB does not form. We further reinvestigated the Yamaguchi et al.<sup>3</sup> neutron diffraction data and argue that the signal-to-noise ratio for the Arg52 deuterium atoms is not high enough to conclusively determine the protonation state. At the Glu46-D location, we observed that also bond lengths below the 1.21 Å reported by Yamaguchi et al.<sup>3</sup> produce very similar  $F_O - F_C$  maps. Based on our calculated Glu46-D nuclear equilibrium densities and the structure factors measured by Yamaguchi et al.,<sup>3</sup> we attribute the increase of the Glu46-D bond length to a shift of the minimum on the underlying potential energy surface without LBHB formation. Both our alternative model and the original Yamaguchi model fit the X-ray and neutron scattering data within experimental uncertainty. However, in contrast to the Yamaguchi model, our model also agrees with NMR solution data,<sup>11–13</sup> and the protonation state of Arg52 in our model is more in line with a pH of 9.

## METHODS

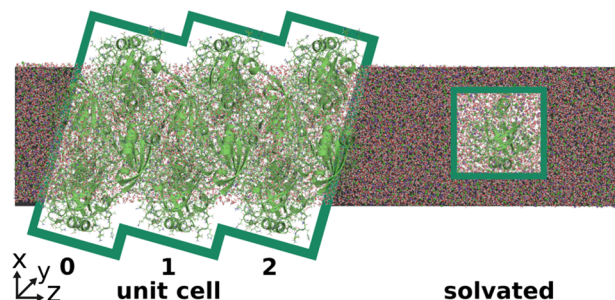
**Preparation Crystal Structure.** The neutron diffraction structure of Yamaguchi et al.<sup>3</sup> (2ZOI) was completed for missing protein and solvent atoms using the GROMACS4.6<sup>19,20</sup> tools. The Yamaguchi et al.<sup>3</sup> structure was post-processed to leave the resolved hydrogen and deuterium positions of protein and water atoms unchanged. We further also prepared the high resolution 1.2 Å X-ray structure of Coureux et al.<sup>21</sup> (2QJ5) in the same way as an independent reference.

From the completed vacuum structures, AMBER03<sup>22,23</sup> protein force field parameters were generated with a previously described AMBER03 extension for the PYP chromophore.<sup>24</sup> In order to also describe the deprotonated Arg52 residue at the molecular dynamics level, charges for the deprotonated arginine residue were parametrized into the AMBER03<sup>22</sup> protein force field. This was done by deriving point charges for the deprotonated state while keeping all other arginine parameters unchanged. To this end, the residue was capped with NME (methylammonium) at the C-terminus and ACE (acetate) at the N-terminus. The structure was optimized at the B3LYP<sup>25</sup>/cc-pVTZ//HF/6-31G\*\* level of theory from which RESP<sup>26</sup> charges were derived, see Duan et al.<sup>22</sup> for details.<sup>24</sup>

The potentials for the Glu46-D nuclear equilibrium densities as well as for the Glu46-D quantum corrections to the Arg52 protonation free energy were calculated using the QM/MM Gaussian09<sup>27</sup> interface in GROMACS 4.6.<sup>19,20,28</sup> A QM region was constructed, which consisted of the pCA chromophore, its hydrogen bonding partners, Pro68, Glu46, Tyr42, and Arg52. In this QM zone, the side chain bonds were capped between the QM and MM zones at the protein backbone using hydrogen link atoms. The link atoms were placed at an effective distance of 4 Å or four bonds away from the Glu46-D atom. At the boundary, direct MM bonding partner charges were omitted<sup>29</sup> during the QM optimization to reduce the effect of overpolarization. We further cross-checked the QM/MM procedure against previously published vacuum calculations.<sup>15</sup> The remaining protein interacted with the QM zone via electronic embedding for which the AMBER03<sup>22,23</sup> partial charges were used. Unless specified otherwise, a cutoff of 5 nm was used, which included a 5 nm sphere of ionic solution and a 5 nm sphere crystal environment into the MM environment.

Three MM environments were considered: The “vacuum environment” consisted of only the protein, the resolved crystal water molecules, and no ions, modeled after the system of Nadal-Ferret et al.<sup>15</sup> No modifications were made to the structure other than completing the structure for missing heavy atoms and protons. The

“crystal environment” consisted of a PYP copy in the center of the crystal slab [Figure 3, left]. The “solvated environment” was constructed by placing a single protein copy of the PYP in solution [Figure 3, right].



**Figure 3.** PYP free energy setup. Solvated simulation setup of three PYP crystal unit cells and a solvated PYP in 5 M NaCl buffer solution.

We created a crystal model in contact with a solvent reservoir that mimics the experimental conditions of high ion concentration as described by Yamaguchi et al.<sup>3</sup> This model [Figure 3] is the first to include space group ( $P6_3$ ) and crystal packaging information from experiment, both of which have not been considered in previous computational studies of the Glu46-D hydrogen bond in PYP. In the crystal model, the unit cells must be neutral on average to prevent a Coulomb explosion of the crystal. Charge neutrality requires modeling of the unknown ion distribution in the crystal as no ion positions were resolved experimentally.<sup>3</sup> Modeling of a charge neutral crystal was achieved through equilibration of solvent and ion molecules into the crystal in a molecular dynamics simulation using the setup described below. The ionic reservoir has approximately the same ionic strength as the experimental buffer solution<sup>3</sup> used in the experiment.

A model of the PYP unit cell was constructed by including all six symmetry mates ( $P6_3$ ) from the symmetry information specified in the 2ZOI structure using Chimera.<sup>30</sup> This unit cell was replicated two times along the  $\pm z$  direction, parallel to the crystal pore [Figure 2]. A reservoir buffer of 5 M NaCl<sup>31</sup> solution was added to the system together with a single PYP protein as shown in Figure 3. The final setup consisted of 19 copies of PYP, 19663 TIP3P water molecules,<sup>32</sup> 1872  $\text{Na}^+$ , and 1776  $\text{Cl}^-$  ions,<sup>31</sup> totaling 99307 atoms. The central unit cell was in contact with the other crystal unit cells and the ionic solution inside the crystal but not with the buffer reservoir (left side of Figure 3). Therefore, this central unit cell, referred to as unit cell 1, was used for analysis. The size of the simulation box was chosen such that the solvated PYP was 5 nm away from the crystal slice and fully embedded in solvent, while the central PYP molecules in the crystal model were embedded in approximately 5 nm of other crystal symmetry mates. This simulation setup was coupled to a thermal bath via the  $v$ -rescale<sup>33</sup> thermostat at 300 K with a 1 ps time constant. Anisotropic Berendsen pressure coupling with compressibility  $1.0 \times 10^{-13} \text{ bar}^{-1}$  in  $xy$  and  $4.5 \times 10^{-5} \text{ bar}^{-1}$  in  $z$  direction was applied with a time constant of 1 ps, corresponding to a flat two-dimensional crystal. Electrostatics were treated by PME using a Fourier grid spacing of 0.12 nm with a direct space cutoff of 1.0 nm for Coulomb and Lennard-Jones interactions. The system was equilibrated for 25 ns using a 1 fs time step and position restraints on the heavy protein atoms with a force constant of  $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$  and no protein bond length constraints. Water molecules were kept rigid with SETTLE.<sup>34</sup>

Initially, randomly selected water molecules were replaced with ions. The equilibration time of the ions was chosen such that the ions permeated through the full depth of the central crystal pore [Figure 2]. The central crystal pore is a feature observed experimentally<sup>3</sup> as a result of the crystal structure and space group information. This way, a charge neutral crystal model was obtained. For the calculations of the nuclear equilibrium densities of Glu46-D, the protein crystal coordinates were fitted back onto the equilibrated model to

compensate for small deviations from the average position in the X-ray structure.

**pK<sub>a</sub> Shift Estimation.** The pK<sub>a</sub> value of solvent exposed arginine residues lies above 12.<sup>35</sup> In PYP, residue Arg52 is exposed to the solvent even under crystal conditions. One would therefore expect Arg52 to be protonated at the experimental pD value of 9. In order to explain the deprotonated Arg52 in the Yamaguchi structure,<sup>3</sup> a large pK<sub>a</sub> shift of −3 pK<sub>a</sub> units has to be induced by the crystal environment over the solution phase.<sup>12,13</sup> A shift of this magnitude should be detectable in free energy calculations.

