Received: 1 April 2022 | Revised: 29 July 2022 | Accepted: 15 August 2022

DOI: 10.1002/pro.4426

FULL-LENGTH PAPER

THE PROTEIN SOCIETY WILEY

Estimating conformational heterogeneity of tryptophan synthase with a template-based Alphafold2 approach

Guillem Casadevall[1] | **Cristina Duran**[1] | **Miquel Estévez-Gay**[1] | **Sílvia Osuna**[1,2]

[1]CompBioLab Group, Institut de Química Computacional i Catàlisi and Departament de Química, Universitat de Girona, Girona, Spain

[2]ICREA, Barcelona, Spain

Correspondence
Sílvia Osuna, CompBioLab Group, Institut de Química Computacional i Catàlisi and Departament de Química, Universitat de Girona, Carrer Maria Aurèlia Capmany 69, 17003 Girona, Spain.
Email: silvia.osuna@udg.edu

Present address
Sílvia Osuna, ICREA, Pg. Lluís Companys 23, 08010, Barcelona, Spain.

Funding information
H2020 European Research Council, Grant/Award Number: ERC-2015-StG-679001; Human Frontier Science Program, Grant/Award Number: RGP0054/2020; Ministerio de Ciencia e Innovación, Grant/Award Number: PGC2018-102192-B-I00

Review Editor: John Kuriyan

Abstract

The three-dimensional structure of the enzymes provides very relevant information on the arrangement of the catalytic machinery and structural elements gating the active site pocket. The recent success of the neural network Alphafold2 in predicting the folded structure of proteins from the primary sequence with high levels of accuracy has revolutionized the protein design field. However, the application of Alphafold2 for understanding and engineering function directly from the obtained single *static* picture is not straightforward. Indeed, understanding enzymatic function requires the exploration of the ensemble of thermally accessible conformations that enzymes adopt in solution. In the present study, we evaluate the potential of Alphafold2 in assessing the effect of the mutations on the conformational landscape of the beta subunit of tryptophan synthase (TrpB). Specifically, we develop a template-based Alphafold2 approach for estimating the conformational heterogeneity of several TrpB enzymes, which is needed for enhanced stand-alone activity. Our results show the potential of Alphafold2, especially if combined with molecular dynamics simulations, for elucidating the changes induced by mutation in the conformational landscapes at a rather reduced computational cost, thus revealing its plausible application in computational enzyme design.

KEYWORDS

Alphafold2, computational enzyme design, conformational heterogeneity, tryptophan synthase

1 | INTRODUCTION

"What are the features that make proteins evolvable?" questioned Tokuriki and Tawfik in their seminal review paper.[1] As opposite to the traditional view of one well-defined structure of proteins, a new "avant-garde view" in which proteins display conformational variability key for their evolvability was proposed. They described that evolution operates by enriching pre-existing diversities, which provide the protein the ability to acquire new functions. These ensembles of pre-existing conformations in thermal equilibrium with the so-called *native state* are

the basis for proteins' evolutionary adaptability[1–5] and provide an explanation for the observed divergency originated from a few common ancestors,[6] as well as their versatility as shown by the enzymatic promiscuous side activities.[6] However, the ability of proteins and enzymes to adopt multiple conformations might a priori seem to counter with their characteristics of being proficient, accurate, and specific. Indeed, the high catalytic activity of enzymes is mostly attributed to their highly pre-organized active site pockets presenting the catalytic machinery well-positioned for efficiently stabilizing the transition state(s) of the reactions.[7,8] However, the importance of conformational flexibility was demonstrated with the design of catalytic antibodies.[9] Their modest efficiencies as compared to enzymes were attributed to the imperfect steric and electrostatic environment, but also to their restricted conformational heterogeneity.[10] This shows that efficient catalysis requires a delicate balance between active site pre-organization for transition state stabilization, and also the optimization of the conformational ensemble along the catalytic itinerary. Following an enzymatic cycle in detail, the major steps that take place along a general catalytic itinerary are the following: (1) binding of the substrate(s) in the catalytic pocket, which often involves the exploration of additional conformational states presenting properly positioned loops and flexible domains gating the active site access,[11,12] (2) activation of the substrate(s) for productive enzyme-substrate (ES) formation; (3) stabilization of the transition state(s) leading to the formation of the multiple reaction intermediates and product(s); (4) release of the product(s), which again is often accompanied by conformational changes to restart the catalytic cycle. All these steps are key players for enhanced catalytic activity.

Since the proposal of Tokuriki and Tawfik of the ensemble-based conformational diversity key for evolvability, many experimental and computational studies have been reported in the literature supporting this idea.[1,3–5,13] The study of the conformational landscape of natural and laboratory-evolved enzymes showed that by introducing mutations at the active site and also at remote positions changes in the stabilities of the pre-existing conformations can be induced. This was experimentally demonstrated in the laboratory evolution of a phosphotriesterase into an arylesterase (AE) enzyme.[2,13] The AE activity was gradually increased by changing the fluctuation of some key active site gating loops, as shown by the B-factors of the multiple x-ray structures obtained along the laboratory-evolution path. NMR and room-temperature x-ray crystallography of several HG3 Kemp eliminases also showed a change in the conformational ensembles along laboratory evolution.[14,15] From a computational perspective, this ensemble view of enzymes can be represented in the so-called free energy landscape (FEL, Figure 1 for FELs at different reaction stages),[4] which can be reconstructed by means of molecular dynamics (MD) simulations and enhanced sampling techniques.[16–18] In the reconstructed FEL, the relative stabilities of the thermally accessible conformations, as well as the kinetic barriers separating them are represented. Depending on the barrier height that separates a given pair of conformational states, the timescale associated to the transition is faster or slower. Conformational changes that can directly impact catalytic function include side-chain conformational changes in the fast timescale, loop motions often playing a key role in substrate binding/product release in slower timescales, and in some cases allosteric transitions that usually correspond to the slowest processes. The reconstruction of the FEL and how this is shifted after mutation provides crucial information for understanding and designing enzyme function.[4] The introduced mutations located at the active site and many times at remote sites induce a long-range effect affecting enzymatic catalysis. Induced by the mutations introduced, catalytically productive conformational states are stabilized, whereas the non-productive ones for the novel functionality are disfavored, thus converting computational enzyme design into a population shift problem.[19] These observations promoted the exploration of enzyme conformational dynamics for enzyme design.[3,4,13] The reconstruction of ancestral enzymes displaying a higher degree of flexibility with respect to the modern counterparts and their use as initial scaffolds for enzyme design yielded interesting new insights.[20] The higher flexibility of many ancestral variants was found to be key for achieving higher levels of catalytic activity with only a few mutations located at the active site. In this direction, several ancestrally reconstructed enzymes have been used as starting points for enzyme design, for instance for enhancing some residual catalytic promiscuity contained in an enzyme family, for altering the allosteric regulation of some heterodimeric enzymes, among others.[20–22]

The recent success of the neural network Alphafold2 (AF2) in predicting the folded structure from the primary sequence with high levels of accuracy has revolutionized the field.[23–26] The novel AF2 neuronal network incorporates information on the evolutionary, physical and geometric constraints of existing protein structures. AF2 is recognized as one of the milestones in protein structure prediction, and has boosted the application of deep-learning methods for many other applications.[26] Despite the impressive performance of AF2 algorithms in predicting the native lowest in energy structure of proteins, application of AF2 for understanding and engineering function directly from the obtained single *static* picture is

