This is the peer reviewed version of the following article:

Maria-Solano, M.A, Iglesias-Fernández, J., Osuna, S. Deciphering the Allosterically Driven Conformational Ensemble in Tryptophan Synthase Evolution. *Journal of the American Chemical Society (JACS)*, 2019, 141 (33), 13049–13056

It has been published in final form at https://doi.org/10.1021/jacs.9b03646

Deciphering the Allosterically-driven Conformational Ensemble in Tryptophan Synthase Evolution

Miguel A. Maria-Solano,[a] Javier Iglesias-Fernández,*[a] and Sílvia Osuna*[a,b]

[a] CompBioLab group, Institut de Química Computacional i Catàlisi (IQCC) and Departament de Química, Universitat de Girona, Girona, Spain

[b] ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain

ABSTRACT: Multimeric enzyme complexes are ubiquitous in Nature and catalyze a broad range of useful biological transformations. They are often characterized by a tight allosteric coupling between subunits, making them highly inefficient when isolated. A good example is Tryptophan synthase (TrpS), an allosteric heterodimeric enzyme in the form of an αββα complex that catalyzes the biosynthesis of L-tryptophan. In this study, we decipher the allosteric regulation existing in TrpS from *Pyrococcus furiosus* (*Pf*TrpS), and how the allosteric conformational ensemble is recovered in laboratory-evolved stand-alone β-subunit variants. We find that recovering the conformational ensemble of a subdomain of TrpS affecting the relative stabilities open, partially-closed and closed conformations is a prerequisite for enhancing the catalytic efficiency of the β-subunit in the absence of its binding partner. The distal mutations resuscitate the allosterically-driven conformational regulation and alter the populations and rates of exchange between these multiple conformational states, which are essential for the multistep reaction pathway of the enzyme. Interestingly, these distal mutations can be *a priori* predicted by careful analysis of the conformational ensemble of the TrpS enzyme through computational methods. Our study provides the enzyme design field with a rational approach for evolving allosteric enzymes towards improved stand-alone function for biosynthetic applications.

INTRODUCTION

Allostery is a central biological process in which two distinct sites within a biomolecule are functionally connected. Allosteric effects play a key role in protein regulation and cell signaling, and their functional significance has fostered many studies for unveiling the underlying forces that drive allostery. [1-3] In enzymatic mechanisms, allosteric interactions often promote enzyme-substrate binding and product release, and directly affect catalytic turnover.[4-6] Some studies suggest that allostery is an intrinsic characteristic of enzymes,[7] given the fact that distal active site mutations often confer improved catalytic properties.[8-9,10-11] The essential role played by remote mutations in tuning enzyme activity also indicates that allostery could be exploited for the engineering of new enzyme variants.[12]

Allosteric regulation present in multimeric enzyme complexes makes the isolated subunits, i.e. in the absence of their protein partner, highly inefficient. This is indeed the case for Tryptophan synthase (TrpS; EC 4.2.1.20). TrpS is a heterodimeric enzyme complex composed of α-subunits (TrpA) and β-subunits (TrpB) in an αββα arrangement that presents an intricate allosteric communication network between TrpA and TrpB.[13-15] TrpA catalyzes the retro-aldol cleavage of indole glycerol phosphate (IGP) producing glyceraldehyde 3-phosphate (G3P) and indole; the latter is able to diffuse through an internal TrpA-TrpB tunnel to reach the TrpB subunit (see Figure 1).

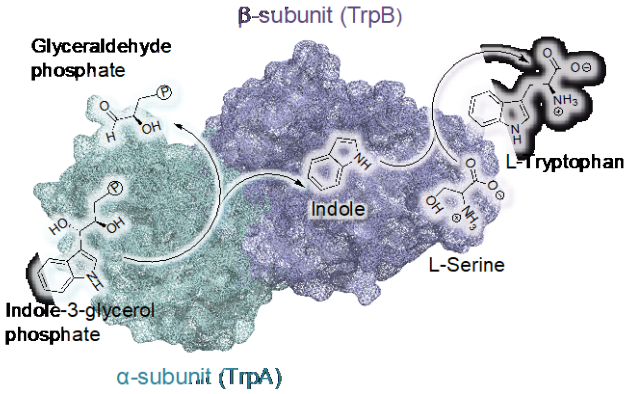


Figure 1. Overview of the Tryptophan synthase (TrpS) mechanism. The enzyme is a heterodimeric complex formed by two subunits: TrpA (shown in teal) and TrpB (in purple).

The TrpB resting state is characterized by a pyridoxal phosphate (PLP)-cofactor covalently linked to the K82 active site residue, forming a Schiff base intermediate (E(Ain)). After transamination with L-serine E(Ser), an external aldimine intermediate E(Aex1) is formed. This intermediate undergoes deprotonation at Cα, assisted by K82, which is followed by a rapid elimination of the Aex1 hydroxyl group to form an electrophilic amino acrylate intermediate E(A-A). In the dimeric complex, indole formed in TrpA reaches the TrpB active site and reacts with E(A-A) to form a quinonoid intermediate E(Q2), which after proton extraction (to recover indole's aromatic character) generates $E(Q_3)$ (not shown in Figure 2a). At this point, protonation at Cα of Q3 by K82 forms the E(Aex2) intermediate, which undergoes a second transamination reaction to finally release the L-tryptophan E(Trp) product and restore the enzyme resting state (see Figure 2).[15]

Previous studies along the catalytic mechanism identified different open and closed conformations of the enzyme in both subunits, which were based on static X-ray structures. These open-to-closed transitions can be defined by the TrpA loop (residues 163-176) that gets ordered and the slow motion of the rigid COMM domain in TrpB (residues 97-184, see Figure 2 and Table S1). Both the TrpA loop and the TrpB COMM domain are part of the active site cavity of each subunit and modulate solvent exposure to prevent substrate loss to the media. Besides, the COMM domain contains the α-helix H6 (residues 174-164) that is directly involved in non-covalent interactions with the indole moiety of the TrpB reactant intermediates. Moreover, E104 located in the COMM domain has been reported to play a role in the stabilization of charge redistribution that takes place during the nucleophilic attack of indole in the A-A intermediate (see Figure 2).[15]

