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

Timo Graen,[†] Ludger Inhester,[‡] Maike Clemens,[†] Helmut Grubmüller,[*,†] and Gerrit Groenhof[*,¶]

[†]*Max Planck Institute for Biophysical Chemistry, Am Faßberg 11, 37077 Göttingen, Germany*

[‡]*Center for Free-Electron Laser Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany*

[¶]*Department of Chemistry and Nanoscience Center, University of Jyväskylä, Finland*

E-mail: hgrubmu@gwdg.de; gerrit.x.groenhof@jyu.fi

Phone: +358 50 3956 020. Fax: +358 14 617 412

Derivation PYP Thermodynamic Cycle

Here, we derive the free energy difference for the protonation of ARG52 including contributions from nuclear delocalization due to quantum and thermal effects on the GLU46 deuterium. This allows us to estimate the effect of a possible correlation between the ARG52 protonation and the formation of a LBHB. We start from a classical thermodynamic integration of the ARG52 protonation using a force field and Newton's equations of motion

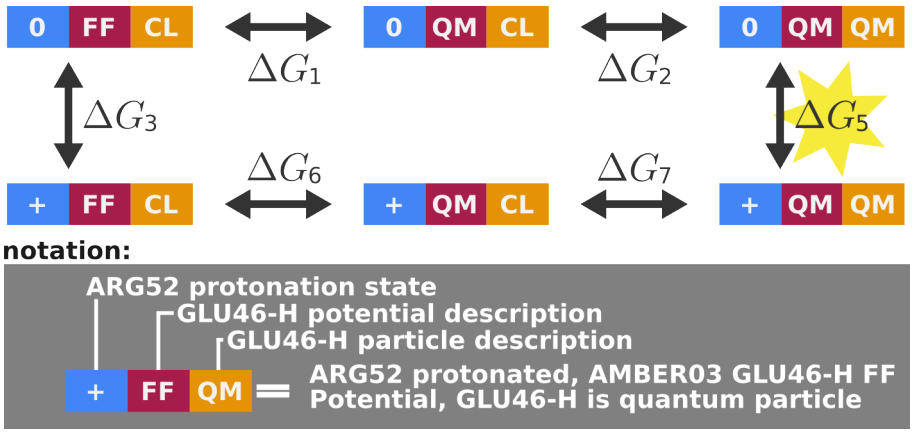

Figure S1: **PYP Free Energy Cycle** The free energy cycle is shown which was used to determine the free engergy difference estimate ΔG_5 between the protonated (+) and deprotonated (0) ARG52 residue including the contributions from the quantum delocalization of the GLU46-D hydrogen. The terms from left to right include the classical AMBER force field potential of GLU46-D, which was then replaced by a QM/MM potential and in the last step described as a quantum mechanical particle instead of a classical one.

for all particles. From this free energy, we first subtract the contribution of the classical GLU46-D atom by integration of the single particle free energy. Next, we added the free energy contribution of moving the particle from the force field to a DFT QM/MM potential description. At this step, GLU46-D contributes as classical particle in a DFT potential. The last remaining step in the cycle removes the free energy contribution of the classical particle and replaces it with a nuclear wave function in a DFT electronic potential. When comparing to the experimental neutron diffraction results, the hydrogen atom was replaced by a deuterium. In the following, the equations used in the cycle are given.

The total free energy difference

$$\Delta G_5 = \Delta G_3 + \Delta G_6 + \Delta G_7 - \Delta G_1 - \Delta G_2 \tag{1}$$

of protonating ARG52 from $0 \rightarrow +$ includes the terms depicted in FIG S1.

First, the force field [FF] potential was replaced by a scanned DFT potential energy surface which requires the following corrections to the the free energy difference for one

environment:

$$\Delta G_1 = G_B - G_A = [G_A - {}_{FF}G^0_{cl} + {}_{DFT}G^0_{cl}] - G_A \tag{2}$$

and

$$\Delta G_6 = G_E - G_D = [G_D - {}_{FF}G^+_{cl} + {}_{DFT}G^+_{cl}] - G_D. \tag{3}$$

with

$${}_{FF}G^{0/+}_{cl} \approx -kT \ln h^{-3} \int d\hat{x}d\hat{p} \exp\left[-\beta(T(\hat{p}) + {}_{FF}V^{0/+}(\hat{x}))\right] \tag{4}$$

and

$${}_{DFT}G^{0/+}_{cl} \approx -kT \ln h^{-3} \int d\hat{x}d\hat{p} \exp\left[-\beta(T(\hat{p}) + {}_{DFT}V^{0/+}(\hat{x}))\right]. \tag{5}$$

The potential ${}_{DFT}V^{0/+}(\hat{x})$ was scanned on a DFT grid around the center of the bond as a function of the ARG52 protonation state (0/+). The Planck constant h and the contribution from the kinetic energy $T(p)$ drop out, leaving only the contribution from the AMBER03 force field potential energy ${}_{FF}V^{0/+}(x)$ and DFT potential ${}_{DFT}V^{0/+}(x)$.

Next, the quantum contribution of describing the GLU46 deuterium as a nuclear wave function instead of a classical particle was included in the terms

$$\Delta G_2 = G_C - G_B = [G_B - {}_{DFT}G^0_{cl} + {}_{DFT}G^0_{qm}] - G_B \tag{6}$$

and

$$\Delta G_7 = G_F - G_E = [G_E - {}_{DFT}G^+_{cl} + {}_{DFT}G^+_{qm}] - G_E \tag{7}$$

with

$${}_{DFT}G^{0/+}_{qm} \approx -kT \ln \left(Tr(\exp\left[-\beta \hat{H}^{0/+}\right]) \right). \tag{8}$$

${}_{DFT}G^{0/+}_{qm}$ was evaluated by calculating the nuclear eigenstates of the discretized stationary Schrödinger Equation with Hamiltonian $\hat{H}^{0/+}$ in the potential ${}_{DFT}V^{0/+}(\hat{x})$. The trace over the eigenstate energies results in the free energy correction. h is the Planck constant, $T(\hat{p})$ is the kinetic energy, $\beta = \frac{1}{kT}$ is the inverse temperature with the Boltzmann factor k.

The last remaining term ΔG_3 was obtained by standard thermodynamic integration ΔG^{TI} using classical particles and a force field description of the potential. The quantum corrected free energy difference for the ARG52 protonation results in

$$\Delta G_5 = \Delta G^{TI} - \left[{}_{FF}G^0_{cl} - {}_{FF}G^+_{cl}\right] + \left[{}_{DFT}G^0_{qm} - {}_{DFT}G^+_{qm}\right] \tag{9}$$

which simplifies to

$$\Delta G_5 = \Delta G^{TI} - \Delta G^{FF}_{cl} + \Delta G^{DFT}_{qm}. \tag{10}$$

The free energy difference ΔG_5 was calculated for the crystal ΔG_5^{cryst} and solution environment ΔG_5^{solv}

$$\Delta\Delta G_5^{cryst-solv} = \Delta G_5^{cryst} - \Delta G_5^{solv} = \Delta\Delta G^{TI} - \Delta\Delta G_{cl}^{AMBER03} + \Delta\Delta G_{qm}^{DFT} \tag{11}$$

Evaluating the effect of the LBHB on the ARG52 pKa (crystal)

LBHB force field

The coordinates of the chromophore and the sidechains of GLU46, and CYS69 were taken from the neutron diffraction structure (2ZOI).[1] Hydrogen atoms were introduced between the C_α and C_β atoms to keep a closed shell electronic structure. The coordinates of light atoms, with the exception of the shared proton of the LBHB were optimized at the HF/6-31G** level of ab initio theory. After the geometry optimization, the electrostatic potential at 10 concentric layers of 17 points per unit area around each atom was evaluated using the electron density calculated at the B3LYP/cc-pVTZ level of theory,[2,3] using the IEFPCM continuum solvent model[4] with a relative dielectric of 4.0 to model the effect of the protein environment.[5] The atomic charges were obtained by performing a two stage RESP fit to the electrostatic potential.[6] With this approximate LBHB force field, the free energy for deprotonating the ARG52 in the crystal was computed.