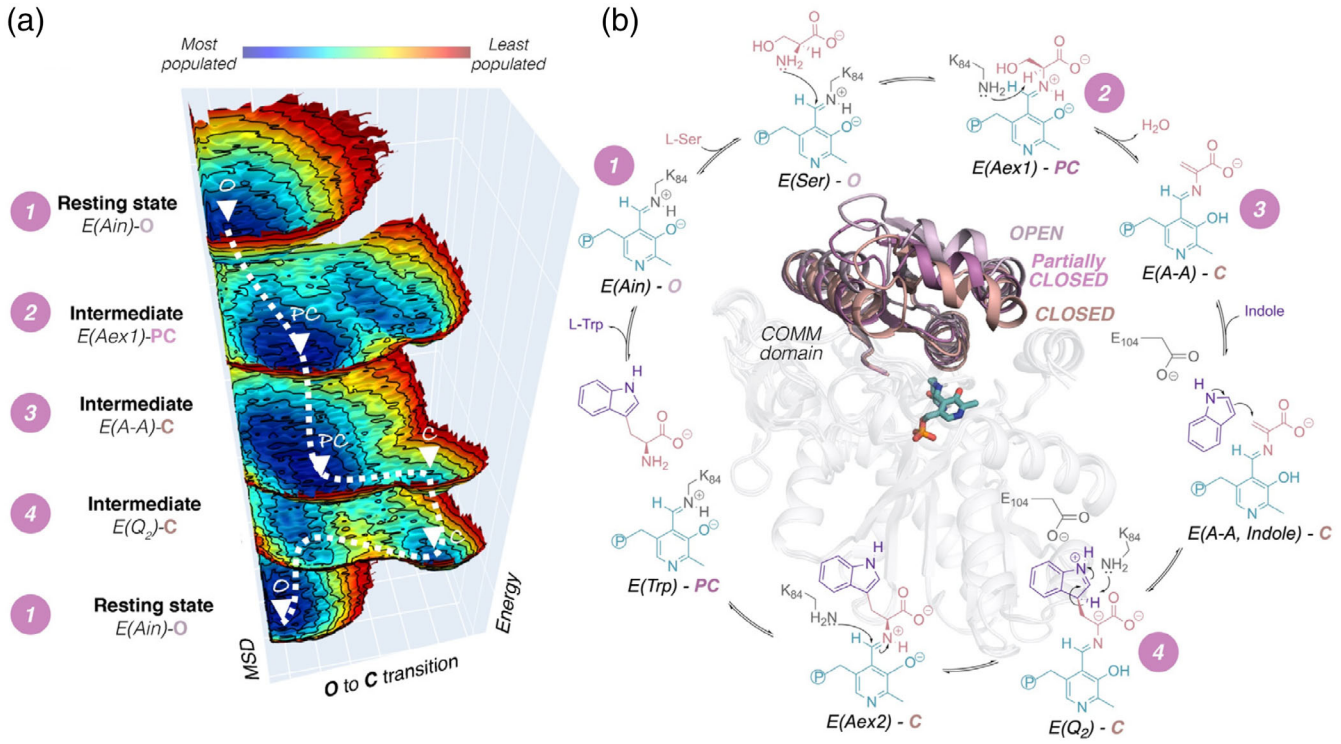

FIGURE 1 (a) Representation of the reconstructed conformational landscape (*Source*: Data from Ref. 17) of tryptophan synthase B (TrpB) at several reaction intermediates along the catalytic itinerary. The enzyme displays a different conformation of the catalytically relevant COMM domain that covers the active site (shown in pink in the structure displayed at the center of panel b): open (**O**) states are adopted in the resting state E(Ain), partially closed (**PC**) at the reaction intermediates E(Aex1) and E(A-A), and closed (**C**) at E(Q_2) states. Most stable conformations are represented in blue, whereas least stable ones in red. (b) Reaction mechanism of TrpB subunit.[29] The conformational states of the COMM domain according to available x-ray data at each reaction intermediate along the catalytic cycle are displayed. Overlay of the different COMM domain conformational states: **O** highlighted in lilac, **PC** in pink, and **C** in brown. Pyridoxal phosphate cofactor is shown in teal, L-Ser in pink and L-Trp in lilac. TrpB, tryptophan synthase B

not straightforward. However, some recent studies have suggested that AF2 can additionally predict multiple conformations of the same protein, and thus it can be potentially used to elucidate the conformational plasticity of biological systems.[27,28] This is exciting as it suggests that AF2 could be applied for assessing the effect of the introduced mutations on the conformational landscape at a rather reduced computational cost, which would boost the development of conformationally driven enzyme designs protocols.[4,19]

In this study, we evaluate the potential of AF2 in assessing the effect of the mutations on the conformational landscape of tryptophan synthase (TrpS). TrpS is a heterodimeric enzyme complex (based on TrpA and TrpB subunits) that performs a multistep reaction mechanism together with a sophisticated allosteric signal communication. TrpA catalyzes the retro-aldol cleavage of indole glycerol phosphate producing glyceraldehyde 3-phosphate and indole, the latter being able to diffuse through an internal TrpA-TrpB tunnel to reach the TrpB subunit (Figure 1). For this enzyme, the allosteric communication between TrpA and TrpB keeps the proper conformations along the cycle optimizing the catalytic steps, thus the absence of the protein partner leads to a deficient conformational ensemble.[17] In the case of TrpB, this involves the change of conformation of a COMM domain that covers the active site, which is known to adapt closed, partially closed, and open conformations (Figure 1).[17,30,31] This fine-tuning of the conformational ensemble induced by the binding partner makes both TrpB and TrpA substantially less efficient when isolated.[32–37] By applying laboratory-evolution, the Arnold lab enhanced the stand-alone activity of *Pyrococcus furiosus* TrpB and generated a new variant named 0B2-*pf*TrpB exhibiting a 2.9-fold increase in k_{cat} with respect to the original complex.[32,33] The reconstruction of the last bacterial common ancestor (LBCA) TrpB by means of ancestral sequence reconstruction showed a high level of stand-alone activity, which was found to be lost along evolution.[22,38] We previously explored the FEL of the ancestrally reconstructed LBCA TrpS in complex and as stand-alone catalyst (LBCA-TrpB), as well as the

wild-type *pf*TrpS complex, isolated *pf*TrpB, and laboratory-evolved stand-alone 0B2-*pf*TrpB enzyme.[17] Our results showed that the low stand-alone activity of isolated *pf*TrpB is due to the restricted conformational heterogeneity of the COMM domain, and the inability to adopt catalytically productive closed conformations. The distal mutations introduced in 0B2-*pf*TrpB recovered the conformational flexibility of the COMM domain to similar levels to those observed for the allosterically regulated *pf*TrpS complex. However, the closed conformation of the COMM domain was substantially more stable in the case of 0B2-*pf*TrpB, which explains its superior catalytic activity with respect to *pf*TrpS. Similar observations were found for LBCA-TrpB whose stand-alone activity was mainly attributed to its conformational heterogeneity and ability to adopt catalytically productive closed conformations of the COMM domain.[39] These two works elucidated the conformational ensemble that a stand-alone catalyst has to display for being efficient, and revealed dramatic changes in the COMM domain conformation, which are important for the multi-step catalytic pathway of TrpB (as shown in Figure 1).[17,39] This information is pivotal for designing new stand-alone TrpB variants, which requires the fine-tuning of the conformational ensemble. In a recent paper,[39] we applied the Shortest Path Map (SPM) methodology[18,19] together with ancestral sequence reconstruction to predict distal activity-enhancing mutations and design a new TrpB variant, that we named SPM6-TrpB. The experimental validation of SPM6-TrpB design demonstrated its superior stand-alone activity in the absence of TrpA to similar levels to those achieved with laboratory evolution (seven-fold increase in k_{cat} with respect to the starting ancestral ANC3-TrpB enzyme).[39] Still the stand-alone activity of the designed SPM6-TrpB was far from that of the reference LBCA-TrpB (k_{cat} of 0.5 and 0.2 s^{-1} for LBCA and SPM6, respectively). This was mostly due again to a restricted conformational heterogeneity and the lack of catalytically productive closed conformations of the COMM domain, as revealed by the FEL reconstruction.

