Supporting Information

Vreede et al. 10.1073/pnas.0908754107

SI Text

SI Methods. ***System preparation and molecular dynamics.*** Preparation of the system and subsequent molecular dynamics simulations were performed with the GROMACS package (version 3.3.1) (S1), employing the Gromos96 43a1 force field [the Gromos96 43a1 force field was used in previous work on photoactive yellow protein (PYP)] (S2, S3) in combination with the Simple Point Charge water model (S4). As a starting point for the folded *pB′* state of PYP, we used the NMR structure of PYP in the dark state [Protein Data Bank (PDB) entry 3PHY, conformation 11 (S5)] with manual alterations to reflect the different chemical topology of the chromophore binding pocket [i.e., protonated para-coumaric acid (pCA) and deprotonated Glu46, and pCA in *cis* configuration, obtained from a cryotrapped crystal structure (PDB entry 1UWP) (S6)]. This structure was placed in a periodic rhombic dodecahedral box of SPC water (S4). The box size allowed accommodation of the protein with a minimum solvation layer of 1.5 nm. Water molecules that resided in the internal hydrophobic cavities were removed. Sodium ions replaced six water molecules to neutralize the charge of −6 on the protein. The system was energy minimized by using 200 steps of the conjugate gradient method and equilibrated to dissipate excess energy and relax the box volume. After relaxation of the water molecules and the protein hydrogens for 20 ps, 1 ns of equilibration of the whole solvated system at constant temperature and pressure resulted in a diameter of 54.83 Å. Van der Waals interactions were treated with a cutoff of 1.4 nm, and particle mesh Ewald handled the electrostatics with a grid spacing of 0.12 nm (S7, S8). Using the linear constraint solver (LINCS) for interactions between protein atoms (S9) and the SETTLE algorithm for water interactions (S10), allowed for a time step of 2 fs. Parameters for pCA were taken from ref. S11. To keep the temperature constant in the transition path sampling (TPS) simulations, we used a modified version of the Andersen thermostat (S12). In subsequent molecular dynamics (MD) simulations we used the Nosé–Hoover thermostat (S13, S14) to maintain a constant temperature. All other settings in these MD simulations were as described above.

TPS. TPS (S15, S16) carries out an unbiased sampling of MD trajectories that connect predefined initial and final states. The transition path sampling is a random walk through trajectory space, where new trajectories are generated from old trajectories by a shooting move algorithm and accepted with a Metropolis rule (S15, S16). A shooting move consists of choosing randomly a time frame (the shooting point) of the old trajectory and creating from it a new trajectory by integrating the equations of motion forward and/or backward in time by using, e.g., conventional MD. Different dynamics and different shooting algorithms require different acceptance rules, as explained in detail in refs. S16 and S17. In this work we use the TPS algorithm as previously implemented, tested, and successfully applied to protein systems (S18, S19). Acceptance of the new paths governed by a Metropolis rule depends on the detailed balance condition and includes two stages. First, the new path must still connect the initial and final regions. If this is not the case, the acceptance probability vanishes, and the trial path can be rejected right away. The second stage in the acceptance rule comes from the fact that the new path can and generally will be of different length than the initial path (S20). At this stage the Metropolis acceptance ratio reduces to $P_{\mathrm{acc}} = \min(1, \frac{N^{(o)}}{N^{(n)}})$, where the min function returns the smaller of its arguments. Instead of wasting CPU time by rejecting paths that do connect initial and final states but are finally too long, we introduce in the second stage of the acceptance-rejection procedure a maximum path length $N_{\max} = N^{(o)}/\xi$, where ξ is a random number $\in [0, 1]$. We determine $N_{\max}$ in advance and stop the generation of the new path whenever the instantaneous value of $N^{(n)}$ exceeds $N_{\max}$.