To estimate the pK<sub>a</sub> shift between crystal and solution, we calculated the free energy difference  $\Delta\Delta G^{\text{cryst-solv}}$  between Arg52 protonation in the crystal and solvated environments; see SI for details of the thermodynamic cycle. This free energy was further also corrected for the quantum delocalization of the Glu46-D deuterium. The pK<sub>a</sub> shift was calculated by thermodynamic integration using the crystal–reservoir simulation box [Figure 3] introduced above. According to the TIP3P water model, the bond lengths of the water molecules were constrained.<sup>34</sup> The simulation parameters of the equilibration phase were used without modifications, including position restraints on the heavy protein atoms with a force constant of 1000 kJ mol<sup>−1</sup> nm<sup>−1</sup>, no constraints on protein bond lengths, and a 1 fs integration time step.

The applied restraints are an approximation in solution to obtain the free energy difference between the correct states. The most important conformational changes contributing to the free energy are the two Arg52 conformations. In order to calculate the free energy differences for (de)protonating Arg52, it is mandatory that the free energy difference is calculated between the two closed conformations as otherwise the wrong difference would be considered. Under the assumption that the Arg52 is indeed deprotonated in the crystal as Yamaguchi et al.<sup>3</sup> have suggested, imposing restraints that favor the crystal structure may artificially upshift the pK<sub>a</sub> of Arg52 for PYP in solution.

The classical contribution to the pK<sub>a</sub> shift,  $\Delta\Delta G^{\text{TI}}$ , was calculated by linearly interpolating the system Hamiltonian between the state at  $\lambda = 0.0$ , in which Arg52 was protonated in the crystal and deprotonated in the solvated PYP, and the state with  $\lambda = 1.0$ , in which Arg52 was protonated in the solvent and deprotonated in the crystal. This setup was designed to maintain charge neutrality, which is essential to avoid artifacts,<sup>36</sup> throughout the free energy cycle. The  $\lambda$  dependent Hamiltonian was integrated in 21 discrete  $\lambda$  windows of 20 ns MD sampling. The last 15 ns of each trajectory was used for the analysis. The same thermodynamic integration procedure was repeated for the inverse process from  $\lambda = 1$  to  $\lambda = 0$ . An error estimate for  $\Delta G^{\text{TI}}$  was bootstrapped from the simulation data using 2500 bootstrap samples from 0.5 ns trajectory windows.

The contribution to the pK<sub>a</sub> shift obtained from these classical MD simulations ( $\Delta\Delta G^{\text{TI}}$ ) does not include contributions from the delocalization of Glu46-D, which may occur if the Arg52 protonation state becomes neutral. We corrected the Arg52 protonation free energy for quantum effects in two steps. First, the free energy contribution of a classical particle Glu46-D deuterium,  $\Delta\Delta G_{\text{cl}}^{\text{AMBER03}}$ , described at the AMBER03 force field level of theory was subtracted [SI eq 4]. We calculated this classical particle contribution by numerical integration over the conformational space, a sphere with a radius corresponding to the AMBER03 O–H distance  $r_{\text{OH}}$ , accessible to the particle given the same fixed oxygen position also used in the determination of the DFT PES. Then, the contribution from the deuterium was reintroduced as a nuclear equilibrium density,  $\Delta\Delta G_{\text{qm}}^{\text{DFT}}$ , described in the next section. The subscripts qm and cl denote whether a quantum mechanical (wave function) or classical (particle) description of the nucleus was used. This process was repeated for the crystal and solution environment, leading to the final quantum corrected estimate,  $\Delta\Delta G^{\text{cryst-solv}}$ . A detailed derivation of the individual terms is given in the SI.

The final pK<sub>a</sub> shift estimate

$$\Delta\text{p}K_{\text{a}} \approx \Delta\Delta G^{\text{cryst-solv}} / (2.303RT) \quad (1)$$

with  $RT_{300\text{K}} = 2.494$  kJ/mol was calculated from the quantum corrected Arg52 protonation free energy

$$\Delta\Delta G^{\text{cryst-solv}} \approx \Delta\Delta G^{\text{TI}} - \Delta\Delta G_{\text{cl}}^{\text{AMBER03}} + \Delta\Delta G_{\text{qm}}^{\text{DFT}} \quad (2)$$

Although our calculations of the Glu46-D nuclear equilibrium densities that are explained below suggest that neither with a protonated nor deprotonated Arg52 a LBHB is formed in the crystal or solution, we nevertheless investigated the effect the LBHB would have on the pK<sub>a</sub> of Arg52. In these simulations, which are described in detail in Supporting Information (S5–S7), we calculated the free energy required to deprotonate Arg52 in a infinitely periodic crystal with and without a LBHB between Glu46 and the chromophore. The former situation was modeled by using atomic charges [see SI S-5] for the Glu46–chromophore pair with a LBHB, whereas for the latter situation, the normal AMBER03 charges were used as described in the Methods section.

**Glu46-D Quantum Corrections and Equilibrium Densities.** In the quantum correction to the free energy and calculations of the nuclear equilibrium densities, the Glu46-D deuterium was treated as a nuclear quantum particle in an electronic potential described by DFT. The corrections were calculated for two QM/MM potential energy surfaces, calculated at the MPW1B95<sup>37</sup> and CAM-B3LYP<sup>17</sup> levels of theory with AMBER03<sup>22</sup> MM point charges and a 6-31+G(d,p) basis set. The nuclear equilibrium densities correspond to the probability density of finding the deuterium in its potential energy landscape at 300 K. Derived bond length expectation values can differ from the potential energy minimum due to zero point and thermal effects. The DFT potential energy for Glu46-D was evaluated on a numerical grid of  $17 \times 16 \times 17$  points around the geometrical center of the bond from the crystal structure coordinates. The grid spacing  $\Delta h = 0.15$  bohr was chosen such that the scanned region included both minima relevant for the delocalization of the particle. Because the deuterium wave function was solved in these calculations, we considered all possible O–D bond lengths and angles within the grid region simultaneously.

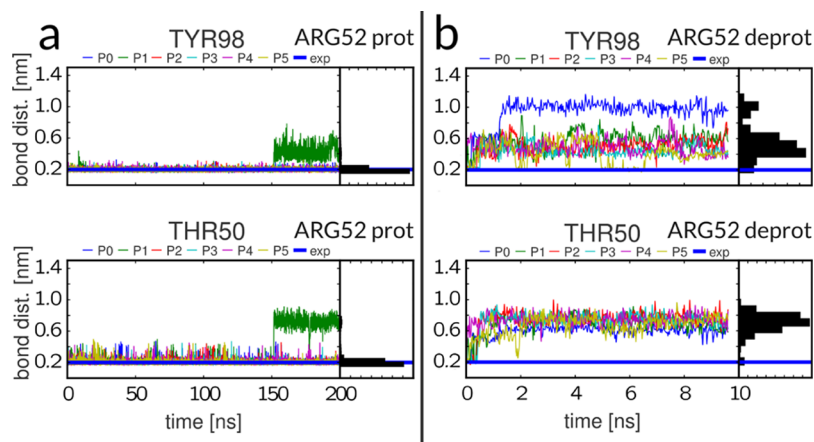
In addition to these single particle nuclear equilibrium densities, also a three particle O–D–O system was solved on a  $19 \times 17 \times 19$  grid. In these calculations, a finer grid spacing of  $\Delta h = 0.05$  bohr was used for the two oxygen nuclei (pCA-O, Glu46-O) due to their larger mass. Each of the three particles was treated as a 1D degree of freedom, which was fully coupled to all others. The two oxygens Glu46-O and pCA-O from the chromophore were scanned along their C=O bonds, whereas the deuterium was scanned along the coordinate connecting the two oxygen rest positions. The three particle potential energy surfaces were calculated in a QM/MM crystal model using the CAM-B3LYP<sup>17</sup>/6-31+G(d,p) level of theory. The QM region was energy minimized following the procedure used in the calculation of the NMR chemical shifts below. The potential energy at each of the 4624 (D), and 6137 (O–D–O) grid points was evaluated with QM/MM single point calculations. The potential energy grids were used as input to solve the discretized Schrödinger equation.

The extent of delocalization of Glu46-D was calculated by solving the 3D stationary discretized nuclear Schrödinger equation

$$\left[ \frac{-\hbar^2}{2m} \mathbf{D}^2 + \mathbf{V}_{\text{N}} \right] \psi_{\text{N}} = E \psi_{\text{N}} \quad (3)$$

for the hydrogen/deuterium atom of mass  $m$ , second derivative operator  $\mathbf{D}^2$ , and potential energy  $\mathbf{V}_{\text{N}}$ . The NuSol<sup>38</sup> program was used, which implements the *sinc*-DVR method as described by Colbert and Miller.<sup>39</sup> This approach can be applied to similar problems as the AI-PIMD methods applied by Markland et al.<sup>40</sup> The calculations were performed in atomic units with  $\hbar = 1$  and deuterium and hydrogen mass as suggested by NIST CODATA 2014,  $m_{\text{deuterium}} = 3670m_{\text{e}}$  and  $m_{\text{hydrogen}} = 1836m_{\text{e}}$ , respectively. For the oxygen nucleus, a mass of 16 protons or  $m_{\text{hydrogen}} = 29376m_{\text{e}}$  was used.

