Predicting the reaction coordinates of millisecond light-induced conformational changes in photoactive yellow protein

Jocelyne Vreede[1], Jarek Juraszek, and Peter G. Bolhuis

Van't Hoff Institute for Molecular Sciences, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV, Amsterdam, The Netherlands

Understanding the dynamics of large-scale conformational changes in proteins still poses a challenge for molecular simulations. We employ transition path sampling of explicit solvent molecular dynamics trajectories to obtain atomistic insight in the reaction network of the millisecond timescale partial unfolding transition in the photocycle of the bacterial sensor photoactive yellow protein. Likelihood maximization analysis predicts the best model for the reaction coordinates of each substep as well as tentative transition states, without further simulation. We find that the unfolding of the α-helical region 43–51 is followed by sequential solvent exposure of both Glu46 and the chromophore. Which of these two residues is exposed first is correlated with the presence of a salt bridge that is part of the N-terminal domain. Additional molecular dynamics simulations indicate that the exposure of the chromophore does not result in a productive pathway. We discuss several possibilities for experimental validation of these predictions. Our results open the way for studying millisecond conformational changes in other medium-sized (signaling) proteins.

molecular dynamics | protein unfolding mechanism | transition path sampling | transition state ensemble | committor analysis

Investigation of the dynamics of large conformational changes in proteins leads to a better understanding of their functioning. Molecular simulation can complement experiments by modeling the dynamical time evolution of biomolecular systems in atomistic detail. Yet, even though straightforward all-atom molecular dynamics (MD) can now access the microsecond regime, simulating the kinetics of folding and other large conformational changes in proteins at ambient conditions has been limited to relatively small (model) proteins (1). Here we show how the transition path sampling technique enables investigation of the atomistic dynamical pathways for the millisecond timescale partial unfolding transition in a biologically relevant signaling protein, the photoactive yellow protein (PYP). Moreover, analysis of the ensembles of unbiased molecular trajectories allows the extraction of the (possibly complex) reaction coordinates.

PYP is a water soluble blue-light photo receptor from *Halorhodospira halophila* (2, 3) composed of 125 amino acids and a covalently bound chromophore (para-coumaric acid, pCA) (4). The protein folds into a Per–Arnt–Sim core capped by an N-terminal domain (5, 6). Upon the absorption of a blue-light photon, PYP undergoes many different conformational rearrangements, culminating in the formation of the signaling state. The first steps in signaling state formation are the isomerization of pCA, on a picosecond timescale (7), followed by a proton transfer, taking place within microseconds (8, 9). These reactions result in an intermediate pB' with pCA in a strained configuration and a negative charge on a buried glutamate (Glu46) (10, 11). As a consequence, the protein partially unfolds to expose pCA and Glu46 to bulk water (12–14), completing the formation of the signaling state pB.

PYP is an excellent model signaling protein because it is easily accessible for the application of a wide variety of experimental techniques, and hence, many of its characteristics have been elucidated. For example, the initial events of the photocycle have become much clearer with the use of ultrafast spectroscopy techniques (15) and time-resolved x-ray crystallography (16). Still, revealing the structural and dynamical details of the poorly understood mechanism of the submillisecond unfolding process at high resolution, e.g., by time-resolved NMR, remains elusive (13). As stated above, molecular simulation can provide atomistic detail for protein conformational change, complementary to experiments (17). For instance, a recent study using QM/MM showed that proton transfer can only occur when the negative charge on Glu46 is sufficiently stabilized (18). Molecular simulation also provided insights into the formation of the signaling state pB. Replica exchange MD (REMD) simulations indicated that the formation of pB involves the unfolding of the α-helical region 43–51 and the solvent exposure of Glu46 and pCA (14). The predicted signaling state showed remarkable agreement with the NMR structure of the pB state of the N-terminally truncated mutant Δ_{25}-PYP (13, 19).

Notwithstanding the success of the signaling state prediction, the room temperature kinetic mechanism by which pB is formed remained hidden due to the nature of the REMD sampling method. In this work we aim to provide mechanistic details at atomistic resolution of the submillisecond unfolding process of PYP by molecular simulation. However, while the conformational transitions occur too fast for high-resolution experimental methods, they are too slow to occur spontaneously in fully atomistic MD simulations. These long timescales originate from relatively high free-energy barriers between (meta)stable states, hampering efficient sampling of conformational space when using conventional MD. Moreover, protein conformational changes in general can occur via a myriad number of pathways, and hence, observation of a single event will not be conclusive. To avoid the long waiting times inherent in conventional MD of activated processes, many rare event simulation techniques have been developed (20). Whereas these methods are well suited for computing free-energy barriers and other thermodynamic properties, they often fail to yield mechanistic insight at ambient conditions. Examples are the artificial acceleration of the dynamics by raising the temperature (e.g., REMD 21) and methods that bias the system toward the barrier region along a predefined reaction coordinate (RC). In the latter case, the RC can be complex, and a poor choice might lead to a wrong reaction mechanism, bad sampling statistics, and a poor estimate of the rate constants (22). The transition path sampling (TPS) (23) algorithm avoids both these drawbacks

by collecting an ensemble of short reactive trajectories (in our case, several ns) connecting a predefined initial and final state, without prior knowledge of the transition state region. In contrast, a straightforward MD trajectory would require many microseconds to achieve the same. The TPS method enables evaluation of the mechanism, the transition state ensemble, and the RC in terms of order parameters (22). A good RC must be able to predict the so-called committor (24). This can be tested by performing a committor analysis (22–24), which requires running multiple additional MD simulations from putative transition state configurations. More efficient committor analysis algorithms include a GNN approach (25), a Bayesian path statistical analysis (26), and a likelihood maximization (LM) method (27, 28). The latter approach has the advantage of requiring only the results from a TPS simulation.

