A Favorable Path to Domain Separation in the Orange Carotenoid Protein

Mahmoud Sharawy [†], Natalia B. Pigni [‡§], Eric R. May [†], and José A. Gascón [‡*]*

[†]Department of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut 06269-3060, USA

[‡]Department of Chemistry, University of Connecticut, Storrs, Connecticut 06269-3060, USA

[§]Instituto de Ciencia y Tecnología de Alimentos Córdoba (ICYTAC-CONICET), Ciudad Universitaria, Córdoba X5000HUA, Argentina

Corresponding Authors

*Eric R. May (eric.may@uconn.edu). *José A. Gascón (jose.gascon@uconn.edu)

ABSTRACT

The Orange Carotenoid Protein (OCP) is responsible for nonphotochemical quenching (NPQ) in cyanobacteria, a defense mechanism against potentially damaging effects of excess light conditions. This soluble two-domain protein undergoes profound conformational changes upon photoactivation, involving translocation of the ketocarotenoid inside the cavity followed by domain separation. Domain separation is a critical step in the photocycle of OCP because it exposes the N-terminal domain (NTD) to perform quenching of the phycobilisomes. Many details regarding the mechanism and energetics of OCP domain separation remain unknown. In this work, we apply metadynamics to elucidate the protein rearrangements that lead to the active, domain-separated, form of OCP. We find that translocation of the ketocarotenoid canthaxanthin has a

profound effect on the energetic landscape and that domain separation only becomes favorable following translocation. We further explore, characterize, and validate the free energy surface (FES) using equilibrium simulations initiated at different states on the FES. Through pathway optimization methods, we characterize the most probable pathway to domain separation and reveal the barriers along that pathway. We find that the free energy barriers are relatively small (<5 kcal/mol), but the overall estimated kinetic rate is consistent with experimental measurements (>1 ms). Overall, our results provide detailed information on the requirement for canthaxanthin translocation to precede domain separation and an energetically feasible pathway to dissociation.

KEYWORDS. OCP, domain dissociation, nonphotochemical quenching, cyanobacteria, metadynamics

INTRODUCTION

Photosynthetic organisms have developed energy dissipation strategies to prevent damage from the potentially harmful effects of excess light conditions. The Orange Carotenoid Protein (OCP) is a soluble protein with a photoprotective function in cyanobacteria, responsible for nonphotochemical quenching (NPQ) after its activation by blue-green light.[1–4] This 35 kDa protein consists of two domains, which are connected by an unstructured linker.[5] OCP can bind different carotenoids, including 3'-hydroxyechinenone (3hECN), echinenone (ECN), canthaxanthin (CAN) and zeaxanthin (ZEA).[1] However, the presence of at least one keto group (i.e. 3hECN, ECN, and CAN) is essential for photoactivation.[6] The C-terminal domain (CTD) is a mixed α-helical/β-sheet structure known to act as a "sensor", while the N-terminal domain (NTD) has an all α-helical secondary structure, and it is known as the "effector" domain.[7,8] In addition, a short α–helix (αA) in the N-terminal extension (NTE) plays a stabilizing role through interactions with the CTD.[9]

Encapsulated between the two domains, the ketocarotenoid interacts with the protein environment through hydrogen bonds and hydrophobic contacts.

The activation mechanism from the inactive orange form (OCP^O) to its red form (OCP^R) involves a sequence of events including a 12 Å translocation of the carotenoid into the NTD,[10] and the separation of the two domains.[11,12] These rearrangements are essential to expose the domain-interface part of the NTD to interact with the phycobilisomes (PBS), the primary light-harvesting antennae complexes of cyanobacteria, in order to quench excess energy.[13–16] Beyond the desire to uncover the biophysical basis of OCP photoactivation, understanding the detailed complexity of this system could aid in the design of high-efficiency materials for solar energy capture and storage.

During the OCP^O photoactivation process, the initial events upon excitation occur in picoseconds at a very low quantum yield (~1%), leading to a first red-shifted photoproduct (P1) in which two key hydrogen bonds between the carotenoid and CTD residues Tyr201 and Trp288 are broken. Recent spectroscopy experiments have revealed a series of additional intermediate, red-shifted photoproducts (P2, P2', P3), in which the carotenoid is translocated into the NTD on a timescale of ~10 µs.[17] The next step of the OCP photocycle is the transition from P3 to the fully active OCP^R in which the separation of the two domains occurs on a slow (~1 ms) timescale.[17]

The only available full-length high-resolution structures of OCP are for the inactive state (OCP^O). While a detailed structural characterization of the dissociated (and other intermediate) state(s) is still elusive, recent X-ray radiolitic labelling analyses of solvent accessibility revealed a series of structural changes, from carotenoid migration to domain separation.[11] Small-angle

neutron and X-ray scattering experiments point to rearrangements leading to a flexible, 'molten globule' structure, of the fully activated OCP[R].[18–20]

The OCP photoactivation process can be divided in three main parts: 1) initial photoactivation of the ketocarotenoid that leads to disruption of key hydrogen bonds, 2) translocation of the ketocarotenoid, and 3) domain dissociation. We (Pigni and Gascón) have recently addressed the second segment of this process by performing translocation of the carotenoid canthaxanthin (CAN) via metadynamics[21] and obtained molecular models of the intermediate photoproducts.[22] This included obtaining a model for OCP-P3, where CAN has fully migrated into NTD, by matching its spectral signatures to experimental measurements. As stated above, spectroscopy studies suggest that carotenoid translocation (formation of OCP-P3) is relatively slow (~10 µs). However, separation of the domains (OCP-P3 to OCP[R]) occurs on a timescale multiple orders of magnitude slower than CAN translocation (> 1 ms).[17] Classical molecular dynamics simulations are well suited to explore configurations within a local energy minimum, but to sample across complex energy landscapes with energy barriers significantly exceeding $k_\mathrm{B}T$ requires extremely high computational costs and is unlikely to be a feasible approach to obtain meaningful (i.e. converged, statistically significant) results. Therefore, the application of enhanced sampling methodologies in order to drive the system toward configurations separated by high energy barriers along the surface, is an advisable approach.[23]

In the present work, we employ well-tempered metadynamics[24] (WT-metaD) to drive OCP domain separation. We explore domain separation from two alternative starting points: the crystal structure of the inactive form, OCP[O] (PDB: 4XB5), and our model of the intermediate OCP-P3.[22] Using both OCP[O] and OCP-P3 allows us to examine the effect of carotenoid translocation on the energetics of domain separation. A careful definition of two collective variables (CVs) led us to

4

achieve a converged, robust estimate of the free energy surface (FES), which we then utilize to identify the most probable pathway for dissociation and to present a putative model of the fully active OCP[R]. While we were conducting our study, a similar study by Bondanza *et al.* was recently reported which also utilized WT-metaD to study OCP domain separation.[25] However, our study stands in contrast to the work from Bondanza *et al.* in a few aspects, which are necessary conditions for a correct interpretation of the process of domain separation. Notably, (i) our WT-metaD results in a converged FES due to resampling of minima and is validated through equilibrium simulations, (ii) we show that dissociation is a thermodynamically favorable process *only* when the carotenoid is translocated, and (iii) we observe physically meaningful energy barriers (< 5 kcal/mol) along the dissociation pathway and the overall estimated kinetics are in the range of experimental measurements. Therefore, we present, for the first time, an energetically favorable path for domain separation in OCP.

RESULTS AND DISCUSSION

The NTD and CTD of OCP interact through a major and a minor interface (Figure 1a, b). The major interface is located along a plane approximately orthogonal to the long axis of the carotenoid. The minor interface is composed of interactions between the short α-helix (αA) in the NTE and the CTD β-sheet.[11] Although there is general agreement that carotenoid translocation precedes domain separation,[1,10,17] there is still some debate about the sequence of events that lead to domain separation at the major and minor interfaces.[11,19,26–28] Here we examine these questions via enhanced sampling and equilibrium MD simulations and provide a detailed analysis of the structural changes and the path to dissociation.

Free energy surfaces of domain separation in OCP^O and OCP-P3

To explore the dissociation of the OCP, we performed two-dimensional WT-metaD simulations. The two CVs were defined to represent the distance between domains at the minor interface (CV1) and at the major interface (CV2). A precise definition of the CVs is shown in Figure 1a, b.

