WIREs Wiley Interdisciplinary Reviews

Article Title: The challenge of predicting distal active site mutations in computational enzyme design

Article Type: Advanced Review

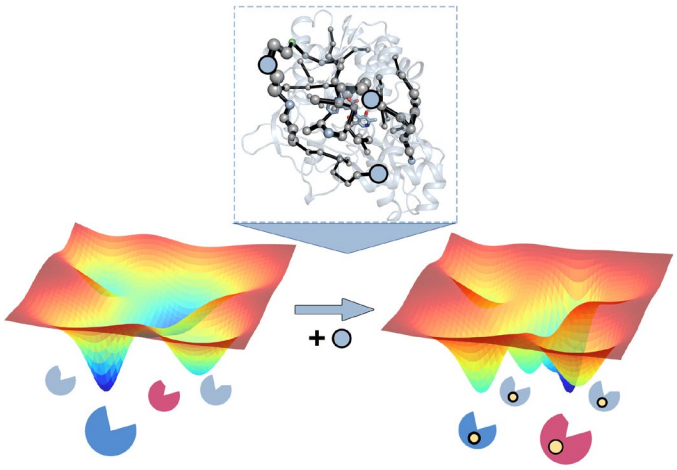
Authors:

Sílvia Osuna[1,2*]
ORCID ID: 0000-0003-3657-6469,
[1]Institut de Química Computacional i Catàlisi (IQCC), and Departament de Química, c/ Maria Aurèlia Capmany 69, 17003 Girona, Spain
[2]ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain

Abstract

Many computational enzyme design approaches have been developed in recent years that focus on a reduced set of key enzymatic features. Initial protocols mostly focused on the chemical steps(s) through transition state stabilization, whereas most recent approaches exploit the enzyme conformational dynamics often crucial for substrate binding, product release and allosteric regulation. The detailed evaluation of the conformational landscape of many laboratory-evolved enzymes has revealed dramatic changes on the relative stabilities of the conformational states after mutation, favouring those conformational states key for the novel functionality. Of note is that these mutations are often located all around the enzyme structure, which contrasts with most of the computational design strategies that reduce the problem into active site alterations. Recent computational strategies have been developed that consider enzyme design as a population shift problem, i.e. redistribution of the relative stabilities of the conformational states induced by mutations. These strategies focus on reconstructing the conformational landscape of the enzyme, and applying correlation-based tools to elucidate the underlying allosteric network of interactions and identify potential mutation hotspots located at the active site, but most importantly at distal positions for the first time.

Graphical/Visual Abstract and Caption



Computational enzyme redesign as a population shift problem: the challenge lies in the computational prediction of distal and active site mutations for altering the relative

stabilities of the conformational ensemble of the enzyme and favouring those conformational states of importance for the novel function.

1. INTRODUCTION

Enzyme design focuses on the generation of new biocatalysts with improved properties by altering their natural constituting sequence of aminoacids. The interest in enzyme design arises from their advantageous characteristics, such as their ability to operate under mild biological conditions achieving high efficiency, selectivity and specificity, but also from the intellectual challenge of creating novel enzymatic activities. Enzyme design is a stringent test to understand enzyme catalysis, evolution, folding and stability.

Enzymes are typically evolved for catalytic activity, enantioselectivity, thermodynamic stability, substrate specificity, stability in non-aqueous solvents and cosolvents, among other properties. Available enzyme design approaches involve the selection of which residues will be subjected to mutagenesis, generation of new variants, followed by screening protocols to evaluate whether an enhancement in the property of interest is reached. How these different points are targeted classify the available approaches into rational design,[1-3] and Directed Evolution (DE).[4, 5] Rational design focuses on a selected set of hotspot positions identified based on multiple sequence alignments, the structural analysis of the active site pocket and potential tunnels for substrate binding, and/or in-depth computational modelling of different enzyme properties by means of Quantum Mechanics (QM), Quantum Mechanics/Molecular Mechanics (QM/MM), Molecular Dynamics (MD) and MonteCarlo (MC) simulations, among others.[2] Given the vast protein sequence space, rational design reduces the enzyme design problem by usually predicting mutations on the active site pocket or substrate binding tunnels. This contrasts with the powerful DE[6-9] technique meritoriously awarded the Nobel prize in Chemistry in 2018. DE was initially based on iterative cycles of random mutagenesis to improve the selected property of interest, but subsequent improvements were lately introduced based on bioinformatic tools[10-15] and sequence analysis[16, 17] to generate smarter libraries, protein engineering techniques,[6, 18, 19] gene synthesis,[20] and sophisticated high-throughput screening techniques.[21] Machine-learning sequence-function models for guiding DE have also been applied.[22, 23] The combination of DE with rational computational enzyme design has been shown to be a powerful strategy to boost the low activities reached by computational designs. [24-26] Similarly, promiscuous enzymatic side-activities can be dramatically enhanced.[27] Impressive examples of laboratory-engineered enzymes have been reported in the literature,[28-30] such as the recent Merck-Codexis in vitro biocatalytic cascade for the synthesis of the anti-HIV drug islatravir.[28]