We employ TPS simulations to investigate the mechanisms underlying the transition from the folded pB' state to the unfolded signaling state pB of PYP (see Fig. 1). Starting in the intermediate state pB', instead of in the receptor state, allows us to use classical force fields, thus avoiding prohibitively costly quantum chemical computations. Any TPS simulation requires definitions for the initial and final states and an initial trajectory connecting these states, which can be provided by, for instance, REMD simulations (14, 19). Starting from the pB' state, we sampled four transitions to end up in the pB state (see Fig. 1). Our findings indicate that productive signaling state formation starts with losing α-helical content in the chromophore binding pocket, followed by solvent exposure of Glu46, and subsequent exposure of pCA. The state with only pCA solvent exposed is an off-pathway intermediate. Whether pCA or Glu46 become solvent exposed first correlates with the presence of a salt bridge between Asp20 an Lys55.

This work shows that TPS enables the sampling of millisecond timescale processes at room temperature and at atomistic detail. Additional LM analysis can predict relevant reaction coordinates for each conformational rearrangement in PYP. For the loss of α-helical content, we find that the best RC involves the overall deviation from the folded helix, but also a helical hydrogen bond and the distance between Pro54 and Ala44. RCs of the exposure of pCA and Glu46 involve the breaking of hydrogen bond interactions in the chromophore binding pocket, between pCA, Glu46 and Tyr42. The reaction coordinates we found for the various transitions do not directly involve (bulk) water order parameters. Nevertheless, bulk water provides additional stabilization of the negatively charged Glu46 and the hydrophilic chromophore, leading to their solvent exposure. Indeed, the formation of pB is slowed down and dramatically reduced at lowered humidity (29).

Our findings provide previously undescribed insights in the mechanism of light-induced unfolding in PYP, which potentially can be tested by experiments. Probing the existence of the predicted off-pathway intermediate experimentally requires sensitive time-resolved measurements that specifically target either pCA or Glu46, and preferably both. By isotopically labeling pCA and/or Glu46, vibrational shifts, due to changes in the local environment of these residues, can be monitored using FTIR (30) or resonance Raman spectroscopy (31). Furthermore, the fluorescence of the single tryptophan in PYP is quenched by energy transfer to pCA and strongly reflects changes in orientation of either pCA or Trp119 (32). Other techniques, such as time-resolved THz spectroscopy (33) and time-resolved optical rotatory dispersion (34), might be sensitive enough to detect the intermediates. The prediction that a salt bridge is correlated with the unfolding route can also be addressed experimentally. Recently, a time-resolved diffusion study suggested that the N-terminal region, including Asp20, loses α-helical content after the proton transfer (35). Varying the ionic strength of the environment of PYP has a profound effect on the formation of pB, as shown by time-resolved absorption and fluorescence spectroscopy (36). These observations agree well with our findings that the favored unfolding pathway starts with the disruption of the Asp20-Lys55 ionic bond. Furthermore, future mutagenesis experiments altering the properties of this salt bridge could test our predictions. Changing Asp20 and Lys55 to alanine could lead to an accelerated photocycle, although we believe this effect to be minor, because other residues can form a similar ionic linkage (36). A promising experiment would be to modify the ionic bond to a disulphide bridge, by mutating Asp20 and Lys55 into cysteine. This would lead to less light-induced unfolding and significant retardation of the photocycle.

In addition to the experiments, future simulation studies might focus on computing the rate constant by for instance transition interface sampling (37) or milestoning (38). Also, the detection of important transitions might be automated, e.g., by means of applying Markov State Model approaches (39).

Results and Discussion

TPS Simulations. Previous REMD simulations showed that the light-induced conformational change of PYP involves the unfolding of helix α3 (region 43–51), followed by the solvent exposure of pCA and Glu46. Also, free-energy profiles from these simulations suggested the presence of metastable states occurring during the formation of pB (19). Fig. 1B displays such a free-energy profile as a function of the distance dXE, between pCA (phenolic oxygen O4′) and Glu46 (atom CD) and $rmsd_\alpha$, the RMSD of helix α3, with respect to the conformation in the NMR structure (19). The figure shows several minima that represent the various (intermediate) states visited by the protein during the unfolding process. The label pB' indicates the fully folded state; I_α indicates a
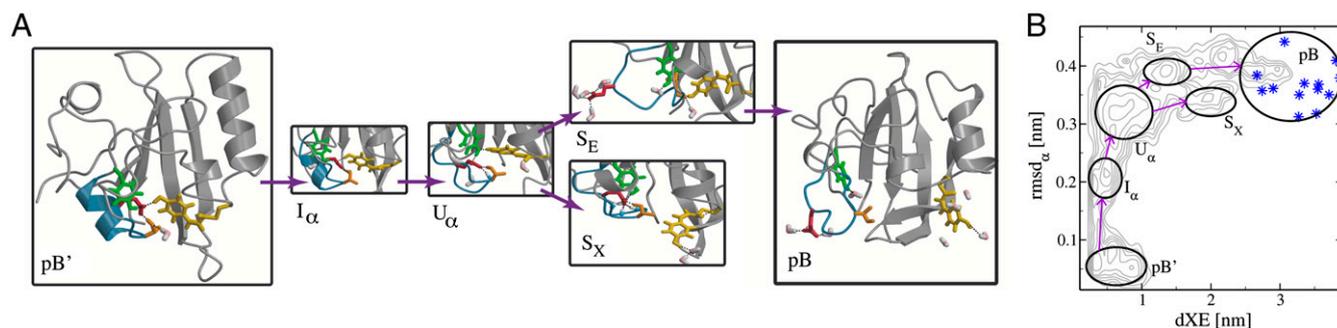