With this approach, the three-dimensional Glu46-D ground and first 10 excited state nuclear wave functions were calculated. Subsequently, the equilibrium density,  $\rho$ , was derived by Boltzmann weighting the eigenstates with their corresponding energies at the experimental



**Figure 4.** Free MD simulations of the Arg52 hydrogen bond network. The two hydrogen bonds lengths to residues Tyr98 (top) and Thr50 (bottom) are shown. (a) Arg52 is protonated and compared to the experimental distance (blue line). (b) Same setup as panel a but for the deprotonated Arg52 (AMBER03). Two sets of free MD simulations for different neutral Arg52 parametrizations are available in the SI. No position restraints were applied.

temperature of 300 K. Throughout the manuscript, all nuclear equilibrium densities are at 300 K.

The two sources of nuclear delocalization considered here are quantum effects of the ground state and thermal effects due to the population of nuclear excited states. A LBHB should have excited states very close to the ground state, which is why at 300 K there could be a significant extent of nuclear excited state occupation. The state populations can be determined from the energies of the calculated eigenstates. This way, we distinguish changes in the location of the potential minimum and the ground state density from changes in the extent of delocalization due to thermal occupation.

**Calculation of Chemical Shifts.**  $^1\text{H}$  NMR chemical shifts are sensitive to small changes in hydrogen bond lengths, and solution values are available for PYP.<sup>11</sup> We calculated the chemical shifts for Glu46-H in the three electrostatic environments introduced above: “vacuum”, “solvent”, and “crystal”, using the same QM/MM region separation. Prior to the chemical shift calculations, the QM region was optimized in the field of the static MM environment at the CAM-B3LYP/6-31+G\*\* level of theory. During the optimization, all heavy atoms of the protein backbone up to the  $\text{C}_\beta$  carbon were frozen, and only the remaining QM zone degrees of freedom were optimized. The  $^1\text{H}$  chemical shifts were calculated in the presence of the full MM environments around the protein from the optimized QM/MM geometries using the GIAO method<sup>41</sup> as implemented in GAUSSIAN09.<sup>27</sup> All shifts are reported relative to the TMS- $d_4$ <sup>11</sup>  $\delta_{\text{H}} = 31.72$  ppm averaged reference, which too was optimized at the CAM-B3LYP/6-31+G\*\* level of theory in combination with a PCM continuum model<sup>42</sup> of water with  $\epsilon_{\text{water}} = 78.3553$ .

## RESULTS AND DISCUSSION

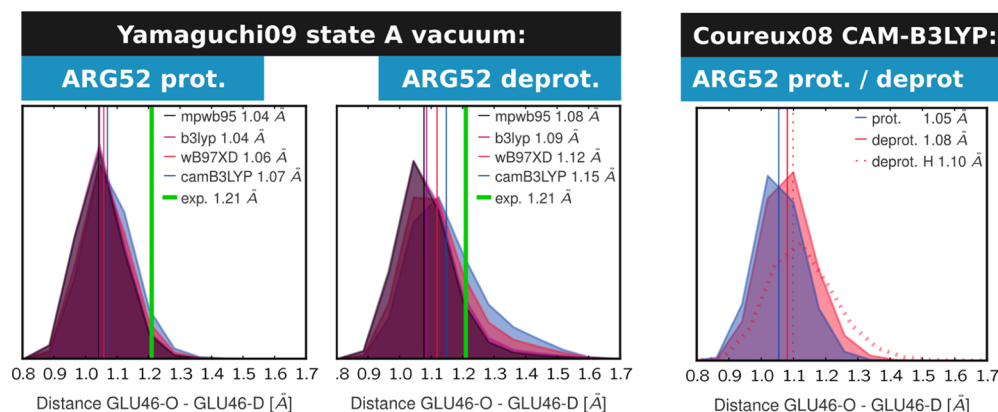
**Stability of the Arg52 Hydrogen Bond Network.** The stability of the Arg52 hydrogen bond network in the crystal was investigated. The Arg52 hydrogen bond network consists of two hydrogen bonds to residues Tyr98 and Thr50 with a bond length of  $d = 0.2$  nm.<sup>3</sup> First, Arg52 was simulated in the protonated state, and the lengths of the two hydrogen bonds were obtained for the six PYP copies in the center of the crystal model from 200 ns of free MD simulation [Figure 4a]. Of the 19 PYP copies in the system, only the six copies in the center of the crystal slab were considered as these most closely represent the crystal environment. Next, the deprotonated Arg52 was simulated for the neutral arginine force fields with charges derived according to AMBER03 specifications [Figure 4b and SI]. In the simulations with a protonated Arg52, the hydrogen bond network remained intact, with the exception of one PYP

copy, in which the bonds broke after  $\sim 150$  ns. In contrast to the protonated Arg52, the deprotonated Arg52 led to rapid breaking of the hydrogen bond network within the first nanosecond of simulation for all six copies of PYP. Two complementary parametrizations for a neutral arginine side chain were tested as well but did not affect the stability of the hydrogen bond network around Arg52. Thus, also in these simulations, the hydrogen bonds between the protein and Arg52 side chain disrupted rapidly. The results of these alternative parametrizations are discussed in the SI [SI-10].

The observed breaking of the hydrogen bond network for the neutral Arg52 caused the residue to move into the solvent and away from the position reported in the experimental X-ray structures 2ZOI, 2ZOH, and 2QJ5. Only the protonated Arg52 remained close to the conformation found in the X-ray structure with the hydrogen bond network in tact. It seems reasonable to assume that the experimental X-ray structures rule out an outward flipped Arg52 conformation.

In the X-ray and neutron diffraction structures<sup>3</sup> at 298 K, there is a well resolved crystal water molecule (ID 1023 in 2ZOI), which is in contact with Arg52. In our crystal MD simulations with a protonated Arg52, we also observe a water at that position, which reversibly exchanged during the simulations. However, in the simulations of the deprotonated Arg52, no water molecules occupied this position, which is another indication that Arg52 is protonated in the crystal. In the solution model, the protonated Arg52 showed a bound water at the same position. Solvent contacts for the deprotonated Arg52 were unspecific and not ordered but present due to the larger number of available neighbors. We speculate that this contributes to the free energy difference of deprotonating Arg52 in the crystal compared to the solvent.

A deprotonated Arg52 was unstable in free MD simulations. In contrast, a protonated Arg52 resulted in a much more stable Arg52 hydrogen bond network over the trajectory length of 200 ns. Additional force field parametrizations, which enforce the increased Glu46-O–D LBHB bond length, were also performed but nevertheless resulted in the same instability of the deprotonated Arg52 hydrogen bond network [SI S-10–S-12]. Furthermore, the position of crystal water 1023 in 2ZOI was occupied in simulations with a protonated Arg52 but not in



**Figure 5.** Comparison of different DFT functionals for the description of the Glu46-D delocalization in the Yamaguchi<sup>3</sup> and Coureux<sup>21</sup> structures (PYP in vacuum). (A) A one-dimensional cut through the three-dimensional equilibrium density  $|\psi|^2$  of Glu46-D is shown for a protonated Arg52 (left) and a deprotonated Arg52 (right). The expectation values for the Glu46-D distance  $\langle x \rangle$  at  $T = 300$  K are shown as vertical lines together with the value inferred by Yamaguchi et al.<sup>3</sup> on the basis of their neutron diffraction data. (B) Same setup as panel A, but calculations were performed on the Coureux et al.<sup>21</sup> [pdb id 2QJ5] crystal structure using the CAM-B3LYP functional. The dotted line shows the result for hydrogen instead of deuterium.

simulations with a deprotonated Arg52 for all three parametrizations tested.

These findings suggest that the charge on the protonated Arg52 is required to preserve the hydrogen bond network with the protein and the crystal water contacts.

**Estimation of Arg52  $pK_a$  Shift Due to the Crystal Environment.** We used the crystal slab model system described in the Methods section [Figure 3] to compute the Arg52  $pK_a$  shift due to the crystal environment. The Arg52  $pK_a$  difference between crystal and solution [Figure 3] was obtained by thermodynamic integration and corrected for QM effects associated with deuteron delocalization (Methods and SI). For these calculations, we exploited prior information about the Arg52 conformation in the crystal from the neutron and X-ray structures.<sup>3</sup> By placing restraints on the heavy protein atoms, the simulation was forced to sample conformations close to the experimental crystal structure, while allowing local fluctuations.