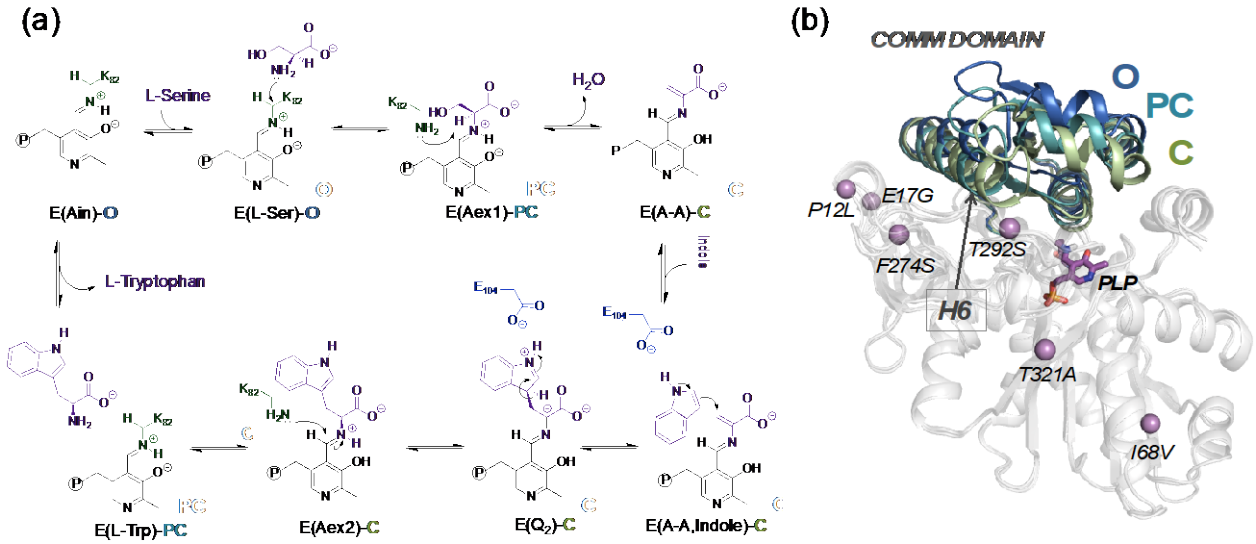


Figure 2. (a) Tryptophan synthase reaction mechanism of TrpB subunit showing the conformational states of the COMM domain according to available X-ray data at each reaction intermediates. The degree of closure of the COMM domain is represented by colored labels in blue (open, **O**), teal (partially closed, **PC**), and green (closed, **C**). (b) Overlay of representative X-ray structures showing the COMM domain(97-184 residues) in **O**, **PC** and **C** states. The pyridoxal phosphate (PLP) cofactor is shown in purple, the *Pf*TrpB^{0B2} stand-alone DE mutation positions are marked with purple spheres, and the COMM domain α-helix H6 (residues 174-164) is highlighted.

Tryptophan synthase has found applications in many fields of synthetic chemistry, in particular, for the production of noncanonical amino acids (NCAAs).[16-18] The use of TrpS for industrial purposes is hampered by its multimeric structure and the low activity of TrpB as stand-alone enzyme. Detailed insights were obtained in studies by Prof. Arnold and co-workers, who applied Directed Evolution (DE) to a thermophilic TrpB from *Pyrococcus furiousus* (*Pf*TrpB). They optimized the enzyme for stand-alone function,[19] and later on for the production of a variety of Trp derivatives.[19-25] The most efficient stand-alone catalyst was achieved by introducing up to six activating mutations, which were located far away from active site positions (see Figure 2b). Note that P12L, E17G and F274S are located close to the TrpA-TrpB protein interface. Analysis of spectroscopic data suggested that *Pf*TrpB stand-alone variants and the *Pf*TrpS complex were better in stabilizing closed conformations of the COMM domain upon substrate binding than isolated *Pf*TrpB.[24] However, despite showing drastic differences in catalytic efficiency, X-ray data failed to find a connection between COMM domain closure and stabilization of the enzyme. In particular, the COMM domain structure is almost identical among different organisms (e.g. *Salmonella typhimurium* and *Pyrococcus furiosus*), isolated *Pf*TrpB enzyme and *Pf*TrpB stand-alone variants, although all of them diverge in functionality. These observations suggest that the origin behind their different catalytic efficiencies could be attributed to alterations in the enzyme conformational ensemble induced by distal active site mutations. Such effects have not been explored yet, although they are crucial to understand how the stand-alone functionality was achieved.

In this work, we elucidate how the different reaction intermediates and distal mutations introduced in laboratory-evolution alter the allosterically-driven conformational ensemble of *Pf*TrpS. Surprisingly, the introduced distal mutations increase the conformational heterogeneity of the COMM domain; hence, the *Pf*TrpB enzyme has the ability to access the different COMM domain conformations, which are essential for efficient catalysis in the absence of its binding partner. Through careful inspection of the conformational ensemble of *Pf*TrpS with our recently developed SPM tool[12] we were able to identify the most important positions to recover the allosterically-driven conformational ensemble, which coincide with the mutations introduced in laboratory-evolution. Our study shows clearly that stand-alone versions of allosterically-regulated enzymes can be rationally designed by targeting the recovery of the allosterically-driven conformational ensemble.

RESULTS AND DISCUSION

Available structural data show that the TrpB COMM domain is able to explore open (**O**), partially closed (**PC**), and closed (**C**) conformations along the multi-step TrpB catalytic pathway, due to the allosteric regulation exerted by TrpA (see Figure 2). Considering only the TrpB subunit, X-ray studies revealed that its resting state (i.e. E(Ain)) is characterized by **O** COMM domain conformations (1V8Z)[26], which are shifted towards **PC** states at the external aldimine intermediate E(Aex1) (5DW0)[19]. All subsequent reaction intermediates (i.e. from the electrophilic amino acrylate E(A-A) to E(Aex2)) were crystallized in **C** states (4HN4[27] and 3CEP[28]). A recent X-ray structure (5DW3)[19] indicated that the **PC** conformation is recovered once Trp is formed at E(Trp), preparing the enzyme for product release and the next turnover (see Figure 1-2 and Table S1 for more structural data).