MD simulations

The starting coordinates of the PYP crystal were taken from the x-ray structure (PDB entry: 2ZOH).[1] Six copies of the protein, including 120 crystal waters, were placed inside the unit cell with periodic boundaries and soaked in 5M NaCl solution by adding 1276 TIP3P waters,[7] 70 Na^+ and 34 Cl^- ions. The total system contained 17,663 atoms and was equilibrated for 40 ns. The equilibration MD simulations were run at constant volume and temperature of 300 K,[8] with a time constant of 0.1 ps for the temperature coupling. The LINCS algorithm was used to constrain bond lengths,[9] allowing a time step of 2 fs in the classical simulations. SETTLE was applied to constrain the internal degrees of freedom of the water molecules.[10] A 1.0 nm cut-off was used for non-bonded Van der Waals interactions, which were modeled by Lennard-Jones potentials. Coulomb interactions were computed with the smooth particle mesh Ewald method,[11] using a 1.0 nm real space cut-off and a grid spacing of 0.12 nm. The

relative tolerance at the real space cut-off was set to 10^{-5}. All force field simulations were performed in single precision with the Gromacs-4.5.3 molecular dynamics program.[12]

The change in free energy upon removing the proton from the ARG52 guanidium group was computed by thermodynamic integration. To avoid artifacts associated with a non-neutral simulation box,[13] we simultaneously changed the charge of 29 randomly selected sodium ions from +1.0 to +1.03448 e. Such small increase of charge did not visibly affect the sodium-water radial distribution functions.

Classical molecular dynamics trajectories of 1.0 nanoseconds each were generated at 21 equidistant points along the λ-interval, and the ensemble average $\langle \partial H/\partial \lambda \rangle_\lambda$ was computed, leaving out the first 200 ps. The ensembles were generated with a stochastic dynamics integrator running at 300 K with a friction coefficient of 0.5 ps^{-1}.

Two sets of thermodynamic integration simulations were preformed. First three simulations were carried out using the atomic charges and a GLU46-$H_\epsilon 2$ proton location corresponding to a normal hydrogen bond between the chromophore and GLU46. The second three simulations were performed using the atomic charges and proton position that correspond to the postulated LBHB configuration. The final free energy for deprotonating ARG52 with and without a LBHB between GLU46 and the chromophore was obtained by averaging over the three thermodynamic integration simulations.

Results and discussion

In the crystal, two effects need to be considered. First, the cost of delocalizing GLU46-D. Second, the free energy a delocalized GLU46-D does provide to stabilize the deprotonated ARG52 through re-polarization of the chromophore.

The first contribution can directly be calculated from the nuclear equilibrium densities [FIG6] of the manuscript and equation 5. In the crystal model, the delocalized GLU46 state is 6.1 kJ/mol higher in energy than the localized one.

To further assess the energy provided by the LBHB to stabilize ARG52, the free energy

of deprotonation in the presense and absense of a GLU46 LBHB was calculated as described above. Here, the free energy to deprotonate ARG52 (and simultaneously increase the charge of 29 Na^+ ions by 0.03448 e) in presence of a normal hydrogen bond between the chromophore and the GLU46 side chain was -1205 ± 2 kJmol^{-1}, while the free energy in the presence of the LBHB was -1213 ± 2 kJmol^{-1} . Thus, formation of the LBHB between GLU46 and the chromophore in the crystal stabilizes the deprotonated state of the Arg52 sidechain by 8 kJmol^{-1}, or about 1.4 pK_a units.

Therefore, in the crystal, it costs around 6 kJ/mol to delocalize the GLU46-D, by which -8 kJ/mol are gained on the ARG52 stabilization. This free energy does not suffice to downshift the ARG52 pKa from above 11.2[14] to the experimental pH of 9.[1]

GLU46-D delocalization under varying levels of theory

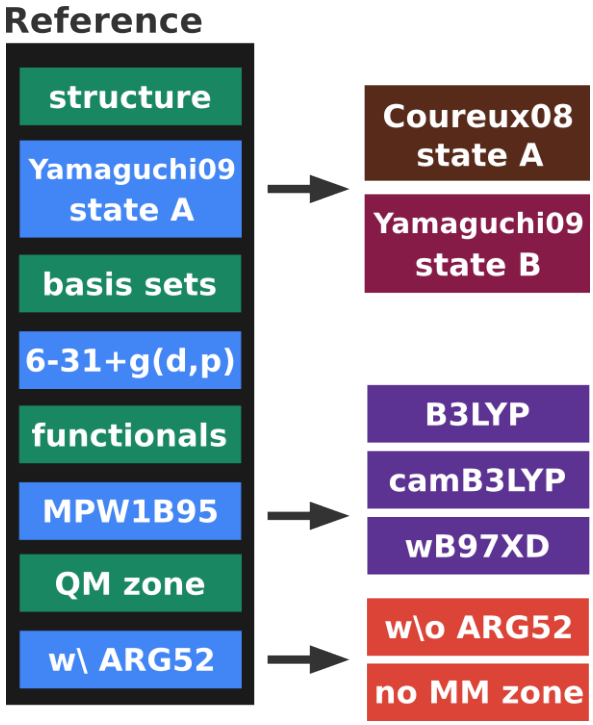

Figure S2: **Model Calculations for Varying Theoretical Setups** Parameters varied in the description GLU46-D of the vacuum crystal structure (2ZOI) for both the protonated and deprotonated states of ARG52. The parameters in the blue boxes on the left are the default values used in the QM/MM model calculations. Changes in this default setup are indicated by black arrows which were performed in separate calculations. Changes include the change of DFT functional and QM region size. Additionally, structural changes were made by performing calculations on the A and B state of the Yamaguchi structure as well as model calculations on the Coureux et al.[15] structure.

Previous studies[16,17] have demonstrated a delocalization of GLU46-D triggered by the deprotonation of ARG52 in agreement with the mechanism proposed by Yamaguchi et al.[1] We identified four main free parameters in these studies, (i) the choice of vacuum structure, (ii) the choice of DFT functional, (iii) the sizes of the QM and MM subsystems, and (iv) the choice of using the isolated protein in vacuum. First, based on the first three parameters, we investigated the robustness of the reported delocalization mechanism by modifying each parameter in individual simulation setups as depicted in Figure 2 for the vacuum model before including also the effect of the crystal and solution environment.

The QM region consisted of the pCA chromophore, its hydrogen bonding partners,

PRO68, GLU46, TYR42 and ARG52. This QM region was perturbed by removing ARG52 from the QM description and adding it to the classical force field description of the protein. The DFT functional was varied from the hybrid MPW1B95[18] developed for non covalent interactions to the less specific B3LYP,[2] the long range attenuated CAM-B3LYP[19] and the fully reparameterized ωB97XD.[20] On the structural level, we calculated the differences in hydrogen delocalization between the crystal A and B state within the Yamaguchi (2ZOI) structure and between the Yamaguchi[1] and Coureux[15] (2QJ5) structure. Further, the QM region, corresponding to the isolated chromophore pocket, without the protein was investigated to study the effect of including the protein environment in addition to also including protein symmetry mates.

