

Rational prediction of distal activity-enhancing mutations in tryptophan synthase

Miguel A. Maria-Solano,^{[a]#*} Thomas Kinateder^{[b]#} Javier Iglesias-Fernández,^[a] Reinhard Sterner,^{[b]*} and Silvia Osuna^{[a,c]*}

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

[b] Institute of Biophysics and Physical Biochemistry, Regensburg Center for Biochemistry, University of Regensburg, Universitätsstrasse 31, 93053 Regensburg, Germany

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

#: These authors contributed equally to the work

ABSTRACT

Allostery is a central mechanism for the regulation of multi-enzyme complexes. The mechanistic basis that drives allosteric regulation is poorly understood, but harbors key information for enzyme engineering. In the present study, we focus on the tryptophan synthase complex that is composed of TrpA and TrpB subunits, which allosterically activate each other. Specifically, we develop a rational approach for identifying key amino acid residues of TrpB distal from the active site. In particular, we predict positions crucial for shifting the inefficient conformational ensemble of the isolated TrpB to a productive ensemble through intra-subunit allosteric effects. The experimental validation of the new conformationally-driven TrpB design demonstrates its superior stand-alone activity in the absence of TrpA, comparable to those enhancements obtained after multiple rounds of experimental laboratory evolution. Our work evidences that the current challenge of distal active site prediction for enhanced function in computational enzyme design can be ultimately addressed.

INTRODUCTION

Enzymes are some of the most sophisticated biomolecules that exist on Earth. They achieve impressive rate accelerations thanks to their highly preorganized active site pocket, while exhibiting remarkable conformational flexibility key for their function, regulation and evolution.¹⁻⁷ Enzymes are dynamic biological entities, being their catalytic activity directly related to their structure and the broad ensemble of conformations they sample in solution.⁴⁻⁶ This conformational equilibrium can be shifted, for example, by the binding of a ligand to a given site. This in turn influences the binding or the turnover of a substrate at the active site of the enzyme, a phenomenon that is called “allostery”.^{8,9} Likewise, the introduction of an amino acid substitution in the protein sequence not only induces an evident structural change but also a redistribution of the conformational ensemble, which in turn can potentially impact catalytic activity.^{4, 6, 10, 11} Indeed, it has been proven that allosteric effects are not restricted to effector binding, but instead single point mutations or covalent attachment (e.g. phosphorylation), among others can induce similar responses.^{9, 12, 13}

Identifying mutations that modulate enzyme activity is the primary goal of enzyme engineering. One approach to enzyme engineering is Directed Evolution (DE), which has been applied to a myriad of enzyme systems successfully identifying active site and distal mutations, providing access to impressive tailor-made enzyme variants at the expense of large and expensive screening efforts.¹⁴⁻¹⁷ Rational design emerged as an attractive alternative to decrease the screening efforts to a reduced number of promising enzyme variants based on prior structural knowledge and computational approaches.¹⁸⁻²¹ Given the sophisticated nature of enzyme catalysis, multiple computational strategies and protocols have been developed in recent years for computational enzyme design.²⁰ The evaluation of the conformational landscape of enzymes along distinct natural and DE evolutionary pathways has evidenced that the introduced mutations progressively tune the conformational ensemble, stabilizing key conformational states for the novel function.^{4, 6, 10, 20} Of note is that the mutations introduced with DE are often located distal from the active site pocket, which given the vast sequence space are computationally challenging to predict.^{20, 22, 23} In addition to that, the computational prediction of which remote mutations can induce the desired population shift to favor the

key conformational ensemble for novel functionality is an extremely difficult task.²⁰ Our group has recently shown that active site and distal positions targeted by DE can be computationally identified through the coupling of MD simulations with cross correlation methods such as the Shortest Path Map (SPM).^{20, 24} SPM has been applied for identifying DE mutations in the retro-aldolase, monoamine oxidase and tryptophan synthase enzymes suggesting its potential application for smart library construction for enzyme design.^{20, 24}

Tryptophan synthase (TrpS) is an excellent model system for studying allosteric properties. TrpS is a heterodimeric enzyme complex formed by α (TrpA) and β (TrpB) subunits in an $\alpha\beta\beta\alpha$ arrangement. The functional unit is formed by a TrpA and an associated TrpB subunit (**Figure 1a**).^{25, 26} TrpA catalyzes the retro-aldol cleavage of indole-3-glycerol phosphate (IGP) producing glyceraldehyde-3-phosphate (G3P) and indole, which diffuses along an internal tunnel towards the TrpB active site.²⁷ TrpB is a pyridoxal phosphate (PLP) cofactor dependent enzyme that catalyzes the production of L-Tryptophan by condensation of indole and L-serine in a multistep reaction mechanism, which mainly comprises: (1) formation of a Schiff base intermediate (Ain) at the resting state by covalent attachment of PLP cofactor to the catalytic lysine, (2) transamination with L-Ser, (3) indole coupling, and (4) formation of several quinonoid intermediates (Q) to finally release L-Trp. This complex multi-step mechanism involves multiple proton donor/abstraction steps assisted by the catalytic lysine (**Scheme S1**).²⁸ Of relevance is the tight allosteric coupling between TrpA and TrpB along the catalytic itinerary.^{29, 30} TrpA and TrpB catalyze different reactions that are synchronized (i.e. TrpA tunes the TrpB conformational ensemble and vice versa). This fine tuning of the conformational ensemble involves open-to-closed transitions of the rigid COMM domain that forms a lid covering the TrpB active site (**Figure 1b**) and an active site loop of TrpA, as shown by X-ray and computational data.^{26, 31, 32} Given the tight allosteric communication exerted between subunits, both TrpA and TrpB are much less efficient when isolated, which hampers TrpB industrial application for non-canonical amino-acids production.³³⁻³⁸ Arnold and coworkers addressed this limitation by applying DE to optimize activity of TrpB from the TrpS of *Pyrococcus furiosus* for stand-alone function (i.e. recovery of the catalytic activity in the absence of the allosteric protein partner TrpA).^{33, 34} Interestingly, the most evolved variant (*pf*TrpB^{OB2}) was even more efficient than the original *pf*TrpS complex (2.9-fold increase in k_{cat}), and contained 5 out of the 6 mutations located distal from the active site. This manifests that the recovery of activity exerted by the distal mutations is induced through allosteric effects.^{33, 34} Intrigued by the allosteric regulation induced by distal mutations, we explored the conformational energy landscape of the *pf*TrpS enzyme complex, the *pf*TrpB isolated enzyme and the stand-alone *pf*TrpB^{OB2} evolved variant.³¹ Free energy calculations revealed that the DE mutations in *pf*TrpB^{OB2} recovered the allosterically driven conformational ensemble of the *pf*TrpS complex, allowing the exploration of open, partially closed and closed conformations of the COMM domain, which is required for the multi-step catalytic pathway. The *pf*TrpB stand-alone activity was thus achieved through the recovery of the conformational ensemble present in the *pf*TrpS complex. In fact, the allosterically driven conformational ensemble was not only recovered but also improved, as a higher stability of catalytically productive closed states was found in the case of *pf*TrpB^{OB2}. This explained the *pf*TrpB^{OB2} superior activity with respect to the *pf*TrpS complex. In contrast, isolated *pf*TrpB showed a restricted COMM domain conformational heterogeneity and catalytically unproductive closed states. Careful analysis of the *pf*TrpS conformational ensemble through SPM correlation-based tools elucidated the enzyme pathways most contributing to the TrpS conformational dynamics, which interestingly included some important DE positions.^{20, 31} This suggests that the identified positions with SPM can potentially alter the enzyme conformational dynamics, and thus its stand-alone activity. However, multiple positions are identified and there is a lack of information on which specific amino-acid substitution should be introduced for achieving an efficient conformational ensemble for stand-alone function.