Fig. 1. The reaction network of the pB' to pB transition. (A) Structures of metastable states in the pB' to pB transition (see main text for a description). The helix α3 is rendered blue, and the stick models represent relevant residues: pCA (yellow), Glu46 (red), Tyr42 (green) and Thr50 (orange). Black dotted lines indicate hydrogen bonds. Water molecules are shown as pink and white stick models. (B) Free-energy profile projected in the $dXE - rmsd_\alpha$ plane. The labeled ellipses roughly indicate the location of the metastable states. The gray contour lines indicate the free-energy profile obtained from a previous REMD simulation, separated by $1k_BT$ (19). The blue stars indicate the conformations obtained by time-resolved NMR on a mutant of PYP (13).

state in which the helical conformation of helix α3 has partially disappeared. U_α denotes a state in which this region is completely unfolded, while Glu46 and pCA are still inside the protein. The labels S_E and S_X correspond to states with a solvent exposed Glu46 and pCA, respectively. Finally, pB indicates the signaling state, with both pCA and Glu46 exposed to solvent. The REMD simulations provided the stable state definitions (see Table S2).

Our first attempt to sample the transition from pB' to pB directly with TPS was severely hampered by the presence of the metastable intermediate states I_α, U_α, S_E, and S_X, which caused very long trajectories (i.e., much longer than 10 ns). Therefore, we split up the unfolding transition of pB' to pB into four separate transitions. These processes are: 1) the loss of α-helical structure in helix α3 ($pB' - I_\alpha$); 2) solvent exposure of Glu46 ($U_\alpha - S_E$); 3) solvent exposure of pCA ($U_\alpha - S_X$); and 4) solvent exposure of pCA following Glu46 ($S_E - pB$). We refrained from performing a TPS simulation of the ($S_X - pB$) transition, as this route turned out to be not productive.

We performed TPS simulations of these four transitions, using the definitions for the initial and final states listed in Table S2 and taking initial trajectories from either REMD simulations (19) or high temperature MD trajectories. Table 1 lists the sampling statistics for the four TPS simulations. The average path length indicates that the molecular timescale for the transitions is on the order of ns. Note that the actual reaction times are much longer.

Unfolding of Helix α3. An initial TPS simulation of the $pB' - U_\alpha$ transition yielded very long pathways due to the presence of intermediate state I_α. Therefore, we performed a TPS simulation of the $pB' - I_\alpha$ transition. Figs. 2A and B show representative pathways, plotted onto the path density (see *SI Text*) from both the $pB' - U_\alpha$ and the $pB' - I_\alpha$ TPS simulations. Once the protein is in intermediate state I_α it relaxes within nanoseconds to the U_α state in which the helix is completely unfolded, indicating that the barrier separating I_α and U_α is lower than the barrier between pB' and I_α.

For PYP, we have identified, based on visual inspection of the pathways and on chemical intuition, about 75 order parameters that could be involved in the signaling state formation (see Table S1). Only a fraction of these are used for the stable state definitions (see Table S2). We computed these order parameters for the entire shooting point ensemble of the $pB' - I_\alpha$ TPS simulation. Application of the LM analysis yielded the linear combination of up to three order parameters that best describe the RC (see Table 2). When considering only single order parameters for the $pB' - I_\alpha$ transition the best model for the RC is $rmsd_\alpha$. Adding more order parameters to the RC model is significant only if the log-likelihood ($\ln L$) increases by at least $\delta L_{min} = \frac{1}{2}\ln(N)$ (28), with N the number of shooting points. For this simulation $N = 4149$ and hence, $\delta L_{min} = 4.17$. Including the number of water molecules around Tyr42, nw_Y improves $\ln L$ by a value of $\delta L = 19$. Using three order parameters, $\ln L$ improves significantly by $\delta L = 13$ when nw_Y is replaced by $dhb2$ and dPA, respectively the length of the backbone hydrogen bond between Ala44 and Asp48 and

the distance between Pro54 and Ala44. Application of the Bayesian path statistical analysis (26) confirms these findings (see Figs. S1 and S2).

The description for the optimal RC obtained with the LM analysis agrees with visual inspection of the $pB' - I_\alpha$ pathways. In pB' Tyr42 is completely shielded from bulk water by Ala44 and Pro54. During the unfolding of helix α3, water molecules form hydrogen bonds to Tyr42, facilitated by Ala44 and Pro54 moving away from each other. This hydration is transient, as dPA decreases again when helix α3 is completely unfolded (see Fig. 2A). dPA is not the only relevant order parameter, because the LM analysis also identified the $dhb2$ as an important contribution to the RC. Interestingly, including the helical hydrogen bonds between Asn43 and Gly47 ($dhb1$), and between Ala45 and Ile49 ($dhb3$) also increased the $\ln L$ significantly but not as much as the Ala44-Asp48 hydrogen bond. These three hydrogen bonds are located at the solvent accessible side of helix α3. The remaining two helical hydrogen bonds, located at the protein side of helix α3, do not contribute at all to the RC.

The LM analysis predicts that structures with $r = 0$ are transition states. From the $pB' - I_\alpha$ path ensemble we extracted the conformations with the 3-variable RC (see Table 2) within the interval $r \in [-0.05, 0.05]$. This TS ensemble is also plotted in Fig. 2A and B, together with the predicted $r = 0$ dividing surface. Visual inspection suggested that this ensemble is quite uniform. Fig. 2E displays a typical configuration from this predicted TS ensemble. The relevant order parameters that contributed most to the RC are highlighted. Helix α3 is clearly still in a helical conformation, indicating that the TS is very close to the folded state. We explicitly computed the committors of the predicted $r = 0$ structures (see *SI Text*), and confirmed that 30% of the predicted structures are true transition states (with committor values between 0.3–0.7). The other conformations had committors close to either zero or unity. We refrained from performing a full committor analysis due to excessive computational costs. The TS ensemble for the $pB' - I_\alpha$ transition is thus very sensitive to small changes in the protein conformation, indicating it involves a relatively narrow barrier.