Evaluating the effect of LBHB on pKa of ARG52 - Second and Third Parametrization.

LBHB force field

A second set of neutral ARG52 charges was derived to address the a possible hydrophobic stabilization of the neutral ARG52. A charged ARG52 can form hydrogen bonds via charge-dipole interactions with residues TYR98 and THR50 in addition to stabilizing the crystal water with ID 1023 (2ZOI). For the neutral ARG52 these interactions are weaker, while the interaction between the neutral ARG52 and the smeared out charge distribution on the GLU46-chromophore system may get stronger upon LBHB formation. To investigate the effect of the latter, we also introduced a parameter set for the GLU46-chromophore pair with a LBHB. The purpose of the second set of parameters is to investigate the influence of a (hypothetical) LBHB on the charge distribution on the GLU46, the chromophore and the interaction with ARG52. We optimized the geometry of the light atoms in the QM region while the GLU46-OH distance was constrained to the proposed 1.21Å LBHB distance.[1] The charges were parametrized using the same setup as before and reflect the polarization in both GLU46 and the chromphore due to the formation of a LBHB.

For reference purposes, a third set of neutral arginine charges were created. The third set included the chromophore, residues GLU46, TYR42 and the ARG52 hydrogen bonding partners THR50 and TYR98 into the ARG52 charge calculation. In these parameters, the polarization of the neutral arginine in the PYP environment is taken into account. For the third parameter set, the positions of the light atoms were optimized at the HF/6-31G* level of theory, before calculating the RESP charges. In this step also the HF charges of all surrounding amino acids were fitted but only the ARG52 charges were used and adopted in the subsequent simulations. The Lennard-Jones parameters were not changed as the nitrogen and the hydrogen parameters for the NH and NH2 chemical groups are identical in the AMBER03 force field.

MD Simulations

We repeated the free energy calculations for the second neutral ARG52 parameter set introduced above. In these calculations, both the protonated and deprotonated ARG52 state also included repolarized chromophore and GLU46 charges in the presence of a 1.21Å GLU46-D bond length. These calculations include the free energy contribution from repolarizing the chromophore pocket residues at the point charge level and increasing the GLU46-D bond length.

Results

The calculated free energy shift estimate for the second parametrization, including a repolarization of the chromophore pocket upon ARG52 deprotonation, was

$$\begin{aligned} \Delta\Delta G_5^{cryst-solv} &= \Delta\Delta G^{TI} - \Delta\Delta G_{cl}^{AMBER03} + \Delta\Delta G_{qm}^{DFT} \qquad (12) \\ &= 36 \pm 2\text{kJ/mol} - 0.10 \text{ kJ/mol} + \text{-}0.13 \text{ kJ/mol}. \end{aligned}$$

The obtained pKa up shift increases to 6.2 pKa units. No down shift was observed. The free energy upshift is the result of adding the two protonation and deprotonation events in solution and the crystal respectively.

In addition to repeating the free energy calculations, also the free MD simulations without position restraints were repeated for both additional parameter sets. Both setups result, again, in a rapid breaking of the ARG52 hydrogen bond network [FIG3,FIG4].

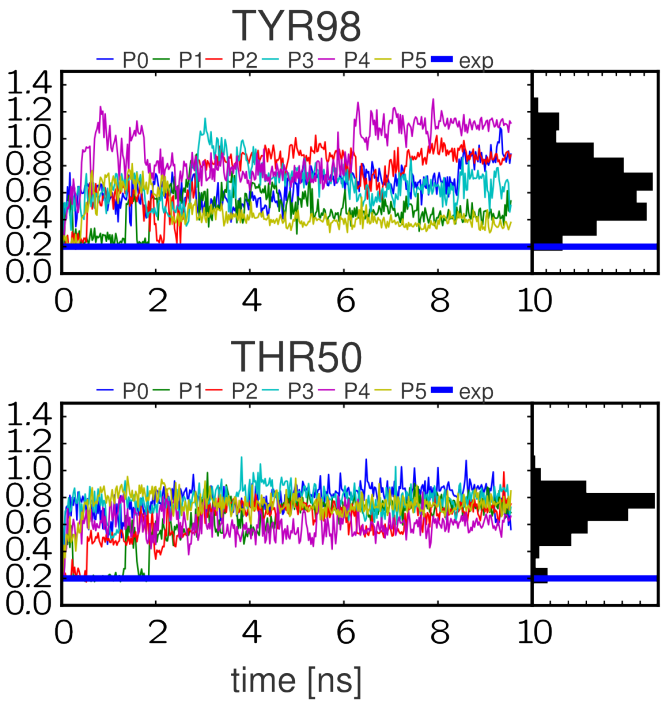

Figure S3: **Free MD simulation of PYP crystal** Simulation of the hydrogen bond network of neutral ARG52 for the second force field parametrization (QM/MM charges). Distances in nm.

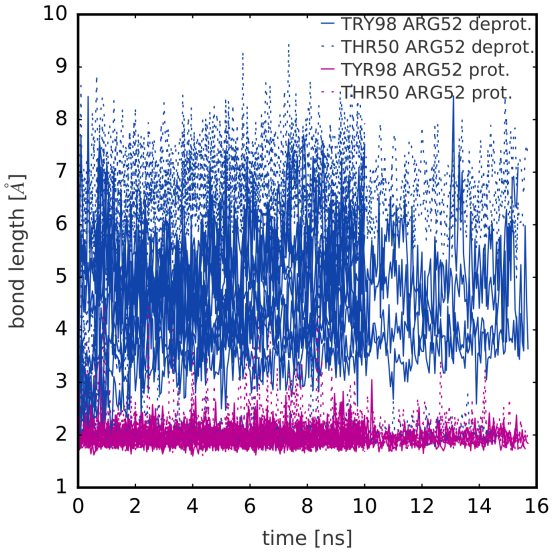

Figure S4: **Free MD simulation of PYP crystal** 10 simulation trajectories of the hydrogen bond network of neutral ARG52 in the crystal (blue) and the protonated ARG52 in solution (red) for the third force field parametrization (QM/MM charges). Distances in Å.

GLU46-D Delocalization in the PYP - Vacuum

We determined the ARG52 protonation state under crystal like conditions and found that it should be considered protonated. Nevertheless, in this section, we assess the influence of the ARG52 protonation state on the LBHB for the protein in vacuum at the QM/MM level of theory. This analysis is independent of the pKa estimate and does not assume knowledge about the protonation state. We assessed the extent of GLU46-D delocalization as a function of ARG52 protonation at varying levels of theory and for different protein structures. The extent of nuclear delocalization is an indication for a low barrier hydrogen bond in PYP.

First, a simple crystal model was created based only on information present in the PDB file. To this end, symmetry mates were added in a sphere of r_c = 3, 5 and 7 nm radius from the GLU46-D. The resulting effect on the delocalization extent is shown for the MPW1B95 functional in figure S5. Within the applied cutoff radii, the protonated and deprotonated ARG52 states result in almost indistinguishable equilibrium densities for the proton/ deuterium and GLU46-D distance expectation values. The occupation numbers of the eigenstates [TAB 1] also support this observation. Here, the delocalization extent even decreases slightly with deprotonation of ARG52 reflected in a small increase of ground state occupation from ≈96% to ≈97%.