An orthogonal *in silico* method to analyze functional transitions in enzyme evolution is ancestral sequence reconstruction (ASR).³⁹⁻⁴¹ In a previous work, we reconstructed the TrpS phylogenetic tree and identified a shift in the allosteric modulation exerted by TrpA on TrpB activity.^{42, 43} The analysis of the steady state kinetic parameters

of the last bacterial common ancestor (LBCA) revealed high stand-alone activity of LBCA-TrpB and its allosteric inhibition in the presence of TrpA. Along the phylogenetic tree, this inhibition was gradually inverted towards allosteric activation existing in modern TrpB (**Figure 1c**).

This inversion of the allosteric effect exerted by TrpA on TrpB between ANC2 and ANC3 provides a perfect starting point for an SPM-based design. Specifically, we wanted to identify residues within the allosteric network of TrpB that are able to rescue the missing allosteric activation from TrpA and predict mutations that convey stand-alone function in the context of the inefficient ANC3 TrpB. To this end, we intended to explore the conformational ensemble of the stand-alone LBCA TrpB enzyme system, and identify key positions by means of our developed SPM correlation-based tool. Sequence comparison of the identified positions along the phylogenetic tree further reduces the number of potential mutations and provides the specific amino-acid substitutions for stand-alone function. This approach decreases the experimental screening to one single mutant and includes the rational prediction of both active site and distal mutations. Our study presents a computational enzyme design approach that is not restricted to active site mutations and demonstrates that the challenge to rationally predict distal mutations can be ultimately addressed by exploring the conformational energy landscape of enzymes in combination with cross correlation and bioinformatic tools.

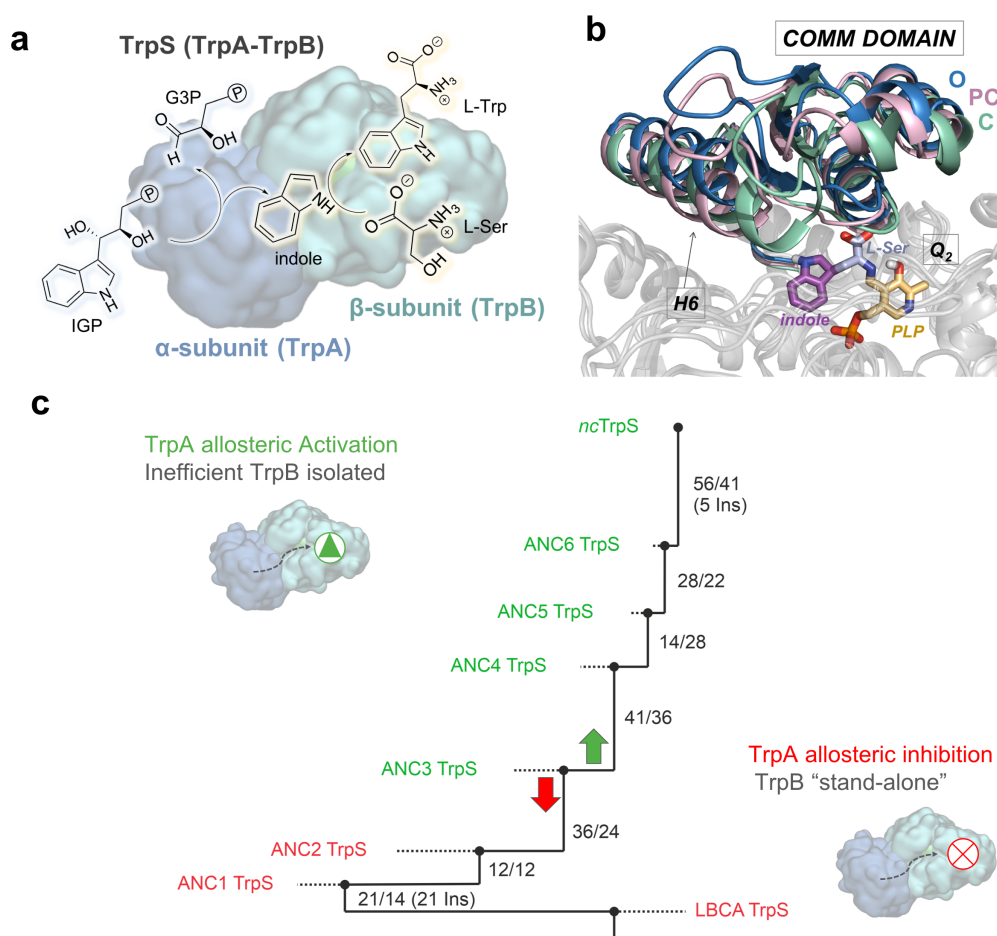


Figure 1. Overview of Tryptophan Synthase (TrpS) enzyme. (a) The functional unit of TrpS consists of a heterodimer, which is formed by TrpA (blue) and TrpB (green). TrpA catalyzes the cleavage of indole-3-glycerol phosphate (IGP) to glyceraldehyde-3-phosphate (G3P) and indole, which in TrpB reacts with activated L-Ser in a multistep mechanism to yield L-Trp (see **Scheme S1**). (b) Overlay of *pf*TrpS metastable conformations from previous computational exploration showing the transition of the COMM domain (residues 97-184) from an open (blue, O), to a partially closed (pink, PC) to a closed conformation (green, C). Highlighted are the α -helix H6 of the COMM domain (residues 174-164) and the reaction intermediate Q_2 in the active site, which is colored as a

function of its origin molecule (PLP cofactor in orange, L-Ser in blue and indole in purple).³¹ c, The phylogenetic tree shows the path from the last bacterial common ancestor (LBCA) TrpS over six intermediate nodes (ANC1 TrpS to ANC6 TrpS) to the extant *Neptuniibacter caesariensis* TrpS.⁴² Numbers next to each edge indicate the number of mutations accumulated in TrpA and TrpB with respect to the previous node. While LBCA-TrpB gets deactivated by TrpA and exhibits stand-alone function, the allosteric effect of TrpA is reverted along the phylogenetic tree with a switch between ANC2 TrpB and ANC3 TrpB to an allosteric activation, as observed in extant *nc*TrpB.