The final state of the thermodynamic integration corresponds to the suggested physical states in solution<sup>12,13</sup> (Arg52 prot.) and in the crystal (Arg52 deprot.).<sup>3</sup> Therefore, going from the unphysical starting conformation to the physical end state should decrease the free energy (i.e., spontaneous process). In this transformation, the proton was not physically transferred from a protein in the crystal to the protein in solution via a chemically feasible pathway. However, because free energy difference between these thermodynamic states is independent of the pathway, we may interpret the calculated free energy difference as the free energy associated with the actual proton (deuteron) transfer reaction.

The calculated free energy shift estimate is

$$\begin{aligned} \Delta\Delta G^{\text{cryst-solv}} &= \Delta\Delta G^{\text{TI}} - \Delta\Delta G_{\text{cl}}^{\text{AMBER03}} + \Delta\Delta G_{\text{qm}}^{\text{DFT}} \\ &= 12.2 \text{ kJ/mol (95\%CI, 11.1 to 13.4 kJ/mol)} \\ &\quad - 0.10 \text{ kJ/mol} + -0.13 \text{ kJ/mol} \quad (4) \end{aligned}$$

This number translates into a  $pK_a$  shift of 2.1  $pK_a$  (95% CI, 1.9  $pK_a$  to 2.3  $pK_a$ ) and -2.2  $pK_a$  (95% CI, -2.0 to -2.4  $pK_a$ ) units for the inverse process. All errors were estimated via bootstrapping as described in the Methods section. The quantum correction,  $\Delta\Delta G_{\text{qm}}^{\text{DFT}}$ , increases from -0.1 kJ/mol (MPW1B95) to 1.0 kJ/mol (CAM-B3LYP) when the DFT

functional is changed. The quantum correction reflects changes in the shape of the Glu46-D nuclear equilibrium densities between crystal and solution. Here, the two reported functionals are the most conservative (MPW1B95) and the most delocalized (CAM-B3LYP). For both qm corrections, the  $pK_a$  is 2.1 and 2.3  $pK_a$  units higher in the crystal than in solution.

The calculated free energy cost, including both enthalpic and entropic contributions, to effectively move the deuteron from the crystal to solution suggests that Arg52 is likely protonated in the crystal if it is protonated in solution.<sup>11-13</sup> We attribute the large upshift mainly to the loss of crystal water ordering for the deprotonated Arg52 described above.

These results contain two important findings. First, the  $pK_a$  shift estimate with restraints placed on the Arg52 hydrogen bond network showed an upshift in  $pK_a$  by at least 2  $pK_a$  units in the crystal compared to the solution PYP. Second, the contribution from the quantum delocalization  $\Delta\Delta G_{\text{qm}}^{\text{DFT}}$  of -0.1 (MPW1B95) to 1.0 kJ/mol (CAM-B3LYP) barely contributes to the total  $pK_a$  shift estimate. This correction is insufficient to cause a down shift in  $pK_a$  for the overall process and suggests that nuclear quantum corrections of the Glu46 proton/deuteron are not responsible for the deprotonation of Arg52.

A deprotonation in the crystal was found to be even more unlikely than in solution due to the increased  $pK_a$ . Also the influence of nuclear quantum effects, at the employed level of theory, is not strong enough to shift the  $pK_a$  noticeably toward a neutral Arg52 and neither can the more delocalized charge distribution on the Glu46 and the chromophore pair with a LBHB [SI S-10-S11].

**Glu46-D Delocalization in PYP: Vacuum.** In this section, we tested the effect that vacuum conditions far from the experimental crystal environment have on the extent of Glu46-D delocalization.

First, calculations were performed for the protein in vacuum. From these calculations, the extent of Glu46-D delocalization was determined. We solved the nuclear Schrödinger equation for a proton/deuteron in the QM/MM potentials calculated with four density functionals also including the functionals that have been used in previous theoretical works: MPW1B95,<sup>37</sup> B3LYP,<sup>25</sup> CAM-B3LYP,<sup>17</sup> and  $\omega$ B97XD.<sup>43</sup> The Glu46-D bond length expectation values,  $\langle x \rangle$ , at 300 K are shown [Figure 5]

for the protonated and deprotonated Arg52. One dimensional cuts through the three-dimensional equilibrium density,  $\rho(x)$ , of Glu46-D between the chromophore oxygen HC4-O4' and Glu46-O atom are shown. For all functionals, the Glu-D bond length increased to 1.08–1.15 Å if Arg52 was neutral. We found this behavior for both the Yamaguchi and the X-ray structure [pdb id 2QJ5] by Coureux et al.<sup>21</sup> The expectation values for the protonated Arg52 all show a smaller increase between 1.04 and 1.07 Å in bond lengths. In the 1.2 Å resolution X-ray structure of Coureux et al.,<sup>21</sup> the long delocalization shoulder into nuclear excited states for the deprotonated Arg52 is not present. This is surprising, as the Coureux et al.<sup>21</sup> structure also has a very short O–O distance, which too should create conditions in favor of a LBHB.

In order to distinguish between increased Glu46-O–D expectation values due to shifts in the potential minimum versus occupation of nuclear excited states, we calculated the occupation of the nuclear excited states [Table 1]. We found

**Table 1. Thermal Occupation Probability (%) of Nuclear Eigenstates for the Protonated and Deprotonated Arg52 State for Different Functionals and Yamaguchi<sup>3</sup> and Coureux<sup>44</sup> Structures in Vacuum<sup>a</sup>**

	Yamaguchi [2ZOI]			Coureux [2QJ5]	
	MPW1B95	B3LYP	$\omega$ B97XD	CAM-B3LYP	CAM-B3LYP
Arg52 protonated					
$S_0$	97.1	97.2	97.0	97.0	96.8
$S_1$	2.1	2.0	2.1	2.0	2.4
$S_2$	0.6	0.5	0.6	0.6	0.6
$S_3$	0.1	0.2	0.2	0.4	0.1
Arg52 deprotonated					
$S_0$	93.0	91.4	89.3	88.3	97.1
$S_1$	4.4	6.1	8.2	9.3	2.0
$S_2$	1.7	1.6	1.6	1.4	0.6
$S_3$	0.5	0.5	0.5	0.5	0.3

<sup>a</sup>Figure 5.

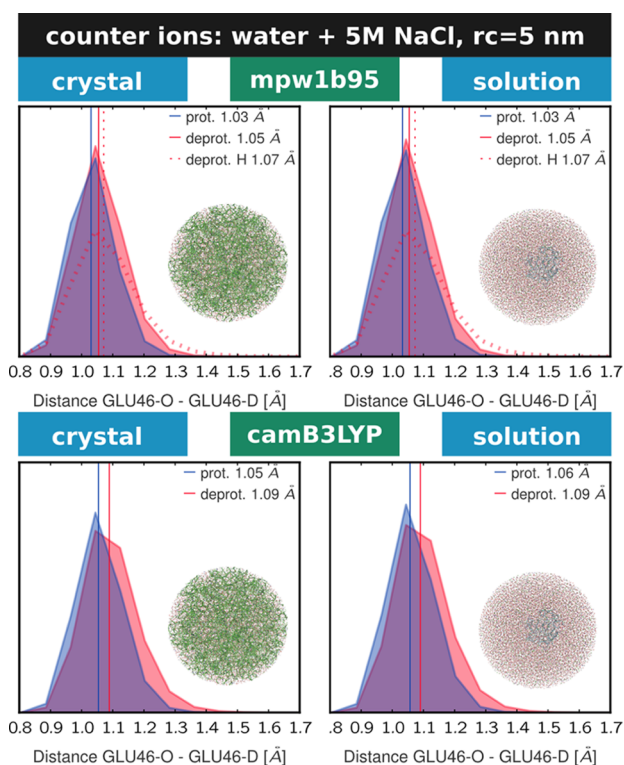
significant occupation of nuclear excited states of over 10% for the Yamaguchi structure when Arg52 was deprotonated but not in the reference structure by Coureux et al.<sup>44</sup> at either Arg52 protonation level.

Combining the expectation values of the equilibrium densities with the occupation values of the excited states, we find that the delocalization of the Glu46-D deuterium due to excited state occupation, is only present for the 2ZOI structure in vacuum with a deprotonated Arg52. No such excited state delocalization was observed in either Arg52 protonation state for the 2QJ5 structure. Despite these differences, all QM/MM setups consistently show a small increase in the Glu46O–D distance independently of the Arg52 protonation state.

#### Glu46-D Delocalization in PYP: Crystal and Solution.

We investigated the delocalization in the charge neutral solvated crystal and for a single PYP in solution. From these solvated systems, the distance expectation values and delocalization extent were calculated.