One of the strongest points of the powerful DE technique is its ability to introduce mutations to the entire protein scaffold, i.e. it is not restricted to active site alterations. In many examples where DE was used to enhance the initially poor catalytic activity of the starting enzyme, dramatic fold increases were obtained after introducing mutations situated far away from the active site.[5, 31] For instance, the 1000-fold increase in activity achieved after applying DE on the acyltransferase LovD enzyme was obtained by introducing 29 mutations, among which 18 were located on the enzyme surface.[32] In fact, the enzyme activity was initially increased by ca. 2-fold with a mutation introduced ca. 15 Å from the catalytic site.[32] The same trend was observed in non-related enzymes such as cytochrome P450, Diels-Alderase, phosphotriesterase, or sitagliptinase, to name a few (see Figure 1 for some selected examples of laboratory-evolved enzymes with mutations introduced at a mean distance from the active site of ca. 15 Å).[5] In fact, there is a lack of direct correlation between the effect of the introduced mutation on the enzyme turnover (k_{cat}) and its proximity to the active site, which contrasts with specificity that is indeed much more determined by active site mutations.[5] The coupling of distal residues impacting the enzyme active site and its catalytic activity suggests that a considerable amount of long-range allostery (i.e. regulation of activity by effector/partner binding) operates in many proteins.[33] The rationalization of how these distal mutations influence the enzyme catalytic activity has been achieved by means of computational tools,[2] mostly through extensive MD simulations.[32, 34] In many cases, these distal mutations alter the network of non-covalent interactions leading to a redistribution of the enzyme pre-existing conformational states that favour catalysis, or altering the conformational dynamics of key structural flexible elements such as loops and lids gating the active site pocket.[27, 35, 36] The rationalization of such changes induced by distal mutations can be satisfactorily obtained by computational modelling, however, the current challenge lies in the prediction of which distal positions can impact and regulate enzymatic activity.[32, 34, 37] We have learned from DE that distal mutations are required for boosting the enzyme catalytic activity, but how feasible it is to computationally predict them? The development of computational tools able to predict remote mutations is highly desirable as the premise of computationally designing efficient enzymes could be in principle reached. This advance review provides a brief overview of the available computational enzyme design strategies, and focuses on highlighting some recent computational enzyme design protocols not restricted to active site mutations whose main focus is the alteration of enzyme conformational dynamics.