To assess the effect of carotenoid translocation on domain separation, we performed the same metadynamics procedure starting from two alternative structures: OCP^O (the inactive OCP) and OCP-P3, a model of the intermediate photoproduct with the carotenoid fully translocated into the NTD (obtained in our previous work).[22] The resulting FESs are shown in Figure 1c, d. The substantial energetic differences between the FESs of OCP^O and OCP-P3 indicate that carotenoid translocation has a profound effect on domain dissociation energetics. Most notably, the global-minima are located at considerably different CV2 values suggesting that domain separation is thermodynamically favored in OCP-P3, and it is not favored in OCP^O (Figure 1c, d). We also projected the 2D FESs onto each CV (Eq. 2 in Methods) to generate 1D profiles (Figure 1e, f, S1). Projections onto CV2 indicate that the global minimum for OCP^O is CV2 ≈ 29 Å (Figure 1e), similar to the initial CV2 value (see Table 1 in Methods). In contrast, the FES for OCP-P3 has a broad minimum in CV2 ranging from ~36–40 Å, indicating a domain separation of ~10 Å (Figure 1f). Thus, our FES calculations clearly show that carotenoid translocation precedes domain separation during OCP photoactivation, consistent with UV-vis and IR spectroscopic experiments.[17] Note that, while the FES in OCP^O presents a global minimum at CV1 ~40 Å (Figure S1), rather than the starting conformation at ~10 Å, such minimum is kinetically inaccesible during the photoactivation cycle (see analysis of barriers below).

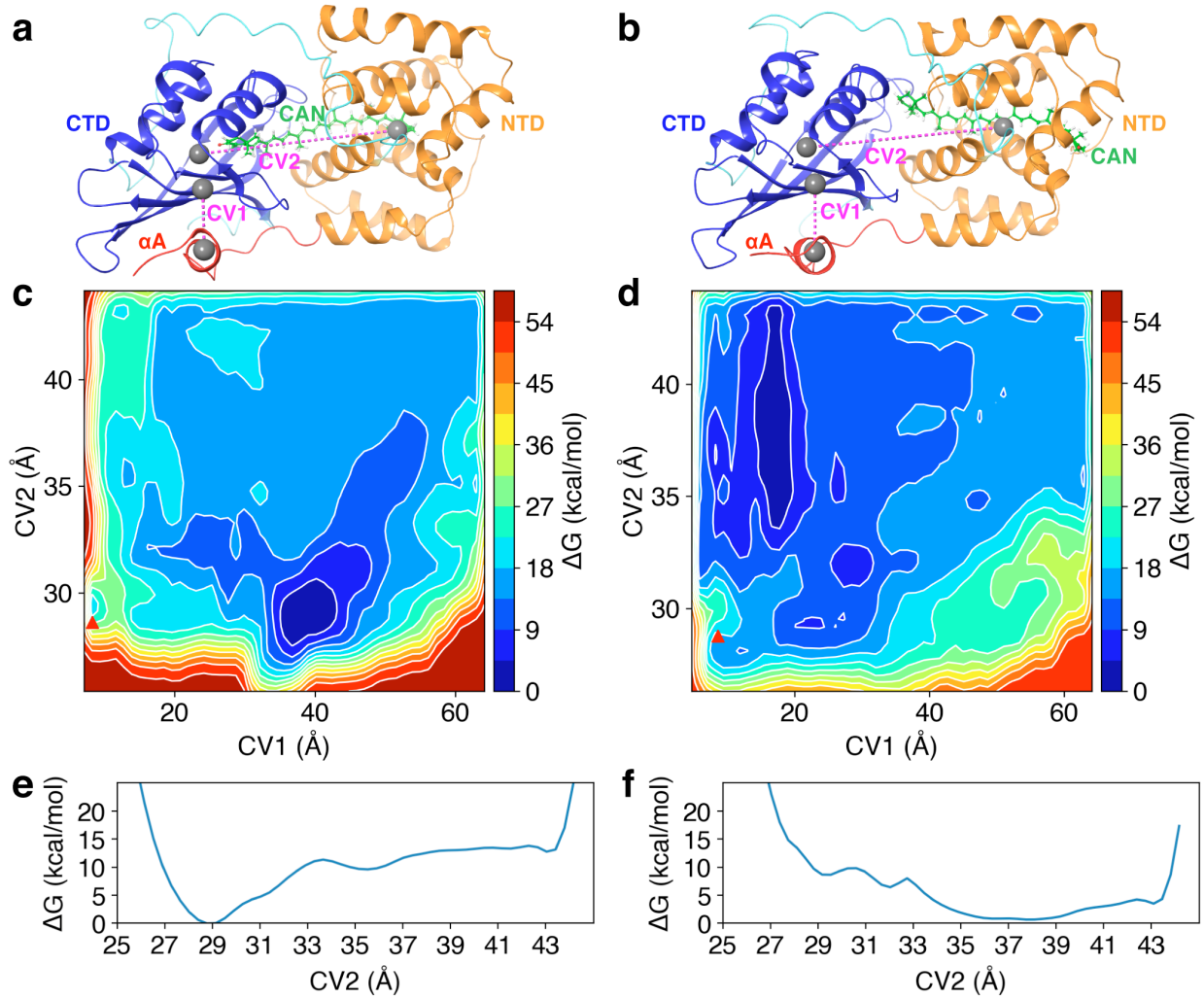Figure 1. FES results of domain separation starting from OCP^O (left panels) and OCP-P3 (right panels). (a, b) Molecular representations of OCP^O (a) and OCP-P3 (b). The CVs are depicted as the distances between the center of mass (COM, grey spheres) of the αA helix and that of the CTD β-strands (CV1), and the distance between the NTD and CTD COMs (CV2). (c, d) 2D FESs for OCP^O (c) and OCP-P3 (d). Triangles in (c) and (d) denote the initial (CV1, CV2) values. (e) Projection of the 2D FES in (c) onto CV2, and (f) projection of the 2D FES in (d) onto CV2. The 2D-to-1D projection is described by Eq. 2 in the Method section.

Recent work by Bondanza *et al*. has also studied OCP domain dissociation prior to and subsequent to carotenoid translocation into the NTD.[25] Although their study produced a domain

separated state, the authors explicitly noted that the relative free energy of the closed and open states could not be obtained from their simulations due to the lack of attainment of the ergodic regime. This lack of convergence resulted in unreasonably high energetic barriers for the dissociation process (~70–90 kcal/mol) even when the carotenoid is translocated. Moreover, regardless of the carotenoid position, their global free energy minimum was the domain-associated state. In marked contrast, the FESs obtained from our simulations exhibit convergence, state-recrossing events (Figure S1), physically reasonable energy barriers (< 5 kcal/mol), and correct prediction of the dissociated state as being thermodynamically favorable only after OCP-P3 formation.

Exploring the free-energy surface at equilibrium

The estimated FESs from WT-metaD can be highly sensitive to the simulation hyper-parameters, the CV choices, and the simulation times. Following our observation that domain separation is only favorable in OCP-P3, we performed a series of equilibrium MD simulations for selected starting states (A–F) from the FES of OCP-P3 (Figure 1d) to test its reliability. We chose low energy states, high energy states, and what we perceived as a saddle point. For each of these states, we performed four MD simulation replicas of 100 ns each. The equilibrium simulations for each state were combined, and their (CV1, CV2) values were projected onto the FES in Figure 2a–f. The distributions indicate that the configurational samplings from MD simulations reflect the FES's topology.