In this work, we evaluate the potential of AF2 for quickly estimating the conformational heterogeneity of different TrpB displaying different levels of stand-alone activity. We first evaluate the effect of using different multiple sequence alignment (MSA) depths in the AF2 predictions for all TrpB systems. We then develop a template-based AF2 approach consisting on providing a set of either x-ray based templates or conformations extracted from MD simulations to estimate the conformational heterogeneity of the different systems. Finally, we run short nanosecond timescale MD simulations from each AF2 prediction to quickly estimate the FEL. Our results show the potential of AF2, especially if combined with MD simulations, for elucidating the changes induced by mutation in the conformational landscapes.

2 | RESULTS

2.1 | Exploring the conformational heterogeneity by altering the multiple sequence alignment depths in AF2

Inspired by recent pre-print papers in which the conformational heterogeneity of some proteins was estimated by reducing the depth of the input MSAs used in AF2 algorithm (as well as the number of recycles),[27,28] we decided to test this methodology in *pf*TrpB, 0B2-*pf*TrpB, LBCA-TrpB, and SPM6-TrpB (see more details in Section 4). For these systems, we have previously reconstructed the FEL and have the experimental characterization of the stand-alone activity.[17,39] In Figure 2, the previously reconstructed FEL of the 0B2-*pf*TrpB variant[17] is shown together with the predictions of AF2 for the different analyzed systems considering different MSAs depths (represented with vertical lines colored from orange to dark blue depending on the MSA depth). The *x* axis denotes the open-to-closed transition of the COMM domain, which ranges from 1–5 (open, **O**), 6–10 (partially-closed, **PC**), and 11–15 (closed, **C**). The predictions obtained by AF2 for *pf*TrpB and 0B2-*pf*TrpB are very similar: in both cases **PC** conformations of the COMM domain are predicted when the MSA depth is higher than 512 (teal lines). Indeed, by increasing the MSA depth more structures closer to the native **PC** state as predicted by the original AF2 are obtained (Table S1). **O** structures are also predicted when a reduced number of MSA (32–64) is used instead (the standard deviation of the **O**-to-**C** path of the predicted structures is ca. 2 for both systems at low MSA depths of 32–64 indicating that several levels of closure of the COMM domain are predicted, Table S1). For these variants no **C** conformations are obtained. This is completely changed in the case of LBCA-TrpB and SPM6-TrpB. In both cases, **C** conformations of the COMM domain are predicted when high MSA depths are used (256–5120), and by reducing the MSA depth to 32–64 only **PC** structures (no **O** structures) are instead obtained (Table S1). Although most of the predicted structures for LBCA and SPM6 TrpB (with high MSAs) fall in the range of **C** conformations (mean value of 11 in the **O**-to-**C** pathway in Figure 2 and Table S1), a higher flexibility of the COMM domain is predicted for the ancestral enzyme: LBCA-TrpB predictions have **O**-to-**C** values in the 11–15 range, whereas SPM6-TrpB in the 13–15 (see *x* axis in Figure 2, and larger deviation in Table S1). It should be emphasized that the native state

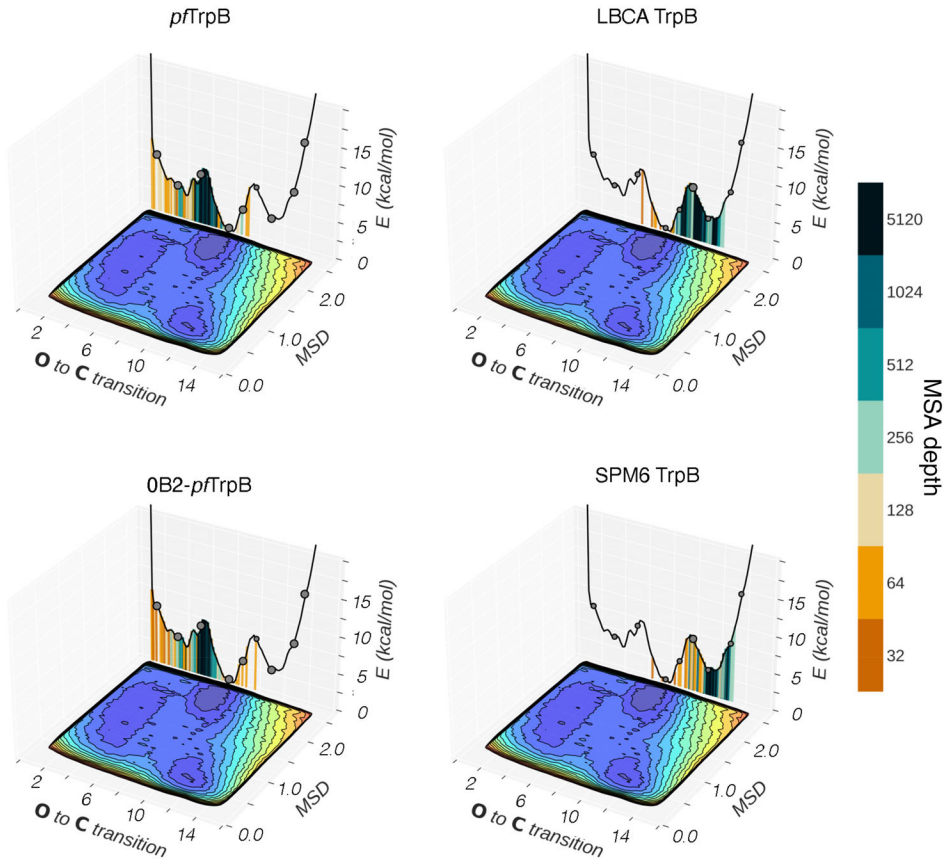

FIGURE 2 Representation of the previously reconstructed FEL of the 0B2-*pf*TrpB variant.[17] The *x* axis denotes the open-to-closed transition of the COMM domain, which ranges from 1–5 (open, **O**), 6–10 (partially closed, **PC**), to 11–15 (closed, **C**), the y axis is the MSD deviation from the path of **O**-to-**C** structures generated. Most stable conformations are shown in blue, whereas higher in energy regions in red.[17] The predictions of AF2 for the different analyzed systems are represented on the 2D-FEL representation using vertical lines colored from orange to dark blue depending on the MSA depth: AF2 predictions obtained with a 32 MSA depth are shown with a vertical orange line, 64 in light orange, 128 in light brown, 256 in light cyan, 512 in cyan, 1024 in teal, and 5120 in dark blue. Black dots indicate some representative available x-ray structures, the size of the spheres is proportional to the sequence identity of the x-ray with respect to the studied TrpB system. FEL, free energy landscape; MSD, mean square deviation

predicted by the default AF2 for both LBCA and SPM6 TrpB has a **C** conformation of the COMM domain. Altogether these results obtained by reducing the MSA depths suggest that 0B2-*pf*TrpB and *pf*TrpB have a higher ability to visit **O** and **PC** conformations, whereas LBCA and SPM6 TrpB **PC** and **C** structures. Interestingly, a higher conformational flexibility is predicted for LBCA-TrpB, especially if compared with SPM6-TrpB, which is in line with our previous reconstructed FELs that show a much-limited conformational heterogeneity of the designed SPM6 variant with respect to LBCA TrpB.[39] Finally, to quantitatively assess the AF2 predicted conformational heterogeneity in the context of structural variance as observed in x-ray data, we applied Principal Component Analysis (PCA, Figure S3). We focused on the carbon alpha distances of the conserved amino acids of the set of x-ray structures used. The first two components describe 74% and 12.5% of the total variance. The projection of the predicted AF2 structures with different depths of MSA shows no major deviations from the space generated with experimentally determined structures, thus providing evidence for the validity of the predictions even with a low MSA depth.