RESULTS

Reconstruction of Ancestral TrpS conformational ensembles

As shown in previous studies, natural evolution has altered the need of TrpS to be allosterically regulated.⁴² As opposed to modern TrpB, the ancestral LBCA TrpB was found to operate less efficiently (in terms of k_{cat}) in the presence of TrpA.⁴³ The allosteric inhibition imparted by TrpA suggests that the ancestral TrpB in complex is less efficient in accessing the catalytically productive conformational states required for enhanced activity.³¹ Interestingly, LBCA TrpB affinity towards L-Ser substrate was enhanced in the heterocomplex form (**Table S1**). To provide the molecular basis for stand-alone activity and higher affinity towards L-Ser, we decided to computationally reconstruct the free energy landscape (FEL) of LBCA TrpB in the presence (i.e. heterocomplex TrpS) and absence of TrpA (**Figure 2**). We employed metadynamics simulations to reconstruct the FEL associated with the open-to-closed transition of the COMM domain (see **Figure S1**) at the resting state (i.e. E(Ain)) and at the Q₂ intermediate (i.e. quinonoid intermediate formed after indole coupling, see **Scheme S1**). The reconstructed FEL of the LBCA TrpB(Ain) in the absence of TrpA, indicates that TrpB(Ain) mostly visits partially closed (PC) conformational states of the COMM domain (**Figure 2a**). This is altered in the presence of TrpA, which clearly induces a shift in the FEL stabilizing open (O) states with similar deviations from the reference path (**Figure 2a** and **2b**, on the left). At the resting state, closed (C) states are inaccessible for both systems. The analysis of the access tunnels to the active site for L-Ser binding through CAVER calculations (see **Figure 2c** and **Figure S2**) indicates that the PC conformational ensemble of the isolated LBCA TrpB has a substantially narrower tunnel bottleneck than the accessible O states of the complex. This finding indicates that the O conformational ensemble improves L-Ser accessibility to the active site, thus explaining the enhanced $K_M^{\text{L-Ser}}$ values displayed by the LBCA TrpS complex.

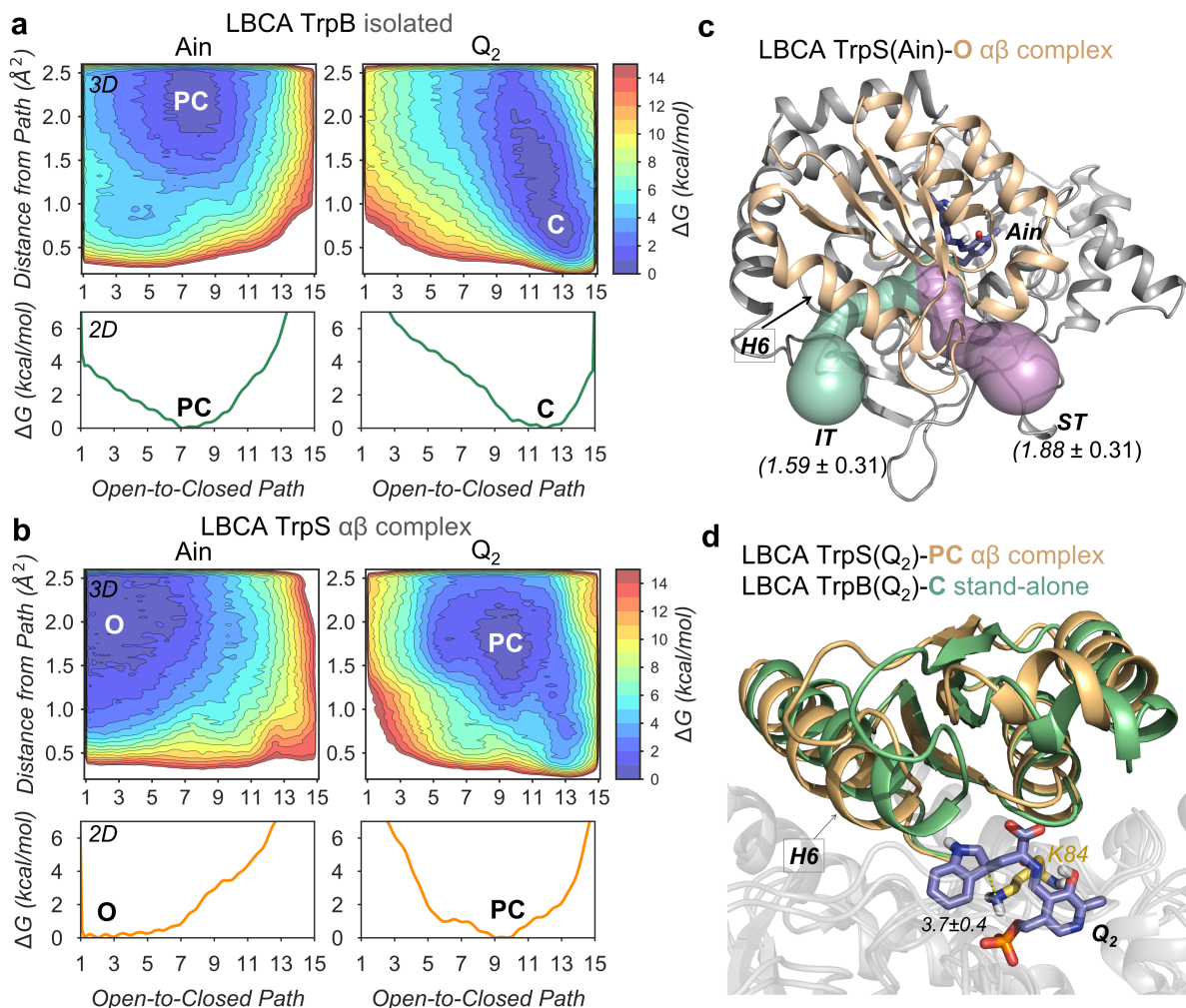


Figure 2. Computational exploration of the LBCA conformational ensemble. Free energy landscape (FEL) associated with the COMM domain open-to-closed (O-to-C) conformational transition of the LBCA TrpB (a) and LBCA TrpS (b) at Ain and Q_2 reaction intermediates. The x-axis corresponds to the progression along the reference O-to-C path generated from X-Ray data, while the y-axis to the mean square deviation (MSD) distance from the reference path c, Tunnels access of LBCA TrpS-O state at Ain reaction intermediate for the L-Ser substrate, computed with CAVER 3.0. The averaged bottleneck radii (in \AA) for the internal TrpA-TrpB tunnel (IT, green) and the secondary tunnel (ST, violet) found are also shown. (d) Overlays of the metastable conformations of the partially closed (PC) state of LBCA-TrpS (orange) and the closed (C) state (green) of LBCA-TrpB at Q_2 reaction intermediate. The catalytic proton transfer distance (in \AA) between the K84 (yellow) residue and the Q_2 reaction intermediate (slate) is also represented.