The calculated Glu46-D distance expectation values [Figure 6] and occupation probabilities [Table 2] are shown for both MPW1B95 and CAM-B3LYP density functionals. In contrast to the vacuum calculations, neither MPW1B95 nor CAM-B3LYP show Glu46-D delocalization to the extent previously observed in the vacuum calculations on the same 2ZOI crystal structure. A shift in the potential energy minimum of the CAM-B3LYP



**Figure 6.** Comparison of the crystal and solvent environment for Glu46-D delocalization. The extent of Glu46-D delocalization was calculated in the crystal (left) and solution (right) phase for the MPW1B95 (top) and CAM-B3LYP (bottom) functionals. The protonated Arg52 state is shown in blue, and the deprotonated state in red. The dotted lines show the result for hydrogen instead of deuterium.

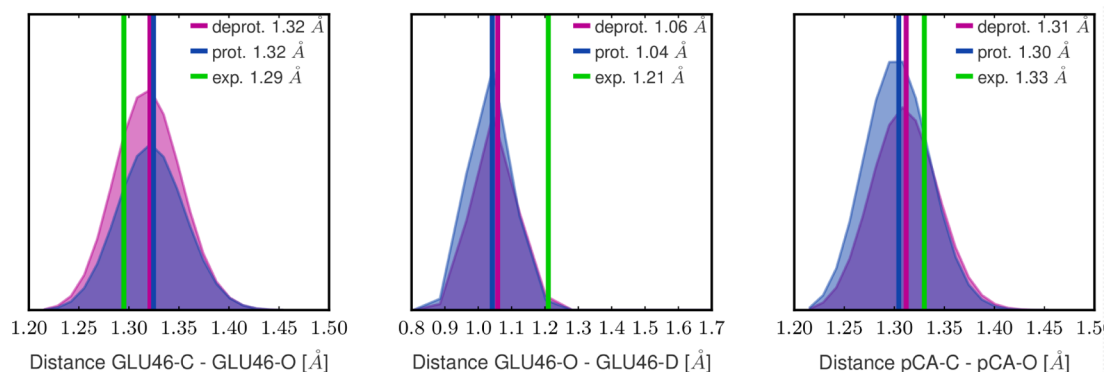
**Table 2. Thermal Occupation Probability of Nuclear Eigenstates for the Protonated and Deprotonated Arg52 State for the Charge Neutral Solvated PYP Crystal and PYP in Ionic Solution<sup>a</sup>**

	MPW1B95		CAM-B3LYP	
	crystal	solution	crystal	solution
Arg52 protonated				
$S_0$	96.8	96.9	96.9	97.0
$S_1$	2.4	2.3	2.3	2.2
$S_2$	0.6	0.6	0.7	0.6
$S_3$	0.1	0.1	0.1	0.2
Arg52 deprotonated				
$S_0$	97.1	96.9	96.2	95.9
$S_1$	1.9	2.0	1.7	1.8
$S_2$	0.5	0.6	1.4	1.6
$S_3$	0.4	0.4	0.5	0.5

<sup>a</sup>Figure 6.

calculations increases the Glu46O–D distance expectation value from 1.05 to 1.09 Å when Arg52 is deprotonated.

In the above nuclear equilibrium density calculations, we restrained all heavy atoms to their X-ray positions. Next, we investigated the effect of relaxing the coordinates on the QM/MM potential energy surface. We therefore used the optimized QM region and the same crystal MM environment from the chemical shift calculation to compute the PES. In addition to the deuterium position, we also included the positions of the donor and acceptor oxygen atoms, Glu46-O and pCA-O,



**Figure 7.** Fully coupled  $3 \times 1D$  nuclear O–H–O equilibrium densities for the QM/MM optimized coordinates (solvated crystal model). (left) Nuclear equilibrium density along the Glu46-C–O distance for the protonated (red) and deprotonated (blue) Arg52. The orthogonal degrees of freedom were integrated out. (center) Same as before but for the Glu46-D distance with respect to the Glu46-C–O equilibrium position. (right) Same as before but for the chromophore oxygen pCA–C–pCA–O. (green lines) Distances proposed by Yamaguchi et al.

respectively, into the nuclear equilibrium density calculations. The results from the three atom densities [Figure 7] agree well with the calculations performed on the crystal coordinates.

While in these calculations, the QM region was relaxed to the DFT potential energy surface, no optimization of the MM environment was performed. Therefore, the same MM environment is sampled throughout all calculations. We investigated the snapshot dependence of different solvent configurations in the crystal model described in the SI but could not find significant differences (data not shown).

The Glu46-D deuterium is located deep inside the chromophore pocket, shielded from direct contacts with the outside environment, a feature that we assume is relevant for protein function. The very similar bond length expectation values in the crystal and solution environment [Table 3] suggest that long-range electrostatics effects are not directed but averaged out, in contrast to the charged vacuum conditions. The solvent and crystal environments cannot be distinguished by the Glu46-D bond length alone. A significant difference

**Table 3.** Glu46-O–D Bond Length Expectation Values for the Investigated Environments and DFT Functionals from Nuclear Equilibrium Density Calculations<sup>a</sup>

structure model	DFT functional	Arg52 prot. bond length [Å]	Arg52 deprot. bond length [Å]
2ZOI vacuum	MPWB95	1.04	1.08
	B3LYP	1.04	1.09
	wB97XD	1.06	1.12
	camB3LYP	1.07	1.15
2QJ5 vacuum	camB3LYP	1.05	1.08
crystal model	camB3LYP	1.05	1.09
	MPWB95	1.03	1.05
solution model	camB3LYP	1.06	1.09
	MPWB95	1.03	1.05
crystal QM/MM opt	camB3LYP	1.04	1.06

<sup>a</sup>The calculated bond lengths are the result of solving the nuclear Schrödinger equation for Glu46-D, while all other coordinates were kept frozen. The O–O distance was 2.56 Å in all calculations. Only in the Glu46-D nuclear equilibrium densities for the QM/MM optimized crystal model were the two neighboring oxygens also considered as quantum mechanical particles. All possible bond lengths were considered within the grid boundary conditions described in the Methods section.

exists only in the vacuum environment. At the same time, the decrease of the nuclear ground state population by 1% also points against a LBHB with thermal quantum delocalization as this value is too low to cause any large shifts in the Glu46-D bond length expectation value. The calculations without bond relaxation can directly be compared to the previous calculations of the protein in vacuum.<sup>15</sup>

**Glu46-H Chemical Shifts: Vacuum, Crystal, and Solution.** The chromophore pocket of the Yamaguchi et al.<sup>3</sup> PYP structure (2ZOI), with Arg52 protonated, was optimized for the three electrostatic environments, solution, crystal, and vacuum. We found a very close agreement between the crystal and solution environments for the O–O (Glu46-pCA) and O–H (Glu46-H) distances of  $d_{\text{OO}}^{\text{sol}} = 2.56$  Å,  $d_{\text{OO}}^{\text{cry}} = 2.56$  Å, and  $d_{\text{OH}}^{\text{sol}} = 1.02$  Å,  $d_{\text{OH}}^{\text{cry}} = 1.02$  Å [Table 4]. The optimized distances

**Table 4.** Glu46-H Chemical Shifts,  $\delta_{\text{H}}$

environment	Glu46-O–H [Å]	Glu46-O–pCA-O [Å]	$\delta_{\text{H}}$ [ppm]
solution	1.02	2.56	15.8
crystal	1.02	2.56	16.0
vacuum	1.05	2.48	18.5
exp solution <sup>11</sup> (NMR)			15.2
exp crystal <sup>3</sup> (neutron)	1.06–1.22 <sup>a</sup>	2.56	

<sup>a</sup>Range (this work) from re-refinement, Figure 10

were identical up to the presented digits. This is in excellent agreement with the experimental crystal<sup>3</sup> distance of  $d_{\text{OO}}^{\text{sol,cry}} = 2.56$  Å. In contrast, the vacuum environment caused a decrease in the O–O distance to  $d_{\text{OO}}^{\text{vac}} = 2.48$  Å and an increase in the O–H distance to  $d_{\text{OH}}^{\text{vac}} = 1.05$  Å.

Solution and crystal geometries also resulted in similar chemical shifts  $\delta_{\text{H}}^{\text{cry}} = 16.0$  ppm and  $\delta_{\text{H}}^{\text{sol}} = 15.8$  ppm in reasonable agreement with the experimental value of  $\delta_{\text{H}}^{\text{sol}} = 15.2$  ppm.<sup>11</sup> In the vacuum environment, the increased O–H distance and decreased O–O distance increased the chemical shifts to  $\delta_{\text{H}}^{\text{vac}} = 18.5$  ppm.