Solvent Exposure of Glu46 and pCA. Once the helix is unfolded and the protein is in the U_α state, the solvent exposure of both Glu46 and pCA constitutes the next step in the formation of pB. Trajectories from the REMD simulations (19) showed the solvent exposure of either Glu46 or pCA only, subsequently followed by the solvent exposure of the other residue. Using REMD trajectories as initial pathways, we performed two TPS simulations, one for the solvent exposure of Glu46 ($U_\alpha - S_E$) and one for the exposure of pCA ($U_\alpha - S_X$). Definitions of the initial and final states are listed in Table S2. Both transitions proved difficult to sample, as the processes of exposing pCA or Glu46 involve long and diffusive paths. Figs. 2C and D show representative, uncorrelated pathways for the $U_\alpha - S_E$ and the $U_\alpha - S_X$ transitions, plotted onto the path density. Note that these room temperature pathways are fully decorrelated from the initial REMD trajectories.

The $U_\alpha - S_E$ transition. LM analysis of the $U_\alpha - S_E$ shooting point ensemble reveals that the best RC for the exposure of Glu46 to the solvent is dXE, the distance between Glu46 and pCA (see Table 2). As pCA is an important hydrogen bond donor, this distance parameter indicates that disruption of the hydrogen bond network around Glu46 inside the protein leads to Glu46's solvent exposure. Adding more order parameters does not significantly improve $\ln L$, although adding the backbone torsions of Ala45 and the distance between Glu9 and Lys110 almost reached the threshold. The latter parameter is a salt bridge that is involved in the formation of pB (36). For the best RC model $r = dXE$, the TS ensemble is plotted as squares in Fig. 2C and D. Fig. 2F shows one of these transition states. This prediction agrees with

Table 1. Statistics of the TPS ensembles. The average path length is a weighted average over the whole ensemble. Decorrelated pathways have lost the memory of the previous decorrelated pathway. The aggregate time is the ensemble aggregate length

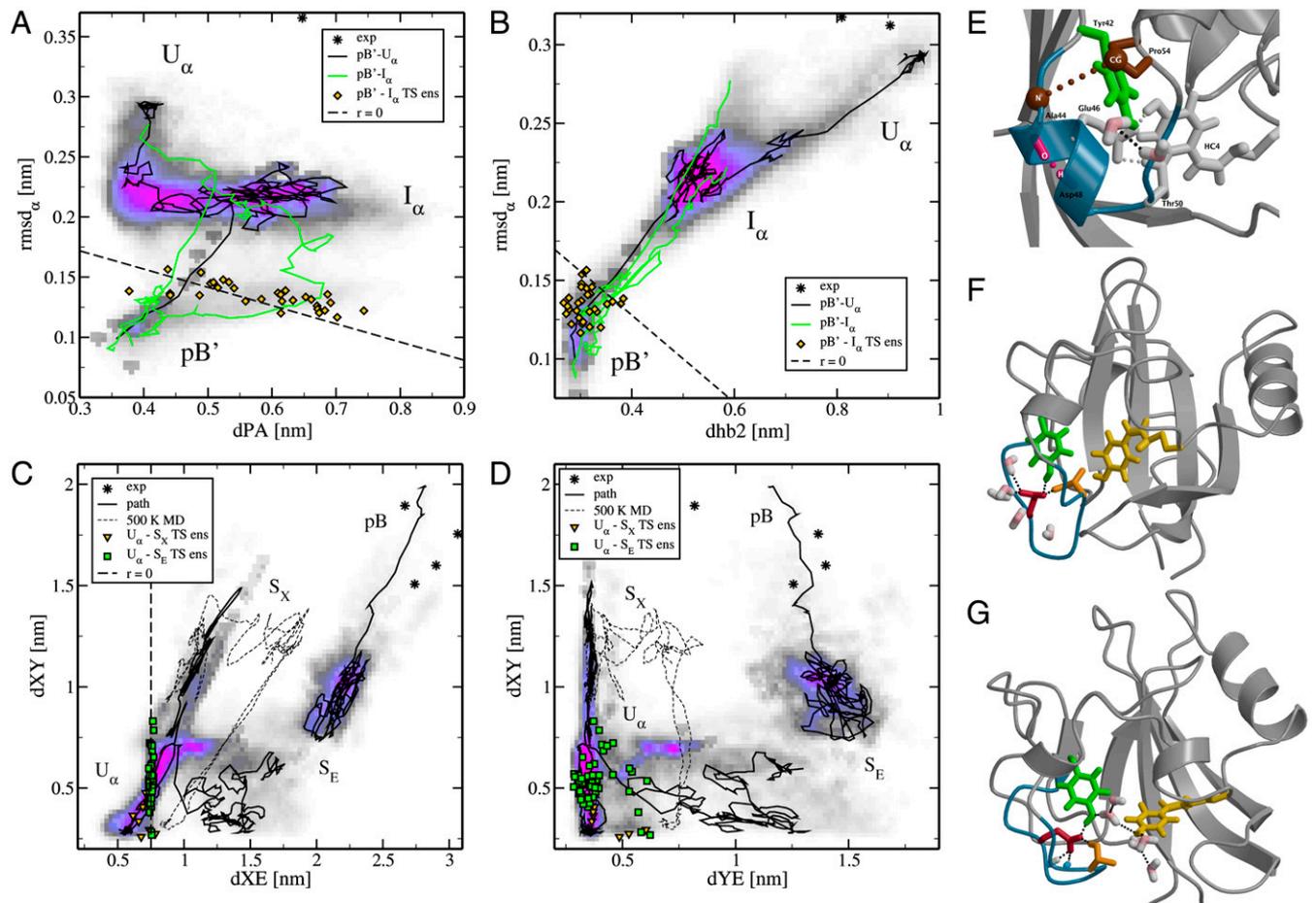|	$pB' - I_\alpha$	$U_\alpha - S_E$	$U_\alpha - S_X$	$S_E - pB$
acceptance	41%	25%	38%	44%
avg. path length	105 ps	1.8 ns	1.5 ns	1.7 ns
accepted paths	3847	305	584	311
decorr. paths	180	18	7	29
aggregate time (μs)	1.0	2.3	2.3	1.2