More interesting is the fact that TrpA was found to inhibit the TrpB catalytic efficiency, as isolated LBCA TrpB displays a ca. 8.4-fold k_{cat} higher value. As we show in our previous study,³¹ the catalytic activity of TrpS can be estimated by evaluating its ability to visit catalytically competent C states of the COMM domain. The catalytically-relevant closed conformational ensemble displays an efficient active site preorganization by means of optimized non-covalent interaction networks and short catalytic distances between the Q_2 intermediate and the conserved catalytic K84 that acts as proton acceptor. In particular, the H6 COMM domain helix was found to play an important role in the closure to form non-covalent interactions with the indole moiety of Q_2 . In the present work, the reconstructed FEL associated to the COMM domain open-to-closed (O-to-C) transition for LBCA TrpB (Figure 2a and d) indicates that at the Q_2 intermediate, the catalytically productive C conformational ensemble is indeed accessible for efficient catalysis. The structural characterization of the visited C conformational states shows LBCA TrpB

adopts catalytically productive COMM domain closure with appropriate K84-Q2 proton transfer distances (**Figure 2d** and **Figure S3** and **S4**), as discussed above. This evidences that LBCA TrpB has stand-alone properties derived from the exploration of stable catalytically competent C conformations in the absence of TrpA. On the contrary, LBCA TrpA alters the conformational landscape of TrpB as it induces a shift towards PC conformations hampering the ability of the COMM domain to complete the O-to-C transition for achieving catalytically productive C states (**Figure 2b**). As expected, PC conformations of LBCA TrpB in the presence of TrpA do not exhibit a competent closure of the COMM domain, in particular this is notorious for the H6 region. Besides, the K84-Q2 proton transfer distances are larger (**Figure 2d** and **Figure S3**). In summary, our results indicate that the destabilization of the competent C LBCA TrpB ensemble is the main responsible of the allosteric inhibition exerted by the LBCA TrpA protein partner. It is worth mentioning that we estimated a similar effect (i.e. destabilization of the competent C ensemble) for the allosteric inhibition exerted by pfTrpA on the laboratory-evolved stand-alone pfTrpB0B2. Another interesting aspect of LBCA conformational dynamics is its limited conformational heterogeneity (i.e. a narrow set of states are sampled), especially if compared with the previously studied allosteric pfTrpS complex and the laboratory-evolved stand-alone pfTrpB0B2 catalyst. A high degree of conformational heterogeneity was observed for the latter cases, which explored the complete O-to-C transition at Q2 intermediate. The lack of O states of the COMM domain at the Q2 intermediate for LBCA-TrpB suggests a more rigid COMM as the reaction evolves, and an infrequent transition towards O state, thus suggesting that product release might be rate limiting.

Computational prediction of distal active site mutations for stand-alone function

The mutations introduced along an evolutionary pathway progressively tune the conformational ensemble of enzymes towards novel function.^{4,6,10,20} In this context, distal active site mutations have been shown to play a crucial role in natural and laboratory evolvability.^{12,22} Their prediction considering the vast protein sequence space that yields a targeted function is, however, an extremely challenging task in computational enzyme design.²⁰ We have recently reported that molecular dynamics coupled to correlation-based tools are promising methodologies for the identification of both active site and distal positions targeted in non-rational laboratory evolution experiments.^{20,24} In particular, we successfully developed and applied the Shortest Path Map (SPM) method for identifying the enzyme pathways that most contribute to the conformational dynamics of the *pfTrpS* enzyme. Of relevance is that the identified positions contained or make persistent non-covalent interactions with residues targeted in the laboratory evolution of the *pfTrpS* enzyme for stand-alone function.³¹ SPM identifies important positions for the enzyme conformational dynamics, thus reducing the potential number of mutational hotspots.

Inspired by the previous work on the TrpS ancestral reconstruction, we focused our computational design on the ancestral ANC3 TrpB scaffold.^{42,43} This enzyme corresponds to the third node of the phylogenetic tree and exhibits reversion of allosteric inhibition towards activation along the evolution pathway (**Figure 1c**). In other words, ANC3 TrpB is the first enzyme that is allosterically dependent on TrpA, thus being highly inefficient as stand-alone catalyst (**Table S2**). The absence of TrpA decreases ANC3 TrpB activity 30.2-fold in terms of k_{cat} , suggesting a reduced conformational O-to-C ensemble. Given the success of SPM in identifying key positions for the enzyme conformational dynamics, we decided to apply our computational methodology to rationally engineer an ANC3 TrpB catalyst towards stand-alone activity. Our initial reference protein was LBCA TrpB, as it exhibits stand-alone properties thanks to its ability to adopt stable and efficient closed states of the COMM domain. The SPM analysis of the LBCA TrpB SPM identified 74 possible hotspots that potentially regulate the enzyme conformational dynamics (74 out of 413 residues, i.e. 18% of the total enzyme). This number is too large for an efficient rational design of ANC3 TrpB, as it is unclear, which positions should be targeted and which substitutions should be introduced to establish stand-alone function. We solved this problem by analyzing the sequence conservation between LBCA TrpB and the targeted ANC3 TrpB system for the 74 SPM positions (see the workflow followed in **Figure 3**). Comparing the residues at each of the 74 SPM positions reduced the number of sites to 6 and specified the nature of the mutation to the amino acid found in LBCA TrpB. Interestingly, 5 out of 6 positions were located far away from the active site.