Our results for the solution structure are comparable to the values of Nadal-Ferret<sup>15</sup> with  $\delta_{\text{H}}^{\text{sol}} = 13.7$ –16.4 ppm and Saito et al.<sup>7</sup> with  $\delta_{\text{H}}^{\text{sol}} = 14.5$ –14.6 ppm. Although it is not clear if continuum electrostatics were used in the Saito et al.<sup>7</sup> calculations, we assumed that here. Small differences are further expected due to the use of a different reference molecule in these two computational studies, while TMSP- $d_4$

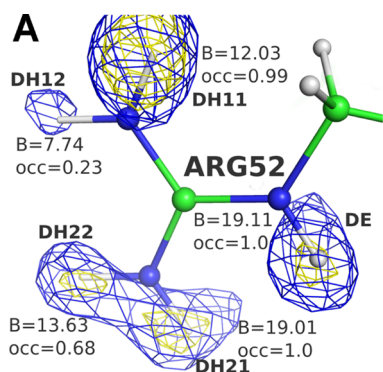


was used in this work and the experiment.<sup>11</sup> Also a different DFT functional, CAM-B3LYP rather than B3LYP, was used here. No agreement between the Glu46-O–H distance in the optimized QM/MM geometry and the distance reported in the 2ZOI neutron structure was found, but we emphasize that the Glu46-O–H distances are minimum positions from the PES and not expectation values from nuclear equilibrium densities.

#### Reexamination of the Yamaguchi Neutron Density.

The results of the  $pK_a$  calculations suggest that Arg52 is protonated in the crystal. Under these conditions, we found no direct evidence of a LBHB but instead a slight increase in the Glu46-O–D bond lengths to 1.05–1.06 Å (CAM-B3LYP). Furthermore, the deprotonation of Arg52 had no major effects on the Glu46-O–D distance in the crystal environment. In this section, we reexamine to what extent the neutron diffraction data by Yamaguchi et al.<sup>3</sup> is also compatible with these findings.

The Arg52 residue was originally suggested to be deprotonated based on a reduced nuclear density at the deuterium position Arg52-DH12. An occupation of 0.23 was assigned in the structure refinement, but no density was visible in Figure 1E of the original publication.<sup>3</sup> Here, we created an  $F_O - F_C$  difference map of residue Arg52 omitting the four deuterium atoms DH11, DH12, DH21, and DH22 and also DE [Figure 8], contour levels are shown at  $2.3\sigma$  and  $3.5\sigma$ . The map



**Figure 8.** Arg52  $F_O - F_C$  deuterium omit maps. (A) An  $F_O - F_C$  difference map is shown of residue Arg52 in which deuteriums DH11, DH12, DH21, DH22, and DE were omitted [pdb code 2ZOI]. The contour levels are shown at  $2.3\sigma$  (blue lines) and  $3.5\sigma$  (yellow). Next to the deuterium positions, the B values and occupation numbers from the original Yamaguchi refinement (pdb code 2ZOI) are shown.

is based on the published structure factors and 2ZOI coordinates following the original refinement procedure<sup>3</sup> in nCNS 1.1.<sup>45–47</sup> We found good agreement with the Yamaguchi data at  $3.5\sigma$  but could also resolve the Arg52-DH12 nuclear density at a lower  $2.3\sigma$  level. At this lower  $\sigma$ , the presence of a small density volume at the DH12 site is visible. The DH22 position was found to also exhibit a reduced nuclear density as reflected by the low occupation value of 0.68 in the 2ZOI structure. The four Arg52 sites, DH11, DH12, DH21, and DH22, exhibit very different shapes and intensities of the neutron diffraction densities and two out of four sites are not fully occupied.

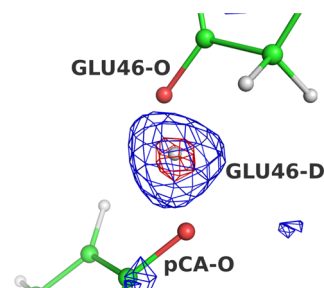
The reduced densities of the DH12 and DH22 deuterons is surprising and cannot, as in the case of a second solvent exposed arginine in the sequence, Arg124, be caused by a high rotational flexibility of the C–N bonds. Although we cannot rule out that the reduced density reflects a superposition of protonated and deprotonated states of the guanidinium moiety,

we consider this unlikely, given the large differences between the protonated and deprotonated conformations of the Arg52 side chain in our simulations. Notwithstanding the cause, the observation that in fact all four densities differ markedly from each other points to considerable uncertainties in these densities, including DH12 and DH22. A well resolved deprotonated Arg52 should have three out of four sites fully occupied and exhibit similar density shapes and intensities.

We note that in the previous refinement by Yamaguchi et al.,<sup>3</sup> neutron scattering data at lower resolution (15283 reflections) were combined with higher resolution (28967 reflections) X-ray data. As a result, the accuracy of the Glu46-D deuterium position and density in question, which are not seen in the X-ray data, are not enhanced by the X-ray data, and, hence, remain low.

The experimental Arg52 density maps do not rule out a protonated Arg52-DH12 and fully occupied Arg52-DH22 masked by noise. This noise can have multiple sources. Incomplete deuteration is one mechanism that contributes to missing densities and large background levels. If hydrogen atoms were present in a small fraction of the crystal cells at the Arg52 position, the incoherent scattering of hydrogen can reduce the observed deuterium density at that position below background levels.<sup>48,49</sup>

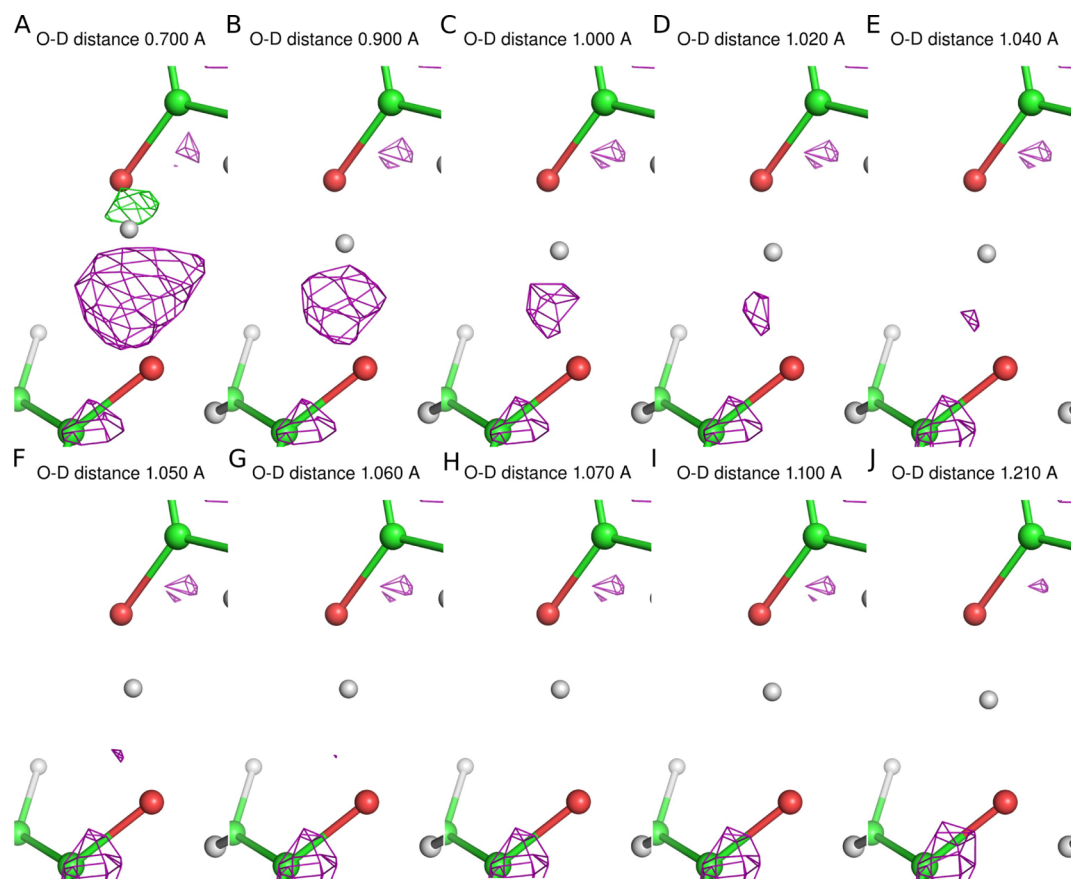
In their original work, Yamaguchi et al.<sup>3</sup> presented an  $F_O - F_C$  map omitting the Glu46-D deuterium. We also generated  $F_O - F_C$  omit maps with the measured structure factors



**Figure 9.**  $F_O - F_C$  omit map of Glu46-D. An  $F_O - F_C$  map omitting the Glu46-D deuterium is shown at the position refined by Yamaguchi et al.<sup>3</sup> [pdb code 2ZOI]. The contour levels are at  $3\sigma$  (blue) and  $6\sigma$  (red).

at  $3\sigma$  and  $6\sigma$  contour levels that are in good agreement with the maps from the original publication.<sup>3</sup> While the Glu46-D omit map may seem to position the atom with high precision, the relative accuracy with respect to alternative positions cannot be determined from this type of map. Further, because phases of the complete structure are used, the  $F_O$  map still contains contributions from the omitted atom, which could lead to uncertainty in the difference density maps. Therefore, the true accuracy of the position is not as high as a  $6\sigma$  contour level might suggest.