Table S1: **Thermal occupation of nuclear eigenstates** Comparison of thermal occupation probability for the protonated and deprotonated ARG52 state for symmetry mates at $r = 5$ nm cutoff. [FIG S5]

$r_c = 5$ nm	ARG52 prot.	ARG52 deprot.
S_0	96.3	97.1
S_1	2.9	2.0
S_2	0.7	0.6
S_3	0.1	0.2

The seemingly absence of changes in the GLU46-D distance expectation value as a function of cutoff distance r_c does not rule out further strong crystal effects from contributions beyond 7 nm. Nonetheless, the absence of changes in the range of r_c =3-7 nm likely speaks

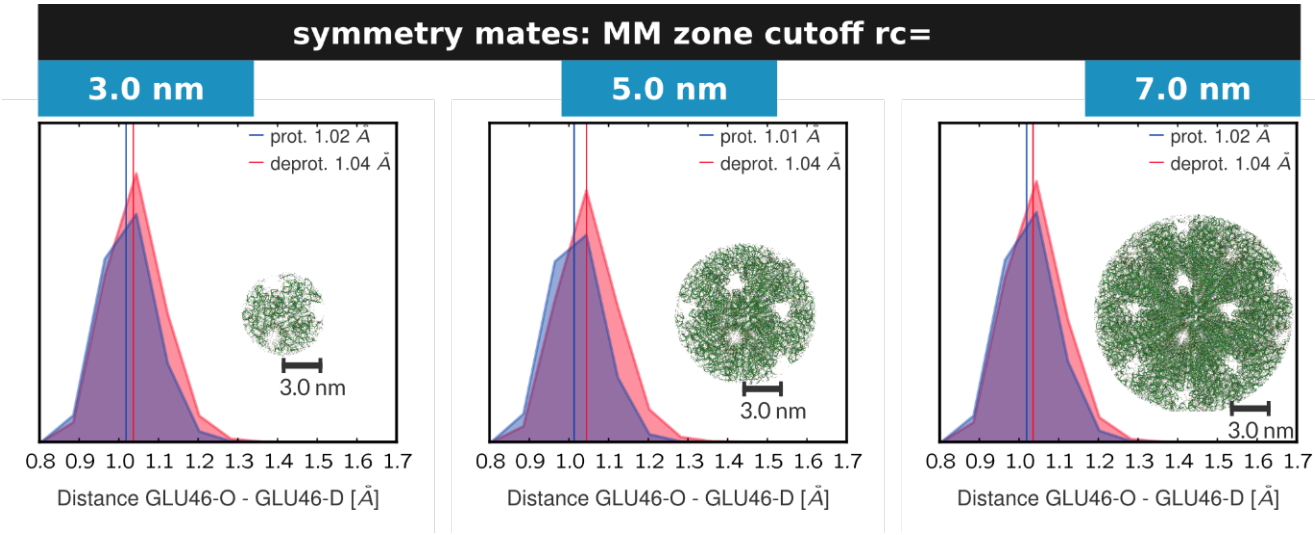

Figure S5: **Inclusion of crystal packing effects and neighboring unit cells at different cutoff ranges** r_c The effect of replacing the vacuum environment with symmetry mates based on information provided in the 2ZOI crystal structure. The resulting crystal model was highly charged due to the absence of counter ions. The protonated ARG52 state is shown in blue and the deprotonated state in red.

against a sudden increase of GLU46-D distance for larger r_c values.

Table S2: **Thermal occupation of nuclear eigenstates** Comparison of thermal occupation probability for the protonated and deprotonated ARG52 state for different functionals.

ARG52 prot.	MPW1B95	B3LYP	ωB97XD	CAM-B3LYP
S_0	97.1	97.2	97.0	97.0
S_1	2.1	2.0	2.1	2.0
S_2	0.6	0.5	0.6	0.6
S_3	0.1	0.2	0.2	0.4
ARG52 deprot.				
S_0	93.0	91.4	89.3	88.3
S_1	4.4	6.1	8.2	9.3
S_2	1.7	1.6	1.6	1.4
S_3	0.5	0.5	0.5	0.5

The Boltzmann weights at room temperature were calculated as the occupation percentile for the lowest nuclear states [TAB 2]. It can be seen that the description of the ARG52 protonated state is very similar between the functionals with around 97% occupation of the ground and about 2% occupation of the first excited state despite of the differences in the location of the potential minima. The results are very different for the deprotonated

ARG52. Here, the ground state occupation varies from 93% for the MPW1B95 functional to around 88% for CAM-B3LYP. Thus, the tested functionals agree qualitatively on a shift of the deuteron mean position and a broadening of its distribution upon deprotonating of ARG52.

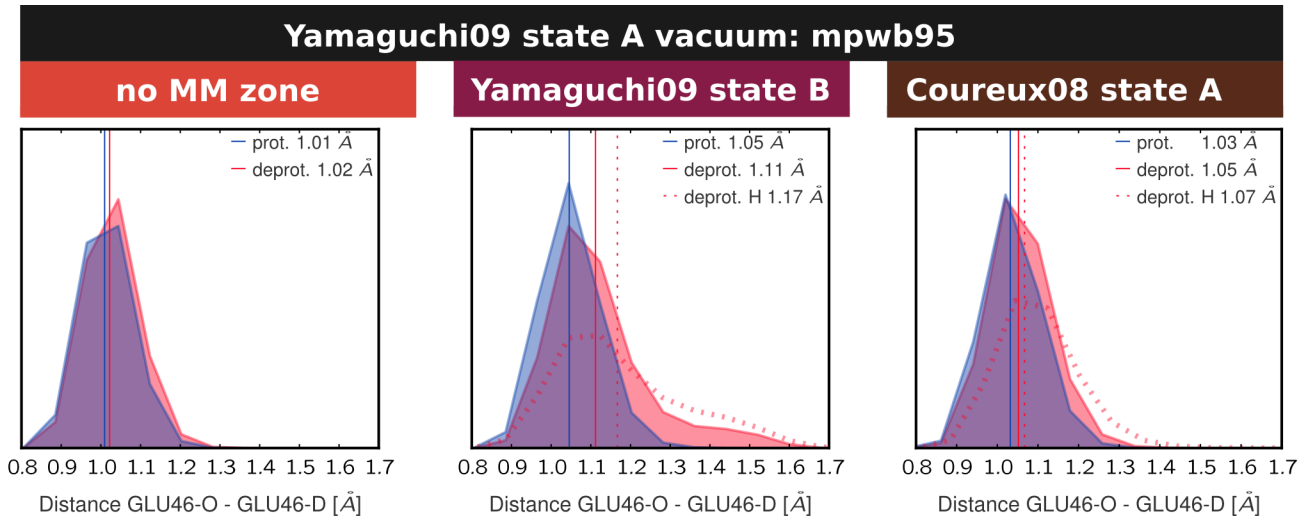

Figure S6: **Comparison of GLU46-D delocalization for structural changes in the protein** The GLU46-D delocalization was calculated for the QMzone of 2ZOI without the MM environment (left). Conformation B of the 2ZOI structure (center) and conformation A of the 2QJ5 structure. A one dimensional cut through the three dimensional probability density $\rho(x)$ of deuterium GLU46-D is shown for protonated ARG52 (blue) and deprotonated ARG52 (red). The expectation values for the GLU46-D distance $< x >$ at $T = 300$ K are shown as vertical lines. Dashed lines indicate the delocalization for hydrogen instead of deuterium.

The crystal model based on the symmetry information in the 2ZOI PDB file did not include counter ions as these were not resolved in the experiment. This resulted in an unphysical net charge of the crystal model system and also local vacuums in the crystal pore.