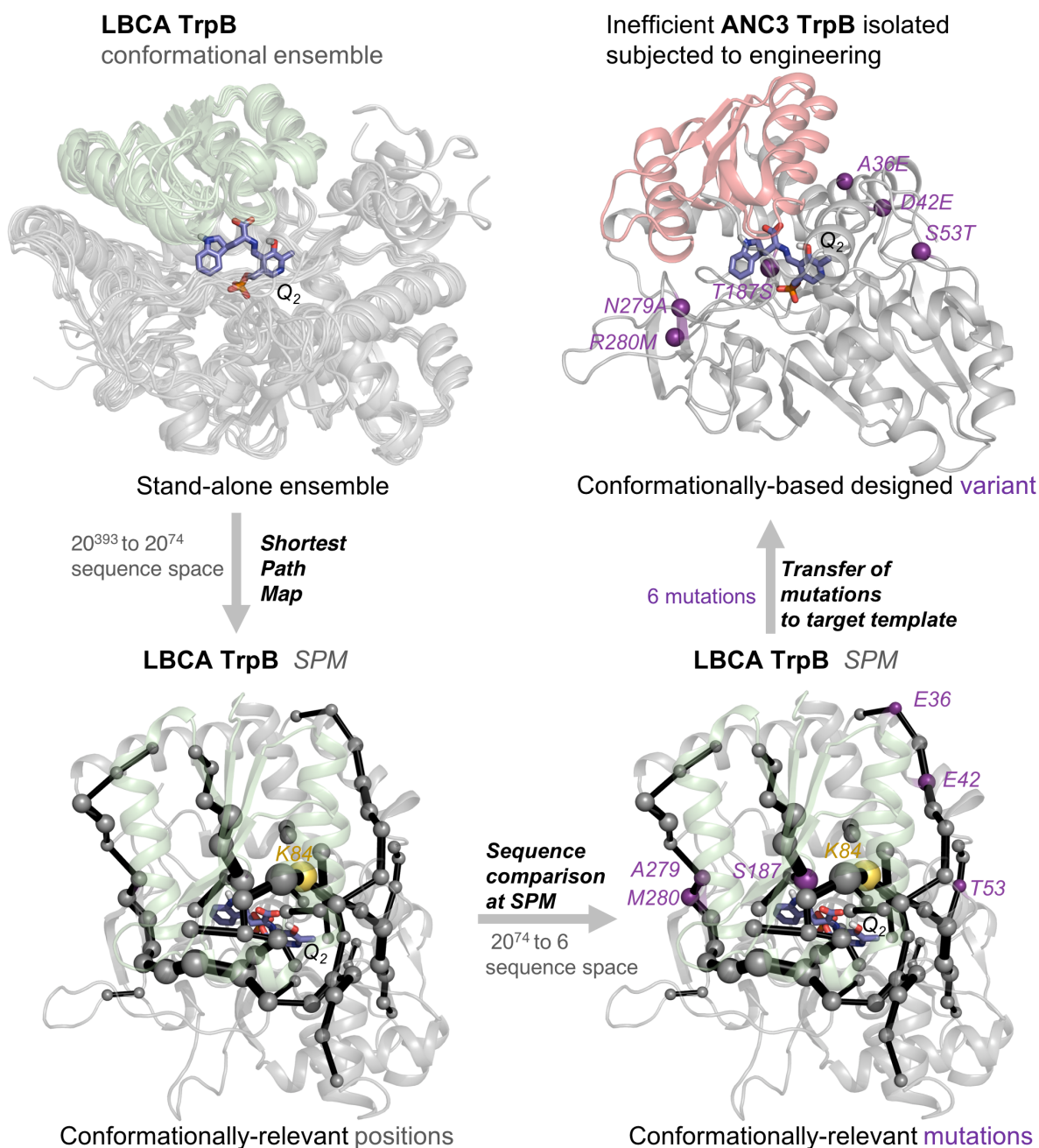


Figure 3. SPM-based computational workflow for the rational design of SPM6 TrpB enzyme variant. By analyzing the conformational ensemble of the stand-alone LBCA TrpB with high catalytic activity (upper left ensemble) through the SPM, we identified positions (grey spheres, lower left structure) within allosteric pathways (black edges) in the enzyme that most contribute to the LBCA TrpB conformational dynamics in the Q2 intermediate. Thereby the size of each edge and node corresponds to the relevance for conformational dynamics; catalytic K84 is highlighted in yellow. Excluding residues that do not participate in an allosteric pathway reduces the sequence space from 20^{393} to 20^{74} possible activity enhancing substitutions. Sequence comparison at the SPM positions between stand-alone LBCA TrpB and inefficient ANC3 TrpB reduces the sequence space to 6 mutations with respect to LBCA TrpB (lower right structure, purple residues), that were introduced into ANC3 TrpB (upper right structure, purple residues) and tested *in-vitro*.

Rational SPM-based ANC3 TrpB designs for stand-alone function

The application of the SPM method coupled to sequence comparison between two variants exhibiting rather high (LBCA TrpB) or low (ANC3 TrpB) stand-alone function reduced the SPM library to only 6 specific mutations in ANC3 TrpB: A56E, D62E, S73T, T207S, N299A and R300M. This ANC3 variant was termed SPM6 TrpB. Interestingly, none of the mutations are located at the COMM domain, 5 out of 6 mutations are located far away from the active site (*ca.* 18-29 Å), among which N299A and R300M are near the TrpA-TrpB protein interface and only S73T is located at the active site pocket (**Figure 3** and **Figure S5**). The computational screening of ANC3 TrpB, the ANC3 TrpS and the SPM6 TrpB enzyme variant by means of conventional molecular dynamics simulations suggested that both SPM6 TrpB and ANC3 TrpS are able to retain the closed conformation of the COMM domain. In contrast, isolated ANC3 TrpB explores additional non-productive conformations (**Figure S6**). This fast screening computational protocol suggests a rather low stability of the C state of the COMM domain in isolated ANC3 TrpB, which explains its low stand-alone catalytic activity (**Figure 4a**). These computational insights encouraged us to experimentally test the SPM6 enzyme variant. As shown in **Figure 4** and **Table S3**, SPM6 TrpB successfully enhances the catalytic activity with respect to ANC3 TrpB *by almost* one order of magnitude (7-fold increase in k_{cat}). The catalytic efficiencies for both, L-Ser and indole are also improved. It is worth emphasizing that a similar fold increase in stand-alone catalytic activity was achieved in *pf*TrpB by means of multiple rounds of laboratory evolution.³³ Our SPM-based computational approach therefore provides the same order of improvement in stand-alone activity but by only testing one single rationally designed variant. Another interesting aspect to evaluate is whether the SPM6 mutations have an impact in the allosteric modulation exerted by TrpA. The catalytic activity of the ancestral ANC3 TrpB increases 30.2-fold in terms of k_{cat} thanks to the TrpA-triggered allosteric activation. Unexpectedly, the introduction of SPM6 mutations to ANC3 TrpB confer increased stand-alone activity while still retaining some TrpA allosteric activation (the activity of SPM6 TrpB is enhanced 5.5-fold in the presence of TrpA). This indicates that the SPM6 distal mutations tune the O-to-C conformational ensemble of SPM6 TrpB through long-range intra-subunit allosteric effects but these changes in the conformational landscape do not prevent TrpA allosteric activation. In fact, the combination of both inter-subunit and intra-subunit allosteric effects yields SPM6 TrpS complex displaying even higher catalytic activity than the ancestral ANC3 TrpS complex (i.e. 1.3-fold increase, **Figure 4a**).

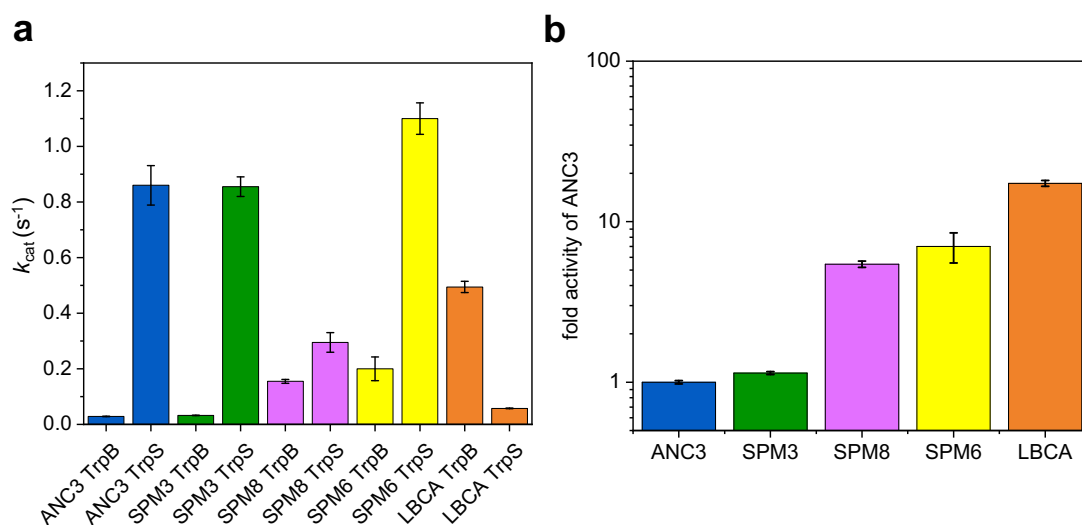


Figure 4. Illustration of the TrpB kinetic characterization. (a) Activity changes of ANC3, SPM3, SPM8, SPM6 and LBCA isolated TrpB enzymes upon complexation with their corresponding TrpA protein partners, in terms of the average values for k_{cat} . (b) TrpB fold activity at logarithmic scale of ANC3 respect to SPM3, SPM8, SPM6 and LBCA. The new TrpB designs SPM6 and SPM8 are 7 and 5.4-fold more active than ANC3 TrpB, while the

reference stand-alone LBCA TrpB 17.4-fold. Errors bars indicate the standard deviation observed in two separate experiments.