To also compare alternative Glu46-D positions, we calculated multiple  $F_O - F_C$  difference maps [Figure 10] in which the Glu46-D deuterium position was changed along the axis connecting the chromophore oxygen pCA-O and the Glu46-O. We found large differences in density for very short Glu46O–D distances  $r_{OD} < 1$  Å [Figure 10A,B], small differences in the range  $r_{OD} = 1.00$ – $1.05$  Å [Figure 10C,D,E,F], and close to identical maps for distances between  $r_{OD} = 1.06$  and  $1.21$  Å [Figure 10G,H,I,J].



**Figure 10.**  $F_O - F_C$  difference maps of the 2ZOI structure. An  $F_O - F_C$  map scan is shown as a function of the Glu46-D deuterium position within the bond at  $3\sigma$  (purple) and  $-3\sigma$  (green) contour levels.

From the calculated  $F_O - F_C$  difference maps, we estimate that a lower bound for the Glu46-O-D distance that is still in agreement with the experimental data, is larger than 1.06 Å. This value is approximately 0.15 Å shorter than the previously reported<sup>3</sup> 1.21 Å but also confirms that the Glu46-pCA hydrogen bond is not “normal” as it, too, does have an increased bond length. Based on our analysis (Figure 10), the experimental structure factors appear not to contradict our findings of a protonated Arg52 with the calculated Glu46-D hydrogen bond length of 1.05 Å (CAM-B3LYP) in the crystal. Instead, the Yamaguchi model and our results both lie within the error margins of the experimental data.

To summarize, both the protonated Arg52 and the increased Glu46-D bond length (1.05 Å) (i) do not conflict with the neutron structure factors, (ii) are in line with an experimental pH of 9, (iii) preserve the Arg52 hydrogen bond network with its crystal waters in free simulations, and (iv) do not require the involvement of nuclear excited states at room temperature.

## CONCLUSIONS

We have carried out several calculations to investigate whether a low barrier hydrogen bond (LBHB) exists in PYP, both in the crystal and in solution. The results of these calculations suggest an alternative for the Yamaguchi model. Table 5 summarizes our results, together with all available experimental data in light of the two models discussed. As can be seen, the presented alternative interpretation in which Arg52 is protonated and no Glu46-D LBHB exists is consistent with all available

**Table 5. Current and Previous Results for the Arg52 Protonation State and the Existence of a Glu46-D LBHB in PYP**

	This work ARG52 prot.	Yamaguchi LBHB Model <sup>3</sup> ARG52 deprot.
neutron diffraction 295 K	✓ <sup>a</sup>	✓ <sup>3</sup>
solution NMR	✓ <sup>12,13</sup>	✗
PYP in solution (DFT)	✓ <sup>8</sup>	✗
PYP in vacuum (DFT)	✗	✓ <sup>15,16,c</sup>
PYP in crystal (MD)	✓ <sup>f,e</sup>	(✗) <sup>d</sup>
no LBHB		LBHB
neutron diffraction 295 K	✓ <sup>a</sup>	✓ <sup>3</sup>
solution NMR	✓ <sup>11</sup>	✗
PYP in solution (DFT)	✓ <sup>15,b</sup>	✗
PYP in vacuum (DFT)	✗	✓ <sup>15,16,c</sup>
PYP in crystal (DFT)	✓ <sup>b</sup>	✗

<sup>a</sup>This work, both models lie within experimental error of the reinvestigated neutron diffraction data. <sup>b</sup>This work, no LBHB irrespective of the Arg52 protonation state in nuclear equilibrium density calculations. <sup>c</sup>This work. <sup>d</sup>This work, Arg52 pK<sub>a</sub> upshift of 12–36 kJ/mol between crystal and solution. <sup>e</sup>This work, deprot. Arg52 unstable in free MD simulations using three independent parametrizations. <sup>f</sup>This work, prot. Arg52 yields stable Tyr98 and Thr50 hydrogen bonds and recovers crystal water with ID 1023 (2ZOI).

experimental and theoretical results, except previous computations based on a single protein in vacuum.

Specifically, we re-examined  $F_O - F_C$  difference and omit maps generated from the original Yamaguchi et al. data.<sup>3</sup> We found that in addition to the previous model with a large O-D distance of  $r_{O-D} = 1.21$  Å, which is characteristic of a LBHB, also a more regular hydrogen bond distance of  $r_{O-D} > 1.06$  Å is

in good agreement with the neutron density and with the  $r_{\text{O-D}} = 1.05 \text{ \AA}$  distance obtained in our CAM-B3LYP nuclear equilibrium density calculations. We conclude that the large differences in experimental neutron densities at the four Arg52 deuterium positions<sup>3</sup> do not exclude a protonated Arg52 nor do they rule out a partially deprotonated Arg52. A possible explanation for the missing density other than partial occupation is incoherent hydrogen scattering.<sup>48</sup> Therefore, additional evidence is required.

To resolve this controversy and distinguish between the two models, we have performed a series of simulations. MD simulations of PYP with a protonated Arg52 under crystal conditions resulted in Arg52 conformations in agreement with experiments. We observed stable hydrogen bonds connecting Arg52 to Thr50 and Tyr98 at the experimentally observed 3 Å distance. Further, the crystal water near Arg52 was occupied in our 300 K MD simulations in agreement with the 298 K Yamaguchi et al.<sup>3</sup> crystal structure (HOH 1023 in 2ZOI). In contrast, all simulations with a deprotonated Arg52 yielded conformations that markedly deviate from the experimental X-ray and neutron structures. Our MD simulations thus only support a protonated Arg52 both in the crystal and in solution. Notably, for the solvated protein, recent NMR data<sup>12,13</sup> has shown Arg52 to be protonated.

Given the remaining uncertainties of whether the force field describes the deprotonated Arg52 accurately enough, we carried out  $\text{pK}_a$  calculations to directly address the protonation state. Only a  $\text{pK}_a$  down shift for the transfer of the Arg52 proton from crystal to solution is compatible with a neutral Arg52 in the crystal. However, we calculated a  $\text{pK}_a$  upshift from our free energy calculations, in which we also considered the contribution of the LBHB and the associated charge distribution in the chromophore pocket [SI]. Although we would not consider these calculations as definite proof, given the inaccuracies inherent to molecular mechanics force fields based on point charges, the combined experimental and simulation results very strongly support a protonated Arg52 and are at variance with a deprotonated Arg52.

We also investigated the hypothesis<sup>15,16</sup> that deprotonation of Arg52 might favor the LBHB, as previously suggested based on calculations of a single PYP protein copy in vacuum. We calculated the quantum mechanical nuclear equilibrium densities of the Glu46 deuterium at various levels of theory. Strikingly, the occupation of nuclear excited states, indicative of a LBHB, was only observed in the vacuum calculations. For crystal and solution environments, however, no LBHB was observed in our calculations. These findings suggest that the observed slight increase of O–D bond length from the typical value of 1.0 Å to about 1.06 Å can be explained by a shift in the potential minimum without involvement of nuclear quantum effects. This is a particularly striking example of the pronounced impact the environment around the protein may have on the properties of a hydrogen bond and also underscores the high sensitivity of this particular H-bond to long-range electrostatics, in contrast to common hydrogen bonds whose structure seems to be much less sensitive to their environment.<sup>50</sup> We attribute the largest effect to the presence of a largely isotropic environment around the protein (crystal, solution) rather than the anisotropic protein environment of an isolated PYP in vacuum. We conclude that the previously reported vacuum calculations do not necessarily support a LBHB in PYP.