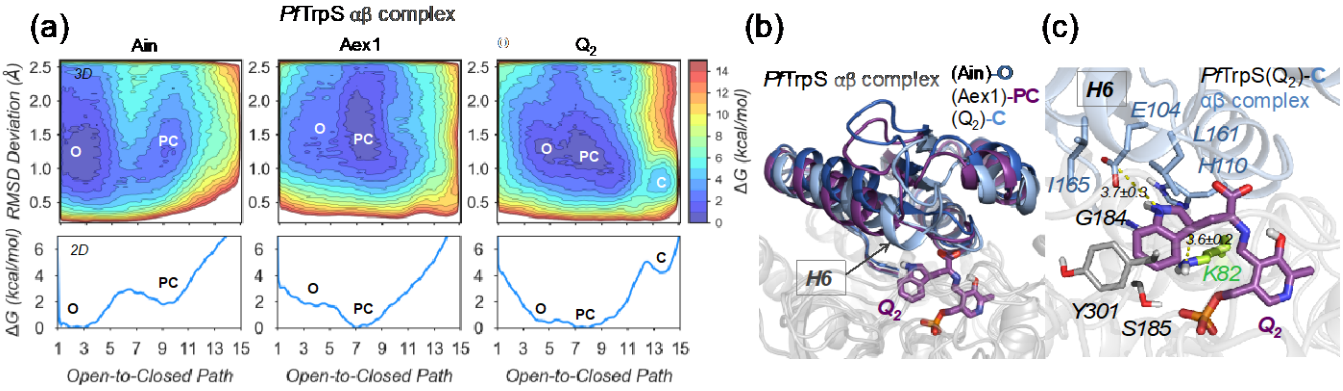


Figure 3. (a) Free energy landscape (FEL) associated to the COMM domain Open-to-Closed (**O**-to-**C**) conformational exchange of the *Pf*TrpS complex enzyme at Ain, Aex1 and Q2 reaction intermediates. (b) Overlay of the *Pf*TrpS metastable conformations of the open (**O**) state at Ain intermediate, partially closed (**PC**) at Aex1 and closed (**C**) at Q_2, respectively showing the entire **O**-to-**C** sampled transition. (c) Detailed active site view of the *Pf*TrpS metastable conformation of the **C** state at Q_2 intermediate (shown in purple). Active site residues are shown in gray, except for those included in the COMM

domain (shown in blue), and the catalytic K82 proton transfer residue (green). The catalytic distances (in Å) between charge-charge stabilization E104-Q2 and proton transfer K82-Q2 are also represented.

Allosteric transitions, such as the TrpA-triggered **O**-to-**C** exchange of the COMM domain in TrpB, are relatively slow domain motions that take place on time scales larger than our currently accessible simulation times.[29] Indeed, initial 500 ns standard MD simulations of the *Pf*TrpS in the αβ complex, the isolated *Pf*TrpB wild-type and the stand-alone *Pf*TrpB^{0B2} enzyme variant in multiple reaction intermediates (Ain, Aex1 and A-A) failed to sample the entire allosteric transition, with no clear RMSD differences observed between the studied systems (Figure S2). To overcome this limitation, we employed enhanced sampling techniques. In particular, we applied the metadynamics approach[30-31] to reconstruct the Free Energy Landscape (FEL) associated with the COMM domain **O**-to-**C** transition of the *Pf*TrpS αβ complex, isolated *Pf*TrpB wild-type, and evolved stand-alone *Pf*TrpB^{0B2} variant (see details in SI). Several intermediates along the catalytic cycle were modeled. In particular, we selected E(Ain), E(Aex1), E(A-A) and E(Q2) to evaluate the **O**-**PC**-**C** conformational exchange of the COMM domain found in X-ray data, but also to reproduce the multi-step mechanism under study (see Figure 2).

Population shift towards closed conformations along the allosteric TrpS catalytic pathway. To elucidate the allosterically-driven conformational ensemble of *Pf*TrpS αβ complex we reconstructed the FEL associated to the conformational dynamics of the COMM domain for each reaction intermediate (see Figure 3a). As expected from X-ray data, in the resting state of the enzyme, *Pf*TrpS-Ain, the **O** conformational state is highly favored, in agreement with its functional role in Ser binding. However, less stable **PC** states (*ca.* 2 kcal/mol higher in energy) are also visited with an associated **O**-to-**PC** transition energy barrier of only *ca.* 3 kcal/mol. As the *Pf*TrpS enzymatic reaction progresses, a population shift occurs towards the stabilization of **PC** states (see Figure 3a). After the reaction with serine in the external aldimine Aex1 intermediate, the open **O** state is destabilized by *ca.* 2 kcal/mol with respect to the **PC** state, which becomes the most stable conformation. In contrast to Ain and Aex1, the quinonoid Q2 intermediate generated after indole coupling samples all possible conformations of the COMM domain: **O** and **PC** states are almost equally stabilized, while the **C** state is *ca.* 5 kcal/mol higher in energy. The associated **PC**-to-**C** barrier is *ca.* 6 kcal/mol. This suggests that the adoption of the fully closed COMM domain conformation is the limiting factor, in agreement with the spectroscopic data for *Pf*TrpS.[19, 24] Such closed active states form an optimized network of hydrophobic interactions between the enzyme and the indole moiety (see Figure 3c). Several side chains residues, including Y301, S185, and COMM domain G184, H110, L161 and I165 define this network.

Comparison of *Pf*TrpS(Q2)-**O** and -**C** metastable structures shows that the helix H6 closure is needed for forming CH··CH and CH··π interactions between L161 and I165 with the indole moiety and also a hydrogen bond with the E104 residue (Figure 3b-c). The **C** state of *Pf*TrpS(Q2) shows a highly pre-organized active site with E104 and the proton acceptor K82 properly positioned for catalysis together with the indole moiety establishing many non-covalent interactions with the active site pocket (see Figure 3c and S8a-b). The high stability of **O** states at the Q2 intermediate suggests that the COMM domain of *Pf*TrpS(Q2) is already prepared for product release and recovery of the native state of the enzyme for the next cycle. Altogether, these findings highlight the crucial role of the allosterically-driven conformational ensemble of the COMM domain of *Pf*TrpS for efficiently optimizing the multiple steps along its catalytic cycle.