Next, we tested the robustness of the predictions made for the A conformation of the Yamaguchi crystal against the isolated QMzone in vacuum, the Yamaguchi B conformation and the independent crystal structure by Coureux et al. (2QJ5). The result for the MPW1B95 functional are shown in figure 6. We observed the B conformation in the Yamaguchi structure to exhibit a very strong delocalization of GLU46-D. This effect was also observed in the

thermal occupation probabilities [TAB3]. While the protonated ARG52 conformation was described very similarly to the A conformation, the B conformation with its 86% ground state and 12% excited state occupation is more delocalized than even the CAM-B3LYP results for the A state.

Table S3: **Thermal occupation of nuclear eigenstates** Comparison of the thermal occupation probability for protonated and deprotonated ARG52. The A and B state from the 2ZOI crystal structure together with state A from the 2QJ5 structure are show. Additionally, the QMzone from the 2ZOI structure without the protein MM environment is shown. [FIG S6]

ARG52 prot.	2ZOI state A	2ZOI state B	QMzone only	2QJ5 state A
S_0	97.1	97.1	96.0	96.8
S_1	2.1	2.1	3.1	2.4
S_2	0.6	0.6	0.7	0.6
S_3	0.1	0.2	0.1	0.1
ARG52 deprot.				
S_0	93.0	85.6	96.6	97.1
S_1	4.4	11.8	2.7	2.0
S_2	1.7	1.5	0.6	0.6
S_3	0.5	0.5	0.1	0.3

In contrast, no significant difference between the protonated / deprotonated ARG52 state occupation or shift in GLU46-D distance was observed for the isolated QM zone and the Coureux structure. Both ARG52 protonation states approximately yielded 96-97% and 2-3% ground state and first excited state occupation, respectively. Interestingly, both the isolated QM zone and the Coureux structure show a slightly larger extent of excited state occupation for a protonated ARG52. The Gaussian shape of the distributions are the default shape of nuclear equilibrium densities, which lack occupation of nuclear excited states.

The large differences in GLU46-D delocalization between the chromophore pocket in vacuum (no MM zone) and in the Yamaguchi 2ZOI[1] A and B side chain conformations suggest that the PYP protein environment is the cause of the deuterium delocalization. However, this delocalization was not found in the Coureux et al. 2QJ5[15] structure. This narrows the conditions necessary for GLU46-D delocalization down to the specific protein

environment found in the 2ZOI structure.

While in principle, it appears that all functionals manage to capture a difference between the protonated and deprotonated ARG52 state in the 2ZOI structure, the influence of the MM protein environment is described very differently across the functionals in addition to the inherent differences between the functionals themselves. It might be argued that MPW1B95 is missing the long range attenuation, as present in ωB97XD or CAM-B3LYP, in order to correctly predict the otherwise large extent of delocalization. However, the observed large delocalization extent for MPW1B95 in the 2ZOI side chain conformation B speaks against this. Also for CAM-B3LYP, a very large increase in delocalization was observed which should give rise to concern as the major change between the A and B states is a lysine flip at ≈ 1.4 nm distance from the GLU46-D. This suggests just how sensitive the observed effect is on the protein environment, especially in vacuum calculations. Therefore, it must be questioned if any accurate prediction at all can be made based on the crystal coordinates in vacuum.

Force Field Charges Neutral Arginine

Table S4: The partial charges for the two neutral arginine parameter sets are given.

Atom	charge param. 1	charge param. 2
N	-0.570775	-0.3670
H	0.384125	0.2260
CA	-0.173900	0.0410
HA	0.134800	0.0420
CB	-0.091000	-0.0420
HB1	0.044300	0.0635
HB2	0.044300	0.0635
CG	-0.040700	-0.3570
HG1	0.028700	0.0805
HG2	0.028700	0.0805
CD	0.524200	0.6020
HD1	-0.039100	-0.0285
HD2	-0.039100	-0.0285
NE	-0.941900	-1.0580
HE	0.373900	0.4150
CZ	0.971500	1.1190
NH1	-0.950000	-1.1090
HH11	0.384300	0.4210
HH12	0.000000	0.0000
NH2	-1.018200	-1.0270
HH21	0.398100	0.4050
HH22	0.398100	0.4050
C	0.742625	0.6370
O	-0.592975	-0.5840

Force Field Charges Neutral Arginine Parameter Set 2

Table S5: The partial charges for the neutral arginine, pCA and GLU46 are given.

GLU46	name	charge
46	CD	0.760122
46	OE1	-0.482652
46	OE2	-0.497237
46	HE2	0.158612
ARG52		
52	N	-0.300879
52	H	0.233693
52	CA	-0.131381
52	HA	0.053266
52	CB	-0.058850
52	HB1	0.0225
52	HB2	0.0225
52	CG	-0.082581
52	HG1	0.018543
52	HG2	0.018543
52	CD	0.389569
52	HD1	0.037183
52	HD2	0.037183
52	NE	-1.013558
52	HE	0.389052
52	CZ	1.116343
52	NH1	-1.080069
52	HH11	0.409922
52	HH12	0.0
52	NH2	-0.997553
52	HH21	0.382298
52	HH22	0.382298
52	C	0.730308
52	O	-0.578332
pCA126		
126	C1	0.070194
126	O91	-0.872871
126	C2	-0.382419
126	H2	-0.000011
126	C3	0.270469
126	H3	0.222438
126	C1a	-0.348029
126	C2a	-0.028420
126	H2a	0.412698
126	C3a	-0.927010
126	H3a	0.543793
126	C4a	1.111450
126	O4a	-0.808343
126	C5a	-0.927010
126	H5a	0.543793
126	C6a	-0.028420
126	H6a	0.412698

Optimized Geometry - CAM-B3LYP/6-31+G** - Solvated, Crystal and Vacuum PYP Environment

Table S6: Optimized QM region used for chemical shift prediction - solvated environment.