Fig. 2. Path density plots and transition state estimates of the *pB'* to *pB* transition. (*A* and *B*) Path density and representative pathways showing the unfolding of helix α3 in two representations, (*A*) the (*dPA*, *rmsd*$_\alpha$) plane and (*B*) the (*dhb2*, *rmsd*$_\alpha$) plane. The three pathways shown originate from the *pB'* − *U*$_\alpha$ (black) and the *pB'* − *I*$_\alpha$ (green) TPS simulations. The pathways are plotted onto the path density (see *SI Text*) from both TPS simulations. Yellow diamonds indicate the transition state ensemble. (*C* and *D*) Path density and representative pathways showing the solvent exposure of Glu46 and pCA in (*C*) the (*dXE*, *dXY*) plane and (*D*) the (*dYE*, *dXY*) plane. The dashed path is taken from a high temperature MD simulation of the *S*$_X$ state at 500 K. The yellow triangles and the green squares indicate the transition state ensembles for *U*$_\alpha$ − *S*$_X$ and *U*$_\alpha$ − *S*$_E$, respectively. The path density ranges from white (no paths) to magenta (many paths) via gray and blue. The labels indicate the different states. Experimental data is shown as stars. The dashed line shows the predicted dividing surface *r* = 0. (*E*) TS estimate of the *pB'* − *I*$_\alpha$ transition. Helix α3 is shown in blue, hydrogen bond 2 is depicted as a pink stick model. Atoms of Pro-54 and Ala-44 used for the calculation of *dPA* are shown as brown spheres. Other residues making up the chromophore binding pocket are shown as light gray sticks. (*F* and *G*) TS estimates of solvent exposure of (*F*) Glu46 and (*G*) pCA. Representation is as described in Fig. 1. In (*G*) two backbone N-H groups are also shown as blue-white stick models.

visual inspection of the pathways, suggesting that the main characteristic of the transition is the shift in the hydrogen bond network of the acidic side-chain of Glu46. Here, the water molecules do not enter the protein, but rather Glu46 moves to the solvent. On the *U*$_\alpha$ side of the barrier Glu46 interacts with Tyr42, Thr50, and the protein backbone. On the *S*$_E$ side water molecules provide stabilization for Glu46. The TS conformation in Fig. 2*F* shows a balance between these two hydrogen bonds networks, in which pCA does not participate.

The *U*$_\alpha$ − *S*$_X$ transition. LM analysis of the *U*$_\alpha$ − *S*$_X$ simulation shows that *dXY*^{com}, the distance between Tyr42 and pCA, is the single order parameter with the highest ln *L*. This RC indicates that the strength of the hydrogen bond between pCA and Tyr42 governs the solvent exposure of pCA. Allowing for multiple order parameters in the LM analysis reveals that including *dXE* significantly improves the log-likelihood. Subsequently adding *dXE*^{com}, the center of mass distance between pCA and Glu46 to the RC is barely deemed significant (see Table 2). The transition state ensemble is plotted as triangles in Fig. 2*C* and *D* and in Fig. S3. Fig. 2*G* displays one of the predicted con-

formations of the TS ensemble for the pCA solvent exposure, showing a partially solvated pCA with a water-mediated hydrogen bond between Tyr42 and pCA. Glu46 is still located inside the protein and does not interact with water molecules at all but instead forms hydrogen bonds to N-H groups in the protein backbone to stabilize its negative charge. Visual inspection of the *U*$_\alpha$ − *S*$_X$ TPS trajectories show Tyr42 and Glu46 interacting, while pCA detaches from this hydrogen bond network and becomes (partially) solvated.

Final Unfolding Step Toward *pB*. Our results so far seem to suggest the existence of two possible pathways, both leading from the *U*$_\alpha$ state to the unfolded signaling state *pB*, each with its own intermediate state, *S*$_E$ and *S*$_X$. However, the final state *pB* has both Glu46 and pCA solvent exposed. To test how accessible *pB* is from the intermediates, *S*$_E$ and *S*$_X$, we performed 10 ns of conventional MD starting from conformations from both *S*$_E$ and *S*$_X$ at ambient and elevated temperatures. At 300 K no additional solvent exposure events occur, indicating that both *S*$_E$ and *S*$_X$ are sufficiently metastable to prevent a fast escape from these states. To reach *pB* thus involves crossing another barrier. We

Table 2. LM analysis of the TPS ensembles sampled in three TPS simulations. The minimal improvement of likelihood required to accept an additional order parameter in the RC is indicated for each analysis as δL_{min}. All distances are given in nm.