In addition to the flexible path length approach (S20), we improve the Monte Carlo acceptance ratio by using a one-way shooting algorithm (S16). The shooting move starts by randomly choosing a point on the initial trajectory containing $N^{(o)}$ time slices. The chosen time slice is denoted τ. Drawing a random number ξ between 0 and 1 determines whether the new trajectory is generated in a backward or forward direction. $\xi > 0.5$ selects to grow a forward trajectory and $\xi \leq 0.5$ a backward trajectory, by reversing the momenta in the chosen time slice. The equations of motion are then integrated by using Andersen coupled MD (see below). Because of the stochastic dynamics, we do not have to change the momentum of the shooting point. During the evolution of the new path, the instantaneous path length as well as several order parameters are monitored. By denoting τ_f the number of time slices in the new trajectory initiated from time slice τ, the length of the new path when shooting forward is $N^{(n)} = \tau + \tau_f$, whereas for backward shooting moves the new path length becomes $N^{(n)} = (N^{(o)} - \tau) + \tau_f$. The MD integration ceases when the path reaches either the initial or the final state.

We use stochastic dynamics by coupling the water molecules in the system to the Andersen thermostat, with a coupling constant of $p = 0.00031$ (S19). This probability corresponds to selecting roughly one water molecule per time step on average. After selecting a water molecule, its momenta are randomly drawn from the Maxwell–Boltzmann distribution, leaving the rotational motion unchanged, resulting in alteration of the center of mass motion of the molecule. Compared to simulations using the Nosé–Hoover thermostat, the diffusion constant of the water does not change with the choice of the coupling constant p. Applying the Andersen thermostat ensures that the trajectories will diverge, even when started from identical initial conditions. By using a relatively low coupling constant p only changing momentum of water molecules, the trajectories will closely resemble true dynamics.

Order parameters. By using various order parameters we aim to characterize the transitions occurring in the formation of the PYP signaling state. These order parameters include distances between various groups and atoms, the number of waters within a distance from specified residues, rmsd of different parts of the protein, and hydrogen bond counts within helices and the chromophore binding pocket. A list of these order parameters is given in Table S1. These parameters were computed by using a combination of GROMACS analysis tools and Perl scripts. Several of these order parameters also describe the initial and final states in the TPS simulations (Table S2). In the likelihood maximization (LM) analysis (see below), all 75 parameters are included as possible candidates for the reaction coordinate.

Sampling the transition path ensembles. A single TPS simulation allows for sampling only one barrier at the time, because intermediate states will lengthen pathways dramatically. Therefore we split the process of signaling state formation into four parts, as indicated in Fig. 1 of the main text: unfolding of the α3-helix (pB'-I_α), exposure of pCA (U_α-S_X), exposure of glutamic acid 46 (U_α-S_E), and exposure of both pCA and Glu46 (S_E-pB).

The definitions of the initial and final states for all four TPS simulations are listed in Table S2.

Starting a TPS simulation requires an input trajectory connecting the initial state to the final state. For three of the four TPS simulations we used high temperature unfolding trajectories, taken from previous replica exchange MD (REMD) simulations (S21, S22). For the S_E-pB transition, we used a high temperature MD simulation starting from a conformation in the S_E state. Initial TPS simulations served to equilibrate these high temperature trajectories, resulting in suitable room temperature input pathways.

Initially, the definitions of the stable states originate from the observations done for REMD simulations of this reaction (S22). To verify that the definitions of the stable states are sufficient, we performed several MD simulations initiated with start and end points of the TPS trajectories. This check also enables a more thorough characterization of the stable states in the light-induced unfolding reaction of PYP. The results of the MD simulations were also used in the Bayesian path statistical analysis (see *SI Results and Discussion*).

During the TPS simulations we store the accepted paths, the rejected paths, and the path statistics. We also store the forward and backward shooting point configurations and the outcome of the trajectory (initial or final state) for the LM analysis.

Decorrelation of pathways was monitored by constructing sampling trees (S19). A path was deemed decorrelated and independent from the previous independent path if it had no configurations in common. The decorrelation takes more than one accepted shot because of the one-way nature of the shooting algorithm. The number of decorrelated pathways is listed in Table 1 of the main text.

Reaction coordinate analysis. When performing the TPS simulations, we monitor about 75 order parameters. Some of them are used to determine whether the initial or final state has been reached. These initial and final state order parameters are chosen on the basis of the previous REMD simulations (S22) and are not likely to describe the reaction coordinate (RC) optimally. As follows from the transition path theory (S23, S24), a good reaction coordinate must be able to predict the commitment probability (committor or p-fold) $p_B(x)$ of a conformation x to the final state; i.e., the reaction coordinate must be able to describe the progress of the reaction. This is exactly what the committor $p_B(x)$ does: Each value of $p_B(x)$ gives the probability of reaching the final state. Committor analysis facilitates judging whether an order parameter is a good reaction coordinate (S16, S25). However, this procedure is very expensive from a computational point of view, to the extent that it easily requires a time comparable with a whole TPS simulation.