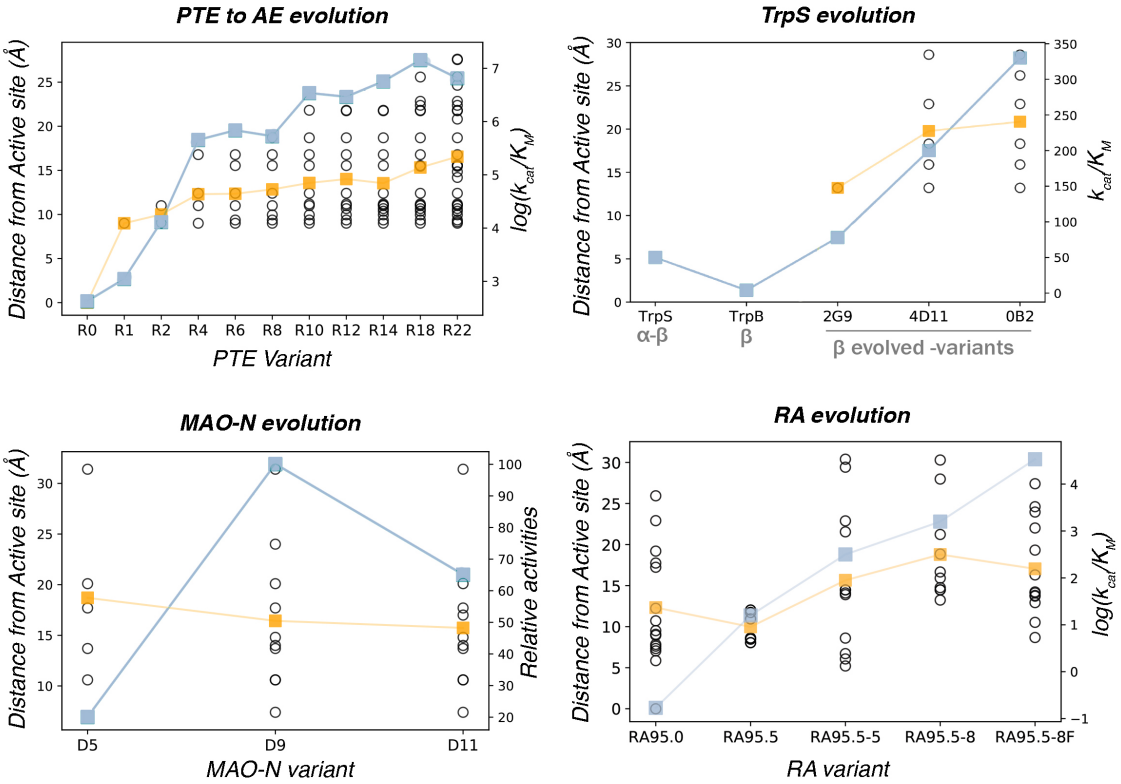


Figure 1. Selected examples of laboratory-evolved enzymes with Directed Evolution: Phosphotriesterase (PTE) to arylesterase (AE) conversion,[27] Tryptophan Synthase (TrpS) for stand-alone activity,[38] Monoamine Oxidase (MAO-N) for broadening substrate scope,[39] and Retro-aldolase (RA) for enhanced activity.[10, 31] The enzyme activity along the rounds of evolution is shown in blue (right y-axis), and the mean distance between the Cα of the introduced mutations with respect to the Cα of the key catalytic residue in orange (and in Å, left y-axis). The distance from the active site of each mutation is marked with a sphere.

2. ENZYME CATALYSIS AND DESIGN: WHICH FEATURES SHOULD BE OPTIMISED TOWARDS NOVEL FUNCTION?

Enzymes present a highly pre-organised active site that precisely arrange the catalytic residues for efficiently stabilizing the transition state(s) of the reaction.[40-44] Apart from the chemical step, the enzyme should also properly bind the substrate in the active site pocket, and release the product for starting again the catalytic cycle. These are essential steps in enzyme catalysis that often require the exploration of additional thermally accessible conformations of the enzyme. These additional conformations sometimes involve changes in loops and flexible domains gating the active site access. [45, 46] Although highly debated, some studies suggest a direct link between active site dynamics and catalysis of the chemical step of the reaction.[40, 47-50] Whether there exists such connection or not, there is multiple evidence to the fact that enzymes change conformation along the catalytic cycle, and in some cases conformational change is what limits the process. [51, 52] Studies also suggest that evolvability of enzymes is related to their inherent conformationally rich dynamic nature.[53] Therefore, enzymes present a precisely arranged active site pocket, but also a high degree of flexibility and versatility, as shown by their ability to catalyse additional side reactions (i.e. promiscuous reactions),[54] and also their tolerance to evolve towards novel functions either by natural or laboratory evolution.[1, 36, 53, 55, 56]

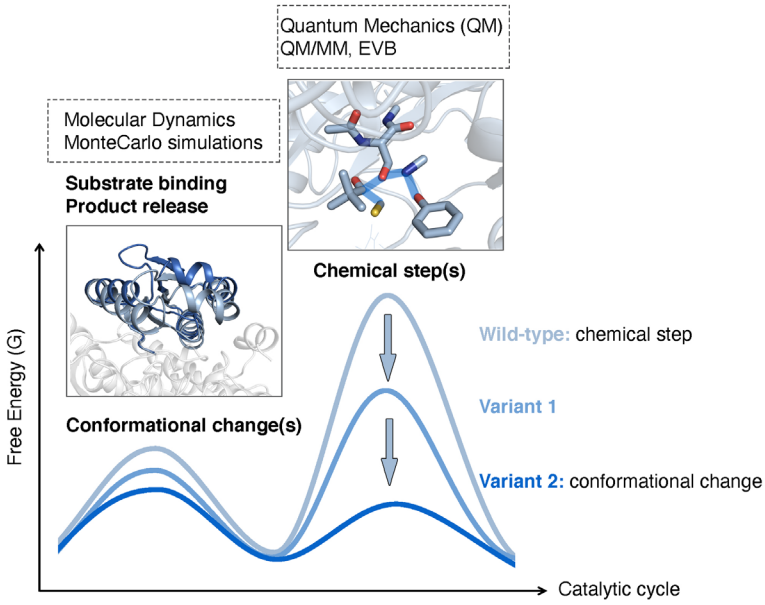


Figure 2. Key features of enzyme catalysis that should be optimised for novel function. In many enzymes, the chemical step(s) are rate-limiting, still conformational dynamics play a crucial role for substrate binding, product release and allosteric regulation. By introducing mutations, conformational change can be rate-limiting (such as the case of variant 2). The main computational tools that can be used for describing the chemical step(s) or the conformational dynamics are also detailed.