n	ln L	reaction coordinate
$pB' - I_\alpha$ transition $\delta L_{min} = 4.17$		
1	−2117	$3.89 - 29.10 \times rmsd_\alpha$
2	−2098	$3.88 - 26.35 \times rmsd_\alpha - 0.19 \times nw_{Y42}$
3	−2085	$5.11 - 16.81 \times rmsd_\alpha - 4.68 \times dhb2 - 2.55 \times dPA*$
$U_\alpha - S_E$ transition $\delta L_{min} = 3.06$		
1	−280	$-2.03 + 2.70 \times dXE*$
2	−277	$-3.94 + 2.88 \times dXE + 1.50 \times dEK$
3	−275	$-4.00 + 2.81 \times dXE + 1.55 \times dEK + 0.40 \times \cos(\phi_{45})$
$U_\alpha - S_X$ transition $\delta L_{min} = 3.33$		
1	−396	$-6.40 + 6.92 \times dXY^{com}$
2	−388	$-5.92 + 4.25 \times dXY^{com} + 3.01 \times dXE$
3	−383	$-5.05 + 5.02 \times dXY^{com} - 2.51 \times dXE^{com} + 4.30 \times dXE*$

*RC used to model the committor.

performed MD simulations starting from the same conformations at higher temperatures (425, 450, 475, and 500 K). One 500 K simulation samples the solvent exposure of Glu46 from the S_X state (indicated as a dashed curve in Fig. 2CD), but instead of reaching pB, pCA goes back into the protein, followed by solvent exposure of Glu46. Because we do not observe the $S_X - pB$ transition even at high temperature we label S_X a "dead end." In contrast, starting from the S_E state, solvent exposure of pCA occurs already at 425 K. Thus, it is easier for pCA to leave the protein once Glu46 is already solvent exposed than vice versa. Using this unfolding MD trajectory as an initial path, we performed a TPS simulation of the $S_E - pB$ transition (see Table 1). This transition proved to be less difficult to sample than the $U_\alpha - S_X$ transition. The representative $S_E - pB$ pathway plotted in Figs. 2C and D indeed reaches pB and the end point overlaps well with the experimental data (plotted as stars) (13). Examination of the pathways reveals that in the barrier region pCA is hydrated and still inside the protein. The chromophore also forms hydrogen bonds with the side chains of Asp97 and Arg52. Experimental studies have indicated that Arg52 and Tyr98 may act as gating residues, facilitating hydration of the chromophore binding pocket (40).

The Most Likely Pathway is Correlated with a Salt Bridge. We have shown that when Glu46 exposes to the solvent first, formation of the signaling state will follow, whereas if pCA goes out first, the protein enters a dead-end route. What triggers the protein to choose either the successful or the unproductive pathway? Comparison of the two TS estimates for $U_\alpha - S_E$ and $U_\alpha - S_X$ (Figs. 2F and G) reveals that the region 43–51 has a different conformation depending on whether Glu46 is exposed. Searching for interac-

tions that may influence the conformation of the α3 region, we found a salt bridge between Asp20 and Lys55, which also exists in the pB' and receptor states. This ionic bond is absent in the $U_\alpha - S_E$ TS but is present in the $U_\alpha - S_X$ TS. To find out whether this salt bridge is broken in U_α we extended several of the TPS trajectories in the backward time direction. Starting from four U_α endpoints from $U_\alpha - S_E$ paths and four U_α endpoints from $U_\alpha - S_X$ trajectories, we performed 10 ns MD simulations, denoted toS_E and toS_X, respectively. Fig. 3A shows a 2-D free-energy plot (negative logarithm of the probability) as a function of $rmsd_\alpha$ and dDK, the salt-bridge distance between CG of Asp20 and NZ of Lys55. While this plot does not represent a converged free-energy landscape, it clearly shows that U_α splits into two substates with different dDK distance. In the toS_X simulations, $dDK = 0.5$ nm, indicating that Asp20 and Lys55 have a strong electrostatic interaction. In contrast, dDK ranges from 0.9 nm to over 2.5 nm in the toS_E simulations, indicating that the salt bridge is broken and unlikely to reform. Figs. 3C and D show configurations with an open and a closed salt bridge. These observations suggest that the absence of the DK salt bridge leads to the Glu46 exposure, while its presence results in solvent exposure of the pCA first. As Asp20 is located in the N-terminal domain of PYP, the conformation of this domain is correlated to the presence of the salt bridge, and thus to whether the protein will choose a dead end or a productive pathway.

To determine when exactly the salt bridge breaks, we also plotted the TPS path density of the $pB' - I_\alpha$ transition in Fig. 3A. While the path density on the pB' side shows already signs of two substates, the path density on the I_α end exhibits an even clearer picture, with peaks corresponding to a formed salt bridge at
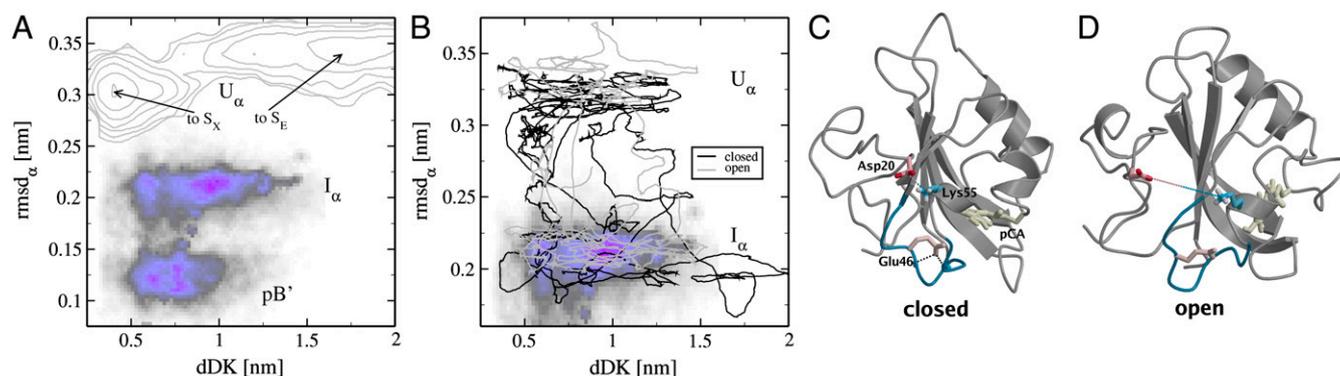