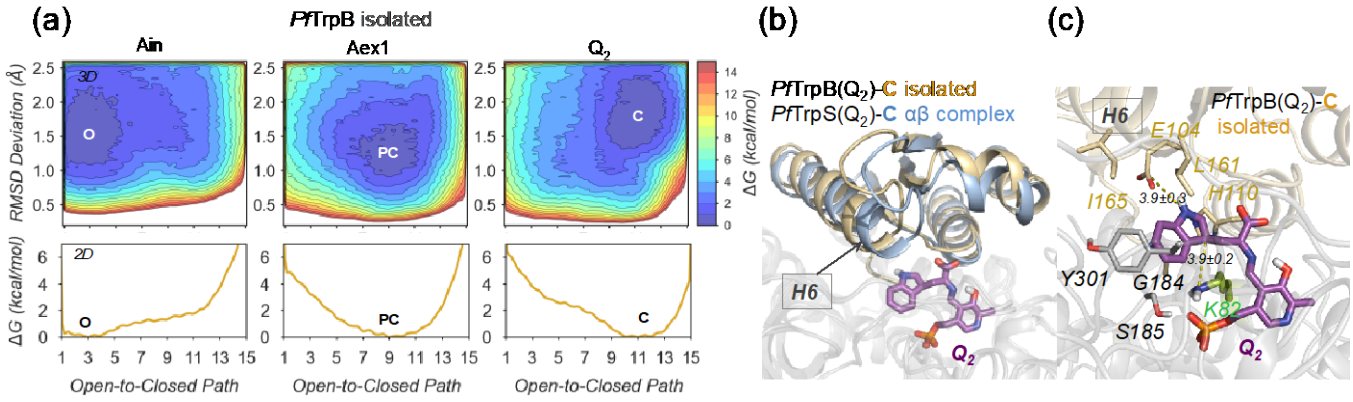


Figure 4. (a) Free energy landscape (FEL) associated to the COMM domain Open-to-Closed (**O**-to-**C**) conformational exchange of the *Pf*TrpB isolated enzyme at Ain, Aex1 and Q2 reaction intermediates. (b) Overlay of the metastable conformations of the closed (**C**) states at Q_2 intermediate for *Pf*TrpB (in orange) and *Pf*TrpS (blue). (c) Detailed active site view of the *Pf*TrpB metastable conformation of the **C** state at Q_2 intermediate (shown in purple). Active site residues are shown in gray, except for those included in the COMM domain (shown in orange), and the catalytic K82 proton transfer residue (green). The catalytic distances (in Å) between charge-charge stabilization E104-Q2 and proton transfer K82-Q2 are also represented.

Isolated *Pf*TrpB displays restricted COMM domain heterogeneity and unproductive closure. Experimental data showed that in the absence of the allosteric partner *Pf*TrpA, *Pf*TrpB activity decreases 3-fold (k_{cat} of 0.31 and 1.0 s^{-1} for isolated *Pf*TrpB and *Pf*TrpS, respectively).[19] Our reconstructed FELs corresponding to isolated *Pf*TrpB display some similarities to the *Pf*TrpS system (see Figures 3a and 4a). Ser binding at the Aex1 intermediate shifts the conformational ensemble from **O** towards **PC** states. Similarly, a population shift towards **C** states at the Q_2 intermediate is observed. Contrary to the situation for the *Pf*TrpS complex, a single energy minimum is found at the Aex1 and Q2 intermediates of *Pf*TrpB. In fact, the COMM domain is not able to escape from **O** states at Ain, **PC** at Aex1, and **C** at Q2 intermediates as the other states are inaccessible. Therefore, *Pf*TrpB in the absence of *Pf*TrpA allosteric regulation has a very limited conformational heterogeneity of the COMM domain, which hampers the multi-step reaction pathway. It is also worth mentioning that the stable **C** states at the *Pf*TrpB-Q2 intermediate are highly deviated from the reference **O**-to-**C** conformational path (i.e. RMSD larger than 1.5 Å, see Figure 4a). A detailed structural analysis of the isolated wild-type *Pf*TrpB(Q2) compared to the *Pf*TrpS(Q2) complex in **C** states, indicates that the isolated *Pf*TrpB enzyme cannot efficiently sample catalytically competent **C** states; this is in particular true for the key COMM H6 closure (Figure 4b-c and S8c). Furthermore, the proton transfer catalytic distance K82-Q2 is also longer than in *Pf*TrpS (3.9±0.3 Å vs. 3.6±0.3 Å). Our simulations have therefore shown that in the absence of its *Pf*TrpA allosteric partner, the *Pf*TrpB COMM domain displays a restricted conformational landscape, which lacks the ability to easily access **O**, **PC**, and **C** states existing in the allosterically-driven conformational ensemble of *Pf*TrpS.

Activating distal mutations for stand-alone function recover COMM domain heterogeneity. *Pf*TrpB was evolved for stand-alone function generating a new variant *Pf*TrpB^{0B2}, which displays a considerably improved catalytic constant with respect to both the isolated wild-type *Pf*TrpB and the *Pf*TrpS complex (k_{cat} of 2.9 s^{-1}, 0.31 s^{-1}, 1.0 s^{-1} for *Pf*TrpB^{0B2}, *Pf*TrpB, *Pf*TrpS, respectively). It is also worth mentioning that the activity of the evolved *Pf*TrpB^{0B2} decays dramatically in the presence of *Pf*TrpA (k_{cat} of 0.04 s^{-1}).[19] Intrigued by the restricted conformational dynamics of the COMM domain as found in isolated *Pf*TrpB, we decided to explore whether distal mutations introduced in laboratory evolution were able to recover the allosterically-driven conformational ensemble of *Pf*TrpS. By comparing the reconstructed FELs for stand-alone *Pf*TrpB^{0B2} and *Pf*TrpS complex along the different reaction intermediates (see Figure 3a and 5a), it becomes evident that the *Pf*TrpB^{0B2} variant recovers the conformational heterogeneity of the COMM domain, characteristic of the allosterically regulated dimeric enzyme. However, interesting differences between both systems are found to be crucial for rationalizing their catalytic activities.