All these evidences emphasise the complex nature of enzyme catalysis and the key features that should be optimised for rationally evolving the enzyme towards new functions: substrate accessibility to the active site pocket, active site pre-organisation for efficient transition state(s) stabilization, product release and optimisation of the enzyme conformational dynamics. Unfortunately, the computational tools used for accurately investigating the chemical steps along the catalytic cycles are too costly to be applied for analysing the (un)binding processes and the conformational dynamics of the enzyme (see Figure 2).[1, 2] This justifies the vast number of computational approaches that exist in the literature. Depending on the enzymatic property of interest, the computational approach used will put the focus only to some of the mentioned features. Most of the computational enzyme design protocols exploit some residual promiscuous activity of the natural enzyme, and thus can be considered as "redesigns". In those cases where none of the known natural enzymes has some initial activity or the required active site pocket, *de novo* computational protocols can be applied. [2, 57] These protocols introduce the desired catalytic pocket into natural or artificial scaffolds achieving some initially low activities that could then be further enhanced, and thus "redesigned" often by DE.[24-26] The reasons for the low activities provided by computational protocols have been related to the non-optimal arrangement of the catalytic residues for transition state stabilization, [58, 59] and that essential conformational changes for substrate binding and product release are usually ignored. [45] The limitation of most of the computational protocols of restricting the mutations to the enzyme active site pocket also hampers the design of highly active variants, as shown by many DE examples.[5]

3. OVERVIEW OF COMPUTATIONAL ENZYME DESIGN APPROACHES

In this section, a brief overview of the available computational enzyme design approaches will be presented based on how they tackle the above-mentioned key features of enzyme catalysis, i.e. whether they put the focus mostly on the chemical step(s) of the process, or on the enzyme conformational dynamics. Additionally, details on how and where they introduce mutations (active site only or distal) will be also provided. A special group of computational tools that are difficult to fit into this classification are those based on sequence alignment and bioinformatic tools that have been developed for generating smart libraries for enhancing the enzyme activity, selectivity, and stability.[3] These strategies do not have the focus on the chemical step of the reaction nor in the

conformational dynamics of the enzyme, but rather compare multiple sequences and/or protein folds to identify potential hotspots. Of note is that these approaches have the potential of identifying not only active site mutations but also distal ones. For instance, JANUS and SigniSite are based on multiple-sequence alignments and identify potential mutations for converting structurally related enzymes that have different functionalities.[16, 60] In the particular case of α/β-hydrolase fold superfamily, 3DM database (ABHDB) is useful for selecting mutational hotspots for altering the enzyme functionalities.[13-15] Similarly, HotSpot Wizard[17] can be used to design smart libraries based on the identified mutational hotspots. Additionally, a variety of web servers can be used to compare the protein fold rather than the sequence to identify homologues and potential positions to mutate, such as Dali, FATCAT, or PDBeFold.[61] FuncLib is an automated method that is based on phylogenetic analysis and Rosetta to predict active site mutations.[62] These computational tools based on web-servers have the great advantage of being fast and easy to use, so that their application is not limited to computational experts only.[61]

3.1 Approaches focused on the chemical step(s)

One of the most successful strategies for *de novo* computational enzyme design is the *inside-out* protocol. It is based on the transition state stabilization concept and applies Quantum Mechanics (QM) to compute which is the ideal geometry of the catalytic residues for efficient catalysis. This ideal arrangement (often called *theozyme*) is then grafted into an existing protein scaffold with *RosettaMatch,*[63] and is further refined by *RosettaDesign*.[64] This protocol introduces mutations only in the active site pocket, and the selection of which positions to mutate can be performed either by letting the program to automatically mutate the set of active site residues in the designable area or by manually selecting the positions based on visual inspection of the X-ray structure and chemical intuition. The latter is usually the preferred and most recommended option. For extensive reviews of the *inside-out* protocol and designed variants, check references [64] and [65]. Further improvements to the protocol were introduced such as the incorporation of MD simulations at the end of the design process to identify and rank the best enzyme mutants based on how well the *theozyme* geometry was maintained in the MD runs. [66, 67] In this direction, the CASCO (CAtalytic Selectivity by COmputational design) framework developed by Janssen couples the Rosetta framework with high-throughput MD simulations to evaluate whether key active site geometrical constraints (distances and angles) are maintained, i.e. whether Near Attack Conformations (NAC) are visited. This strategy has been used to engineer enzyme stereoselectivity based on active site mutations, and can replace most of the experimental screening assays.[68, 69] The Janssen lab also developed FRESCO (Framework for Rapid Enzyme Stabilization by COmputational libraries) for identifying potential point mutations, whose effect is evaluated by means of short MD simulations.[70] Rosetta was also further improved by including the so-called 'backrub' move on the protein backbone to include some additional flexibility to the designable area.[71] Other frameworks for flexible protein design were introduced by Baker in the RosettaReModel. [72] In some cases, instead of focusing the design on a single X-ray structure, an ensemble of conformations generated via normal mode analysis,[73] discrete Molecular Dynamics (DMD) simulations,[74] or from the introduction of small Φ/Ψ moves is considered.[75] This strategy of using an ensemble of closely related conformations of the enzyme is usually referred as multi-state design.[76, 77] The coupling of Rosetta and MD simulations was also found to overcome some of the Rosetta conformational sampling limitations.[67, 78] In references [79] and [80], further techniques for including some flexibility to the design process can be found. Finally, there are some alternative approaches to RosettaMatch that include SABER program,[81] Scaffold-Selection,[82] OptGraft,[83] and PRODA_MATCH,[84] among others.