Ma and Dinner employed the concept of the committor to develop a more efficient and systematic procedure for identifying reaction coordinates on the basis of genetic neural networks (S26). This approach still requires an extensive committor analysis. Best and Hummer developed a reaction coordinate analysis technique that combines the equilibrium probability distribution with path ensemble distributions (S27, S28). The basis is the following Bayesian relation for the conditional probability $p(\mathrm{TP}|r)$ that a configuration with a certain value of reaction coordinate r lies on a transition path (TP) connecting the initial and final states:

$$p(\mathrm{TP}|r) = \frac{p(r|\mathrm{TP})p(\mathrm{TP})}{p_{\mathrm{eq}}(r)}. \qquad \textbf{[S1]}$$

Here $p_{\mathrm{eq}}(r)$ denotes the equilibrium distribution as a function of r, $p(r|\mathrm{TP})$ is the distribution of configurations with a certain r visited along transition pathways in the TPS ensemble, and the normalizing factor $p(\mathrm{TP})$ is the overall likelihood to be on a TP (S28). The conditional probability $p(\mathrm{TP}|r)$ is large for r values that are common to transition pathways but are rarely visited in equilibrium. Thus, one can identify transition states with the maximum of $p(\mathrm{TP}|r)$ or, in other words, with the configurations that have the largest probability that trajectories passing through them are reactive (S27). For diffusive dynamics $p(\mathrm{TP}|r) = 2p_B(r)(1-p_B(r))$, where $p_B(r)$ is the committor averaged over all configurations with r. For a good reaction coordinate, all transition states—which have a maximum $p(\mathrm{TP}|r)$—should correspond to approximately the same value of the reaction coordinate, and thus $p(\mathrm{TP}|r)$ should be peaked around the transition state value of r. When r is a poor reaction coordinate, $p(\mathrm{TP}|r)$ will be more or less featureless, because of the lack of correlation between r and $p(\mathrm{TP}|r)$. The Best–Hummer approach requires a good estimate of the equilibrium distribution in the transition region, which might be difficult to obtain.

LM analysis. Peters et al. (S29, S30) recently devised a committor analysis on the basis of a LM estimation that requires only data from a TPS simulation itself. The basic idea is that each trial shot in the TPS simulation is a realization of a committor computation (see below). The LM approach extracts a linear combination of order parameters that best describes the reaction coordinate r. Required as input for this procedure is the ensemble of N forward (or backward) shooting point configurations x_{sp} belonging to the accepted trajectories ending in the final state B ($x_{\mathrm{sp}} \to B$) and the rejected shooting points ending in the initial state A ($x_{\mathrm{sp}} \to A$). By using these configurations, the LM optimizes the likelihood

$$L = \prod_{x_{\mathrm{sp}} \to B} p_B(r(x_{\mathrm{sp}})) \prod_{x_{\mathrm{sp}} \to A} (1 - p_B(r(x_{\mathrm{sp}}))), \qquad \textbf{[S2]}$$

where the committor $p_B(r)$ as a function of the reaction coordinate r is modeled by a tanh function (S30):

$$p_B(r(x)) = \frac{1}{2} + \frac{1}{2}\tanh(r[q(x)]). \qquad \textbf{[S3]}$$

The reaction coordinate $r(x)$ in turn is approximated as a linear combination of order parameters $q(x)$:

$$r(q(x)) = \sum_{i=1}^{n} a_i q_i(x) + a_0. \qquad \textbf{[S4]}$$

The LM analysis method facilitates the screening for many (combinations of) order parameters as candidate reaction coordinates. The LM then gives the linear combination of order parameters which reproduce the shooting point data best.

For each shooting point in the ensemble we computed all the parameters listed in Table S1. Employing the LM approach for the forward and backward shooting point ensembles, we tested all possible linear combinations of up to three order parameters and kept the one with the maximum likelihood as our best model for the reaction coordinate (S30).

According to a Bayesian criterion, adding more variables to the RC model is significant only if the log-likelihood $\ln L$ increases by at least $\delta L_{\mathrm{min}} = \frac{1}{2}\ln(N)$ (S29), with N the total number of shooting points in the ensemble.