In order to further test the SPM predictive power and the robustness of the strategy followed so far, we additionally targeted two other SPM based approaches. In the first one, we followed the same workflow as for the SPM6 design but used instead of LBCA TrpB the isolated ANC2 TrpB as stand-alone reference protein for the SPM pathway analysis. After identifying the shortest path map in ANC2 TrpB and subsequent sequence comparison between ANC2 TrpB and ANC3 TrpB we identified 3 SPM positions and predicted 3 mutations in ANC3 TrpB: S73T, N299S and R300M. The corresponding variant was termed SPM3. This reduced number of non-conserved SPM positions makes sense since ANC2 and ANC3 are closer in the phylogenetic tree than LBCA and ANC3.

In the second approach, we conducted a SPM analysis on the ANC3 TrpS complex. The rationale behind taking this complex as a reference was that, while isolated ANC3 TrpB is poorly active and its allosteric communication is likely truncated, complexation with TrpA leads to high activity and likely a restored allosteric network. After identifying allosterically relevant SPM positions within ANC3 TrpS, we again compared the ANC3 TrpB sequence to LBCA TrpB in order to predict mutations that lead to a stand-alone catalyst. Following this protocol, we identified two extra positions as compared to SPM6 (SPM8): R53N and M187I, where R53N is far away the active site and M187I is located at the H6 helix of the COMM domain. It should be noted that the 3 positions of SPM3 and 6 out of 8 of SPM8 were previously identified in SPM6. Following the same computational MD-based screening protocol SPM3 and SPM8 TrpB variants were analyzed, which suggested a rather high stability of the C state of the COMM domain for enhanced activity (**Figure S6**). The experimental validation of the computational predictions for both SPM3 and SPM8 variants revealed enhanced catalytic activities of ANC3 TrpB when isolated. SPM8 TrpB exhibited a similar activity enhancement (5.4-fold in terms of k_{cat}) as SPM6 TrpB, whereas a quite modest enhancement was obtained for SPM3 TrpB in line with the reduced number of mutations (1.1-fold). Regarding the inter allosteric effects exerted by TrpA, SPM3 and SPM8 variants also showed TrpA allosteric activation. In particular, SPM3 showed a similar degree of the k_{cat} increase in complex as ANC3 (26.3-fold), while SPM6 (5.5-fold) and SPM8 (1.9-fold) enzyme variants present a reduced predisposition to the TrpA allosteric activation (**Figure 4**).

DISCUSSION

Allosteric regulation is a central biological process focused on the functional connection between distinct sites on either a single biological entity or among complex multimeric structures.^{9, 12, 44} This regulation of enzymatic function is not limited to effector or protein partner binding, as similar effects have been observed by covalent attachment or by introducing mutations located at distal positions of the enzyme active site.^{9, 12} The elucidation of the underlying mechanism and forces that drive allosteric regulation has the enormous potential of identifying key positions for regulating enzymatic function, which could be exploited in enzyme design.²⁰ The present study indeed demonstrates that distal active site positions, regulating the allosterically-driven conformational ensemble and thus the enzyme activity, can be successfully identified by means of correlation-based tools and sequence comparison. Given the vast sequence and conformational space, the rational prediction of mutations, especially those located at remote positions from the active site impacting enzymatic function is an extremely difficult task in the computational enzyme design field. Apart from that, the identification of the amino acid substitutions that optimize the enzyme conformational ensemble for a targeted enzyme function is extremely challenging. Our study focuses on the engineering of stand-alone function taking advantage of the substantial allosteric contributions that distal mutations were exerting on the laboratory-evolved variants.^{31, 33, 34} The exploration of the free energy landscape of the ancestrally reconstructed LBCA TrpS in complex and as stand-alone catalyst (LBCA TrpB), together with our previous findings³¹ on the wild-type *pf*TrpS complex, isolated *pf*TrpB and laboratory-evolved *pf*TrpB^{OB2} has elucidated the conformational ensemble that a stand-alone catalyst has to display for being efficient. This information is pivotal for fine-tuning the conformational ensemble and progressing towards the targeted enzyme design goal. We find that

LBCA TrpB naturally adopts a stable catalytically productive COMM domain closure, which is hampered by the presence of the LBCA TrpA protein partner. LBCA TrpA therefore induces an allosteric inhibition of LBCA TrpB activity, which contrasts with the TrpA allosteric activation usually found in modern TrpB enzymes. In this study we exploit the intrinsic ability of LBCA TrpB to efficiently stabilize catalytically competent COMM domain closed conformations when isolated (i.e. crucial for stand-alone properties), and develop a novel computational enzyme design approach for achieving stand-alone function. In particular, we apply our SPM method to identify the enzyme pathways and positions that most contribute to the LBCA TrpB conformational dynamics. We hypothesized that these conformationally-relevant SPM positions could be potential hotspots for tuning the conformational ensemble of TrpA-dependent TrpB enzymes. The reconstruction of the phylogenetic tree from LBCA TrpS to the modern *nc*TrpS provided an intermediate variant ANC3 TrpB, which exhibits a high allosteric activation from ANC3 TrpA (i.e. ANC3 TrpB is highly inefficient when isolated). The application of SPM into LBCA TrpB reduced the sequence space from 393 to 74 hotspots, suggesting that *ca.* 18% of the positions play a conformationally-relevant role. However, this still leads to a massive amount of enzyme variants to screen. Interestingly, the analysis of sequence conservation at the identified SPM positions between LBCA and ANC3 TrpB templates reduced this large number to only 6 positions. This approach assumes that the transfer of the non-conserved conformationally-relevant SPM mutations from the LBCA to the targeted ANC3 TrpB template will alter the enzyme conformational dynamics and induce the stabilization of the catalytically relevant closed state of the COMM domain. It is worth mentioning that among these 6 mutations 5 are distal from the active site and none is included in the COMM domain.