2.2 | Exploring the conformational heterogeneity by altering the multiple sequence alignment depths and using x-rays as templates

The inputs for AF2 calculation are the primary sequence of the enzyme, a MSA generated with information of evolutionary related proteins, and the three-dimensional (3D) coordinates of a small number of homologous structures named templates. In the previous section, we reduced the depth of the input MSAs used in AF2

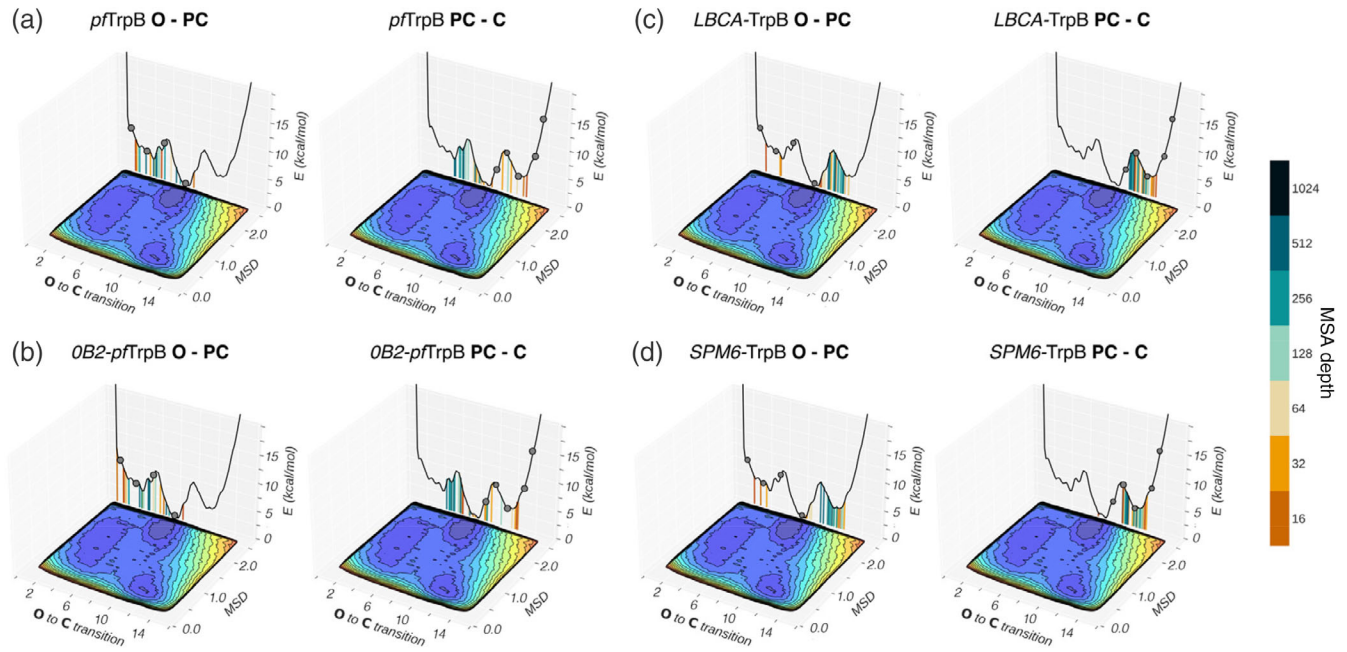

FIGURE 3 Representation of the previously reconstructed FEL of the 0B2-*pf*TrpB variant,[17] and the predictions of the x-ray template-based AF2 approach for the different analyzed systems: *pf*TrpB (a), (b) 0B2-*pf*TrpB, (c) LBCA-TrpB, (d) SPM6-TrpB. The *x* axis denotes the open-to-closed transition of the COMM domain, which ranges from 1–5 (open, **O**), 6–10 (partially closed, **PC**), to 11–15 (closed, **C**), the *y* axis is the MSD deviation from the path of **O**-to-**C** structures generated. Most stable conformations are shown in blue, whereas higher in energy regions in red.[17] The predictions of the x-ray template-based AF2 for the different analyzed systems are represented on the 2D-FEL representation using vertical lines colored from orange to dark blue depending on the MSA depth: AF2 predictions obtained with a 16 MSA depth are shown with a vertical orange line, 32 in light orange, 64 in light brown, 128 in light cyan, 256 in cyan, 512 in teal, and 1024 in dark blue. For each studied case, the predictions obtained by AF2 using x-ray templates with **O**-**PC** conformations of the COMM domain are shown in the left, whereas the results with x-ray templates presenting **PC**-**C** conformations in the right. Black dots indicate the used x-ray structures as input templates, and the size of the spheres is proportional to the sequence identity of the x-ray with respect to the studied TrpB system. FEL, free energy landscape; MSD, mean square deviation

algorithm to evaluate the conformational heterogeneity of four TrpB enzymes displaying different levels of stand-alone activity, and found mostly **C** conformations for the ancestral and designed TrpB, whereas **PC** for *pf*TrpB based variants. Our hypothesis is that by additionally fine-tuning the other input parameter of AF2, that is, the set of templates used to extract the 3D information, more information regarding the enzyme ability to adopt **O**, **PC**, and **C** conformations can be derived. In fact, in the original formulation of AF2 five models are provided that use different number of MSA depths and template structures to encourage diversity in the predictions.[25] It should be also noted that in a recent paper the effect of including different x-ray templates on AF2 predictions was tested only for a specific protein target that was exclusively modeled in only one of the conformations (even with low MSA depths).[27] Different decoy structures were used as templates in a recent pre-print paper based on assessing the coevolution dependency of the AF2 learned potential function for scoring protein structures.[40] In this study, we have tested the hypothesis that by altering the set of AF2 templates the conformational heterogeneity of the target enzymes can be estimated. We have used a reduced number of template structures based on the available x-ray structures presenting a sequence identity larger than 70% with respect to all systems (Table S2 and Figure S1). The side-chain conformation was kept in the template, however, as done in a previous study[40] we also run the simulations by hiding side-chain information from the template and providing only the coordinates of the carbon beta (or carbon alpha in the case of glycine, Figures S4 and S5 for the results without side-chain). In Figures 3 and S6, the results from this template-based AF2 strategy are displayed. For the same primary sequence, the prediction from this template-based AF2 approach suggests a different level of closure of the COMM domain depending on the x-ray template structures used (either in a **C**, **PC**, or **O** conformation, Figures 3 and S6). When **C** and **PC** x-ray templates are used, AF2 prediction for *pf*TrpB and 0B2-*pf*TrpB mostly suggests **PC** conformations of the

COMM domain, especially at high MSA depths (ca. 72% of **PC** and **C** structures are predicted, Table S3). The differences between the predicted structures for both systems are small (they only differ in six mutations), although a slightly higher number of **C** conformations are suggested for 0B2-*pf*TrpB (74 vs. 71% for 0B2-*pf*TrpB and *pf*TrpB, respectively, Table S3). When **O** and **PC** templates are used instead, **PC** conformations are predicted similarly for both cases (79 and 21% of **O-PC** and **PC-C**, respectively, for both cases). The same strategy when applied in the case of LBCA and SPM6 TrpB shows that **C** conformations of the COMM domain are most frequently predicted irrespectively of the x-ray template structure used and the MSA depth applied, as shown in the previous section (Table S3). As discussed in the previous section, the conformational variance of the obtained predictions with this x-ray template-based approach is in line with the structural variance observed in x-ray data (Figure S3).