Committor calculation. We explicitly computed the committor for 29 configurations that were identified as putative transition states by the LM analysis for the I_α-U_α transition, with $r \in [-0.05, 0.05]$. The committor for a configuration is the fraction of MD trajectories, initialized with random momenta, that reach the final state. The number of trajectories is determined by the required accuracy of the committor value (S16). We calculated the committor with an accuracy of 0.1 with 95% confidence. The

committor computation typically required 100–500 trajectories per configuration.

Path density. The path density plots show the fraction of pathways in the TPS ensemble that pass through the given values of the order parameters at least once (S19). These plots were prepared as follows. Each pathway in the ensemble was smoothed by taking a running average with a window of 100 ps. Next, by discretizing each of the order parameter intervals in about 30–40 bins, we constructed for each trajectory in the ensemble a binary matrix, in which 1 means that the path visits the bin at least once and 0 means no visitations at all. These matrices were subsequently ensemble-averaged, resulting in the path density maps presented in Figs. 2 and 3 of the main text and Fig. S3.

Results and Discussion. ***Bayesian path statistics analysis.*** As described above, the Best–Hummer analysis can test the quality of proposed reaction coordinates by computing the $p(\mathrm{TP}|r)$ from Eq. **S1**. We performed this analysis for the pB'-I_α, U_α-S_E, and U_α-S_X transitions to confirm the results of the LM method. For this analysis we require both the equilibrium distribution as a function of the reaction coordinate $p_{\mathrm{eq}}(r)$ and the TP ensemble distribution $p(r|\mathrm{TP})$. The first can be constructed from the REMD data or, in the case of pB'-I_α, also from the MD trajectories in pB'. $p(r|\mathrm{TP})$ can be constructed from the trajectories sampled in the TPS simulation. The $p(\mathrm{TP})$ factor is considered an arbitrary scaling constant.

In Fig. S1 we plot the different distributions for the pB'-I_α transition. The scaling factor $p(\mathrm{TP})$ is chosen such that all distributions fall within the same range. Fig. S1(*a*) and (*b*) show $p_{\mathrm{eq}}(r)$, $p(r|\mathrm{TP})$, and $p(\mathrm{TP}|r)$ for the optimized reaction coordinate obtained from the LM analysis, $r = 5.11 - 16.81 * \mathrm{rmsd}_\alpha - 4.68 * dhb2 - 2.55 * dPA$ (see Table 2 in the main text). Both the MD and REMD results show a peak at the location of the transition state $r = 0$. The spurious peak around $r = -2$ arises from accumulation of TPS trajectories in the stable states. The main reason for the additional peak is that the optimized free energy barrier for this transition is around $4k_{\mathrm{B}}T$, leading to relatively high $p_{\mathrm{eq}}(r)$ at the transition state $r = 0$. Fig. S1(*c*) and (*d*) show the three distributions for the best single order reaction coordinate, $r = 3.89 - 29.1 * \mathrm{rmsd}_\alpha$. Also here a sharp peak occurs at the transition state value $r = 0$, although the difference between the REMD and MD results is slightly larger. Again, spurious peaks arise from trajectory accumulation in the stable states. On the basis of this analysis it is hard to conclude which RC is better. To compare with an arbitrary order parameter we show in Fig. S1(*e*) and (*f*) the distributions for *dDK*. Here, the distributions are very broad and there is no clear peak. Thus, the Best–Hummer analysis agree with the results of the LM method for pB'-I_α presented in the main text. We note that there are convergence problems with obtaining the equilibrium distribution from MD and REMD that complicate a proper comparison. Application of umbrella sampling or other biasing methods might improve the sampling but require an a priori chosen order parameter.

Fig. S2 shows the result of a similar analysis for the U_α-S_E and U_α-S_X transitions, using only REMD data. Fig. S2(*a*) shows the $p_{\mathrm{eq}}(r)$, $p(r|\mathrm{TP})$, and $p(\mathrm{TP}|r)$ for the optimized RC $r = -2.03 + 2.70 * dXE$, as obtained from the LM analysis for the U_α-S_E transition. A peak occurs at the proposed transition state $r = 0$ but turns out to be caused by accumulation of the TPS distribution, and not by a low equilibrium distribution. Again, this effect is probably caused by poor REMD convergence. The distributions for the U_α-S_X transition are shown in Fig. S2(*b*) as a function of $r = -5.05 + 5.02 * dXY^{\mathrm{com}} - 2.51 dXE^{\mathrm{com}} + 4.30 * dXE$, which according to the LM analysis should be a reasonable RC. However, the situation is even worse than for the U_α-S_E transition. In this case, there no maximum at the proposed transition state $r = 0$, but a minimum instead. Again this effect is due to an unconverged equilibrium distribution.