The MD simulations for state A (the associated state), which corresponds to the early stages of OCP-P3 formation, show a narrow distribution of CV1 and CV2 values, indicating that the system is trapped in a narrow well (Figure 2a). In contrast, simulations that started near what we perceive as a transition state on the FES (state B) produce wider sampling and show the system

"falling back" into state A (Figure 2b). When the input structure represents a configuration from a point after the transition state (state C), the simulations did not return to state A; instead, the simulations sampled configurations from the free-energy basin corresponding to the dissociated state (Figure 2c). The MD samplings for state D (Figure 2d) resemble those at state C in that both simulations sample configurations near the global minimum at the dissociate state (state D). Interestingly, when simulations are initiated from the flat region to the right (higher CV1 values)
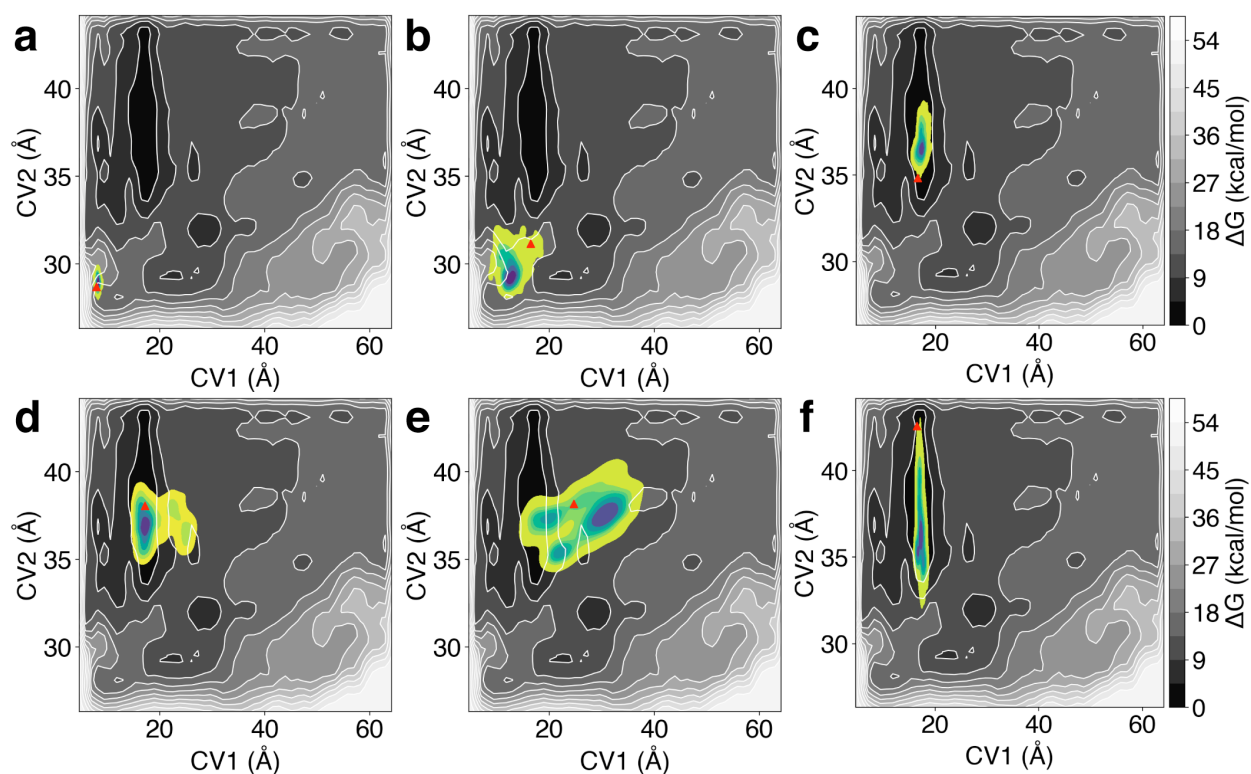

Figure 2. Probability densities of the CV values of the OCP-P3 domains separation. (a)–(f) Distributions of the CV1 and CV2 values obtained by unbiased equilibrium MD simulations. The distributions (colored) overlay the free energy surface of OCP-P3's domain separation (grey); dark colors denote relatively high probability densities. A red triangle denotes the coordinates (CV1, CV2) associated with the initial state used for the MD simulations. Each distribution was calculated from a set of four 100-ns simulations.

of the dissociated state basin, the equilibrium MD simulations sample a broad range of the configurational space (Figure 2e), reproducing the shallow, entropically favorable, region of the FES. Finally, a set of MD simulations were started from a structure near the upper limit of CV2 (Figure 2f) which produced a distribution that extends along the narrow valley of the dissociated state. This observation led us to confirm that the wall imposed on CV2 in WT-metaD (see Methods) does not affect the estimation of the FES since none of the equilibrium simulations extended beyond the wall location (the wall was not present in the equilibrium MDs).

Overall, these results show that equilibrium MD simulations replicate the features of the FES. Thus, our WT-metaD simulation produced a valid and accurate FES for estimating OCP domain dissociation energetics and dynamics.

Path to domain separation

After the validation of the FES obtained from our WT-metaD simulations, we can now address mechanistic questions such as: What is the most probable path on the FES that leads to domain separation in OCP? What are the structural features of the states along this path? To answer these questions, we estimated the most probable transition path (i.e., minimum free energy path, MFEP) between basins in the 2D FES using the Dijkstra algorithm.[29] We calculated the MFEP between the initial OCP-P3 state (state A) and the global minimum (state D). Our results indicate that the MFEP transits through three metastable states, M1, M2, and M3 (Figure 3). The MFEP shows that αA rearrangements (the minor interface, CV1) precede separation of the CTD-NTD (the major interface, CV2), as suggested by several experiments.[17,19,26,27]
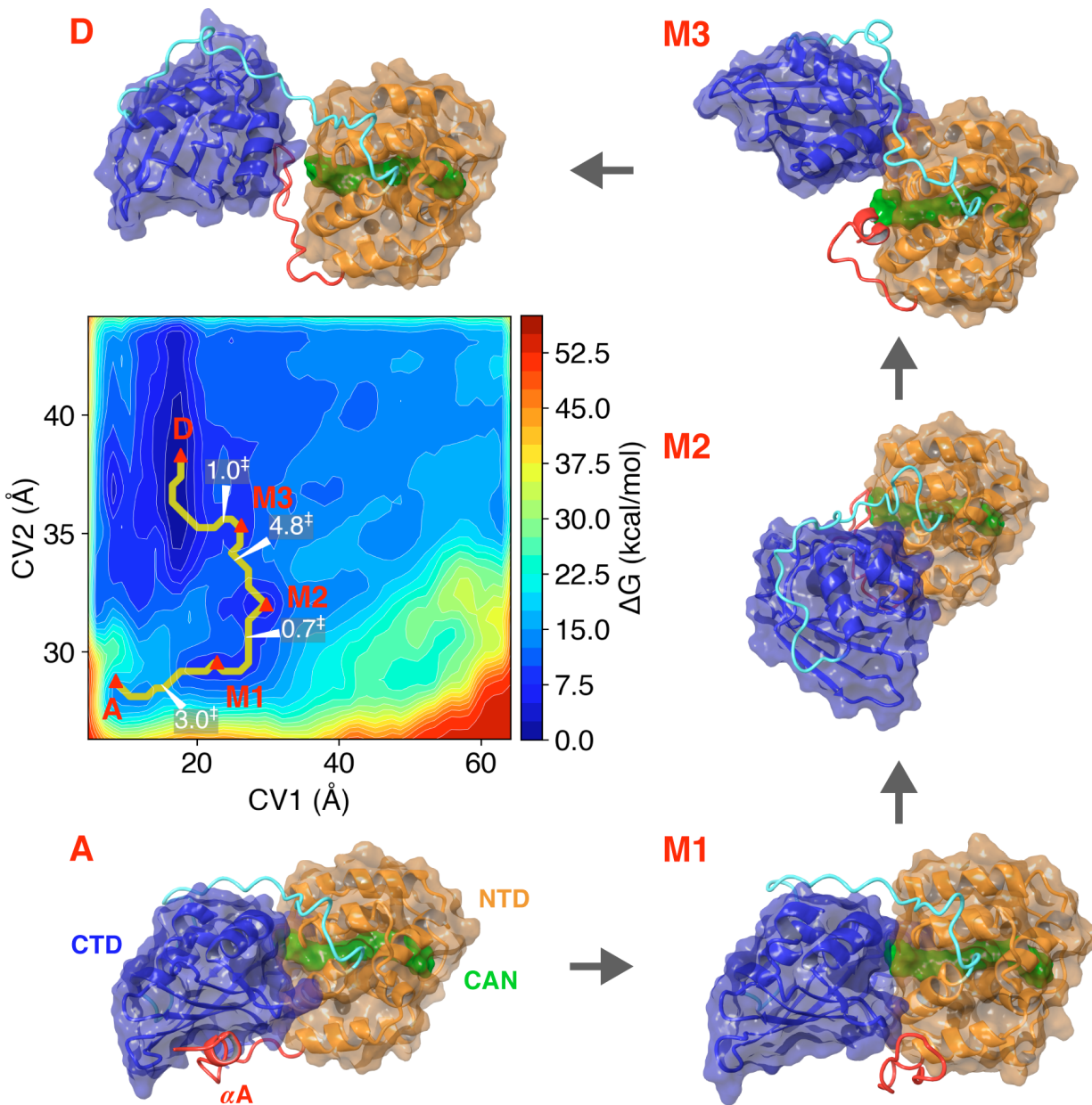
Figure 3. Minimum free energy path (MFEP) of the OCP-P3 domain dissociation. The molecular-surface representations correspond to the common configuration in the states labeled on the free energy surface. On the FES, the wedges denote free energy barriers in units of kcal/mol; the yellow line represents the MFEP.