2.3 | Exploring the conformational heterogeneity by altering the multiple sequence alignment depths and using molecular dynamics conformations as templates

In the specific case of TrpB multiple x-ray structures displaying different conformations of the COMM domain are actually available (Figure S1). However, this is not the case for most of the systems. In this section, we assessed and compared the outcome of the template-based AF2 using conformations extracted from MD simulations instead of x-ray structures. As far as we know, this was not tested in any of the previously mentioned studies based on using AF2 to extract information of the conformational landscape.[27,28] In particular, we used as input a reduced number of conformations displaying either **C**, **PC**, or **O** conformations of the COMM domain extracted from our recently published FELs at the Q_2-bound state

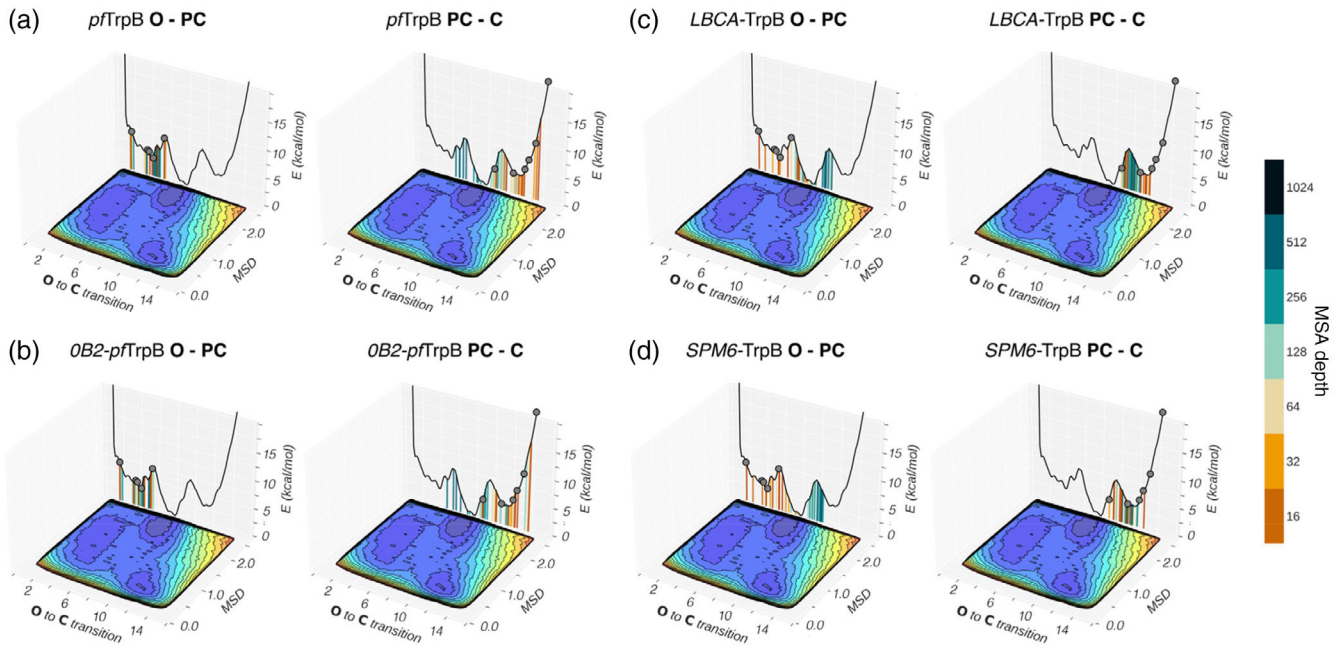

FIGURE 4 Representation of the previously reconstructed FEL of the 0B2-*pf*TrpB variant,[17] and the predictions of the MD extracted template-based AF2 approach for the different analyzed systems: *pf*TrpB (a), (b) 0B2-*pf*TrpB, (c) LBCA-TrpB, and (d) SPM6-TrpB. The *x* axis denotes the open-to-closed transition of the COMM domain, which ranges from 1–5 (open, **O**), 6–10 (partially closed, **PC**), to 11–15 (closed, **C**), the *y* axis is the MSD deviation from the path of **O**-to-**C** structures generated. Most stable conformations are shown in blue, whereas higher in energy regions in red.[17] The predictions of the MD template-based AF2 for the different analyzed systems are represented on the 2D-FEL representation using vertical lines colored from orange to dark blue depending on the MSA depth: AF2 predictions obtained with a 16 MSA depth are shown with a vertical orange line, 32 in light orange, 64 in light brown, 128 in light cyan, 256 in cyan, 512 in teal, and 1024 in dark blue. For each studied case, the predictions obtained by AF2 using as templates conformations extracted from MD simulations with **O-PC** conformations of the COMM domain are shown in the left, whereas the results with MD templates presenting **PC-C** conformations in the right. (a–d) With vertical lines colored from yellow to teal depending on the MSA depth. The *x* axis denotes the open-to-closed transition of the COMM domain, which ranges from 1–5 (open, **O**), 6–10 (partially closed, **PC**), to 11–15 (closed, **C**). Black dots indicate the used representative MD conformations as input templates. FEL, free energy landscape; MD, molecular dynamics; MSD, mean square deviation; MSA, multiple sequence alignment

of the most evolved 0B2-*pf*TrpB variant (gray dots in Figure 4).[17] Similarly to what was done with x-ray-based templates, the side-chain conformation was included in the 3D template (Figures S9 and S10 for the results without side-chain information). When MD-extracted **C** conformations are used as input templates, the predicted structures for *pf*TrpB and 0B2-*pf*TrpB present **PC** conformations of the COMM, especially if high MSA depths are used (green lines in Figures 4 and S11). This is in line with the results obtained in the previous sections where either the MSA depths were only altered or different x-ray templates combined with different levels of MSA were used. Interestingly, by comparing the range of predicted structures at high MSA depths for both *pf*TrpB and 0B2-*pf*TrpB systems, which differ only in six distal active site mutations, a slightly higher ability to adapt **C** conformations of the COMM domain is predicted for the stand-alone 0B2-*pf*TrpB variant (74% of the predicted structures adopt **PC**-**C** conformations, whereas 67% in the case of *pf*TrpB, Table S3). Indeed, some structures present **O**-to-**C** values in the 12–15 range for 0B2 instead of the 10–13 for *pf*TrpB with MSA depths higher than 256 (teal vertical lines in Figures 4 and S11). As expected, the use of **O** and **PC** conformations of the COMM domain as input templates generate mostly **PC** conformations for both *pf*TrpB systems irrespective of the MSA depth (ca. 85% of the predicted structures present **O**-**PC** conformations for both systems, Table S3). Altogether these results suggest that AF2 predicts as the lowest in energy conformation **PC** structures for *pf*TrpB and 0B2-*pf*TrpB. However, by altering the MSA depths and providing as input templates different conformations of the COMM domain taken from MD simulations some hints about the conformational heterogeneity can be extracted: a higher ability to adopt the catalytically productive **C** conformation is predicted for the stand-alone 0B2-*pf*TrpB, if compared to *pf*TrpB.

The same analysis was performed on the ancestral LBCA-TrpB and SPM6-TrpB design. When **O** conformations are used as input templates, **PC** and **C** structures are predicted for LBCA at high MSA levels, whereas for SPM6 more **C** conformations are obtained (Table S3). At low MSA (brown colored lines in Figure 4), **O**, **PC**, and **C** structures are similarly predicted for both cases. These results are again suggesting a higher conformational flexibility for the ancestral LBCA-TrpB as compared to the SPM6 design, which is accordance with our previously computed FELs. By using **C** conformations as templates, the predictions for both systems at either high and low levels of MSA depths yield **C** structures of the COMM domain (ca. 91% of the predicted structures present **PC**-**C** conformations of the COMM domain in both systems, Table S3). As found for the x-ray template-based AF2 predictions, the conformational variance of the structures generated with this MD template-based approach is in line with the structural variance observed in x-ray data (Figure S3).