Atom	x	y	z
C	-1.0299	4.246551	5.356269
H	-1.51279309699	5.2253025246 4	5.4049159718
H	-1.59209001936	3.56341739659	5.9973209933
C	-0.974445	3.727372	3.935099
C	-0.792102772426	4.59321988147	2.85177641732
H	-0.616467747261	5.64918041266	3.04276838629
C	-0.814433528298	4.14104573356	1.53643631057
H	-0.654303933929	4.82108593466	0.70594897054
C	-1.02290510752	2.78981662968	1.26822987149
O	-1.02481700667	2.34881896764	-0.020359079453
H	-0.713856149557	1.38626827119	-0.023440314876 4
C	-1.19940871806	1.9061940798	2.33621983368
H	-1.3428754076 1	0.85036878609 2	2.13187710248
C	-1.17321499736	2.37725625273	3.64542498146
H	-1.32220021859	1.67466245454	4.46151768753
C	3.053626	3.438098	1.351969
H	2.39162324874	4.16048110996	1.83679674151
H	3.86342160423	3.20900854678	2.04710718192
C	2.27345	2.169501	0.980477
H	1.50774423626	2.38333360562	0.23105900513
H	1.73772396436	1.78998575077	1.8560245058
C	3.17249743113	1.06439070771	0.46176001714
O	4.39182211743	1.1512225686	0.443525620441
O	2.55865096174	-0.028059778818 8	0.039441820018
H	1.5398010851	0.028318700489 2	0.064426969127
C	-4.436284	2.554406	-2.074998
H	-5.4936205038	2.8822194261	-2.2145293411 3
H	-4.3553916664	1.8418724631 1	-1.2503778699 7
C	-3.84767	1.994499	-3.37458
H	-2.7882905340 4	1.75677256219	-3.2380391126
H	-3.9440766373 2	2.74543349428	-4.1656643309 1
N	-4.5782001844	0.78166021668 2	-3.7405184934 8
H	-5.4486411121 2	0.57992100315 9	-3.2616571107 7
C	-4.2792605488	-0.025265224632	-4.7553035430 1
N	-3.2060079745 6	0.21272261727 7	-5.5166168295 6
H	-2.7011751306	1.09827546518	-5.4837845144 3
H	-2.9388167629 1	-0.432873515846	-6.2501404206 5
N	-5.0608639778 4	-1.1055647237 2	-4.9383903569 4
H	-5.9397379994 4	-1.145277384 8	-4.3974951586 5
H	-4.9646344344 5	-1.6375603085 8	-5.8019163054 8
C	-6.3323	-8.599356	-2.169266
H	-6.49923386101	-8.9842900003 8	-1.1630570134
C	-5.245858	-9.428024	-2.84595
O	-5.17992129834	-10.644526515 6	-2.6380215961 9
N	-4.360268	-8.804555	-3.643597
H	-4.3854200736 4	-7.8016945756 4	-3.8268938631 4
C	-3.292927	-9.573533	-4.286665
H	-3.7394788688	-10.4703334322	-4.7278410381 1
C	-2.6219208510 7	-8.7487497027 5	-5.3838533520 5
H	-3.3511348901 7	-8.2056682413	-5.9891298999 6
H	-2.0820066735 7	-9.4002008800 4	-6.0743627462
S	-1.3627590694 6	-7.5647254547 7	-4.8149907602
C	-2.3619434312	-6.1877683818 2	-4.2202978413 4
O	-3.5912456813 1	-6.1966529696	-4.3248371516 4
C	-1.5682329858 4	-5.1541726828 9	-3.6299283202
H	-0.5122378737 5	-5.3488510116	-3.4792574029 3
C	-2.1433206772 3	-3.9852443124 1	-3.2329198837 3
H	-3.1888543517 9	-3.8437949169 7	-3.5032255779 8
C	-1.5411735958 5	-2.9396127001 9	-2.4704917983
C	-2.2830988089 7	-1.7726667216 9	-2.1997198807 6
H	-3.2756248291 4	-1.6733188715 9	-2.6319366935 6
C	-1.7817077487	-0.774456898274	-1.397310977
H	-2.3637812416	0.14672741099 8	-1.1856112398
C	-0.4952065802 89	-0.901813817437	-0.8097917650 72
O	-0.0008944567 57768	0.0273771929931	-0.0491370050 742
C	0.2494569784 07	-2.0819626885 3	-1.0850393113 3
H	1.2287193406 8	-2.1887936468	-0.6323293801 75
C	-0.2549245876 38	-3.0590582988 1	-1.9005108006 9
H	0.3464916552 44	-3.9433507114 4	-2.0853652403 9
H	-0.0972969028 1	-7.5290109118 5	-2.1412297987 1
H	-7.2593574283 6	-8.6868725331 7	-2.7416375474 5
H	-2.5739489382 2	-9.8865534647	-3.5160468921 4
H	-3.8804913854 9	3.4602294784 2	-1.8202550992
H	-0.0755784538961	4.3453854882 9	5.8925250351 1
H	3.53546093114	3.94513508104	0.507484506677

Table S7: Optimized QM region used for chemical shift prediction - crystal environment.

Atom	x	y	z
C	-1.028165	4.249931	5.353637
H	-0.98073174186	5.3469398638	5.3723662595
H	-1.97849084993	3.97232063972	5.82406537028
C	-0.973361	3.729756	3.933286
C	-0.8543284024344	4.59981978634	2.85028056403
H	-0.717818285113	5.66241484421	3.03660424609
C	-0.881127802764	4.14199761002	1.53532191853
H	-0.764647751666	4.82596617483	0.704449364449
C	-1.03181180714	2.7863402131	1.26370721215
O	-1.045328611043	2.3538282114	-0.0272277838903
H	-0.682425010573	1.4109558316	-0.065171334856
C	-1.1464674612	1.90104584063	2.34032113241
H	-1.24930471724	0.841282658251	2.14305450168
C	-1.11736548659	2.36938719277	3.64865086217
H	-1.21624116649	1.66172612318	4.46965437672
C	3.0536	3.42912	1.348953
H	2.38300228154	4.1903861327	1.77741769276
H	3.82025431905	3.2225157178	2.09443836702
C	2.273087	2.170283	0.978171
H	1.52330759336	2.3711044208	0.209334186646
H	1.73201918612	1.78349481754	1.84689284778
C	3.21807426449	1.08939281872	0.470652600361
O	4.43277464747	1.19969449807	0.52804243487
O	2.65857373266	0.010124393714 6	-0.042575481152 4
H	1.64154845641	0.055421224217 7	-0.070352276256
C	-4.436778	2.55394	-2.073876
H	-5.49492381651	2.80706616665	-2.19562801539
H	-4.33371011448	1.85332503177	-1.24153407531
C	-3.84975	1.991885	-3.374002
H	-2.8105065137	1.692118302 12	-3.20129765565
H	-3.83819321716	2.79839338732	-4.11769232183
N	-4.5705635583	0.848641152392	-3.95485960015
H	-5.448993474 2	0.497363282253	-3.54999345177
C	-4.10783097665	0.230284953554	-5.05325098744
N	-2.99491501789	0.640483567512	-5.68189802136
H	-2.57461970831	1.5827312149 6	-5.67547981699
H	-2.6825561405	0.061776178998 4	-6.44850040619
N	-4.73445283659	-0.888222157403	-5.47029667479
H	-5.58818722593	-1.12821018393	-4.98507687764
H	-4.58057115534	-1.34163695255	-6.37848298336
C	-6.334158	-8.601104	-2.160685
H	-6.01812971913	-8.51000693652	-1.10452417311
C	-5.248388	-9.430222	-2.837279
O	-5.16477919492	-10.6569402868	-2.64800336044
N	-4.362781	-8.80657	-3.635592
H	-4.40034930858	-7.80815695302	-3.83198980852
C	-3.295527	-9.576442	-4.277949
H	-3.75133412998	-10.4729173745	-4.71205973899
C	-2.65284918529	-8.75993680477	-5.40042176115
H	-3.40441100174	-8.23880285786	-5.99719006627
H	-2.12667629064	-9.4123707253	-6.09918756949
S	-1.40409956529	-7.548899172	-4.87072297233
C	-2.41066588426	-6.20415163374	-4.21314958251
O	-3.64111128474	-6.21571075051	-4.30496400625
C	-1.61397258687	-5.18746975535	-3.59572947947
H	-0.548212166227	-5.36448139677	-3.50036310685
C	-2.19768037804	-4.05212774412	-3.12020179991
H	-3.26372777913	-3.94088768719	-3.30848783452
C	-1.57749791181	-2.9950158777	-2.3985604773
C	-2.34389212519	-1.86831566853	-2.03090381935
H	-3.39167568121	-1.82536013458	-2.32752327476
C	-1.79197711556	-0.831999555326	-1.3146132938
H	-2.38472034099	0.032716803076 7	-1.03598580373
C	-0.430784787287	-0.871667068992	-0.910288447102
O	0.111081365659	0.104300968864	-0.246398580429
C	0.333466057008	-2.018664184 02	-1.26675734443
H	1.3666890741	-2.06551283495	-0.944211411922
C	-0.221772812686	-3.03683260597	-1.99333367923
H	0.394729580521	-3.89285360487	-2.25598975922
H	-6.46286212149	-7.61830893707	-2.62960597081
H	-7.2725242 4005	-9.15756954 3	-2.17404813201
H	-2.56257032622	-9.87754189417	-3.51603383448
H	3.58735162386	3.91030148704	0.514940259416
H	-0.27041818864	3.87132540016	6.05030182178
H	-3.88340372572	3.46344635612	-1.83172548125