In the previous computational protocols, the effect of the introduced mutation(s) on the activation barrier is not computed, but rather estimated by means of some key geometrical parameters taken from the *theozyme* (or NAC) geometry. Another group of computational approaches focused on the chemical step are those that specifically estimate the effect on the activation barrier by means of Quantum Mechanics/Molecular Mechanics (QM/MM).[85-88] Kamerlin proposed the CADEE approach (Computer-Aided Directed Evolution of Enzymes) that is based on assessing the effect on the activation barrier with Empirical Valence Bond (EVB) calculations.[85] This approach uses valence-bond (VB) theory for describing the chemical reactivity, which allows the simulation of a large number of mutants as it is inherently a classical approach. The protocol performs some short MD simulations to extract different conformationally-related structures (that interconvert in the fast ps-ns

timescales, see Figure 3), which are then used for estimating the mean activation barriers. Similarly, Warshel recently proposed a combinatorial approach for computer aided enzyme design that uses EVB to compute the effect on the activation barrier and introduces some flexibility to the surrounding residues by generating a set of side-chain rotamers.[86] Similar to DE, a large number of variants can be *in-silico* generated with these approaches, but opposite to DE, the libraries have the restriction of active site mutations as the potential number of distal positions to be targeted is too massive to be computationally achievable.

Alternative approaches have been reported for computer aided enzyme redesign. For instance, strategies based on the use of transition state analogues (TSA) to find active site mutations for minimizing the interaction energy of the enzyme active site with TSA,[89] or for minimizing the interaction energy of the substrate with the enzyme pocket by means of Protein Energy Landscape Exploration (PELE) calculations.[90] The latter approach has been recently used to engineer an enzyme with two active site pockets for improved natural and new unnatural activity.[91] PELE was also used to guide DE experiments for evolving fungal high-redox-potential laccases towards high activity and stability.[92] POVME can be used for computing the active site pockets and volumes on the X-ray structure or on an ensemble of MD snapshots, thus identifying some potential hotspots.[93]

3.2 Approaches focused on the enzyme conformational dynamics

Free Energy Landscape and timescales of conformational changes:
Enzymes adopt multiple conformations in solution key for its natural function (see Figure 3). The ensemble of thermally accessible conformations can be represented in the so-called free energy landscape (FEL). In this FEL, the relative stabilities of the different conformations are displayed, as well as the kinetic barriers that separate the different conformational states. The free energy (G) can be defined as the negative logarithm of the population distribution (P) in k_BT units, and thus highly populated states will correspond to the most stable energy minima. Those conformational states that are separated by small energy barriers will be easily interconverted (in ps or ns). These fast transitions usually correspond to side-chain conformational changes (ps to µs). If the conformational transition involves loop motions that are usually key for substrate binding or product release, the process is usually slower and occurs in the nanosecond up to millisecond timescales. In case of allosteric transitions and domain motions, such transitions are the slowest ones and take place in the microsecond to second timescales. All these mentioned conformational changes might have a direct role in the enzyme catalytic activity.