2.4 | Exploring the conformational heterogeneity by short nanosecond timescale molecular dynamics simulations from the x-ray template-based AF2 predicted structures

The previous sections have shown that by altering the MSA depth and providing different sets of templates (either based on x-ray structures or conformations extracted from MD simulations) the conformational landscape of different TrpB systems can be estimated. To further validate AF2 predictions and assess its potential application for rapidly estimating the conformational heterogeneity, we decided to run multiple replica short nanosecond timescale MD simulations starting from the set of AF2 structures predicted in Section 2 (2 replicas of 10 ns MD simulations starting from the ca. 60 different AF2 outputs obtained in the x-ray template based AF2 approach, that is, ca. 1200 ns of accumulated MD simulation time for each TrpB variant). In Figure 5, the reconstructed FEL from the set of MD simulations performed starting from AF2 output structures is shown on top of the previously reconstructed FELs of *pf*TrpB, 0B2-*pf*TrpB, LBCA-TrpB, and SPM6-TrpB.[17,39] As discussed in the previous sections, a higher conformational heterogeneity is predicted for 0B2-*pf*TrpB and *pf*TrpB with respect to LBCA and SPM6-TrpB that mostly adopt **C** conformations of the COMM domain. Interestingly, larger differences are observed when comparing 0B2-*pf*TrpB and *pf*TrpB that only differ in six mutations (98.4% of sequence identity): although there is only one minimum at **C** conformations of the COMM domain in both cases, the estimation of the FEL for 0B2-*pf*TrpB suggest the existence of an additional minima at **O** conformations. The **O**-to-**C** value of the COMM domain at the **C** minima is ca. 9 in *pf*TrpB, whereas ca. 10.5 in 0B2-*pf*TrpB, thus suggesting a higher ability for adopting the catalytically productive **C** conformation in 0B2-*pf*TrpB, as found in our previous study.[17] In the case of LBCA and SPM6-TrpB a much more restricted conformational heterogeneity is found, in line with the previously reconstructed FELs.[39] In fact, in our previous study we found that at the Q_2 intermediate LBCA-TrpB has a wide energy minima at **C** conformations (mostly presenting a larger deviation along the *y* axis),

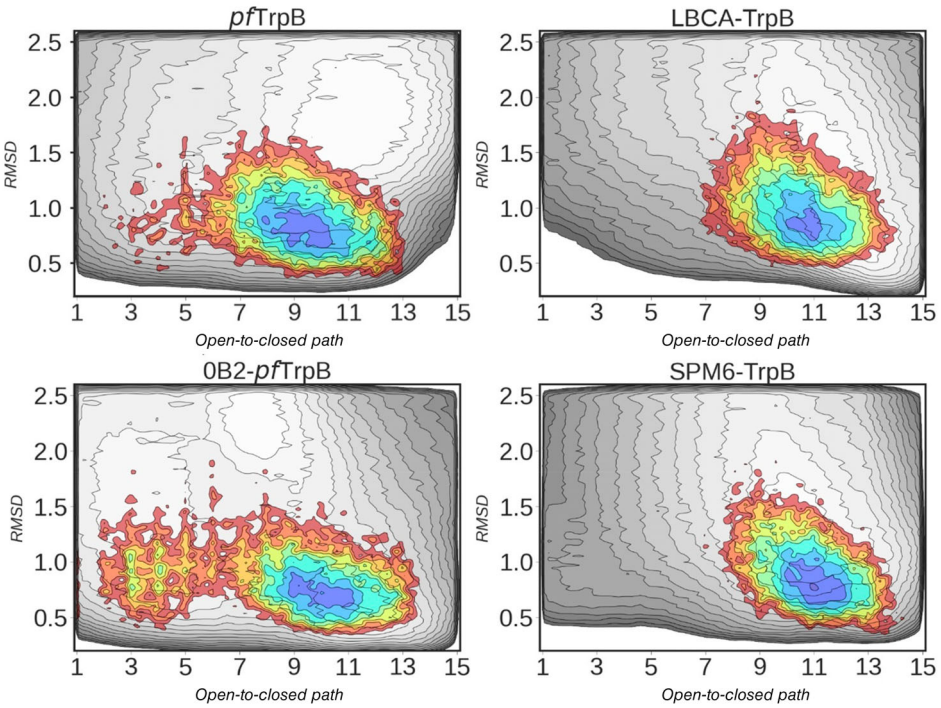

FIGURE 5 Representation of the previously reconstructed FELs of the *pf*TrpB, 0B2-*pf*TrpB, LBCA-TrpB, and SPM6-TrpB (shown in gray scale).[17,39] The estimated FEL from multiple replica short nanosecond timescale MD simulations performed starting at the x-ray template-based AF2 predictions for the different analyzed systems is shown in color on top of the previously reconstructed FELs. The *x* axis denotes the open-to-closed transition of the COMM domain, which ranges from 1–5 (open, **O**), 6–10 (partially closed, **PC**), to 11–15 (closed, **C**), the *y* axis is the MSD deviation from the path of **O**-to-**C** structures generated. Most stable conformations are shown in blue, whereas higher in energy regions in red.[17] FEL, free energy landscape; MD, molecular dynamics; MSD, mean square deviation

which confers the enzyme the ability to visit both catalytically productive and unproductive conformations of the COMM domain. The estimated FEL obtained here from the ensemble of short MD simulations starting at the different AF2 structures also suggest a higher deviation along the *y* axis, and the ability to visit **C** conformation of the COMM domain (**O**-to-**C** values of ca. 11) in line with its high stand-alone activity. The estimated FEL for LBCA and SPM6-TrpB are similar (Figure 5), although the **C** minima for SPM6-TrpB is wider along the **O**-to-**C** axis as it ranges between 9.5 and 12. Although the estimated FELs from this rather short MD simulations present some deviations from the previously reconstructed FELs based on well-tempered multiple-walker metadynamics simulations, the conformational heterogeneity of the different systems can be estimated, which suggests its potential application for rapidly evaluating the effect of mutations into the conformational landscape of enzymes. Finally, the projection of the accumulated MD dataset into the principal component space generated based on the ensemble of available x-ray structures shows no major deviations with experimentally determined structures (Figure S12).

3 | DISCUSSION AND CONCLUSIONS

Tryptophan synthase is a heterodimeric enzyme that features a mechanistically complex reaction mechanism (as shown in Figure 1), together with a fine-tuned conformational ensemble that needs to be optimized for enhanced function.[17,32–37,39] Our previously reconstructed conformational landscapes of the heterodimeric complex as well as several isolated TrpB enzymes showed that by altering the relative stabilities of the open, partially closed, and closed conformations of the catalytically relevant COMM domain, the reaction steps along the catalytic itinerary are optimized.[17] Such conformational changes play an important role in pre-organizing the active site for efficient catalysis, for the binding of the two substrates and for product release. By computationally analyzing multiple TrpBs displaying different stand-alone activities, we found that enhanced stand-alone activity requires the ability to adopt closed conformations of the COMM domain in the absence of the binding partner, as well as a high conformational flexibility to allow substrate binding and product release.[39] The high computational cost associated to FEL reconstruction limits the

exploration of such conformational changes to only a few selected enzyme systems, which is a clear limitation for the routine computational design of new stand-alone variants.[19] In this study, we aimed to test the ability of AF2 to quickly estimate the conformational heterogeneity and changes in the conformational landscapes induced by mutations. We focused on the following systems: the allosterically regulated *pf*TrpB enzyme that in the absence of its binding TrpA partner has restricted conformational heterogeneity and thus low catalytic activity; the laboratory-evolved 0B2-*pf*TrpB that presents stand-alone activity thanks to the six distal mutations introduced that recover the allosterically regulated conformational ensemble; the ancestrally reconstructed LBCA TrpB that does not require TrpA to operate efficiently as it presents the ability adopt different levels of closed conformations of the COMM; and our recently designed SPM6 TrpB variant displaying some stand-alone activity.[17,39] By changing the depth of the MSA used and altering the input template structures (either from x-ray data or conformations taken from MD simulations), the conformational heterogeneity of the systems can be estimated. This is particularly evidenced by running multiple short nanosecond timescale MD simulations from the provided set of structures by this tuned AF2 approach and reconstructing the associated conformational landscapes.