Additional Scans of the Vacuum 2ZOI Structure

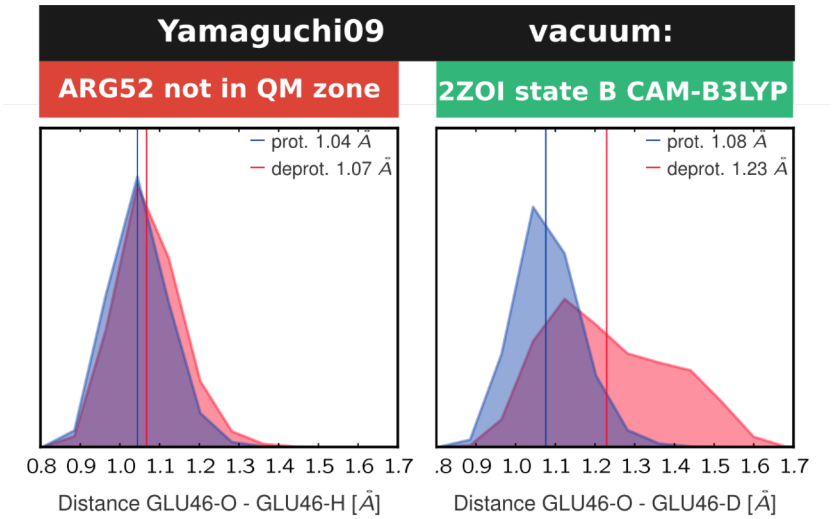

Figure S7: **Additional scans.** (left) ARG52 not part of the QM zone MPW1B95 functional. (right) CAM-B3LYP results for the B state of 2ZOI.

Table S8: Optimized QM region used for chemical shift prediction - vacuum environment.

Atom	x	y	z
C	24.74	3.622	5.538
H	25.2057352589	4.50306992962	5.98645515588
H	25.162493399	2.73500528005	6.01694445234
C	23.236	3.65	5.711
C	22.5462211574	4.86017941724	5.83908843935
H	23.0921413453	5.79261420747	5.72313151813
C	21.1758399635	4.90544127413	6.08199375319
H	20.6522177807	5.85355840598	6.1520886804
C	20.4523857175	3.72083161446	6.20244579636
O	19.1014147085	3.76868190702	6.41490031784
H	18.6740381555	2.97786695801	5.9553491186
C	21.1241927798	2.50407678665	6.08311661589
H	20.5627935041	1.58087289421	6.17122259128
C	22.4938075637	2.4738328918	5.83958385145
H	22.995969434	1.51122220471	5.75835033426
C	20.174	4.939	2.255
H	20.9585908772	5.35640352916	2.8922655052
H	20.6476779986	4.60378742142	1.33059700983
C	19.474	3.773	2.966
H	18.9655280127	4.10875497041	3.87256596763
H	20.2020315432	3.02538171676	3.29415729515
C	18.4584291737	3.10162786269	2.05059117879
O	18.3199106833	3.41122120476	0.86314229326
O	17.7280133714	2.15577760239	2.58897827232
H	17.8763768552	2.05108221108	3.62804062035
C	17.798	4.116	10.147
H	17.9446683667	4.23403645385	11.2244703892
H	18.2819525712	3.19606570848	9.81355658972
C	16.312	4.149	9.772
H	16.2085260487	3.98664326207	8.6947241283
H	15.9108713057	5.14194118276	10.00723777
N	15.4768918517	3.14772972456	10.4672012566
H	15.8784771448	2.66519661396	11.268804017
C	14.1978788371	2.9070993497	10.1710623803
N	13.5290127501	3.6615569794	9.2909922477
H	13.8705595373	4.53871394889	8.88411225716
H	12.6818605115	3.26462586314	8.90014730835
N	13.5372961857	1.90080387192	10.8008660854
H	14.0645117756	1.0664018676	11.0237732307
H	12.5236414967	1.75424278935	10.6956516168
C	13.822	-6.448	10.937
H	14.6807267247	-7.09664552745	10.7545834204
C	12.734	-6.8	9.929
O	12.4625525703	-7.97622887305	9.72988916708
N	12.107	-5.802	9.28
H	12.2865486206	-4.83379677073	9.50156713308
C	11.075	-6.114	8.289
H	10.3691507299	-6.79872997714	8.7772014859
C	10.2762716198	-4.85416891673	7.98284758319
H	10.010909391	-4.27223248703	8.8683180173
H	9.33125076749	-5.08728544867	7.49215342626
S	11.0713835029	-3.70932668802	6.80265021577
C	12.2253074734	-2.80940330102	7.77566192939
O	12.3346514268	-2.89034927162	8.99820817207
C	13.0812987013	-1.9525397241	6.94248311514
H	12.9956743434	-2.02987847827	5.86320076019
C	13.9986923095	-1.1547577651	7.51219408038
H	14.0074405574	-1.12976300057	8.59860546507
C	15.0020425063	-0.345115008706	6.8592630049
C	15.8171605504	0.506231193661	7.61732965636
H	15.6819230517	0.538373783303	8.69748208138
C	16.8037514104	1.27916043348	7.03007568198
H	17.4440070517	1.91498848831	7.63386817231
C	17.0299051218	1.2133211189	5.63629539208
O	17.9520489161	1.94261614986	5.05390573223
C	16.2292260614	0.330151021017	4.87937227738
H	16.4196893375	0.245747756514	3.81473844353
C	15.2433837036	-0.417900209233	5.47736125459
H	14.6587383451	-1.08995779645	4.86945074586
H	19.5126844809	5.76266197174	1.96255104969
H	25.136259409	3.55932219793	4.51602016142
H	18.271705887	4.95974666277	9.64332895372
H	13.4551930581	-6.63566158922	11.9486630716
H	14.1251398138	-5.3880345589	10.9044700805
H	11.3641757561	-6.66945796838	7.34728743757

ARG52 - Water distance at crystal water ID 1023 (2ZOI) site

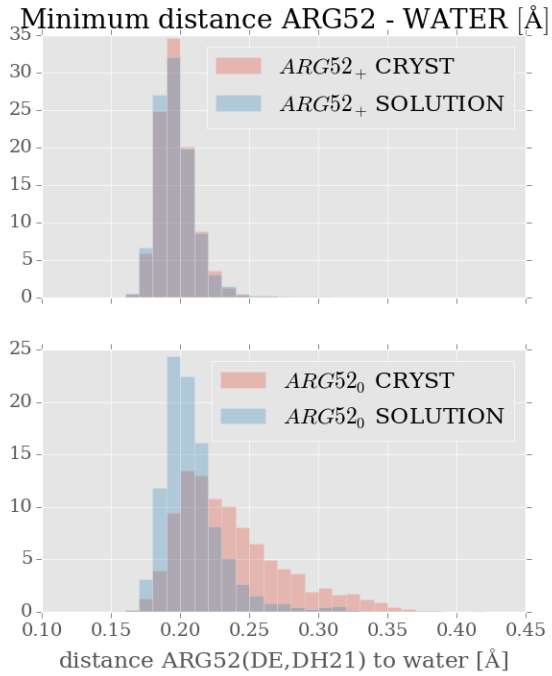

Figure S8: Minimum distance of ARG52 to a water molecule at the binding site with crystal ID 1023 (2ZOI) for a protonated ARG52 (top) and a deprotonated ARG52 (bottom). Molecular dynamics data from production sampling for $\lambda = 0$ and $\lambda = 1$