The LM analysis does not suffer from unwanted contributions from the equilibrium distributions because it focuses only on shooting points on TPs. Hence, the reaction coordinates found by LM analysis are valid only around the transition state regions. If we would sample configuration space orthogonal to $r = 0$, performing a committor analysis would very likely fail. As stated in the main text, a full committor analysis would be beyond the scope of this work.

To complement the figures from the main text, Fig. S3 shows the path density and the transition state ensembles for the solvent exposure of Glu46 and pCA. In this figure it is clear that the predicted transition state ensembles lie close to the predicted $rc = 0$ line, as they should naturally. Note that the transition state in both transitions lies close to the initial states in the projections.

1. van der Spoel D, et al. GROMACS: Fast, flexible and free. (2005) *J Comput Chem* 26:1701–1718.
2. van Gunsteren W, et al. (1996) *Biomolecular Simulation: The GROMOS96 Manual and User Guide* (Vdf Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland).
3. Daura X, Mark AE, van Gunsteren WF (1998) Parametrization of aliphatic CH_n united atoms of GROMOS96 force field. *J Comput Chem* 19:535–547.
4. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J, Pullman, B, ed. (1981) *Intermolecular Forces* 331–342.
5. Dux P, et al. (1998) Solution structure and backbone dynamics of the photoactive yellow protein. *Biochemistry* 37:12689–12699.
6. Kort R, Hellingwerf KJ, Ravelli RB (2004) Initial events in the photocycle of photoactive yellow protein. *J Biol Chem* 279:26417–26422.
7. Darden T, York D, Pedersen L (1993) Particle mesh Ewald—An N.log(N) method of Ewald sums in large systems. *J Chem Phys* 98:10089–10092.
8. Essman U, et al. (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103:8577–8593.
9. Hess B, Bekker B, Berendsen HJC, Fraaije J (1997) LINCS: A linear constraint solver. *J Comput Chem* 18:1463–1472.
10. Miyamoto S, Kollman P (1992) SETTLE–An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J Comput Chem* 13:952–962.
11. Groenhof G, Lensink MF, Berendsen HJC, Mark AE (2002) Signal transduction in the photoactive yellow protein. II. Proton transfer initiates conformational changes. *Proteins Struct Func Bioinf* 48:212–219.
12. Andersen HC (1980) Molecular-dynamics simulations at constant pressure and/or room temperature. *J Chem Phys* 72:2384–2393.
13. Hoover WG (1985) Canonical dynamics—Equilibrium phase-space distributions. *Phys Rev A* 31:1695–1697.
14. Nosé S (1986) An extension of the canonical ensemble molecular-dynamics method. *Mol Phys* 57:187–191.
15. Dellago C, Bolhuis PG, Csajka FS, Chandler D (1998) Transition path sampling and the calculation of rate constants. *J Chem Phys* 108:1964–1977.
16. Dellago C, Bolhuis PG, Geissler PL (2002) Transition path sampling. *Adv Chem Phys* 123:1–78.
17. Bolhuis PG (2003) Transition path sampling on diffusive barriers. *J Phys Condens Matter* 15:S113–S120.
18. Juraszek J, Bolhuis PG (2006) Sampling the multiple folding mechanisms of Trp-cage in explicit solvent. *Proc Natl Acad Sci USA* 103:15859–15864.
19. Juraszek J, Bolhuis PG (2008) Rate constant and reaction coordinate of Trp-cage folding in explicit water. *Biophys J* 95:4246–4257.
20. van Erp TS, Moroni D, Bolhuis PG (2003) *J Chem Phys* 118:7762–7774.
21. Vreede J, Crielaard W, Hellingwerf KJ, Bolhuis PG (2005) Predicting the signaling state of photoactive yellow protein. *Biophys J* 88:3525–3535.
22. Vreede J, Hellingwerf KJ, Bolhuis PG (2007) Helix formation is a dynamical bottleneck in the recovery reaction of photoactive yellow protein. *Proteins Struct Func Bioinf* 72:136–149.
23. E W, Vanden-Eijnden E (2006) Towards a theory of transition paths. *J Stat Phys* 123:503–523.
24. Metzer P, Schütte C, Vanden-Eijnder E (2006) Illustration of transition path theory on a collection of simple examples. *J Chem Phys* 125:084110.
25. Bolhuis PG, Dellago C, Chandler D (2000) Reaction coordinates of biomolecular isomerization. *Proc Natl Acad Sci USA* 97:5877–5882.
26. Ma A, Dinner A (2005) Automatic method for identifying reaction coordinates in complex systems. *J Phys Chem B* 109:6769-6779.
27. Hummer G (2004) From transition paths to transition states and rate coefficients. *J Chem Phys* 120:516–523.
28. Best RB, Hummer G (2005) Reaction coordinates and rates from transition paths. *Proc Natl Acad Sci USA* 102:6732–6737.
29. Peters B, Trout BL (2006) Obtaining reaction coordinates by likelihood maximization. *J Chem Phys* 125:054108.
30. Peters B, Beckham GT, Trout BL (2007) Extensions to the likelihood maximization approach for finding reaction coordinates. *J Chem Phys* 127:034109.