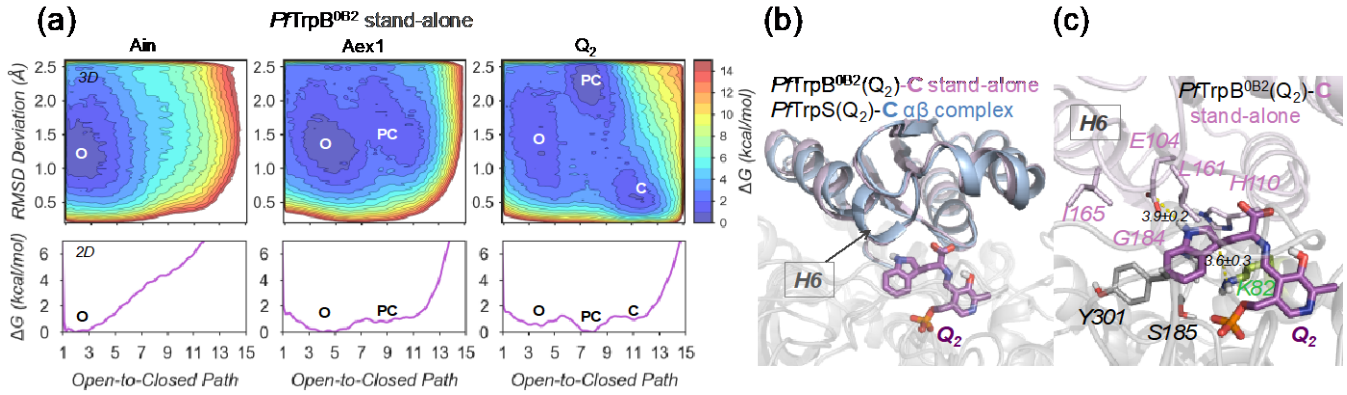


Figure 5. (a) Free energy landscape (FEL) associated to the COMM domain Open-to-Closed (**O**-to-**C**) conformational exchange of the *Pf*TrpB^{0B2} enzyme at Ain, Aex1 and Q2 reaction intermediates. (b) Overlay of the metastable conformations of the closed (**C**) states at Q_2 intermediate for *Pf*TrpB^{0B2} (in pink) and *Pf*TrpS (blue). (c) Detailed active site view of the *Pf*TrpB^{0B2} metastable conformation of the **C** state at Q_2 intermediate (shown in purple). Active site residues are shown in gray, except for those included in the COMM domain (shown in violet), and the catalytic K82 proton transfer residue (green). The catalytic distances (in Å) between charge-charge stabilization E104-Q2 and proton transfer K82-Q2 are also represented.

In the resting state (Ain), *Pf*TrpB^{0B2} has only the **O** state accessible. However, as the reaction progresses, a population shift towards **PC** and **C** states occurs, as also observed in the dimeric *Pf*TrpS complex. At Aex1, **O** and **PC** states have comparable relative stabilities and are separated by a small energy barrier of *ca.* 1 kcal/mol, which allows a fast **O**-to-**PC** conformational exchange. Similar to the *Pf*TrpS system, at the Q2 state the allosterically-driven conformational ensemble containing **O**, **PC**, and **C** states is recovered. Nevertheless, a substantially lower barrier is observed for the **O**-to-**PC**-to-**C** transition of *ca.* 2 kcal/mol as compared to *Pf*TrpS complex. This rather small energy barrier allows *Pf*TrpB^{0B2} to easily adopt the catalytically competent **C** conformation from **PC** and **O** states. This high stability of the catalytically relevant **C** state contrasts with the *Pf*TrpS system where the closed state is *ca.* 5 kcal/mol higher in energy. Such a difference in the stability of the **C** state explains the improved catalytic efficiency of the evolved stand-alone variant. The **C** state of stand-alone *Pf*TrpB^{0B2}(Q2) has an almost identical degree of closure of the COMM domain as the *Pf*TrpS catalytically competent conformation, and a similar catalytic K82-Q2 proton transfer distance (see Figure 5b-c and S9). This indicates that the **C** state of the stand-alone *Pf*TrpB^{0B2}(Q2) variant is properly preorganized for the reaction.

A remarkable difference between the dimeric *Pf*TrpS(Q2) complex and stand-alone *Pf*TrpB^{0B2}(Q2) is found at the **PC** state, which in the case of the evolved variant is highly digressed from the original path (i.e. RMSD > 1.5 Å, see Figure 5a); therefore, we denote this as the novel **PC** state. The large deviation arises from an unexpected large-scale conformational change of 14 Å that positions R159, adjacent to the H6 helix of the COMM domain, towards the active site (see Figure 6a). This novel **PC** conformation has not been previously observed by means of X-ray crystallography. Interestingly, R159 takes over the position previously occupied by L161 at H6 of *Pf*TrpB^{0B2}(Q2), establishing a cation-π interaction with the indole moiety (Figure 6b). We hypothesize that this novel conformation of R159 may play a role in the catalytic cycle, most probably in properly positioning serine and/or indole for the reaction (see SI for a detailed discussion and Figure S10-11).

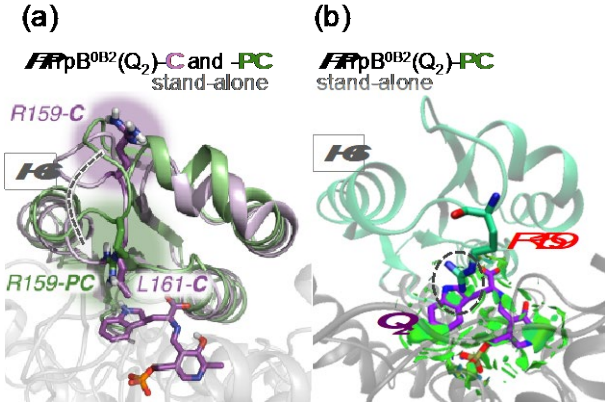


Figure 6. (a) Overlay of the *Pf*TrpB^{0B2} metastable conformations of the closed (**C**, in purple) and novel partially closed (**PC**, in green) states at the Q_2 reaction intermediate. The novel PC state revealed by MD simulations presents R159 from the COMM domain located in the active site close to the position previously occupied by L161 in the **C** state. (b) Representation of the non-covalent interactions (computed with NCIplot[32] at the novel **PC** state of *Pf*TrpB^{0B2}, highlighting (green surfaces) cation-π interaction between R159 and the indole moiety of Q_2.

Our findings indicate that stand-alone *Pf*TrpB^{0B2} recovers the COMM domain heterogeneity, thus making **O** and active **C** states accessible again. The great stabilization of active **C** states together with the novel **PC** conformation displayed by the PfTrpB^{0B2} enzyme variant unravel the increase in catalytic efficiency with respect to the allosterically regulated *Pf*TrpS complex.

Another relevant aspect is the experimentally observed inactivation of the evolved *Pf*TrpB^{0B2} in the presence of *Pf*TrpA. To study this inactivation, we reconstructed the FEL for the *Pf*TrpA-*Pf*TrpB^{0B2}(Q2) complex. Surprisingly, our simulations show that the presence of *Pf*TrpA does not restrict the *Pf*TrpB^{0B2} COMM domain heterogeneity, as it is also able to sample the **O**-to-**C** exchange. However, the formation of the dimeric complex with *Pf*TrpA induces a population shift towards unproductive closed states (i.e. highly deviated from the reference path), similar to those observed in the isolated *Pf*TrpB system (see Figure S12). Thus, *Pf*TrpA truncates the efficient conformational ensemble of the stand-alone *Pf*TrpB^{0B2} yielding non-productive closed conformational states of the COMM domain.