FO-FC maps at 2.3 sigma instead of 3 sigma

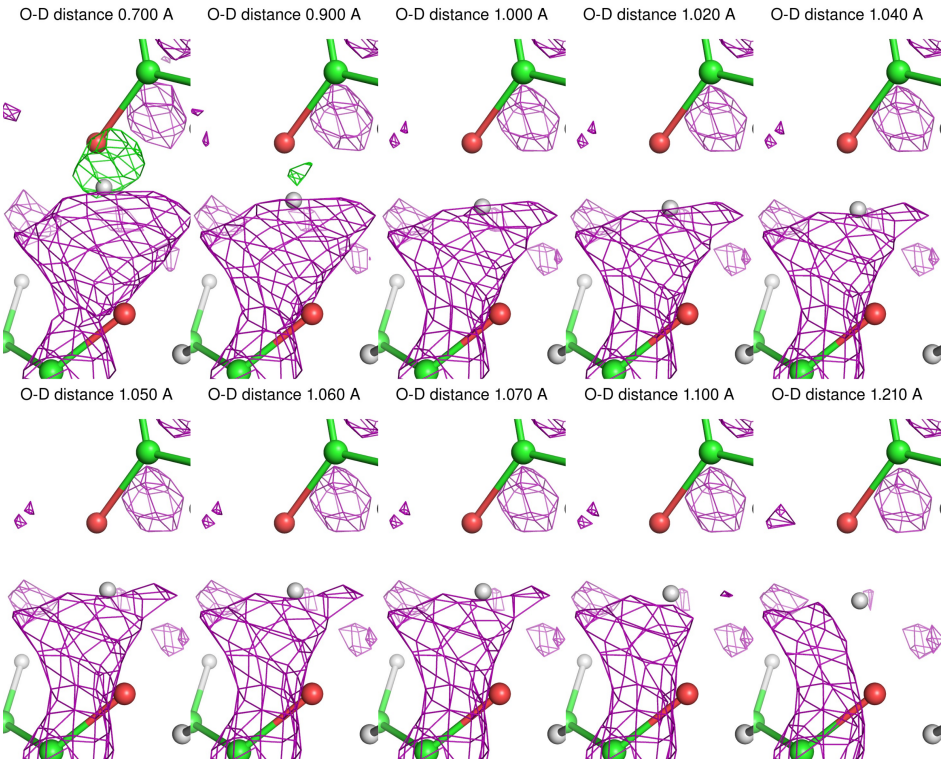

Figure S9: **FO-FC maps at 2.3 sigma instead of 3 sigma** The background noise becomes visible at random positions throughout the structure.

FO-FC maps at 6 sigma instead of 3 sigma

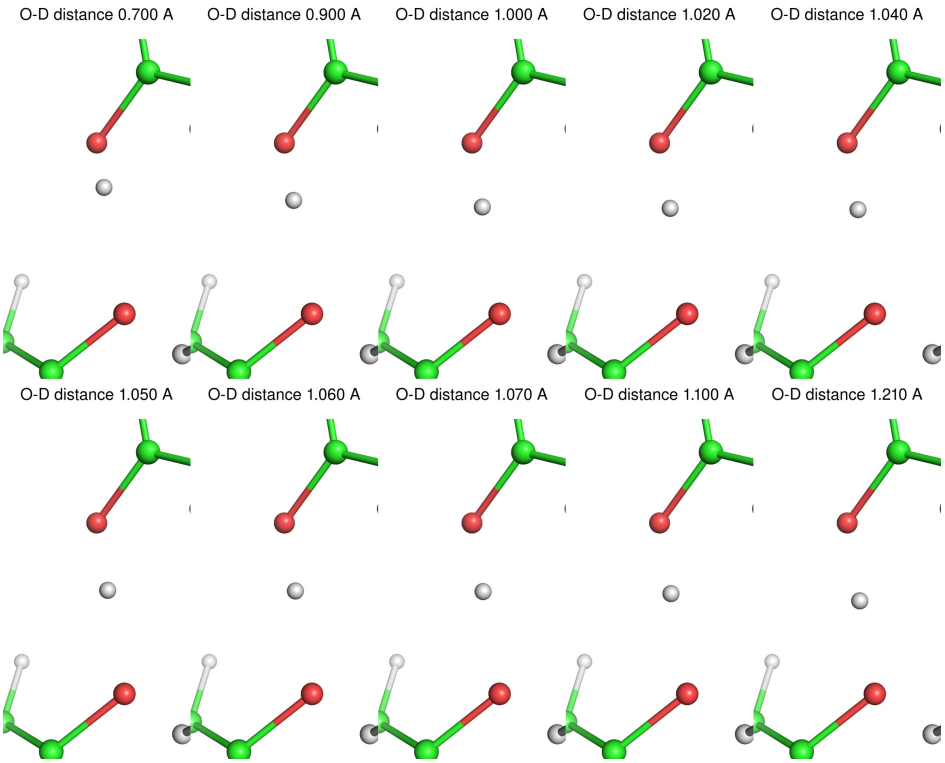

Figure S10: **FO-FC maps at 6 sigma instead of 3 sigma** No density is resolved at this level.

References

(1) Yamaguchi, S.; Kamikubo, H.; Kurihara, K.; Kuroki, R.; Niimura, N.; Shimizu, N.; Yamazaki, Y.; Kataoka, M. *Proceedings of the National Academy of Sciences* **2009**, *106*, 440–444.

(2) Becke, A. D. *The Journal of Chemical Physics* **1993**, *98*, 5648–5652.

(3) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

(4) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–2093.

(5) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *The Journal of Physical Chemistry* **1993**, *97*, 10269–10280.

(6) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.;

Cieplak, P.; Luo, R.; Lee, T. *Journal of Computational Chemistry* **2003**, *24*, 1999–2012.

(7) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of chemical physics* **1983**, *79*, 926–935.

(8) Bussi, G.; Donadio, D.; Parrinello, M. *The Journal of chemical physics* **2007**, *126*, 014101.

(9) others,, et al. *Journal of computational chemistry* **1997**, *18*, 1463–1472.

(10) Miyamoto, S.; Kollman, P. A. *Journal of computational chemistry* **1992**, *13*, 952–962.

(11) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *The Journal of chemical physics* **1995**, *103*, 8577–8593.

(12) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *Journal of Chemical Theory and Computation* **2008**, *4*, 435–447.

(13) Hub, J. S.; de Groot, B. L.; Grübmuller, H.; Groenhof, G. *Journal of Chemical Theory and Computation* **2014**, *10*, 381–390.

(14) Oktaviani, N. NMR studies of folded and unfolded proteins: method developments and biological insight. Ph.D. thesis, 2014.

(15) Coureux, P.-D.; Fan, Z. P.; Stojanoff, V.; Genick, U. K. *Structure* **2008**, *16*, 863–872.

(16) Nadal-Ferret, M.; Gelabert, R.; Moreno, M.; Lluch, J. M. *Journal of the American Chemical Society* **2014**, *136*, 3542–3552.

(17) Kanematsu, Y.; Tachikawa, M. *The Journal of chemical physics* **2014**, *141*, 185101.

(18) Zhao, Y.; Truhlar, D. G. *The Journal of Physical Chemistry A* **2004**, *108*, 6908–6918.

(19) Yanai, T.; Tew, D. P.; Handy, N. C. *Chemical Physics Letters* **2004**, *393*, 51–57.

(20) Chai, J.-D.; Head-Gordon, M. *Physical Chemistry Chemical Physics* **2008**, *10*, 6615–6620.