Fig. 3. Characterization of the Asp20-Lys55 salt bridge. (A and B) Path density of the $pB' - I_\alpha$ transition plotted in the $(dDK, rmsd_\alpha)$ plane. (A) The gray lines show a free-energy profile obtained from eight MD simulations of the U_α state at 300 K. (B) The black and gray lines show 100 ps smoothed MD trajectories initiated from the I_α state. (C and D) Structures from U_α with (C) a small and (D) a large value for dDK. Residues Asp20, Lys55, Glu46 and pCA are shown as stick models.

$dDK = 0.6$ nm and broken salt bridge with $dDK = 9$ nm and $dDK = 1.1$ nm. We performed 32 MD simulations of 15 ns at 300 K of the I_α state. Starting conformations were taken from end points in the $pB' - I_\alpha$ TPS simulation. Half of these had a value below 0.6 nm for dDK ("closed") and the other half had a value higher than 0.9 nm for dDK ("open"). Eight systems reached the U_α state within 15 ns, four of which started with an open salt bridge and four with a closed salt bridge. Fig. 3B shows the (100 ps smoothed) trajectories of the simulations that crossed the low $I_\alpha - U_\alpha$ barrier. All these simulations sample both open and closed conformations for the salt bridge, suggesting that fluctuations in U_α determine which unfolding pathway the protein will take.

Methods
Preparation of the system and all MD simulations were performed with the GROMACS 3.3.1 package (41), employing the Gromos96 43a1 force field (42, 43) in combination with the SPC water model (44). Transition path sampling (22, 23) carried out an unbiased sampling of MD trajectories that connect predefined initial and final ensembles of states, in the current case the different metastable states in PYP as derived from previous REMD simulations (19). TPS is a random walk through trajectory space, where new trajectories are generated from old trajectories by a shooting algorithm and accepted with a Metropolis rule (22, 23). Here, we use the one-way, flexible path length TPS algorithm as previously implemented and applied to protein systems (37). The likelihood maximization (LM) analysis (27, 28) screens many linear combinations of order parameters as candidate RCs and predicts the one that best describes the data. Required as input for this procedure is the ensemble of shooting point configurations belonging to accepted trajectories and the rejected shooting points. The Bayesian path statistical analysis (26) is used to confirm the predicted reaction coordinates. In addition, we computed the committor for certain configurations by starting many MD trajectories initialized with random momenta and measuring the fraction of paths that reaches the final state. Details on these methods can be found in *SI Text*.