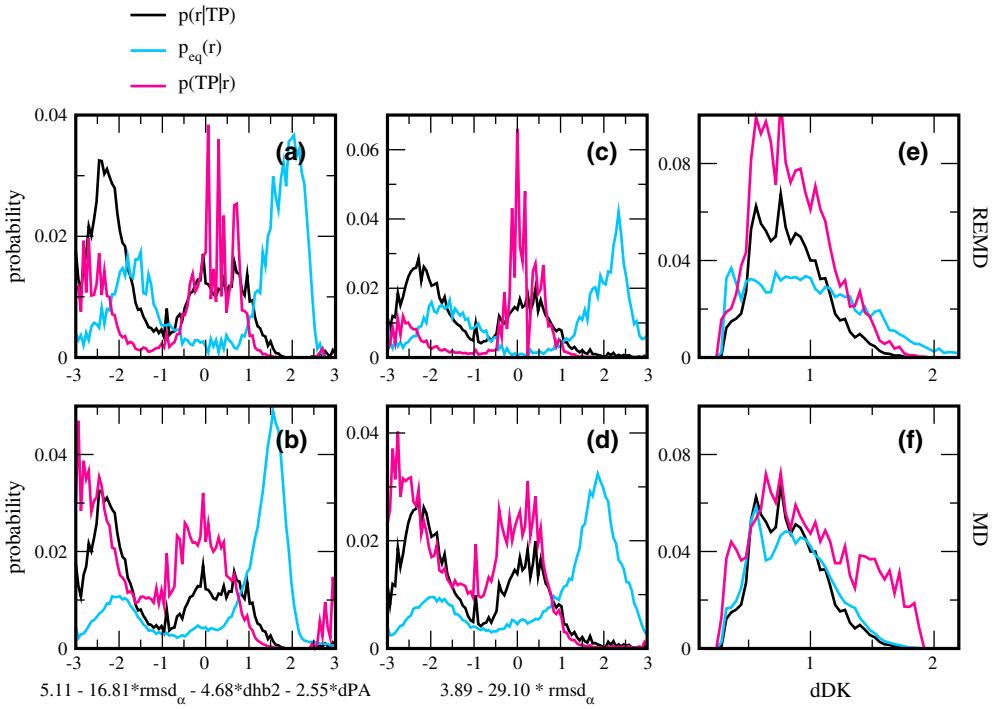

Fig. S1. Best–Hummer analysis of the pB'-I_α transition. The equilibrium distribution p_{eq} is obtained from either replica exchange simulations (*a*, *c*, and *e*) or MD simulations of the pB' state (*b*, *d*, and *f*). The probability distribution of TPs $p(r|\mathrm{TP})$ is obtained from the path ensemble of the pB'-I_α TPS simulation. The distributions $p(\mathrm{TP}|r)$ result from applying Eq. **S1** and are scaled arbitrarily for clarity. We performed this analysis for three order parameters: (*a* and *b*) the reaction coordinate resulting from the LM analysis, (*c* and *d*) the rmsd_α, and (*e* and *f*) the distance between Asp20 and Lys55 (*dDK*).