COMM domain heterogeneity as an essential factor in indole active-site accessibility. Available X-ray structures after E(Aex1) formation display **C** conformations of the COMM domain. However, our analysis of the allosterically-driven conformational ensemble of *Pf*TrpS complex and that of stand-alone *Pf*TrpB^{0B2} provided evidence for a high flexibility of the COMM domain and the ability to visit **O**, **PC**, and the catalytically relevant **C** states. The question that remains is: what is the specific role of **O** conformational states of the COMM domain after Ser binding? One possibility would be to assist in either indole binding or Trp release after the reaction. Experimentally, the Michaelis constant for indole binding in the stand-alone *Pf*TrpB^{0B2} enzyme variant (8.7 µM) was improved with respect to isolated *Pf*TrpB (77 µM), but also compared to the enzyme complex *Pf*TrpS (20 µM). [19]

To elucidate the changes in indole binding and the role played by the COMM domain **O** states, we reconstructed the FELs at the electrophilic amino acrylate E(A-A) intermediate (Figure 7a-b and S13), and analyzed the available indole substrate access tunnels with the CAVER software.[33] At this A-A intermediate, *Pf*TrpS complex and stand-alone *Pf*TrpB^{0B2} can easily access both **O** and **C** states, which are separated by relatively small energy barriers. The analysis of indole access tunnels in both **O** and **C** states reveals two different entry pathways (Figure 7c-d and S13c): the previously-described internal tunnel (IT) that connects TrpA and TrpB subunits in *Pf*TrpS complex (shown in blue in Figure 7c-d and S13c), and a secondary tunnel (ST) connecting the active site with a novel entry path not described before (shown in green). **C** states of the COMM domain yield a narrow bottleneck tunnel radius hampering indole diffusion outside the active site, thus capturing it for efficient catalysis (see Figure 7-d). Therefore, the differences in indole binding should be related to **O** COMM domain states. Interestingly, the isolated wild-type *Pf*TrpB is not able to sample the **O** state, which results in indole access through **PC** conformations that have a much narrower bottleneck radius (see Figure S13c). This leads to less favorable K_M values for *Pf*TrpB, as observed experimentally.

At the **O** state of *Pf*TrpS complex, indole diffusion occurs along the internal TrpA-TrpB tunnel, suggesting that the secondary tunnel (green in Figure 7c-d and S13C) may play a role in Ser binding and/or Trp release. For the stand-alone *Pf*TrpB^{0B2} variant, both tunnels show a large bottleneck radius, thus no tunnel preference for indole entrance to the active site is found (see Figure 7c). Altogether, these calculations indicate that the recovery of the allosterically-regulated COMM conformational ensemble of *Pf*TrpS, especially **O** state accessibility, is also key for indole binding.

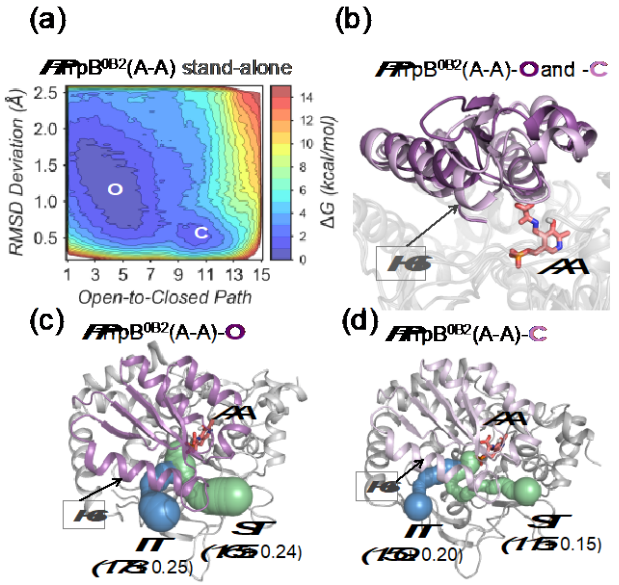


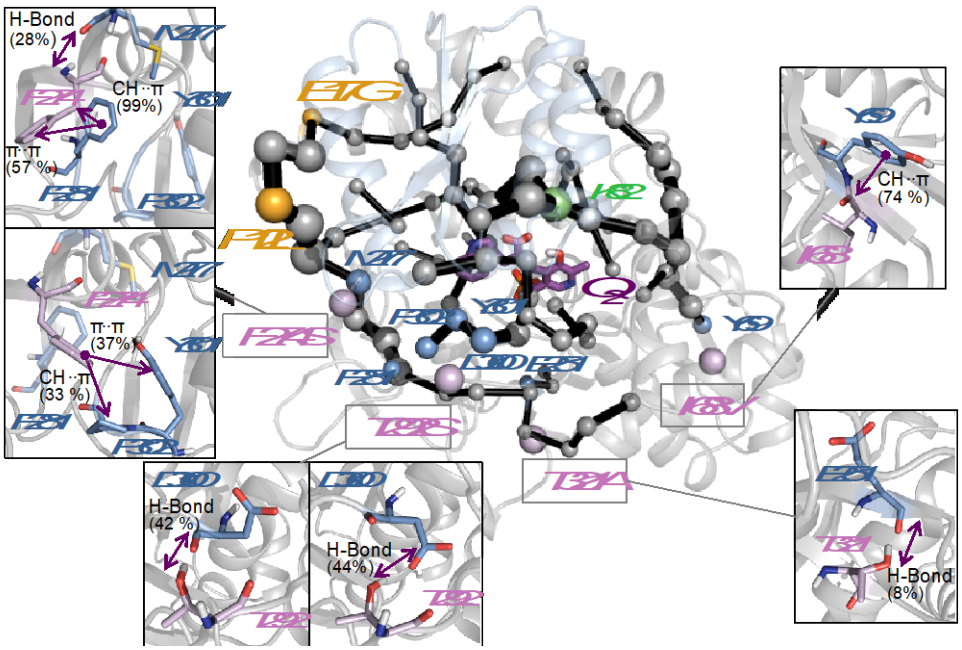
Figure 7. (a) Free energy landscape (FEL) associated to the COMM domain Open-to-Closed (**O**-to-**C**) conformational exchange of stand-alone *Pf*TrpB^{0B2} enzyme at A-A reaction intermediate. (b) Overlay of meta-stable conformations of the **O** (dark purple) and **C** state (light purple) at A-A intermediate of *Pf*TrpB^{0B2}. (c) and (d) Internal (IT, in blue) and secondary (ST, in green) tunnels of *Pf*TrpB^{0B2} at the **O** and **C** states at A-A reaction intermediate computed with CAVER 3.0.[33] The averaged bottleneck radii (in Å) are also shown.