The MFEP determined by the Dijkstra algorithm was independently verified by performing 100 Monte Carlo (MC) walks on the FES. Using Monte Carlo, we found that most paths followed

the Dijkstra MFEP (Figure S2). Of the 100 MC walks, 86 followed a "rightward" path of initially increasing CV1, consistent with the MFEP, while 14 walks followed a more direct path connecting A with D. It should be noted that of the 86 "rightward" paths, 35 walks did not reach the state D basin but remained trapped in an intermediate state (M2). Therefore, MC generated 65 "reactive" trajectories transited from state A to state D, and 78% of those paths followed the MFEP, while 22% followed the alternative path.

The changing interactions at the domain interfaces along the MFEP were visualized through contact map analysis of states A, M1, M2, M3 and D and are presented in Figure S3. The first transition along the MFEP from the initial state (A) involves a 3.0 kcal/mol barrier to reach M1, a state in which the NTD-CTD interface remains intact while the αA helix unfolds and separates from the CTD β-strands (Figure 3). This initial transition is primarily a change along CV1, representing dissociation at the minor interface. A small barrier of 0.7 kcal/mol separates M1 from M2, but drastic changes in the orientation of the CTD relative to the NTD occur in this transition as CV2 evolves toward higher values. The change in the contact maps (Figure S3) from state M1 to M2 is also dramatic, showing virtually all interactions present in the A and M1 states at the major interface are no longer present in M2. The largest barrier along the path (4.8 kcal/mol) separates M2 from M3. Contact map analysis of state M3 (Figure S3) shows interactions between the loop, linking helices αC and αD, and the CTD that are not observed in other states, as well as restoration of contacts between a helix aligned perpendicular to the major interface (αG) and the CTD. As well, during the M2 to M3 transition, the pivoting of the CTD exposes most of the NTD-ligand complex to the solvent. The final transition from M3 to state D requires crossing a small barrier (1.0 kcal/mol), reaching a state with a high CV2 value where the interactions between the NTD and the CTD at the major interface are disrupted.

To compare barrier heights on the FESs between OCP^O and OCP-P3, an MFEP analysis was performed on the OCP^O FES as well (Figure S4). The OCP^O MFEP shows that the energy barriers for dissociation are prohibitive, including a 9.4 kcal/mol barrier to dissociate the minor interface and then an 11.8 kcal/mol barrier along the major interface dissociation leg of the pathway (Figure S4). The major factor contributing to the alteration of the FESs and MFEP barriers between OCP^O and OCP-P3 is the nature in which the carotenoid is interacting with the OCP domains. In OCP^O, the interactions of the ligand are perfectly balanced between the NTD and CTD (Figure S5). In contrast, the translocation of the ligand in OCP-P3 lead to the ligand interactions being almost exclusively with the NTD (Figure S5), which indirectly decreases the barriers for dissociation. Thus, we view the carotenoid ligand as acting as a reinforcing "bolt" between the CTD and NTD in OCP^O, increasing the barriers for domain separation.

With a detailed characterization of the barriers at hand, we can now examine the kinetics of domain separation for OCP-P3. The MFEP for domain dissociation has two significant free energy barriers $\Delta G_1^{\ddagger} = 3.0$ kcal/mol and $\Delta G_2^{\ddagger} = 4.8$ kcal/mol (Figure 4). A simple kinetic analysis based upon a consecutive reaction scheme

$$ A \xrightarrow[\Delta G_1^{\ddagger}]{k_1} M2 \xrightarrow[\Delta G_2^{\ddagger}]{k_2} D, \qquad \text{Sch. 1} $$

results in the overall rate $k_{\mathrm{ov}} = k_1 k_2 / (k_1 + k_2)$. By applying transition state theory, we can estimate the rate constants from the free energy barriers as $k_i = \omega_i e^{-\Delta G_i^{\ddagger}/k_B T}$; and by assuming that the pre-exponential factors are equivalent ($\omega_1 = \omega_2 = \omega$), the overall rate is given by

$$k_{ov} = \omega \frac{e^{-\frac{\Delta G_1^{\ddagger} + \Delta G_2^{\ddagger}}{k_B T}}}{e^{-\frac{\Delta G_1^{\ddagger}}{k_B T}} + e^{-\frac{\Delta G_2^{\ddagger}}{k_B T}}} . \qquad \text{Eq. 1}$$

While the value of ω is well-defined for chemical reactions, it is considerably more ambiguous for

biomolecular configurational changes, where it represents a barrier crossing attempt frequency.

Protein folding literature indicates that $\omega \sim 10^6$ s^{-1} provides a reasonable general estimate.[30,31] Using

this estimate for ω, at a temperature of 300 K, the timescale for domain dissociation from OCP-

P3 ($\tau = 1/k_{ov}$) along the MFEP is $\tau \sim$3 ms. This estimate for domain dissociation occurring at a

millisecond timescale is in excellent agreement with a previous study using mid-IR spectroscopy.[17]
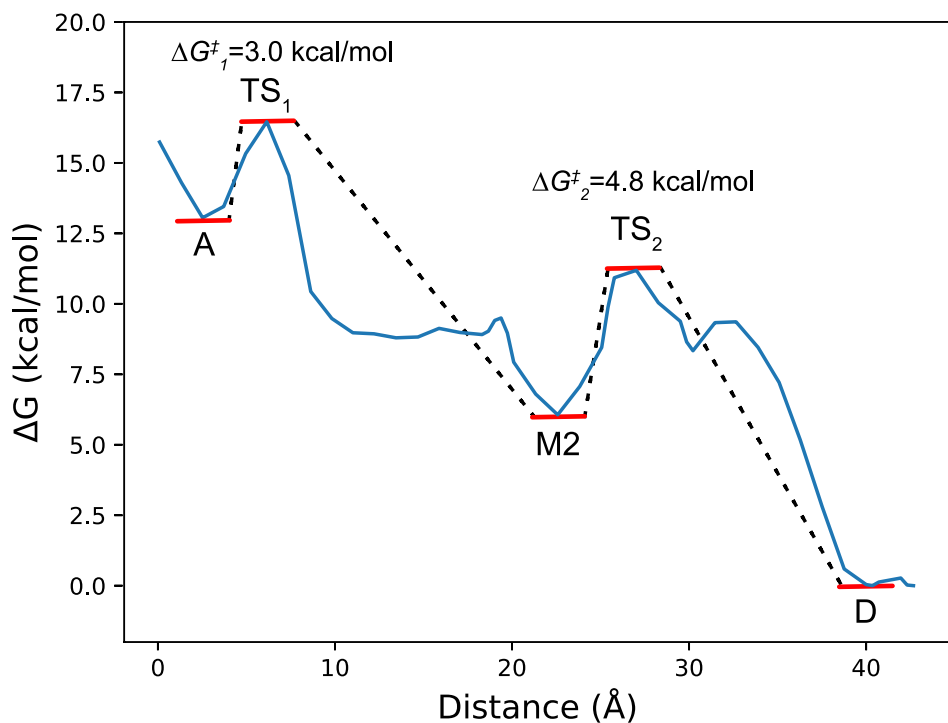

Figure 4. Free energy barriers along the MFEP of the OCP-P3 domain dissociation. A, the

associated state; M2, a metastable state; D, the dissociated state; TS$_1$ and TS$_2$, transition states.