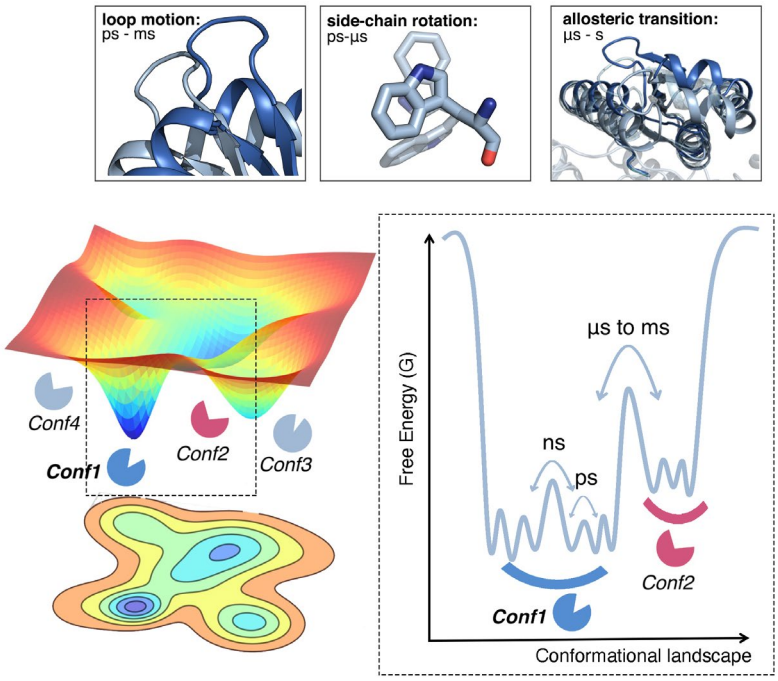


Figure 3. Enzyme Free Energy Landscape (FEL) and timescales of conformational changes key for enzymatic activity. In the scheme, energy minima named Conf 1 (blue) and Conf 2 (red) are composed by multiple related conformations that can be interconverted in the ps-ns timescale, for instance by side-chain

rotations. The conformational transition from Conf 1 to Conf 2 has a higher barrier, and thus a slower associated timescale (μs-ms), for instance allosteric transitions or domain motions.

The engineering of conformationally flexible loops and domains gating the substrate access to the active site has been shown to be crucial for achieving novel enzymatic function.[94, 95] The keyhole-lock-key model of enzymatic catalysis proposes that the access pathway, i.e. the keyhole, is particularly important for substrate entry and product release.[96] Useful bioinformatic tools such as CAVER and CAVERdock applied to either a single X-ray structure or on multiple enzyme conformations taken from MD simulations can yield relevant information on which positions might be subjected to mutagenesis.[95, 97] Of note is that these positions identified are located in the bottleneck (i.e. narrower) regions of the keyhole gating the substrate access to the active site, and thus they correspond to a great strategy to target distal positions at a reduced computational cost. These strategies are based on identifying the enzyme tunnels and channels to introduce mutations in the narrower regions, but they do not strictly focus the design process on the alteration of the conformational dynamics of the loops and domains. The process of evaluating the substrate binding process and introducing mutations in those residues that might prevent substrate binding to the active site has also been achieved with the PELE software.[90] Other computationally more demanding strategies to study substrate entrance or egress from the active site cavity can also be applied, such as Steered MD (SMD), RAM or accelerated MD (aMD).[2, 35, 61] These strategies introduce a computational bias to enhance the crossing of the high barriers associated to these binding events, which take place in the μs to ms timescales. These methods can be used for elucidating the preferred channels for substrate binding, and identifying hotspots in the tunnel and active site pocket for smart library generation.

The *inside-out* approach mentioned in the previous section was applied for rationally designing new Kemp eliminases.[11] The computational designs were then further evolved by DE and iterative design protocols providing new variants with higher activities that include 10-17 mutations situated all around the enzyme structure.[98-103] In HG3.17, one of the most evolved Kemp eliminases reported, 17 mutations were introduced with DE: 11 close to the active site and 6 at distal positions.[103] In a recent study, Chica and coworkers showed that by removing all HG3.17 distal mutations (except two located at the substrate entrance channel) high catalytic efficiencies were also obtained.[104] Many other Kemp eliminases have been designed, [105, 106] however, the most impressive design was achieved by Kamerlin, Sanchez-Ruiz and coworkers. Interestingly, the new design presented only 1-2 mutations at the active site pocket, but still exhibited some of the highest Kemp eliminase activity achieved so far. The main reason for this higher catalytic activity is due to the use of a conformationally flexible ancestral enzyme. This example again shows how important is to consider conformational heterogeneity, and in this case a conformationally rich scaffold, for designing novel enzymatic functions.[36] 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.[107-109] In the so-called multi-state enzyme design, rather than designing the best sequence based on a single static enzyme structure, multiple conformations taken from short MD simulations are used instead.[76] This strategy has also been expanded to consider the different substrates and intermediates that exist in multi-step enzymatic processes to change the enzyme specificity towards non-natural substrates.[77] This approach, similarly to the previously mentioned computational protocols that incorporate flexibility to the enzyme backbone through changes to the Φ/Ψ angles or by applying MD simulations in the iterative designs, introduces some conformational diversity in the design process. However, the ensemble of conformations used can be easily interconverted in the short timescales (*i.e.* they come from the same energy minima in FEL, see Figure 3), and thus often are not representative of the high in energy conformational states that may play a role in the enzyme function. In this line, the use of enzyme conformations taken from room-temperature X-ray crystallography for multi-state design has been recently found to be a promising strategy for computationally designing efficient enzymes.[104] These approaches also have the limitation that the introduced mutations are always restricted to active site alterations, and the positions targeted are selected based on visual inspection of the starting structure(s).