1. Snow CD, Sorin EJ, Rhee YM, Pande VS (2005) How well can simulation predict protein folding kinetics and thermodynamics?. *Annu Rev Biophys Biomol Struct* 34:43–69.
2. Meyer TE, Tollin G, Hazzard JH, Cusanovich MA (1989) Photoactive Yellow Protein from the purple phototrophic bacterium, Ectothiorhodospira halophila—Quantum yield of photobleaching and effects of temperature, alcohols, glycerol, and sucrose on kinetics of photobleaching and recovery. *Biophys J* 56:559–564.
3. Sprenger WW, Hoff WD, Armitage JP, Hellingwerf KJ (1993) The eubacterium Ectothiorhodospira halophila is negatively phototactic, with a wavelength dependence that fits the absorption-spectrum of the Photoactive Yellow Protein. *J Bacteriol* 175:3096–3104.
4. van Beeumen JJ, et al. (1993) Primary structure of a Photoactive Yellow Protein from the phototrophic bacterium Ectothiorhodospira halophila, with evidence for the mass and the binding-site of the chromophore. *Protein Sci* 2:1114–1125.
5. Borgstahl GE, Williams DR, Getzoff ED (1995) 14 ångstrom structure of Photoactive Yellow Protein, a cytosolic photoreceptor—Unusual fold, active-site, and chromophore. *Biochemistry* 34:6278–6287.
6. Pellequer JL, Wager-Smith KA, Kay SA, Getzoff ED (1998) Photoactive Yellow Protein: A structural prototype for the three-dimensional fold of the PAS domain super-family. *Proc Natl Acad Sci USA* 95:5884–5890.
7. Gensch T, et al. (2002) The primary photoreaction of photoactive yellow protein (PYP): Anisotropy changes and excitation wavelength dependence. *Chem Phys Lett* 356:347–354.
8. Meyer TE, Cusanovich MA, Tollin G (1993) Transient proton uptake and release is associated with the photocycle of the Photoactive Yellow Protein from the purple phototrophic bacterium Ectothiorhodospira halophila. *Arch Biochem Biophys* 306:515–517.
9. Hendriks J, Hoff WD, Crielaard W, Hellingwerf KJ (1999) Protonation/deprotonation reactions triggered by photoactivation of Photoactive Yellow Protein from Ectothiorhodospira halophila. *J Biol Chem* 274:17655–17660.
10. Hendriks J, van Stokkum IH, Hellingwerf KJ (2003) Deuterium isotope effects in the photocycle transitions of the Photoactive Yellow Protein. *Biophys J* 84:1180–1191.
11. Pan D, Philip A, Hoff WD, Mathies RA (2004) Time-resolved resonance Raman structural studies of the pB' intermediate in the photocycle of Photoactive Yellow Protein. *Biophys J* 86:2374–2382.
12. Xie A, et al. (2001) Formation of a new buried charge drives a large-amplitude protein quake in photoreceptor activation. *Biochemistry* 40:1510–1517.
13. Bernard C, et al. (2005) The solution structure of a transient photocycle intermediate: Δ25 Photoactive Yellow Protein. *Structure* 13:953–962.
14. Vreede J, Crielaard W, Hellingwerf KJ, Bolhuis PG (2005) Predicting the signaling state of Photoactive Yellow Protein. *Biophys J* 88:3525–3535.
15. Groot ML, et al. (2003) Initial steps of signal generation in Photoactive Yellow Protein. *Biochemistry* 42:10054–10059.
16. Ihee H, et al. (2005) Visualizing reaction pathways in Photoactive Yellow Protein from nanoseconds to seconds. *Proc Natl Acad Sci USA* 102:7145–7150.
17. Groenhof G, et al. (2004) Photoactivation of the Photoactive Yellow Protein: Why photon absorption triggers a trans-to-cis isomerization of the chromophore in the protein. *J Am Chem Soc* 126:4228–4233.
18. Leenders EJM, et al. (2007) Protonation of the chromophore in the Photoactive Yellow Protein. *J Phys Chem B* 111:3765–3774.
19. 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.
20. Frenkel D, Smit B (2002) *Understanding molecular simulation* (Academic Press, San Diego, CA), 2nd Ed.
21. Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 314:141–151.
22. Dellago C, Bolhuis PG, Geissler PL (2002) Transition path sampling. *Adv Chem Phys* 123:1–78.
23. Dellago C, Bolhuis PG, Csajka FS, Chandler D (1998) Transition path sampling and the calculation of rate constants. *J Chem Phys* 108:1964–1977.
24. E W, Vanden-Eijnden E (2006) Towards a theory of transition paths. *J Stat Phys* 123:503–523.
25. Ma A, Dinner A (2005) Automatic method for identifying reaction coordinates in complex systems. *J Phys Chem B* 109:6769–6779.
26. Best RB, Hummer G (2005) Reaction coordinates and rates from transition paths. *Proc Natl Acad Sci USA* 102:6732–6737.
27. Peters B, Trout BL (2006) Obtaining reaction coordinates by likelihood maximization approach for finding reaction coordinates. *J Chem Phys* 125:054108.
28. Peters B, Beckham GT, Trout BL (2007) Extensions to the likelihood maximization approach for finding reaction coordinates. *J Chem Phys* 127:034109.
29. van der Horst MA, van Stokkum IHM, Dencher NA, Hellingwerf KJ (2005) Controlled reduction of the humidity induces a shortcut recovery reaction in the photocycle of Photoactive Yellow Protein. *Biochemistry* 44:9160–9167.
30. Nie B, Stutzman J, Xie A (2005) A vibrational spectral marker for probind the hydrogen-bonding status of protonated Asp and Glu residues. *Biophys J* 88:2833–2847.
31. El-Mashtoly SF, et al. (2005) Structural changes during the photocyce of Photoactive Yellow Protein monitored by ultraviolet resonance Raman spetra of tyrosine and tryptophan. *J Phys Chem B* 109:23666–23673.
32. Hoersch D, Otto H, Cusanovich MA, Heyn MP (2008) Distinguishing chromophore structures of photocycle intermediates of the photoreceptor PYP by transient fluores-cence and energy transfer. *J Phys Chem B* 112:9118–9125.
33. Castro-Camus E, Johnston MB (2008) Conformational changes of Photoactive Yellow Protein monitored by terahertz spectroscopy. *Chem Phys Lett* 455:289–292.
34. Chen E, et al. (2003) Dynamics of protein and chromophore structural changes in the photocycle of Photoactive Yellow Protein monitored by time-resolved optical rotatory dispersion. *Biochemistry* 42:2062–2071.
35. Hoshihara Y, Imamoto Y, Kataoka M, Tokunaga F, Terazima M (2008) Conformational changes in the N-terminal region of Photoactive Yellow Protein: A time-resolved diffusion study. *Biophys J* 94:2187–2193.
36. Hoersch D, et al. (2007) Role of a conserved salt bridge between the PAS core and the N-terminal domain in the activation of the photoreceptor Photoactive Yellow Protein. *Biophys J* 93:1687–1699.
37. Juraszek J, Bolhuis PG (2008) Rate constant and reaction coordinate of Trp-cage folding in explicit water. *Biophys J* 95:4246–4257.
38. Vanden-Eijnden E, Venturoli M (2009) Markovian milestoning with Voronoi tessellations. *J Chem Phys* 130:194101.
39. Huang X, Bowman GR, Bacallado S, Pande VS (2009) Rapid equilibrium sampling initiated from nonequilibrium data. *Proc Natl Acad Sci USA* 106:19765–19769.
40. Genick UK, et al. (1997) Active site mutants implicate key residues for control of color and light cycle kinetics of Photoactive Yellow Protein. *Biochemistry* 36:8–14.
41. van der Spoel D, et al. (2005) GROMACS: Fast, flexible and free. *J Comput Chem* 26:1701–1718.
42. van Gunsteren W, et al. (1996) *Biomolecular simulation: The GROMOS96 manual and user guide* (Vdf Hochschulverlag AG an der ETH Zurich, Zurich).
43. 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.
44. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) *Interaction Forces*, ed Pullman B (D. Reidel Publishing Company, Dordrecht), pp 331–342.