Distal mutations for stand-alone function can be predicted computationally. Our group has recently shown that distal mutations found by DE in the case of multi-step retro-aldolase enzymes can be identi-

fied with residue-by-residue correlation and proximity analysis tools.[12] Intrigued by the possibility of predicting distal positions for stand-alone function, we applied our Shortest Path Map (SPM) method.[12] This computational tool identifies those pairs of residues that have a higher contribution to the conformational dynamics of the enzyme (see Figure 8 and computational details). We focused our analysis on the *Pf*TrpS(Q2) metadynamics trajectory because of the complete **O**-to-**C** conformational exchange sampled in it (see Figure S14 for SPM analysis at other reaction intermediates). *Pf*TrpB^{0B2} presents 6 mutations: P12L, E17G, I68V, T292S, F274S and T321A, from which 2 were directly predicted by the SPM tool, and 3 were directly interacting with a SPM position. The specific effect of each isolated mutation on the enzyme activity is not known, except for two of them: T292S and P12L. The most beneficial mutation T292S (3-fold increase in k_{cat} with respect to *Pf*TrpB)[19] was previously suggested to modulate COMM domain closure based on the T292-D300 interaction observed in X-ray data.[15] In our metadynamics simulations, the hydrogen bond between T292 and D300 is maintained 86% of the time (see Figure 8). Although position 292 is not directly included in our computed SPM path, position 300 is predicted as key for the COMM domain conformational dynamics. This indicates that by altering position D300 interactions (for instance, the T292-D300 interaction) the COMM domain closure can be modulated. Interestingly, P12L distal mutation P12L, which was found to have a slight impact on the k_{cat} of the enzyme, is directly identified with high contributions in our SPM analysis (see orange spheres in Figure 8). Similarly, the distal site E17G is also predicted by SPM, suggesting a role on COMM domain conformational heterogeneity. DE positions F274, and I68, although not strictly included in the SPM path, make direct and stable non-covalent interactions with already SPM predicted positions (see blue and pink spheres in Figure 8). For instance, F274 highly forms CH··π and π ··π interactions with the SPM residues F281 (maintained 99% of the simulation time), Y301 (37%), and P302 (33%), as well as a hydrogen bond with M277 (28%). Similarly, DE position I68 makes CH··π interactions with Y69 (74%), included in SPM. The only DE position that has a minor role in the COMM domain conformational dynamics and makes negligible interactions with SPM residues is T321.

Our new proposed methodology makes use of metadynamics simulations to enforce the sampling of the allosterically-regulated **O**-to-**C** transition, and identifies which residues present a higher contribution to the **O**-to-**C** COMM domain conformational exchange through inter-residue correlation calculations. With this new computational approach, distal positions involved in the allosteric transition can be identified, thus providing a set of key positions for the generation of smart libraries for stand-alone function. This new proposed protocol can be applied to any allosterically-regulated system of interest. This study also provides further evidence for the key role played by the enzyme conformational dynamics in the evolution of enhanced catalytic activities, especially in challenging multi-step mechanisms such as the one catalyzed by TrpS.[34]

Figure 8. Identification of the amino-acids that contribute to the Open-to-Closed (**O**-to-**C**) conformational exchange in *Pf*TrpS at (Q_2) intermediate through Shortest Path Map (SPM) analysis.[12] The size of the spheres and black edges are indicative of the importance of the position for the *Pf*TrpS conformational dynamics. Positions mutated via DE are marked in orange (if they are included in the SPM), or in pink (if they directly interact with SPM residues). SPM residues that interact with the DE positions are marked with blue spheres. For each mutation that interacts with SPM residues, a zoom is provided to show the type of non-covalent interaction and the percentage of interacion time during the simulation.



CONCLUSIONS

Recovering the allosterically-driven conformational ensemble existing in multimeric en-

zymes such as TrpS for stand-alone function is strikingly similar to the dramatic effect induced by distal mutations on the catalytic efficiency of some enzymes. Only those ensembles of conformations that are pre-activated for catalysis are selected and stabilized along the evolutionary process. Understanding the differences between both processes is highly appealing for the rational design of enzymes. The present study demonstrates that fine-tuned control of the allosterically-driven conformational ensemble of *Pf*TrpS plays a key role along its catalytic cycle. By altering the relative stabilities of open, partially-closed and closed conformational states of the COMM domain, each reaction step along the catalytic pathway can be efficiently optimized. Our free energy calculations on the conformational exchange of the COMM domain indicate that the rate for the open-to-closed conformational transition is relatively fast (in the nanosecond to microsecond timescale) in comparison with the reaction steps and turnover time scale (millisecond to second). However, such transitions are essential for pre-organizing the active site pocket to accommodate the different substrates, and efficiently catalyzing Ser and indole coupling for Trp production. Distal mutations, introduced experimentally for converting *Pf*TrpB into an efficient stand-alone variant, recover the allosterically-regulated conformational ensemble of *Pf*TrpS. This enables access to open, partially closed and closed states of the COMM domain. In the absence of such mutations, the isolated *Pf*TrpB lacks the COMM domain conformational heterogeneity, which is required for the challenging multi-step catalytic pathway. By careful analysis of the open-to-closed conformational exchange of the COMM domain, and the residues that contribute more to the exchange, distal mutations introduced via Directed Evolution can be predicted with our recently developed SPM tool. This study shows that by evaluating the native allosterically-regulated conformational ensemble, and the residues that have a higher contribution to the allosteric conformational transition, proficient stand-alone enzyme variants could be rationally designed. The hypothesis that many enzymes are intrinsically regulated allosterically[7] is inspiring, as it also suggests that our novel computational approach proposed here might be of general use in the computational enzyme design field.

ASSOCIATED CONTENT

Supporting information

Computational methods as well as additional tables and figures.

AUTHOR INFORMATION

Corresponding Author

*silvia.osuna@udg.edu

*javier.iglesias@udg.edu

ORCID

Sílvia Osuna: 0000-0001-9458-1114

Javier Iglesias-Fernandez: 0000-0001-7773-2948

ACKNOWLEDGMENT

We thank the Generalitat de Catalunya for the emerging group CompBioLab (2017 SGR-1707) and Spanish MINECO for project PGC2018-102192-B-I00. We are grateful for the computer resources, technical expertise, and assistance provided by the Barcelona Supercomputing Center - Centro Nacional de Supercomputación. M. A. M. S. was supported by the Spanish MINECO for a PhD fellowship (BES-2015-074964), J. I. F. was supported by the European Community for Marie Curie fellowship (H2020-MSCA-IF-2016-753045). S.O. is grateful to the funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ERC-2015-StG-679001). We thank Prof. Swart, Dr. Feixas and Dr. Kress for the helpful discussions and comments on the manuscript.