4. ENZYME DESIGN: A POPULATION SHIFT PROBLEM

The reconstruction of FEL of natural and evolved enzymes provides interesting insights into their conformational landscapes and how these are shifted by mutations. The importance of FEL reconstruction in enzyme design lies in the recognition of enzyme redesign as a population shift problem. The idea behind is that among the ensemble of conformations that the enzyme can adopt in solution, there might be a few of them probably higher in energy that could be of importance for the promiscuous reaction and/or for binding the non-natural substrate. These higher in energy conformational states could then be further stabilised by mutations, thus leading to a population shift in the conformational landscape that translates into an increase in the desired activity. This was indeed found to be the case in the laboratory evolution of a phosphotriesterase (PTE) enzyme into an arylesterase (AE).[27, 55] By applying DE, a gradual change in function was achieved mainly due to distal mutations that shifted the populations of the existing conformational states of the enzyme and stabilised those states key for arylesterase activity (see Figure 1).[27] The change in the conformational landscape was characterized by collecting X-ray structures and associated B-factors of the multiple variants generated along the laboratory evolutionary pathway. This change in the conformational ensembles along laboratory evolution has also been recently observed by means of room-temperature X-ray crystallography of the HG3 Kemp eliminases.[104] The same evidence can be provided by means of computational methods. For instance, long timescale MD simulations on the acyltransferase LovD enzyme whose natural activity depends on LovF binding partner, indicated that the increase in stand-alone activity of LovD achieved by DE was due to the gradual stabilization of some conformational states that presented the catalytically competent arrangement of active site residues in the absence of LovF.[32]

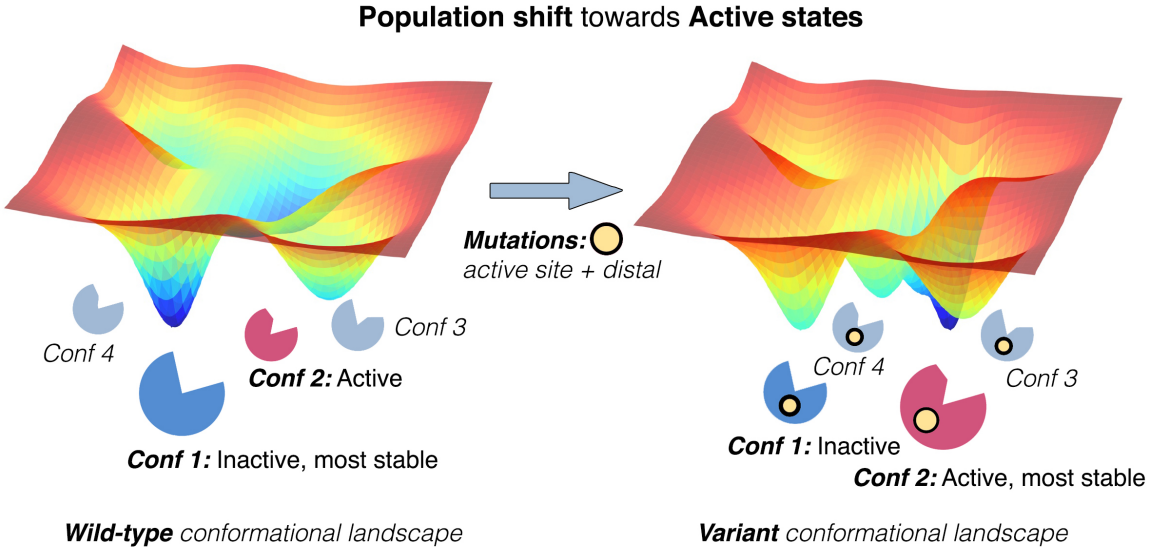


Figure 4. Schematic representation of the population shift concept on the free energy landscape induced by mutations towards novel functionalities. The conformational landscape of wild-type (left) and variant (right) differ in the relative stabilities of the conformational states, being the ones that are functionally relevant for the novel activity stabilised in the variant due to the introduced mutations (located at the active site and/or distal).

Computational enzyme design seen as a population shift problem requires the reconstruction of the FEL of the starting enzyme and also that of the evolved variants, which is computationally expensive as it requires extensive MD simulations and/or enhanced sampling techniques.[1] Most importantly is the fact that the reconstructed FELs provide relevant information on the additional conformational states that could be further stabilised by mutations, but do not provide any clue on which positions either located at the active site or distal might be responsible for stabilizing the desired conformational state. The high computational cost associated with FEL reconstruction together with the existing limitation of identifying key residues for altering the enzyme conformational landscape hampers the development of conformationally-driven enzyme evolution protocols. In this direction, we have analysed the FEL of several natural and evolved enzymes and proposed new computational tools for rationally predicting which positions might be targeted to favour the desired population shift in the conformational landscape (see section 6).