Interestingly, by altering the MSA depth and including either x-ray or MD-based structures as templates, AF2 predicts mostly partially closed (**PC**) conformations of the COMM domain for 0B2 and *pf*TrpB, whereas closed (**C**) structures for the ancestral LBCA and SPM6 TrpB design. By further analyzing the output structures provided by using either **C** or **O** templates in the different systems, one can estimate a higher conformational heterogeneity for *pf*TrpB-based systems, as **O**, **PC**, and **C** conformations can be generated at both higher and lower MSA depths. In contrast, LBCA and SPM6 TrpB predictions are much more restricted to the **PC** and **C** ensemble. The lack of **O** structures for LBCA-TrpB and SPM6-TrpB is also in accordance with our previous calculations at the Q_2 intermediate that suggested an infrequent transition towards **O** states as the reaction progresses, thus suggesting that product release might be rate-limiting.[39] A high similitude in the predictions is observed when comparing 0B2 and *pf*TrpB systems, which is attributed to the high sequence identity between both systems (only six mutations are introduced in 0B2, that is, 98.4% of sequence identity). However, AF2 predictions suggest a higher number of **C** conformations of the COMM domain for 0B2-*pf*TrpB variant (74% vs. 67% of **PC**-**C** structures when using **C** MD templates). This increased number of **C** structures for the evolved variant is in accordance with the reconstructed FELs that show that at the Q_2 intermediate the **C** conformation of the COMM is much more accessible for 0B2 than for *pf*TrpB in the presence of TrpA (i.e., for *pf*TrpS complex).[17] The estimation of the conformational landscapes from multiple replica short nanosecond timescale MD simulations starting at the different x-ray template-based AF2 predictions are in line with the previously reconstructed computationally expensive FELs obtained from well-tempered multiple-walker metadynamics simulations.[39] This suggests that the developed tuned AF2 approach combined with short MD simulations could be potentially applied for rapidly estimating changes on the conformational landscape at a rather reduced computational cost.

Altogether, the distribution of the AF2 predictions in the reconstructed FEL highlights how AF2 learned to locate the global minimum for the input sequence. In this regard, the increase of co-evolutionary information from the MSA forces the network to predict structures close to the global minimum, even if a deviated template is provided. As it can be rationalized from Figures S4 and S6, AF2 predicted as the most probable FEL regions those containing the templates with the highest sequence identities (see largest spheres in Figures S4 and S6). Thus, including co-evolutionary information limits the conformational exploration, which highlights the importance of developing tuned template-based AF2 approaches for assessing the conformational heterogeneity of protein structures. The results provided in this study indicate that by altering the MSA depth and using either x-ray structures or conformations taken from MD simulations, the conformational heterogeneity of related TrpB variants can be quickly estimated. This is specially the case if AF2 predictions are then further evaluated by means of multiple short nanosecond timescale MD simulations. Although much drastic differences are observed when comparing systems presenting lower sequence identities, subtle conformational changes induced by a small number of mutations can also be potentially captured. This is exciting as it suggests that AF2 could be applied for assessing the effect of the introduced mutations on the conformational landscape at a rather reduced computational cost, and opens the door to new AF2-based computational enzyme design approaches.

4 | MATERIALS AND METHODS

AF2 structure prediction starts with a FASTA sequence as an input that is used to generate the MSA and find structural templates, with which AF2 was trained. Five models are obtained as a result, which come from different combinations of random seeds, and considering a

different number of structural templates and extra sequences. This blend of strategies leads to a larger diversity in the predictions.[41] Even so, the obtained structures are often very similar as they are exploring the lowest in energy conformations of the protein. In recent papers, different strategies have been proposed to overcome this *static* picture provided by AF2: (a) del Alamo et al.[27] the number of sequences of the MSA provided to AF2 was modified to contain as few as 16 sequences, they also reduced the number of recycles to 1, and avoid the final MD simulation to reduce the computational cost of the pipeline; (b) Stein and coworkers proposed to replace some specific residues within the MSA to alanine or another residue to potentially manipulate the distance matrices leading to alternate conformations.[28] To investigate whether AF has learned a coevolution-independent potential function for scoring protein structures Roney and Ovchinnikov evaluated the effect of using decoy structures as templates with missing amino acids.[40] In this work, we fine-tune several parameters that differ from the default AF2: (1) MSA depths, as described in the previously mentioned papers,[27,28] giving less coevolution information and thus leading to an increase of the conformation diversity and (2) the set of templates used, that come either from a subset of x-ray structures (as done in del Alamo et al.[27]), or from conformations taken from previous MD simulations[17] (this was not tested in any of the previously mentioned papers).

In particular, we used the following protocol: Starting from the MSA depth alteration, we reduced the number of recycles to one ("*num_recycle* = 1") because of performance reasons. Similarly, Amber minimization was also deactivated, so no structure relaxation was requested ("*amber_relaxed*" = none). Each of the five models were run with 10 different MSA depths. This value is controlled by "*max_extra_msa*" and "*max_msa_clusters*" parameters. The first parameter is described as the number of extra sequences used; and the latter determines the number of the sequence clusters used for the AF2 neural network. Here, the first parameter was comprised between 5120 and 32. Note that, as described in a previous paper,[27] we set the latter parameter as the half of the former parameter except when 5120 sequences are used. In that particular case, "*max_msa_clusters*" parameter was set to 512.[27] In order to include templates, two strategies were performed: (1) considering x-ray structures and (2) conformations extracted from our previously reconstructed free energy landscape of 0B2-*pf*TrpB.[17] Nine x-ray structures (Figure S2) and 11 MD structures (Figure S8) were used as templates presenting different levels of closure of the COMM domain. The parameters: "*num_recycle*" and "*amber_relaxed*" were also maintained as described before. We focused on "*model_ptm_2*" as it presented the most confident structure results in terms of the predicted LDDT-Cα score (pLDDT) and TM-score (pTM) values. In the case of template-based calculations, the MSA depth was altered in the 1024–16 range ("*max_extra_msa*" comprised between 1024 and 16, and "*max_msa_clusters*" was set at its half). Finally, the parameter "*reduce_msa_clusters_by_max_templates*" was deactivated. Also, AF2 calculations considering only the targeted sequence in the MSA was done to ignore co-evolution information (Figures S5 and S10).

4.1 | Molecular dynamics simulations

The starting structures for the four enzymes (*pf*TrpB, 0B2-*pf*TrpB, LBCA-TrpB, and SPM6-TrpB) were generated with the predictions of the x-ray template-based AF2 approach. We performed two replicas of 10 ns MD simulations at the Q_2 intermediate starting from a total of 60, 59, 62, and 59 AF2 structures for *pf*TrpB, 0B2-*pf*TrpB, LBCA-TrpB, and SPM6-TrpB systems, respectively. All calculations were performed using a modification of the amber99 force field (ff14SB) using AMBER 20 (see Appendix S1 for a complete description of the methods).[42]

AUTHOR CONTRIBUTIONS

Sílvia Osuna: Conceptualization (equal); funding acquisition (lead); supervision (lead); writing – original draft (lead); writing – review and editing (lead). **Guillem Casadevall:** Conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); writing – original draft (equal); writing – review and editing (equal). **Cristina Duran:** Data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); writing – original draft (equal); writing – review and editing (equal). **Miquel Estévez-Gay:** Data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); writing – original draft (equal); writing – review and editing (equal).

ACKNOWLEDGMENTS

We thank the Generalitat de Catalunya for the emerging group CompBioLab (2017 SGR-1707) and Spanish MINECO for project PGC2018-102192-B-I00. Cristina Duran was supported by the Spanish MINECO for a PhD fellowship (PRE2019-089147), and Guillem Casadevall and Miquel Estévez-Gay by a research grant from ERC-StG (ERC-2015-StG-679001). Sílvia Osuna 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 Dr. Miguel Á. Maria-Solano for sharing the jupyter-notebook of the free energy landscape of the different TrpB systems analyzed.