REFERENCES

(1) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. *Nature* **2014,** *508* (7496), 331-9.

(2) Nussinov, R.; Tsai, C. J.; Ma, B. *Annu. Rev. Biophys.* **2013,** *42*, 169-89.

(3) Freiburger, L. A.; Baettig, O. M.; Sprules, T.; Berghuis, A. M.; Auclair, K.; Mittermaier, A. K. *Nat. Struct. Mol. Biol.* **2011,** *18* (3), 288-94.

(4) Lisi, G. P.; Loria, J. P. *Curr. Opin. Struct. Biol.* **2017,** *47*, 123-130.

(5) Tzeng, S. R.; Kalodimos, C. G. *Nature* **2012,** *488* (7410), 236-40.

(6) Nussinov, R. *Chem. Rev.* **2016,** *116* (11), 6263-6.

(7) Gunasekaran, K.; Ma, B.; Nussinov, R. *Proteins: Struct., Funct., and Bioinf.* **2004,** *57* (3), 433-443.

(8) Obexer, R.; Godina, A.; Garrabou, X.; Mittl, P. R. E.; Baker, D.; Griffiths, A. D.; Hilvert, D. *Nat. Chem.* **2017,** *9* (1), 50-56.

(9) Currin, A.; Swainston, N.; Day, P. J.; Kell, D. B. *Chem. Soc. Rev.* **2015,** *44* (5), 1172-1239.

(10) Morley, K. L.; Kazlauskas, R. J. *Trends Biotechnol. 23* (5), **2005,** 231-237.

(11) Jiménez-Osés, G.; Osuna, S.; Gao, X.; Sawaya, M. R.; Gilson, L.; Collier, S. J.; Huisman, G. W.; Yeates, T. O.; Tang, Y.; Houk, K. N. *Nat. Chem. Biol.* **2014,** *10* (6), 431-436.

(12) Romero-Rivera, A.; Garcia-Borràs, M.; Osuna, S. *ACS Catal.* **2017,** *7* (12), 8524-8532.

(13) Hyde, C. C.; Ahmed, S. A.; Padlan, E. A.; Miles, E. W.; Davies, D. R. *J. Biol. Chem.* **1988,** *263* (33), 17857-71.

(14) Lee, S. J.; Ogasahara, K.; Ma, J. C.; Nishio, K.; Ishida, M.; Yamagata, Y.; Tsukihara, T.; Yutani, K. *Biochemistry* **2005,** *44* (34), 11417-11427.

(15) Dunn, M. F. *Arch. Biochem. Biophys.* **2012,** *519* (2), 154-166.

(16) Barry, S. M.; Kers, J. A.; Johnson, E. G.; Song, L. J.; Aston, P. R.; Patel, B.; Krasnoff, S. B.; Crane, B. R.; Gibson, D. M.; Loria, R.; Challis, G. L. *Nat. Chem. Biol.* **2012,** *8* (10), 814-816.

(17) Kieffer, M. E.; Repka, L. M.; Reisman, S. E. *J. Am. Chem. Soc.* **2012,** *134* (11), 5131-5137.

(18) Patel, R. N. *Biomolecules* **2013,** *3* (4), 741-777.

(19) Buller, A. R.; Brinkmann-Chen, S.; Romney, D. K.; Herger, M.; Murciano-Calles, J.; Arnold, F. H. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112* (47), 14599-14604.

(20) Herger, M.; van Roye, P.; Romney, D. K.; Brinkmann-Chen, S.; Buller, A. R.; Arnold, F. H. *J. Am. Chem. Soc.* **2016,** *138* (27), 8388-8391.

(21) Murciano-Calles, J.; Romney, D. K.; Brinkmann-Chen, S.; Buller, A. R.; Arnold, F. H. *Angew. Chem. Int. Ed.* **2016,** *55* (38), 11577-11581.

(22) Buller, A. R.; van Roye, P.; Murciano-Calles, J.; Arnold, F. H. *Biochemistry* **2016,** *55* (51), 7043-7046.

(23) Romney, D. K.; Murciano-Calles, J.; Wehrmuller, J. E.; Arnold, F. H. *J. Am. Chem. Soc.* **2017,** *139* (31), 10769-10776.

(24) Buller, A. R.; van Roye, P.; Cahn, J. K. B.; Scheele, R. A.; Herger, M.; Arnold, F. H. *J. Am. Chem. Soc.* **2018,** *140* (23), 7256-7266.

(25) Boville, C. E.; Scheele, R. A.; Koch, P.; Brinkmann-Chen, S.; Buller, A. R.; Arnold, F. H. *Angew. Chem. Int. Ed.* **2018,** *57* (45), 14764-14768.

(26) Hioki, Y.; Ogasahara, K.; Lee, S. J.; Ma, J.; Ishida, M.; Yamagata, Y.; Matsuura, Y.; Ota, M.; Ikeguchi, M.; Kuramitsu, S.; Yutani, K. *Eur. J. Biochem.* **2004,** *271* (13), 2624-35.

(27) Niks, D.; Hilario, E.; Dierkers, A.; Ngo, H.; Borchardt, D.; Neubauer, T. J.; Fan, L.; Mueller, L. J.; Dunn, M. F. *Biochemistry* **2013,** *52* (37), 6396-6411.

(28) Barends, T. R. M.; Domratcheva, T.; Kulik, V.; Blumenstein, L.; Niks, D.; Dunn, M. F.; Schlichting, I. *ChemBioChem* **2008,** *9* (7), 1024-1028.

(29) Henzler-Wildman, K.; Kern, D. *Nature* **2007,** *450* (7172), 964-972.

(30) Barducci, A.; Bonomi, M.; Parrinello, M. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011,** *1* (5), 826-843.

(31) Laio, A.; Gervasio, F. L. *Rep. Prog. Phys.* **2008,** *71* (12), 22.

(32) Contreras-Garcia, J.; Johnson, E. R.; Keinan, S.; Chaudret, R.; Piquemal, J. P.; Beratan, D. N.; Yang, W. *J. Chem. Theory. Comput.* **2011,** *7* (3), 625-632.

(33) Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Damborsky, J. *PLoS Comput. Biol.* **2012,** *8* (10), e1002708.

(34) Maria-Solano, M. A.; Serrano-Hervas, E.; Romero-Rivera, A.; Iglesias-Fernandez, J.; Osuna, S. *Chem. Commun.* **2018,** *54* (50), 6622-6634.