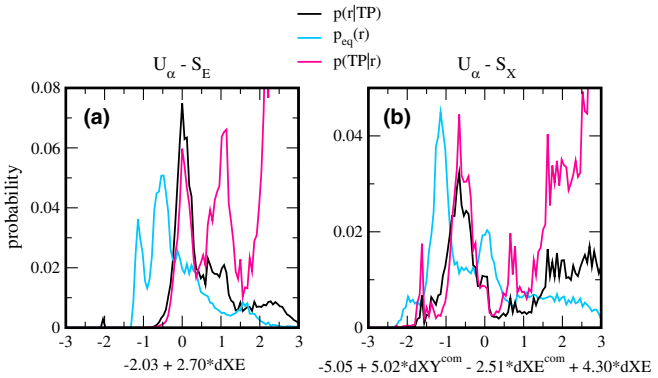

Fig. S2. Best–Hummer analysis of the solvent exposure of Glu46 and pCA. The equilibrium distribution p_{eq} is obtained from replica exchange simulations. The probability distribution of TPs $p(r\mathrm{TP})$ is obtained from the path ensemble of (*a*) the U_α-S_E and (*b*) the U_α-S_X TPS simulations. The distributions $p(\mathrm{TP}|r)$ result from applying Eq. **S1** and are scaled arbitrarily for clarity. We performed this analysis for the reaction coordinates found with the LM analysis.

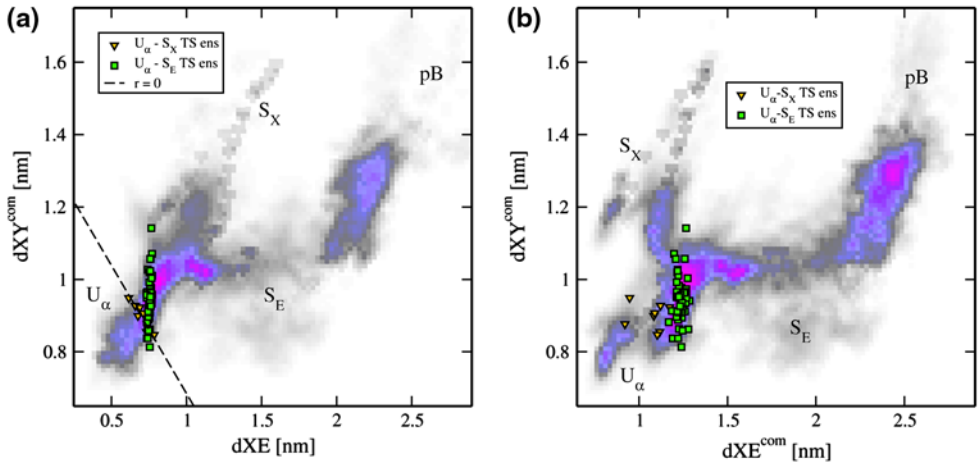

Fig. S3. Path density and transition state ensembles of the solvent exposure of Glu46 and pCA. The path density is plotted in (*a*) the (dXY^{com}, dXE) plane and (*b*) the (dXY^{com}, dXE^{com}) plane. The yellow triangles and green squares indicate the transition state ensembles for, respectively, the U_α-S_X and the U_α-S_E transitions. The dashed line represents the dividing surface $r = 0$ for the solvent exposure of pCA. Note that in both cases the dividing surface is rather close to the initial state U_α.