We have also shown that the investigated properties of PYP, in particular the potential energy surface, the nuclear density,

and the chemical shift of Glu46-D, are largely the same for PYP in solution and in the crystal, but differ for PYP in vacuum. Moreover, the differences observed by Nadal-Ferret et al.<sup>15</sup> between the crystal, which was approximated by a single protein in vacuum, and solution conditions are no longer observable if a more realistic crystal model is included into the calculations. Based on these results, we conclude that the crystal and solution environment cannot be distinguished by the Glu46-D position alone. Taken together, these results also suggest that solution NMR results in favor of a protonated Arg52<sup>12,13</sup> and the absence of a large chemical shift<sup>11</sup> that would indicate a LBHB, are transferrable to the crystal. ssNMR experiments will be needed to confirm this prediction.

On more general grounds, our results reveal an unexpectedly strong sensitivity of LBHB on the electrostatic environment especially in QM/MM calculations of proteins. In particular, simple approximations of the protein crystal environment by vacuum may not, as frequently assumed, yield a sufficiently accurate model to investigate LBHB candidates also in other systems.<sup>51</sup> Finally, our results underscore the need to accurately model the chemical environment when calculating properties that are highly sensitive to the electrostatic interaction, such as LBHBs.

## ■ ASSOCIATED CONTENT

### 📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b05609.

Derivation PYP thermodynamic cycle, evaluation of the effect of the LBHB on the Arg52  $\text{pK}_a$  (crystal), Glu46-D delocalization under varying levels of theory, evaluation of the effect of LBHB on  $\text{pK}_a$  of Arg52 (second and third parametrization), Glu46-D delocalization in the PYP in vacuum, force field charges in neutral arginine, optimized geometry at CAM-B3LYP/6-31+G\*\* for solvated, crystal, and vacuum PYP environments, additional scans of the vacuum 2ZOI structure, Arg52–water distance at crystal water ID 1023 (2ZOI) site, and  $F_{\text{O}} - F_{\text{C}}$  maps at  $2.3\sigma$  and at  $6\sigma$  instead of  $3\sigma$  (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

\*hgrubmu@gwdg.de

\*gerrit.x.groenhof@jyu.fi

### ORCID

Timo Graen: 0000-0002-6552-8477

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

GG is supported by the Academy of Finland grant 292820. TG was supported by the MPG (International Max Planck Research School - Physics of Biological and Complex Systems). LI was supported by the DFG SFB755. MC was supported by the Volkswagen Foundation grant 83940. We thank Dr. Heikki Takala for valuable discussions. We thank Dr. Mehdi Davari for his contributions at the start of the project.

## ■ REFERENCES

(1) Sprenger, W. W.; Hoff, W. D.; Armitage, J. P.; Hellingwerf, K. J. *Bacteriol.* **1993**, *175*, 3096–3104.

- (2) Anderson, S.; Crosson, S.; Moffat, K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 1008–1016.
- (3) Yamaguchi, S.; Kamikubo, H.; Kurihara, K.; Kuroki, R.; Niimura, N.; Shimizu, N.; Yamazaki, Y.; Kataoka, M. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 440–444.
- (4) Schotte, F.; Cho, H. S.; Kaila, V. R.; Kamikubo, H.; Dashdorj, N.; Henry, E. R.; Graber, T. J.; Henning, R.; Wulff, M.; Hummer, G.; et al. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 19256–19261.
- (5) Ko, C.; Virshup, A. M.; Martinez, T. J. *Chem. Phys. Lett.* **2008**, *460*, 272–277.
- (6) Groenhof, G.; Schäfer, L. V.; Boggio-Pasqua, M.; Grubmüller, H.; Robb, M. A. *J. Am. Chem. Soc.* **2008**, *130*, 3250–3251.
- (7) Saito, K.; Ishikita, H. *Biochemistry* **2012**, *51*, 1171–1177. PMID: 22263543.
- (8) Saito, K.; Ishikita, H. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 167–172.
- (9) Saito, K.; Ishikita, H. *Biochim. Biophys. Acta, Bioenerg.* **2013**, *1827*, 387–394.
- (10) Ishikita, H.; Saito, K. *J. R. Soc., Interface* **2014**, *11*, 20130518.
- (11) Sigala, P. A.; Tsuchida, M. A.; Herschlag, D. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 9232–9237.
- (12) Oktaviani, N. NMR studies of folded and unfolded proteins: method developments and biological insight. Ph.D. thesis, 2014.
- (13) Yoshimura, Y.; Oktaviani, N. A.; Yonezawa, K.; Kamikubo, H.; Mulder, F. A. A. *Angew. Chem., Int. Ed.* **2016**, DOI: 10.1002/anie.201609605.
- (14) Harms, M. J.; Schlessman, J. L.; Sue, G. R.; García-Moreno E., B. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 18954–18959.
- (15) Nadal-Ferret, M.; Gelabert, R.; Moreno, M.; Lluch, J. M. *J. Am. Chem. Soc.* **2014**, *136*, 3542–3552.
- (16) Kanematsu, Y.; Tachikawa, M. *J. Chem. Phys.* **2014**, *141*, 185101.
- (17) Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51–57.
- (18) Kanematsu, Y.; Kamikubo, H.; Kataoka, M.; Tachikawa, M. *Comput. Struct. Biotechnol. J.* **2016**, *14*, 16–19.
- (19) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (20) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; et al. *Bioinformatics* **2013**, *29*, 845–854.
- (21) Coureux, P.-D.; Fan, Z. P.; Stojanoff, V.; Genick, U. K. *Structure* **2008**, *16*, 863–872.
- (22) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; et al. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
- (23) Sorin, E. J.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2472–2493.
- (24) Pande, K.; Hutchison, C. D.; Groenhof, G.; Aquila, A.; Robinson, J. S.; Tenboer, J.; Basu, S.; Boutet, S.; DePonte, D. P.; Liang, M.; et al. *Science* **2016**, *352*, 725–729.
- (25) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (26) Cornell, W. D.; Cieplak, P.; Bayly, C. L.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620–9631.
- (27) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision D.01; Gaussian, Inc.: Wallingford, CT, 2009.
- (28) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
- (29) Groenhof, G. *Methods Mol. Biol.* **2013**, *924*, 43–66.
- (30) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- (31) Joung, I. S.; Cheatham, T. E., III. *J. Phys. Chem. B* **2008**, *112*, 9020–9041.
- (32) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (33) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.
- (34) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952–962.
- (35) Fitch, C. A.; Platzer, G.; Okon, M.; Garcia-Moreno E., B.; McIntosh, L. P. *Protein Sci.* **2015**, *24*, 752–761.
- (36) Hub, J. S.; de Groot, B. L.; Grubmüller, H.; Groenhof, G. *J. Chem. Theory Comput.* **2014**, *10*, 381–390.
- (37) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6908–6918.
- (38) Graen, T.; Grubmüller, H. *Comput. Phys. Commun.* **2016**, *198*, 169–178.
- (39) Colbert, D. T.; Miller, W. H. *J. Chem. Phys.* **1992**, *96*, 1982–1991.
- (40) Wang, L.; Fried, S. D.; Boxer, S. G.; Markland, T. E. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 18454–18459.
- (41) Wolinski, K.; Hinton, J. F.; Pulay, P. *J. Am. Chem. Soc.* **1990**, *112*, 8251–8260.
- (42) Scalmani, G.; Frisch, M. J. *J. Chem. Phys.* **2010**, *132*, 114110.
- (43) Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.
- (44) Corry, B.; Jayatilaka, D. *Biophys. J.* **2008**, *95*, 2711–2721.
- (45) Afonine, P. V.; Mustyakimov, M.; Grosse-Kunstleve, R. W.; Moriarty, N. W.; Langan, P.; Adams, P. D. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 1153–1163.
- (46) Yoshimura, Y.; Oktaviani, N. A.; Yonezawa, K.; Kamikubo, H.; Mulder, F. A. A. *Angew. Chem., Int. Ed.* **2016**, DOI: 10.1002/anie.201609605.
- (47) Brünger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J.-S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; et al. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998**, *54*, 905–921.
- (48) Shu, F.; Ramakrishnan, V.; Schoenborn, B. P. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 3872–3877.
- (49) Kurihara, K.; Tanaka, I.; Chatake, T.; Adams, M. W.; Jenney, F. E.; Moiseeva, N.; Bau, R.; Niimura, N. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 11215–11220.
- (50) Sigala, P. A.; Ruben, E. A.; Liu, C. W.; Piccoli, P. M. B.; Hohenstein, E. G.; Martinez, T. J.; Schultz, A. J.; Herschlag, D. *J. Am. Chem. Soc.* **2015**, *137*, 5730–5740.
- (51) Oltrogge, L. M.; Boxer, S. G. *ACS Cent. Sci.* **2015**, *1*, 148–156.