The activated state (OCP[R])

Our proposed path from the OCP-P3 intermediate to a putative model of OCP[R] is in agreement with several experimental reports that show that the overall secondary structure of each domain is stable during the transition from OCP[O] to OCP[R].[10,19,27,28] In particular, Gupta *et al*. carried out circular dichroism (CD) studies that indicated only a small decrease (~2 %) in the α-helical content for OCP[R] compared to OCP[O], a change that is generally attributed to the unfolding of αA.[19] However, while the structures of the domains undergo minimal perturbations during activation, the interdomain interactions and relative orientation experience significant changes. Interdomain changes have been revealed by small-angle X-ray scattering (SAXS), showing a marked increase in the radius of gyration (R_g) for the activated form of OCP,[19] and chemical cross-linking studies have indicated dramatic changes in the relative orientation of the domains along the photoactivation process.[12] Our simulations are consistent with these experimental findings in that we observe drastic changes in the relative NTD-CTD orientation (e.g., domain pivoting). At the same time, the low intrinsic RMSD values (< 2.5 Å) of each domain along the entire WT-metaD simulation (Figure S6) indicate that the structural integrity of the domains is maintained throughout the path, consistent with CD measurements. Furthermore, despite the low RMSD values of individual domains during activation, our model of the activated state (state D) displays greater structural fluctuations compared with OCP[O] (Figure S7a). The average RMSF per residue in the entire OCP[O] is 0.84 Å; in contrast, the average in OCP[R] (state D) is 2.82 Å. This observation can be attributed to multiple factors, but the RMSF increase in OCP[R] is likely dominated by the dissociation of the domains. The activated state of OCP has been referred to as a molten globule-like structure,[20,32] indicating a lack of well-packed tertiary contacts and increased internal dynamics. In addition to the overall increased RMSF values of the activated state we also observe

increased internal dynamics in the CTD in OCP[R] compared to OCP[O] (Figure S7b). These higher internal domain RMSF values for the CTD may result from ligand translocation into the NTD, and correlate with the experimental observations of OCP[R] displaying increased dynamics.

According to our FEP results, the activated OCP can be characterized as a dynamic structure fluctuating among different orientations in solution. The finding of an extended global minimum in our FES, together with the wide distribution of configurations explored in the equilibrium MD simulations, particularly starting at advanced states of domain separation (Figure 2c-f), indicate that OCP[R] is highly flexible. Once the main interactions between the two domains are disrupted, the flexible nature of the linker (25–30 residues) sustains this enhanced conformational freedom. Indeed, from results of chemical cross-linking studies comparing the inactive and the active forms of OCP, Liu *et al.*[12] proposed that the loop between the two domains allow multiple orientations of NTD and CTD in solution. In turn, interactions with other proteins, such as the PBS or the fluorescence recovery protein (FRP), stabilize biologically relevant configurations *in vivo*.

A detailed analysis of the dissociated state can be obtained from the equilibrium MD simulations at the global minimum of OCP-P3 (state D). Figure 5a shows a representative configuration of state D, highlighting the presence of water between the separated domains at the major interface. In general, solvent molecules are known to play an important role in the OCP photoactivation process.[11,19,28] Once OCP domains are dissociated, the carotenoid and the internal face of the NTD are expected to be more exposed to solvent to interact with the PBS complex. Experimental evidence indicates a larger number of water molecules bound to the protein surface in OCP[R].[32] In our simulations, OCP[R] (represented by state D) shows increased SASA values for the protein and the ligand when compared to the "closed" state, OCP-P3 (state A). Figure 5c shows
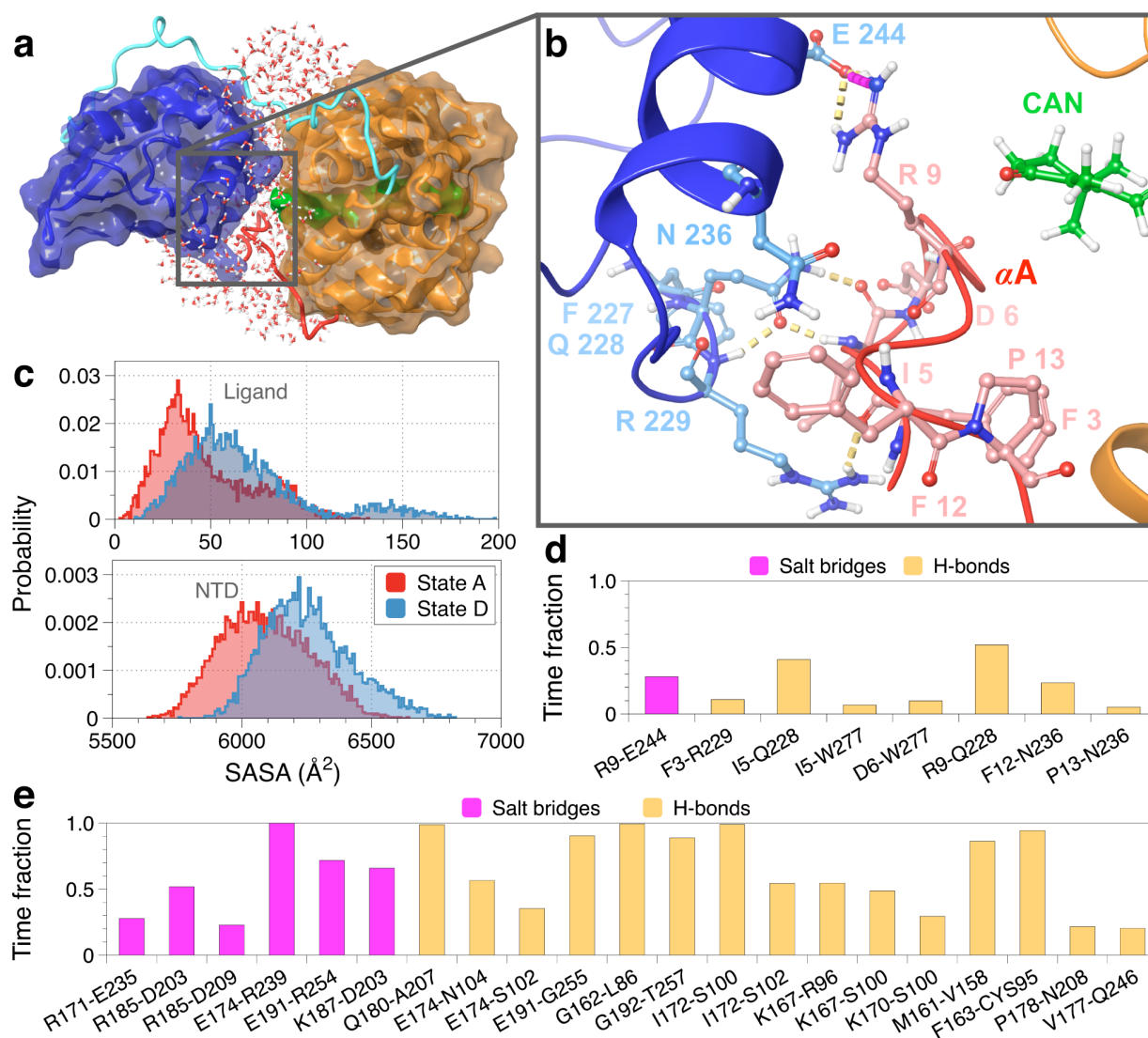
Figure 5. Interactions in OCP-state D and solvent-accessibility comparisons. (a) Molecular-surface representation of state D (dissociated state) and molecular representation of water between domains. (b) Molecular representation of the non-covalent interactions between the NTE and CTD in state D. (c) Probability distributions of the solvent-accessible surface area (SASA) for the ligand (CAN) and the NTD in state A (associated state) and state D. SASA values were calculated using a probe with a radius of 1.4 Å. Data for states A and D were sampled from two sets of MD simulations, each set comprising four 100-ns simulations. (d) Time fractions of the major-noncovalent interactions between the NTE and CTD; (e) between the linker and domains.

that state D reaches ligand SASA values of ~140–150 Å[2] during the equilibrium MD simulations. Interestingly, simulations from our previously published work,[22] revealed ligand SASA values of ~39 Å[2] for OCP[O], similar to the values observed here for OCP-P3. On the other hand, ligand SASA values for the red carotenoid protein (RCP), which consists only of the isolated NTD, were around 147 Å[2].[22] Altogether, these results suggest that our model of OCP[R] (state D) can reach levels of ligand solvent exposure similar to the isolated NTD during unbiased equilibrium MD simulations.