ORCID

Guillem Casadevall https://orcid.org/0000-0003-4442-1600
Cristina Duran https://orcid.org/0000-0003-3094-8823
Miquel Estévez-Gay https://orcid.org/0000-0002-8576-8777
Silvia Osuna https://orcid.org/0000-0003-3657-6469

REFERENCES

1. Tokuriki N, Tawfik DS. Protein dynamism and evolvability. Science. 2009;324:203–207.
2. Campbell E, Kaltenbach M, Correy GJ, et al. The role of protein dynamics in the evolution of new enzyme function. Nat Chem Biol. 2016;12:944–950.
3. Crean RM, Gardner JM, Kamerlin SCL. Harnessing conformational plasticity to generate designer enzymes. J Am Chem Soc. 2020;142:11324–11342.
4. Maria-Solano MA, Serrano-Hervás E, Romero-Rivera A, Iglesias-Fernández J, Osuna S. Role of conformational dynamics in the evolution of novel enzyme function. Chem Commun. 2018;54:6622–6634.
5. Petrović D, Risso VA, Kamerlin SCL, Sanchez-Ruiz JM. Conformational dynamics and enzyme evolution. J R Soc Interface. 2018;15:20180330.
6. Khersonsky O, Tawfik DS. Enzyme promiscuity: A mechanistic and evolutionary perspective. Annu Rev Biochem. 2010;79:471–505.
7. Warshel A, Sharma PK, Kato M, Xiang Y, Liu H, Olsson MHM. Electrostatic basis for enzyme catalysis. Chem Rev. 2006;106:3210–3235.
8. Marti S, Roca M, Andres J, et al. Theoretical insights in enzyme catalysis. Chem Soc Rev. 2004;33:98–107.
9. Winkler CK, Schrittwieser JH, Kroutil W. Power of biocatalysis for organic synthesis. ACS Cent Sci. 2021;7:55–71.
10. Hilvert D. Critical analysis of antibody catalysis. Annu Rev Biochem. 2000;69:751–793.
11. Boehr DD, Nussinov R, Wright PE. The role of dynamic conformational ensembles in biomolecular recognition. Nat Chem Biol. 2009;5:789–796.
12. Hammes GG, Benkovic SJ, Hammes-Schiffer S. Flexibility, diversity, and cooperativity: Pillars of enzyme catalysis. Biochemistry. 2011;50:10422–10430.
13. Campbell EC, Correy GJ, Mabbitt PD, Buckle AM, Tokuriki N, Jackson CJ. Laboratory evolution of protein conformational dynamics. Curr Opin Struct Biol. 2018;50:49–57.
14. Broom A, Rakotoharisoa RV, Thompson MC, et al. Ensemble-based enzyme design can recapitulate the effects of laboratory directed evolution in silico. Nat Commun. 2020;11:4808.
15. Otten R, Pádua RAP, Bunzel HA, et al. How directed evolution reshapes the energy landscape in an enzyme to boost catalysis. Science. 2020;370:1442–1446.
16. Curado-Carballada C, Feixas F, Iglesias-Fernández J, Osuna S. Hidden conformations in aspergillus Niger monoamine oxidase are key for catalytic efficiency. Angew Chem Int Ed. 2019;58:3097–3101.
17. Maria-Solano MA, Iglesias-Fernández J, Osuna S. Deciphering the allosterically driven conformational ensemble in tryptophan synthase evolution. J Am Chem Soc. 2019;141:13049–13056.
18. Romero-Rivera A, Garcia-Borràs M, Osuna S. Role of conformational dynamics in the evolution of retro-aldolase activity. ACS Catal. 2017;7:8524–8532.
19. Osuna S. The challenge of predicting distal active site mutations in computational enzyme design. Wiley Interdiscip Rev Comput Mol Sci. 2021;11:e1502.
20. Gardner JM, Biler M, Risso VA, Sanchez-Ruiz JM, Kamerlin SCL. Manipulating conformational dynamics to repurpose ancient proteins for modern catalytic functions. ACS Catal. 2020;10:4863–4870.
21. Devamani T, Rauwerdink AM, Lunzer M, et al. Catalytic promiscuity of ancestral esterases and hydroxynitrile lyases. J Am Chem Soc. 2016;138:1046–1056.
22. 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;117:346–354.
23. Senior AW, Evans R, Jumper J, et al. Improved protein structure prediction using potentials from deep learning. Nature. 2020;577:706–710.
24. Ourmazd A, Moffat K, Lattman EE. Structural biology is solved—Now what? Nat Methods. 2022;19:24–26.
25. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583–589.
26. Callaway E. 'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures. Nature. 2020;588:203–204.
27. del Alamo D, Sala D, McHaourab HS, Meiler J. Sampling alternative conformational states of transporters and receptors with AlphaFold2. Elife. 2022;11:e75751.
28. Stein RA, Mchaourab HS. Modeling alternate conformations with Alphafold2 via modification of the multiple sequence alignment. bioRxiv. 2021;470469.
29. Dunn MF. Allosteric regulation of substrate channeling and catalysis in the tryptophan synthase bienzyme complex. Arch Biochem Biophys. 2012;519:154–166.
30. Hioki Y, Ogasahara K, Lee SJ, et al. The crystal structure of the tryptophan synthase beta subunit from the hyperthermophile Pyrococcus furiosus. Investigation of stabilization factors. Eur J Biochem. 2004;271:2624–2635.
31. Lee SJ, Ogasahara K, Ma JC, et al. Conformational changes in the tryptophan synthase from a hyperthermophile upon alpha (2)beta(2) complex formation: Crystal structure of the complex. Biochemistry. 2005;44:11417–11427.
32. Buller AR, Brinkmann-Chen S, Romney DK, Herger M, Murciano-Calles J, Arnold FH. Directed evolution of the tryptophan synthase beta-subunit for stand-alone function recapitulates allosteric activation. Proc Natl Acad Sci U S A. 2015;112:14599–14604.

33. Buller AR, van Roye P, Cahn JKB, Scheele RA, Herger M, Arnold FH. Directed evolution mimics allosteric activation by stepwise tuning of the conformational ensemble. J Am Chem Soc. 2018;140:7256–7266.
34. Romney DK, Murciano-Calles J, Wehrmuller JE, Arnold FH. Unlocking reactivity of TrpB: A general biocatalytic platform for synthesis of tryptophan analogues. J Am Chem Soc. 2017; 139:10769–10776.
35. Buller AR, van Roye P, Murciano-Calles J, Arnold FH. Tryptophan synthase uses an atypical mechanism to achieve substrate specificity. Biochemistry. 2016;55:7043–7046.
36. Herger M, van Roye P, Romney DK, Brinkmann-Chen S, Buller AR, Arnold FH. Synthesis of beta-branched tryptophan analogues using an engineered subunit of tryptophan synthase. J Am Chem Soc. 2016;138:8388–8391.
37. Murciano-Calles J, Romney DK, Brinkmann-Chen S, Buller AR, Arnold FH. A panel of TrpB biocatalysts derived from tryptophan synthase through the transfer of mutations that mimic allosteric activation. Angew Chem Int Ed Engl. 2016;55:11577–11581.
38. 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: 709–715.
39. Maria-Solano MA, Kinateder T, Iglesias-Fernández J, Sterner R, Osuna S. In silico identification and experimental validation of distal activity-enhancing mutations in tryptophan synthase. ACS Catal. 2021;11:13733–13743.
40. Roney JP, Ovchinnikov S. State-of-the-art estimation of protein model accuracy using AlphaFold. bioRxiv. 2022;484043. https://doi.org/10.1101/2022.03.11.484043.
41. Hegedűs T, Geisler M, Lukács GL, Farkas B. Ins and outs of AlphaFold2 transmembrane protein structure predictions. Cell Mol Life Sci. 2022;79:73.
42. Case DA, Belfon K, Ben-Shalom IY, et al. AMBER 2020. San Francisco, CA: University of California, San Francisco, 2020.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Casadevall G, Duran C, Estévez-Gay M, Osuna S. Estimating conformational heterogeneity of tryptophan synthase with a template-based Alphafold2 approach. Protein Science. 2022;31(10):e4426. https://doi.org/10.1002/pro.4426