The fast-computational screening of the rationally designed enzyme including these 6 mutations indicated that SPM6 better stabilizes the closed conformational ensemble than the parent ANC3. Indeed, the experimental evaluation of SPM6 indicated the introduced mutations boosted the stand-alone catalytic activity of the inefficient isolated ANC3 TrpB enzyme near one order of magnitude. This enhancement by only testing a single variant is comparable to that observed for the laboratory evolved *pf*TrpB^{OB2} after three rounds of DE, that involved the screening of *ca.* 3,080 variants.³³ The observed enhancement of ANC3 TrpB stand-alone activity still does not completely recover the activity displayed by the ANC3 TrpS complex. The new SPM6 designed variant enhances the low initial 3% activity displayed by ANC3 TrpB up to *ca.* 23%. It should be also mentioned that the SPM6 design is based on the template scaffold LBCA-TrpB, whose catalytic activity is lower than that of ANC3 TrpS complex (LBCA TrpB activity is *ca.* 58% that of ANC3 TrpS). In the case of the DE *pf*TrpB^{OB2} enzyme variant, a 300 % of activity recovery was observed.³³

The partial recovery observed for SPM6, is in part due to the dramatic loss of activity displayed by ANC3 TrpB in the absence of TrpA (97% of activity loss), which is more moderate in *pf*TrpB (69%). These numbers indicate that the total recovery of ANC3 activity is more demanding from an engineering point of view, and suggest that the new generated SPM6 variant still presents some predisposition towards TrpA regulation. This evidences that the distal mutations introduced in SPM6 variant successfully enhanced the stand-alone activity of ANC3 TrpB activity through intra-subunit allosteric effects, however, they did not completely free TrpB from the inter-subunit allosteric regulation exerted by TrpA. To our surprise, SPM6 in complex with TrpA showed the most efficient turnover tested in this work, which indicates that the combination of intra- and inter-allosteric effects can operate synergistically to successfully tune the O-to-C conformational ensemble and achieve high catalytic efficiencies.

Another secondary insight gained from this work comes from the analysis of how the TrpS conformational landscape is altered and conserved along the natural evolutionary pathway. The exploration of the conformational ensemble and the identification of the key conformationally-relevant SPM positions of LBCA, ANC2 and ANC3 phylogenetic nodes and their comparison with the previously studied modern *pf*TrpS revealed that the main allosteric pathways are not significantly altered along evolution. Indeed, the comparison of the generated SPM paths for the different enzymes reveals a rather high number of shared positions, thus suggesting similar TrpB correlated motions among variants. The conservation of the conformationally-relevant positions also explains the common

positions targeted in the different SPM-based strategies for SPM3 and SPM8 designs. Our findings reinforce the original approach based on the LBCA SPM analysis as a robust computational strategy that could be exploited for the rational engineering of TrpB enzyme variants either for improved stand-alone or in complex function. It also evidences that conformational heterogeneity and, in particular, the use of ancestral conformationally-rich scaffolds corresponds to a successful strategy for designing desired enzymatic functions.^{41, 45}

CONCLUSIONS

The approach presented in this work highlights that the exploration of the enzyme conformational ensemble is essential for successful computational enzyme design. The detection of the key conformationally-relevant positions and the combined analysis of its conservation along ancestral phylogenetic trees harbors meaningful information for solving the current challenge in computational enzyme design of distal active site prediction for enhanced function.

ASSOCIATED CONTENT

Supporting information

Computational and experimental methods as well as additional tables and figures are provided in SI.

AUTHOR INFORMATION

Corresponding Author

*miguelangel.maria@udg.edu

*silvia.osuna@udg.edu

[* Reinhard.Sterner@biologie.uni-regensburg.de](mailto:*Reinhard.Sterner@biologie.uni-regensburg.de)

ORCID

Miguel A. Maria-Solano: 0000-0002-7837-0429

Silvia Osuna: 0000-0001-9458-1114

Reinhard Sterner: 0000-0001-8177-8460

ACKNOWLEDGMENT

We thank the Generalitat de Catalunya for the emerging group CompBioLab (2017 SGR-1707) and Spanish MINECO for project PGC2018-102192-B-I00. 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) and Juan de la Cierva-Incorporación fellowship (IJCI-2017-34129). 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), and the Human Frontier Science Program (HFSP) for project grant RGP0054/2020. We thank Sonja Fuchs, Sabine Laberer, Christiane Endres and Jeannette Ueckert for excellent technical assistance.

REFERENCES

1. Benkovic, S. J.; Hammes-Schiffer, S., A Perspective on Enzyme Catalysis. *Science* **2003**, *301* (5637), 1196-1202.
2. Hammes, G. G.; Benkovic, S. J.; Hammes-Schiffer, S., Flexibility, Diversity, and Cooperativity: Pillars of Enzyme Catalysis. *Biochemistry* **2011**, *50* (48), 10422-10430.

3. Marti, S.; Roca, M.; Andres, J.; Moliner, V.; Silla, E.; Tunon, I.; Bertran, J., Theoretical insights in enzyme catalysis. *Chem. Soc. Rev.* **2004**, *33* (2), 98-107.
4. Maria-Solano, M. A.; Serrano-Hervas, E.; Romero-Rivera, A.; Iglesias-Fernandez, J.; Osuna, S., Role of conformational dynamics in the evolution of novel enzyme function. *Chem. Commun.* **2018**, *54* (50), 6622-6634.
5. Boehr, D. D.; Nussinov, R.; Wright, P. E., The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **2009**, *5* (11), 789-796.
6. Petrović, D.; Risso, V. A.; Kamerlin, S. C. L.; Sanchez-Ruiz, J. M., Conformational dynamics and enzyme evolution. *J. R. Soc. Interface* **2018**, *15* (144).
7. Tokuriki, N.; Tawfik, D. S., Protein Dynamism and Evolvability. *Science* **2009**, *324* (5924), 203-207.
8. Gunasekaran, K.; Ma, B.; Nussinov, R., Is allostery an intrinsic property of all dynamic proteins? *Proteins* **2004**, *57* (3), 433-443.
9. Lisi, G. P.; Loria, J. P., Allostery in enzyme catalysis. *Curr. Opin. Struct. Biol.* **2017**, *47*, 123-130.
10. Campbell, E. C.; Correy, G. J.; Mabbitt, P. D.; Buckle, A. M.; Tokuriki, N.; Jackson, C. J., Laboratory evolution of protein conformational dynamics. *Curr. Opin. Struct. Biol.* **2018**, *50*, 49-57.
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., The role of distant mutations and allosteric regulation on LovD active site dynamics. *Nat. Chem. Biol.* **2014**, *10* (6), 431-436.
12. Nussinov, R.; Tsai, C. J.; Ma, B., The underappreciated role of allostery in the cellular network. *Annu. Rev. Biophys.* **2013**, *42*, 169-89.
13. Lee, J.; Goodey, N. M., Catalytic contributions from remote regions of enzyme structure. *Chem. Rev.* **2011**, *111* (12), 7595-624.
14. Qu, G.; Li, A.; Acevedo-Rocha, C. G.; Sun, Z.; Reetz, M. T., The Crucial Role of Methodology Development in Directed Evolution of Selective Enzymes. *Angew. Chem. -Int. Ed. Engl.* **2020**, *59* (32), 13204-13231.
15. Zeymer, C.; Hilvert, D., Directed Evolution of Protein Catalysts. *Annu. Rev. Biochem.* **2018**, *87* (1), 131-157.
16. Arnold, F. H., Innovation by Evolution: Bringing New Chemistry to Life (Nobel Lecture). *Angew. Chem. -Int. Ed. Engl.* **2019**, *58* (41), 14420-14426.
17. Hauer, B., Embracing Nature's Catalysts: A Viewpoint on the Future of Biocatalysis. *ACS Catal.* **2020**, *10* (15), 8418-8427.
18. Damborsky, J.; Brezovsky, J., Computational tools for designing and engineering enzymes. *Curr. Opin. Chem. Biol.* **2014**, *19*, 8-16.
19. Ebert, M. C.; Pelletier, J. N., Computational tools for enzyme improvement: why everyone can – and should – use them. *Curr. Opin. Chem. Biol.* **2017**, *37*, 89-96.
20. Osuna, S., The challenge of predicting distal active site mutations in computational enzyme design. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2020**, e1502.
21. Świderek, K.; Tuñón, I.; Moliner, V., Predicting enzymatic reactivity: from theory to design. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4* (5), 407-421.
22. Currin, A.; Swainston, N.; Day, P. J.; Kell, D. B., Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chem. Soc. Rev.* **2015**, *44* (5), 1172-1239.
23. Morley, K. L.; Kazlauskas, R. J., Improving enzyme properties: when are closer mutations better? *Trends Biotechnol.* **2015**, *23* (5), 231-237.
24. Romero-Rivera, A.; Garcia-Borràs, M.; Osuna, S., Role of Conformational Dynamics in the Evolution of Retro-Aldolase Activity. *ACS Catal.* **2017**, *7* (12), 8524-8532.
25. Hyde, C. C.; Ahmed, S. A.; Padlan, E. A.; Miles, E. W.; Davies, D. R., Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from *Salmonella typhimurium*. *J. Biol. Chem.* **1988**, *263* (33), 17857-71.