Finally, a closer look to the protein interactions in our model of OCP[R] shows that, besides the disruption of the secondary helical structure of the NTE at the early stages of domain separation observed in our results (M1, Figure 3), the NTE establishes close contacts with different parts of the protein along the pathway of domain dissociation. In particular, a network of hydrogen bonds maintains several NTE-CTD interactions in the activated state (Figure 5b). Some of these interactions are highly conserved during the MD simulations (Figure 5d). Frequently explored conformations of this dynamic state show the NTE interacting with residues at the major interface, as shown by the contacts of ARG9, which interacts with GLU244 through a salt bridge (~30% frequency) and with GLN228 through a hydrogen bond with an occurrence exceeding 50% (Figure 5d). In addition, the structural proximity between the loop and the domains is also observed in our model (Figure 5e). Interestingly, there is a general agreement between our findings and chemical cross-linking experiments, which have also demonstrated structural proximity between the NTE, the loop, and the CTD in the activated OCP[R].[12] Liu *et al*. have attributed a key role to the linker loop and the NTE, which act as a "structural buffer system" to facilitate the back and forth conversion between OCP[O] and OCP[R] during the photocycle.[12] The crucial role of the linker in the back-conversion has also been suggested by phylogenetic and mutagenesis analyses.[33] Thus, our

observations of the dynamic interactions between the loop and the NTE, keeping the proximity between the two domains, is in line with experimental observations.

CONCLUSIONS

Through the use of equilibrium MD and metadynamics simulations, we have revealed, for the first time, an energetically favorable path to domain separation in OCP. We find that the translocation of the ketocarotenoid canthaxanthin into the NTD is a prerequisite for a thermodynamically favorable domain dissociation pathway. Furthermore, our calculated free energy surfaces show that the translocation of the ketocarotenoid not only affects the location of the global minimum it also considerably decreases the energy barriers for domain dissociation. The energy barriers we observe along the minimum free energy pathway are surmountable at physiological temperatures, and kinetic analysis predicts a rate of domain dissociation of ~3 ms, in excellent agreement with experimental measurements. The most probable path to domain separation reveals that the detachment of the αA from the β-strands at the minor interface precedes the disruption of interactions at the major interface. In agreement with experimental data, we propose a dynamic and flexible model for OCP^R that can fluctuate among different conformations in solution. It is highly probable that the interaction with other proteins *in vivo* (i.e., PBS, FRP) could help to stabilize the biologically relevant conformations.

While our observations are well supported by agreements with known experimental data, this work also provides predictions about both the mechanism of dissociation and the molecular details of the activated state (OCP^R), which could be further tested experimentally. For example, we predict αA dissociation should precede domain separation; however, we do observe an alternative, less frequent, pathway which does not require αA dissociation. We propose that if the

αA-CTD interactions were stabilized through mutations or crosslinking, there would be a decrease in the rate of photoswitching under saturated light conditions. Additionally, we identify key interactions between αA and the CTD, as well as interactions between the domain linker and the domains. In particular, high probability salt bridges are identified (e.g., E174-R239, R9-E244), which could be disrupted through charge reversal mutations. We predict that disruption of these interactions would also affect the efficiency of the OCP photocycle. In summary, our work provides resolution of long standing questions regarding mechanistic details of OCP activation, and generates new avenues for experimental and theoretical exploration.

METHODS

System preparation

Two systems were constructed using Maestro (Schrödinger 2019-3) to model the OCP^O and OCP-P3 states, both containing the carotenoid ligand canthaxanthin. The protein and ligand coordinates for OCP^O were taken from the X-ray crystal structure of OCP^O (PDB: 4XB5), and the coordinates for OCP-P3 were taken from our previously published work.[22] The Protein Preparation application was used to preprocess the protein-ligand complex, assigning bond orders and adding hydrogens. For the alternative positions of side chains reported in the crystal structures, the most populated ones were used. Optimization of hydrogen bond assignment was performed with default parameters (pH 7.0), and an energy minimization for hydrogens only. The protein was solvated in a cubic box of dimensions 126 Å × 126 Å × 126 Å using the TIP3P water model[34]; the solvated system contained 63,434 molecules. The system was neutralized and modeled with a physiological concentration of salt (0.15 M NaCl), by adding 184 Na$^+$ and 177 Cl$^-$ ions. All minimizations and MD simulations used the force field OPLS3e.[35]

Before performing production MD or WT-metaD simulations, the systems were relaxed using a multi-stage protocol that included a minimization and short MD simulations. After energy minimizations, the systems were relaxed using the following protocol. (1) NVT with Brownian dynamics at $T = 10$ K for 100 ps with solute (protein and ligand) atoms restrained. (2) NVT using Langevin thermostat with $T = 10$ K for 12 ps with solute atoms restrained. (3) NPT using a Langevin thermostat and a Langevin barostat with $T = 10$ K and a pressure of 1 atm for 12 ps with solute atoms restrained. (4) NPT using a Langevin thermostat and a Langevin barostat with $T = 300$ K and a pressure of 1 atm for 12 ps with solute atoms restrained. (5) Lastly, simulation in the NPT ensemble using a Langevin thermostat and a Langevin barostat with $T = 300$ K and a pressure of 1 atm for 24 ps without restrains.

MD and WT-metaD simulation details

Simulations were run on Nvidia GPU hardware with Desmond[36] (Schrödinger 2019-3). All simulations were run in the NPT ensemble (300 K, 1 atm) using Nosé-Hoover chain and Martyna-Tobias-Klein methods.[37] The time step for the simulations was 2.0 fs. The particle mesh Ewald (PME) method[38] was used for long-range electrostatic interactions with a real space cutoff radius of 9.0 Å. WT-metaD simulations were performed for OCP^O and OCP-P3 systems, using two CVs to drive domain separation. CV1 is defined as the distance between the centers of mass (COM) of the αA helix (residues 5–9) and the CTD β-strands (residues 222–224, 230–232, 289–291, 299–300. CV2 is defined as the distance between the COM of the NTD (residues 21–160) and the CTD (residues 195–306). The CVs are graphically depicted in Figure 1a, b. An upper limit wall of 65 Å was set for CV1, and an upper wall of 45 Å was set for CV2. The width of the Gaussian potential was $\sigma = 0.8$ Å, the initial height $W_0 = 1.0$ kcal/mol, the deposition time was $\Delta t = 0.1$ ps, and the WT-metaD term $k_B \Delta T = 5.0$ kcal/mol. The total time for each WT-metaD simulation was 500 ns.

Based on the free energy surface of OCP-P3, selected regions of the free energy surface were explored using 100 ns equilibrium MD simulations performed in quadruplicate. The same parameters that were used for WT-metaD were used for MD (excluding the WT-metaD specific parameters). A list of all-simulations performed is provided in Table 1.

Table 1. List of simulations

Name	Type	Initial CV1	Initial CV2	Length	# Trials
OCP^O	WT-metaD	8.3 Å	28.6 Å	500 ns	1
OCP-P3	WT-metaD	8.6 Å	28.8 Å	500 ns	1
OCP-P3 A	Equilibrium MD	7.9 Å	28.7 Å	100 ns	4
OCP-P3 B	Equilibrium MD	16.4 Å	31.2 Å	100 ns	4
OCP-P3 C	Equilibrium MD	16.5 Å	34.8 Å	100 ns	4
OCP-P3 D	Equilibrium MD	17.2 Å	38.1 Å	100 ns	4
OCP-P3 E	Equilibrium MD	24.7 Å	38.2 Å	100 ns	4
OCP-P3 F	Equilibrium MD	16.5 Å	42.6 Å	100 ns	4

FES and dissociation path calculations

The FESs for OCP^O and OCP-P3 were computed using 51 bins for each CV producing a grid spacing of 1.2 Å along CV1 and 0.4 Å along CV2. Following 500 ns WT-metaD simulations, FESs were produced by Desmond. We monitored the convergence of the WT-metaD simulations by several metrics presented in Figure S1. These convergence metrics include resampling of local and global minimum (i.e. ergodicity) (Figure S1a, b), $\Delta\Delta G$ between initial and global minimum states (Figure S1c, d) and 1-D FES projections (Figure S1e–h). 1D FES profiles were computed by the Boltzmann weighted averaging according to Eq. 2

$$G(CV_i) = -k_\text{B}T\ln \frac{\int e^{-G(CV_i,\,CV_j)/k_\text{B}T}dCV_j}{\iint e^{-G(CV_i,\,CV_j)/k_\text{B}T}dCV_i dCV_j} \qquad \text{Eq. 2}$$

where k_B is Boltzmann's constant, T is the system temperature and $G(CV_i, CV_j)$ is the two-dimensional FES.