Table S1. List of order parameters

Number of waters around	
pCA	nw_X
Tyr42	nw_Y
Glu46	nw_E
Distance between atoms	
pCA-O4′ – Tyr42-OH	*dXY*
pCA-O4′ – Glu46-CD	*dXE*
pCA-O1 – Cys69-N	*dOaC*
Glu46-CD – Thr50-OG1	*dET*
Glu46-CD – Tyr42-OH	*dYE*
Arg52-CZ – Asp97-CG	*dRD*
Lys64-NZ – Thr70-OG1	*dKT*
pCA-O1 – Asp97-N	*dOaN*
pCA-O4′ – Ile49-N	*dXI*
pCA-O4′ – Thr50-N	*dXT*
pCA-O4′ – Arg52-N	*dXR*
pCA-O4′ – Asp97-N	*dXN*1
pCA-O4′ – Asp97-CG	*dXN*2
Ala44-N – Pro54-CG	*dPA*
Gly47-CA – Arg52-O	*dGR*
Glu46-CD – Asn43-ND2	*dEN*
Glu46-CD – Gly51-N	*dEG*
Asn43-O – Gly47-H	*dhb*1
Ala44-O – Asp48-H	*dhb*2
Ala45-O – Ile49-H	*dhb*3
Glu46-O – Thr50-H	*dhb*4
Gly47-O – Gly51-H	*dhb*5
Asp20-CG – Lys55-NZ	*dDK*
Asp24-CG – Lys55-NZ	*dDK*2
Glu9-CD – Lys110-NZ	*dEK*
Glu12-CD – Lys110-NZ	*dEK*2
K111-NZ – Glu116-CD	*dKE*
Distance between center of mass of side chains	
pCA – Tyr42	dXY^{com}
pCA – Glu46	dXE^{com}
pCA – Phe62	$dXF1^{com}$
pCA – Phe96	$dXF2^{com}$
pCA – Ile49	dXI^{com}
Lys64 – Thr70	dKT^{com}
Distance between center of mass of groups of residues	
13–17 – 114–116	*dN* – loop
35–37 – 98–101	*d*loops1
35–37 – 114–116	*d*loops2
rmsd	
11–15	$rmsd_{N1}$
19–23	$rmsd_{N2}$
43–51	$rmsd_{\alpha}$
62–68	$rmsd_{C1}$
75–86	$rmsd_{C2}$
111–116	$rmsd_{loop}$
Dihedral angles in pCA	
N-CA-CB-SG	dih_{CACB}
CA-CB-SG-C1	dih_{CBSG}
Other	
Number of hydrogen bonds in α3	*nhb*
Cosines of dihedral angles ϕ in α3	$\phi_{42} - \phi_{53}$
Cosines of dihedral angles ψ in α3	$\psi_{42} - \psi_{53}$

The order parameters are sorted according to their type. Atom names used in the description of the distances between atoms follow PDB nomenclature. All water molecules within a radius of 3.5 Å around a residue are counted for the *nw* order parameters. For the d^{com} order parameters involving pCA, only the center of mass of the heavy atoms in the phenol group is used. For the rmsd order parameters, the rmsd of the indicated region is calculated with respect to the crystal structure of the receptor state by using C^{α} positions.

Table S2. Stable state definitions

pB'-I_α—unfolding of helix $\alpha 3$

OP	$pB'^{\min}$	$pB'^{\max}$	$I_\alpha^{\min}$	$I_\alpha^{\max}$
nhb	4	5	0	1
rmsd_α	0	0.10	0.23	$+\infty$

U_α-S_E—exposure of Glu46

OP	$U_\alpha^{\min}$	$U_\alpha^{\max}$	$S_E^{\min}$	$S_E^{\max}$
nw_X	0	3	5	$+\infty$
dXY^{com}	–	–	1.0	$+\infty$
dXY	0	0.35	–	–
dYE	0	0.40	0.85	$+\infty$
dXE	0	0.60	–	–
dET	0	0.45	–	–
nw_E	–	–	10	$+\infty$

U_α-S_X—exposure of pCA

OP	$U_\alpha^{\min}$	$U_\alpha^{\max}$	$S_X^{\min}$	$S_X^{\max}$
nw_X	0	3	10	$+\infty$
dXY	0	0.35	0.9	$+\infty$
dYE	0	0.4	–	–
dXE	0	0.45	0.9	$+\infty$
dET	0	0.45	–	–

S_E-*pB*—exposure of pCA after exposure of Glu46

OP	$S_E^{\min}$	$S_E^{\max}$	$pB^{\min}$	$pB^{\max}$
nw_X	0	5	10	$+\infty$
dXY^{com}	0	1	–	–
dXY	–	–	0.9	3.5
dYE	0.85	3.0	0.85	3.0
nw_E	10	$+\infty$	10	$+\infty$

Order parameters (OP) defining the upper (max) and lower (min) boundaries of the stable states for the four TPS simulations. All distances are given in nanometers. A dash demarks order parameters that are not used to define a state. $+\infty$ indicates that there is no upper boundary; all values above a certain threshold are included as belonging to the state.