26. Hioki, Y.; Ogasahara, K.; Lee, S. J.; Ma, J.; Ishida, M.; Yamagata, Y.; Matsuura, Y.; Ota, M.; Ikeguchi, M.; Kuramitsu, S.; Yutani, K., The crystal structure of the tryptophan synthase beta subunit from the hyperthermophile *Pyrococcus furiosus*. Investigation of stabilization factors. *Eur. J. Biochem* **2004**, *271* (13), 2624-35.
27. Fleming, J. R.; Schupfner, M.; Busch, F.; Baslé, A.; Ehrmann, A.; Sterner, R.; Mayans, O., Evolutionary Morphing of Tryptophan Synthase: Functional Mechanisms for the Enzymatic Channeling of Indole. *J. Mol. Biol.* **2018**, *430* (24), 5066-5079.
28. Barends, T. R. M.; Domratcheva, T.; Kulik, V.; Blumenstein, L.; Niks, D.; Dunn, M. F.; Schlichting, I., Structure and mechanistic implications of a tryptophan synthase quinonoid intermediate. *ChemBioChem* **2008**, *9* (7), 1024-1028.
29. Dunn, M. F., Allosteric regulation of substrate channeling and catalysis in the tryptophan synthase holoenzyme complex. *Arch. Biochem. Biophys.* **2012**, *519* (2), 154-166.
30. Niks, D.; Hilario, E.; Dierkers, A.; Ngo, H.; Borchardt, D.; Neubauer, T. J.; Fan, L.; Mueller, L. J.; Dunn, M. F., Allostery and Substrate Channeling in the Tryptophan Synthase Holoenzyme Complex: Evidence for Two Subunit Conformations and Four Quaternary States. *Biochemistry* **2013**, *52* (37), 6396-6411.
31. Maria-Solano, M. A.; Iglesias-Fernández, J.; Osuna, S., Deciphering the Allosterically Driven Conformational Ensemble in Tryptophan Synthase Evolution. *J. Am. Chem. Soc.* **2019**, *141* (33), 13049-13056.
32. Lee, S. J.; Ogasahara, K.; Ma, J. C.; Nishio, K.; Ishida, M.; Yamagata, Y.; Tsukihara, T.; Yutani, K., Conformational changes in the tryptophan synthase from a hyperthermophile upon alpha(2)beta(2) complex formation: Crystal structure of the complex. *Biochemistry* **2005**, *44* (34), 11417-11427.
33. Buller, A. R.; Brinkmann-Chen, S.; Romney, D. K.; Herger, M.; Murciano-Calles, J.; Arnold, F. H., Directed evolution of the tryptophan synthase beta-subunit for stand-alone function recapitulates allosteric activation. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (47), 14599-14604.
34. Buller, A. R.; van Roye, P.; Cahn, J. K. B.; Scheele, R. A.; Herger, M.; Arnold, F. H., Directed Evolution Mimics Allosteric Activation by Stepwise Tuning of the Conformational Ensemble. *J. Am. Chem. Soc.* **2018**, *140* (23), 7256-7266.
35. Romney, D. K.; Murciano-Calles, J.; Wehrmuller, J. E.; Arnold, F. H., Unlocking Reactivity of TrpB: A General Biocatalytic Platform for Synthesis of Tryptophan Analogues. *J. Am. Chem. Soc.* **2017**, *139* (31), 10769-10776.
36. Buller, A. R.; van Roye, P.; Murciano-Calles, J.; Arnold, F. H., Tryptophan Synthase Uses an Atypical Mechanism To Achieve Substrate Specificity. *Biochemistry* **2016**, *55* (51), 7043-7046.
37. Herger, M.; van Roye, P.; Romney, D. K.; Brinkmann-Chen, S.; Buller, A. R.; Arnold, F. H., Synthesis of beta-Branched Tryptophan Analogues Using an Engineered Subunit of Tryptophan Synthase. *J. Am. Chem. Soc.* **2016**, *138* (27), 8388-8391.
38. Murciano-Calles, J.; Romney, D. K.; Brinkmann-Chen, S.; Buller, A. R.; Arnold, F. H., A Panel of TrpB Biocatalysts Derived from Tryptophan Synthase through the Transfer of Mutations that Mimic Allosteric Activation. *Angew. Chem.-Int. Edit.* **2016**, *55* (38), 11577-11581.
39. Hochberg, G. K. A.; Thornton, J. W., Reconstructing Ancient Proteins to Understand the Causes of Structure and Function. *Annu. Rev. Biohys.* **2017**, *46* (1), 247-269.
40. Merkl, R.; Sterner, R., Ancestral protein reconstruction: techniques and applications. *Biol. Chem.* **2016**, *397* (1), 1.
41. Gardner, J. M.; Biler, M.; Risso, V. A.; Sanchez-Ruiz, J. M.; Kamerlin, S. C. L., Manipulating Conformational Dynamics To Repurpose Ancient Proteins for Modern Catalytic Functions. *ACS Catal.* **2020**, *10* (9), 4863-4870.
42. Schupfner, M.; Straub, K.; Busch, F.; Merkl, R.; Sterner, R., Analysis of allosteric communication in a multienzyme complex by ancestral sequence reconstruction. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (1), 346-354.
43. Busch, F.; Rajendran, C.; Heyn, K.; Schlee, S.; Merkl, R.; Sterner, R., Ancestral Tryptophan Synthase Reveals Functional Sophistication of Primordial Enzyme Complexes. *Cell Chem. Biol.* **2016**, *23* (6), 709-715.

44. Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J., The ensemble nature of allostery. *Nature* **2014**, *508* (7496), 331-9.
45. Crean, R. M.; Gardner, J. M.; Kamerlin, S. C. L., Harnessing Conformational Plasticity to Generate Designer Enzymes. *J. Am. Chem. Soc.* **2020**.