Minimum free energy path (MFEP). The MFEP connecting the associated (A) and dissociated (D) states on the OCP-P3 FES was computed using the graph theoretical Dijkstra algorithm[29], which has been shown to be an efficient algorithm to find the true MFEP.[39] The Dijkstra algorithm was implemented in an in-house python script based on the code by G. Bouvier.[40]

Monte Carlo sampling. The MFEP obtained via de Dijkstra algorithm was independently verified by performing 100 Monte Carlo walks using the Metropolis criterion at T = 300 K. The FES was interpolated and discretized into a grid of 0.1 Å in CV1 and 0.03 Å in CV2. Each MC walk consisted of 10,000 moves over this grid with a step size of 0.2 Å in the CV1 direction and 0.06 Å in the CV2 direction.

Trajectory analysis

Root-mean-squared displacement (RMSD). To determine the representative configurations for different states on the FES, structures from WT-metaD were pooled based on proximity to specific CV1, CV2 locations. Pairwise RMSDs were calculated between the pooled structured and the configuration with the lowest RMSD sum with all other configurations, which was defined as the state's representative configuration (Figure S8). To evaluate the degree of configurational changes within the domains of OCP-P3 during the WT-metaD simulations, the RMSD of each domain was calculated over the WT-metaD trajectory, based on a domain level alignment to the initial

configuration (Figure S8). All RMSD alignments and calculations were performed using the backbone atoms of the NTD and CTD.

Root-mean-squared fluctuations (RMSF). The RMSFs were calculated over the equilibrium MD simulations for the α-carbon atoms using the initial configuration as the reference. The dynamics of the OCP^O and OCP-P3 (state D) were compared by calculating the RMSFs over the entire protein (Figure S7a) and for the NTD and CTD separately, (Figure S7b). 400 ns of MD data were used for all RMSF calculations.

Solvent accessible surface area (SASA). To compare the solvation in states A and D, we calculated the probability distributions of the SASA values for the ligand (CAN) and the NTD in both states. The SASA values were calculated over 400 ns of MD simulations using a probe of radius of 1.4 Å.
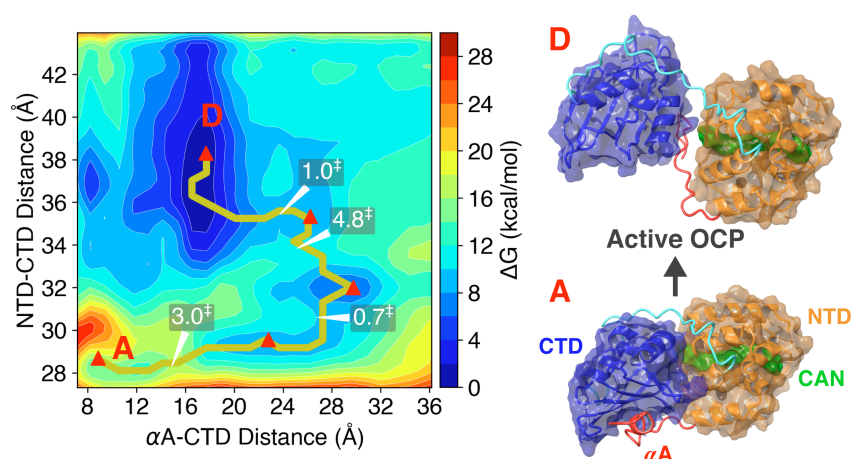
Contact Maps. Contact maps for the stable and metastable states on the OCP-P3 FES were determined for the state representative configurations (Figure S3). Residues pairs which had a minimum heavy atom distance of 6 Å or less were labeled as forming a contact.

ACKNOWLEDGMENTS

TOC FIGURE

REFERENCES

(1)   Kerfeld, C. A.; Melnicki, M. R.; Sutter, M.; Dominguez-Martin, M. A. Structure, Function and Evolution of the Cyanobacterial Orange Carotenoid Protein and Its Homologs. *New Phytol.* **2017**, *215* (3), 937–951.

(2)   Kirilovsky, D.; Kaňa, R.; Prášil, O. Mechanisms Modulating Energy Arriving at Reaction Centers in Cyanobacteria. In *Non-Photochemical Quenching and Energy Dissipation in Plants, Algae and Cyanobacteria*; Demmig-Adams, B., Garab, G., Adams III, W., Govindjee, Eds.; Advances in Photosynthesis and Respiration; Springer Netherlands: Dordrecht, 2014; Vol. 40, pp 471–501.

(3)   Muzzopappa, F.; Kirilovsky, D. Changing Color for Photoprotection: The Orange Carotenoid Protein. *Trends Plant Sci.* **2020**, *25* (1), 92–104.

(4)   Wilson, A.; Ajlani, G.; Verbavatz, J.-M.; Vass, I.; Kerfeld, C. A.; Kirilovsky, D. A Soluble Carotenoid Protein Involved in Phycobilisome-Related Energy Dissipation in Cyanobacteria. *Plant Cell* **2006**, *18* (4), 992–1007.

(5)   Kerfeld, C. A.; Sawaya, M. R.; Brahmandam, V.; Cascio, D.; Ho, K. K.; Trevithick-Sutton, C. C.; Krogmann, D. W.; Yeates, T. O. The Crystal Structure of a Cyanobacterial Water-Soluble Carotenoid Binding Protein. *Structure* **2003**, *11* (1), 55–65.

(6)   Punginelli, C.; Wilson, A.; Routaboul, J.-M.; Kirilovsky, D. Influence of Zeaxanthin and Echinenone Binding on the Activity of the Orange Carotenoid Protein. *Biochim. Biophys. Acta BBA - Bioenerg.* **2009**, *1787* (4), 280–288.

(7) Leverenz, R. L.; Jallet, D.; Li, M.-D.; Mathies, R. A.; Kirilovsky, D.; Kerfeld, C. A. Structural and Functional Modularity of the Orange Carotenoid Protein: Distinct Roles for the N- and C-Terminal Domains in Cyanobacterial Photoprotection. *Plant Cell* **2014**, *26* (1), 426–437.

(8) Wilson, A.; Kinney, J. N.; Zwart, P. H.; Punginelli, C.; D'Haene, S.; Perreau, F.; Klein, M. G.; Kirilovsky, D.; Kerfeld, C. A. Structural Determinants Underlying Photoprotection in the Photoactive Orange Carotenoid Protein of Cyanobacteria. *J. Biol. Chem.* **2010**, *285* (24), 18364–18375.

(9) Sluchanko, N. N.; Slonimskiy, Y. B.; Moldenhauer, M.; Friedrich, T.; Maksimov, E. G. Deletion of the Short N-Terminal Extension in OCP Reveals the Main Site for FRP Binding. *FEBS Lett.* **2017**, *591* (12), 1667–1676.

(10) Leverenz, R. L.; Sutter, M.; Wilson, A.; Gupta, S.; Thurotte, A.; Carbon, C. B. de; Petzold, C. J.; Ralston, C.; Perreau, F.; Kirilovsky, D.; Kerfeld, C. A. A 12 Å Carotenoid Translocation in a Photoswitch Associated with Cyanobacterial Photoprotection. *Science* **2015**, *348* (6242), 1463–1466.

(11) Gupta, S.; Sutter, M.; Remesh, S. G.; Dominguez-Martin, M. A.; Bao, H.; Feng, X. A.; Chan, L.-J. G.; Petzold, C. J.; Kerfeld, C. A.; Ralston, C. Y. X-Ray Radiolytic Labeling Reveals the Molecular Basis of Orange Carotenoid Protein Photoprotection and Its Interactions with Fluorescence Recovery Protein. *J. Biol. Chem.* **2019**, *294* (22), 8848–8860.

(12) Liu, H.; Zhang, H.; Orf, G. S.; Lu, Y.; Jiang, J.; King, J. D.; Wolf, N. R.; Gross, M. L.; Blankenship, R. E. Dramatic Domain Rearrangements of the Cyanobacterial Orange Carotenoid Protein upon Photoactivation. *Biochemistry* **2016**, *55* (7), 1003–1009.

(13) Harris, D.; Tal, O.; Jallet, D.; Wilson, A.; Kirilovsky, D.; Adir, N. Orange Carotenoid Protein Burrows into the Phycobilisome to Provide Photoprotection. *Proc. Natl. Acad. Sci.* **2016**, *113* (12), E1655–E1662.

(14) Jallet, D.; Thurotte, A.; Leverenz, R. L.; Perreau, F.; Kerfeld, C. A.; Kirilovsky, D. Specificity of the Cyanobacterial Orange Carotenoid Protein: Influences of Orange Carotenoid Protein and Phycobilisome Structures. *Plant Physiol.* **2014**, *164* (2), 790–804.

(15) Stadnichuk, I. N.; Krasilnikov, P. M.; Zlenko, D. V.; Freidzon, A. Ya.; Yanyushin, M. F.; Rubin, A. B. Electronic Coupling of the Phycobilisome with the Orange Carotenoid Protein and Fluorescence Quenching. *Photosynth. Res.* **2015**, *124* (3), 315–335.

(16) Wilson, A.; Gwizdala, M.; Mezzetti, A.; Alexandre, M.; Kerfeld, C. A.; Kirilovsky, D. The Essential Role of the N-Terminal Domain of the Orange Carotenoid Protein in Cyanobacterial Photoprotection: Importance of a Positive Charge for Phycobilisome Binding. *Plant Cell* **2012**, *24* (5), 1972–1983.

(17) Konold, P. E.; van Stokkum, I. H. M.; Muzzopappa, F.; Wilson, A.; Groot, M.-L.; Kirilovsky, D.; Kennis, J. T. M. Photoactivation Mechanism, Timing of Protein Secondary Structure Dynamics and Carotenoid Translocation in the Orange Carotenoid Protein. *J. Am. Chem. Soc.* **2019**, *141* (1), 520–530.

(18) Golub, M.; Moldenhauer, M.; Schmitt, F.-J.; Feoktystov, A.; Mändar, H.; Maksimov, E.; Friedrich, T.; Pieper, J. Solution Structure and Conformational Flexibility in the Active State of the Orange Carotenoid Protein: Part I. Small-Angle Scattering. *J. Phys. Chem. B* **2019**, *123* (45), 9525–9535.

(19) Gupta, S.; Guttman, M.; Leverenz, R. L.; Zhumadilova, K.; Pawlowski, E. G.; Petzold, C. J.; Lee, K. K.; Ralston, C. Y.; Kerfeld, C. A. Local and Global Structural Drivers for the Photoactivation of the Orange Carotenoid Protein. *Proc. Natl. Acad. Sci.* **2015**, *112* (41), E5567–E5574.

(20) Maksimov, E. G.; Shirshin, E. A.; Sluchanko, N. N.; Zlenko, D. V.; Parshina, E. Y.; Tsoraev, G. V.; Klementiev, K. E.; Budylin, G. S.; Schmitt, F.-J.; Friedrich, T.; Fadeev, V. V.; Paschenko, V. Z.; Rubin, A. B. The Signaling State of Orange Carotenoid Protein. *Biophys. J.* **2015**, *109* (3), 595–607.

(21) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys Rev Lett* **2008**, *100* (2), 020603.

(22) Pigni, N. B.; Clark, K. L.; Beck, W. F.; Gascón, J. A. Spectral Signatures of Canthaxanthin Translocation in the Orange Carotenoid Protein. *J. Phys. Chem. B* **2020**, *124* (50), 11387–11395.

(23) Bernardi, R. C.; Melo, M. C. R.; Schulten, K. Enhanced Sampling Techniques in Molecular Dynamics Simulations of Biological Systems. *Biochim. Biophys. Acta BBA - Gen. Subj.* **2015**, *1850* (5), 872–877.

(24) Bussi, G.; Branduardi, D. Free-Energy Calculations with Metadynamics: Theory and Practice. In *Reviews in Computational Chemistry*; Parrill, A. L., Lipkowitz, K. B., Eds.; John Wiley & Sons, Inc: Hoboken, NJ, USA, 2015; pp 1–49.

(25) Bondanza, M.; Cupellini, L.; Faccioli, P.; Mennucci, B. Molecular Mechanisms of Activation in the Orange Carotenoid Protein Revealed by Molecular Dynamics. *J. Am. Chem. Soc.* **2020**, *142* (52), 21829–21841.

(26) Liu, H.; Zhang, H.; King, J. D.; Wolf, N. R.; Prado, M.; Gross, M. L.; Blankenship, R. E. Mass Spectrometry Footprinting Reveals the Structural Rearrangements of Cyanobacterial Orange Carotenoid Protein upon Light Activation. *Biochim. Biophys. Acta BBA - Bioenerg.* **2014**, *1837* (12), 1955–1963.

(27) Bandara, S.; Ren, Z.; Lu, L.; Zeng, X.; Shin, H.; Zhao, K.-H.; Yang, X. Photoactivation Mechanism of a Carotenoid-Based Photoreceptor. *Proc. Natl. Acad. Sci.* **2017**, *114* (24), 6286–6291.

(28) Mezzetti, A.; Alexandre, M.; Thurotte, A.; Wilson, A.; Gwizdala, M.; Kirilovsky, D. Two-Step Structural Changes in Orange Carotenoid Protein Photoactivation Revealed by Time-Resolved Fourier Transform Infrared Spectroscopy. *J. Phys. Chem. B* **2019**, *123* (15), 3259–3266.

(29) Dijkstra, E. W. A Note on Two Problems in Connexion with Graphs. *Numer. Math.* **1959**, *1* (1), 269–271.

(30) Schuler, B.; Lipman, E. A.; Eaton, W. A. Probing the Free-Energy Surface for Protein Folding with Single-Molecule Fluorescence Spectroscopy. *Nature* **2002**, *419* (6908), 743–747.

(31) Li, M. S.; Klimov, D. K.; Thirumalai, D. Thermal Denaturation and Folding Rates of Single Domain Proteins: Size Matters. *Polymer* **2004**, *45* (2), 573–579.

(32) Golub, M.; Moldenhauer, M.; Schmitt, F.-J.; Lohstroh, W.; Maksimov, E. G.; Friedrich, T.; Pieper, J. Solution Structure and Conformational Flexibility in the Active State of the Orange Carotenoid Protein. Part II: Quasielastic Neutron Scattering. *J. Phys. Chem. B* **2019**, *123* (45), 9536–9545.

(33) Muzzopappa, F.; Wilson, A.; Kirilovsky, D. Interdomain Interactions Reveal the Molecular Evolution of the Orange Carotenoid Protein. *Nat. Plants* **2019**, *5* (10), 1076–1086.

(34) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935.

(35) Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.; Mondal, S.; Chen, W.; Wang, L.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory Comput.* **2019**, *15* (3), 1863–1874.

(36) Bowers, K. J.; Chow, D. E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In *ACM/IEEE SC 2006 Conference (SC'06)*; IEEE: Tampa, FL, 2006; pp 43–43.

(37) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé–Hoover Chains: The Canonical Ensemble via Continuous Dynamics. *J. Chem. Phys.* **1992**, *97* (4), 2635–2643.

(38) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An Nlog(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092.

(39) Fu, H.; Chen, H.; Wang, X.; Chai, H.; Shao, X.; Cai, W.; Chipot, C. Finding an Optimal Pathway on a Multidimensional Free-Energy Landscape. *J. Chem. Inf. Model.* **2020**, *60* (11), 5366–5374.

(40) Bouvier, G. Compute the shortest path on a grid using python https://bougui505.github.io/2016/08/31/compute_the_shortest_path_on_a_grid_using_python.html (accessed Mar 8, 2021).