5. CONFORMATIONALLY DRIVEN ENZYME EVOLUTION

Many enzymes present long-range allosteric effects, thus binding of an effector to a remote site of the protein impacts its functional activity by either increasing the affinity towards the substrate or by regulating the catalytic activity.[37] This effector binding induces a change in the conformational landscape of the enzyme, which alters both the thermodynamic and dynamic properties. Different types of allostery have been defined depending on how the conformational landscape is altered after effector binding: conformationally or entropically driven. The first term is used when a redistribution of the existing conformational states takes place, whereas the second when the wideness of the free energy minima is modified.[110] The enzyme design paradigm seen as a population shift problem has therefore a high similarity to the conformationally driven allosteric regulation.[27, 111, 112] In the particular case of enzyme evolution, the shift in the conformational landscape is driven by mutations rather than effector binding, and such mutations should stabilise those conformational states more competent for the novel functional activity. In promiscuous enzymes or computational designs with poor initial catalytic activities (which could also be considered promiscuous), those conformational states presenting either the catalytic residues well positioned for catalysis and/or with the appropriate shape and tunnel for binding the desired substrate should be selected for further stabilization. Many natural enzymes present complex multimeric structures and require the regulation imparted by the other subunits, however, for industrial purposes it is much more appealing the use of simpler stand-alone versions of these enzymes. In these cases, those functional conformational states that are stabilised in the presence of the binding partner and that are often destabilised in stand-alone versions should be populated after mutation.

Given the high similarity between conformationally driven allosteric regulation and enzyme design, we hypothesized that computational tools used in the characterization of allosteric proteins could be also successfully applied in computational enzyme evolution.[34] Indeed, we have recently shown that MD simulations coupled to correlation-based tools can be used for identifying key positions that shift the conformational landscape of the enzyme and thus influence its catalytic activity.[2] Of importance is the fact that the identified positions are not restricted to the catalytic pocket, and thus could potentially solve the current challenge of computationally predicting distal active site mutations.

6. PREDICTING DISTAL MUTATIONS THROUGH CORRELATION-BASED TOOLS

Allosteric regulation is often computationally characterized with methods taken from graph theory.[110] In these graph-based approaches, the enzyme is represented by a weighted graph, where each node corresponds to either a single residue or to a community of residues. Graphs present two important properties: the shortest path between two different nodes and the merge of some of the nodes into larger communities (see next section). Additional computational approaches to reveal allosteric effects based on Elastic Network Model (ENM) have also been developed, such as the dynamic flexibility index (DFI).[37] DFI has been used to establish a relationship between the structural dynamics of each residue of the protein and their evolutionary conservation.[37] Allosteric interactions have also been characterized with the dynamic coupling index (DCI) that measures the dynamic coupling between two residues *i* and *j* upon perturbation to residue *j*.[37] Interestingly, the application of DCI in the Tokuriki and Jackson DE pathway for converting a PTE into an AE[27] (see Figure 1) indicated that the targeted positions in DE presented a high dynamic coupling to the catalytic site.[37] This analysis reinforces the idea that distal mutations often observed in laboratory evolution are a major indication of some allosteric regulation operating in the enzyme, and most importantly, it evidences that the tools used to map allosteric interactions could be of use in the enzyme design field.

6.1 Correlation-based methods used in allosteric systems

The first attempts to use graph theory for investigating allosteric regulation focused on the static X-ray structure of the enzyme.[110] In the generated graph, two pairs of nodes (residues) are connected with an edge if the distance between two representative atoms is below a certain threshold.[113] The most important improvement in the

graph construction was provided by Sethi and coworkers as they used short MD simulations (of a few ns) to define which nodes were connected and the weights of the edges used.[114] An edge was drawn between a pair of residues (nodes) if the representative atom of each residue was below a certain distance threshold for a specified fraction of the MD simulation time. The edge connecting each pair of residues *(i,j)* was weighted according to their correlation values ($C_{i,j}$, see equations 1 and 2). Those residues that during the MD simulation undergo highly correlated conformational changes (i.e. $C_{i,j} \rightarrow 1$), the edge connecting both of them will be rather short ($l_{i,j} \rightarrow 0$). In the other extreme, those pair of residues whose movement is non-correlated ($C_{i,j} \rightarrow 0$) will be connected through rather long edges ($l_{i,j} \rightarrow \infty$).

Key equations for converting a protein into a graph:

The protein is converted into a graph, where each node corresponds to a residue, and the edges connecting each pair of nodes (residues) are weighted following equation 1:

$$\boldsymbol{l_{i,j} = -\log(|C_{i,j}|)}$$ **Equation 1**

where $C_{i,j}$ corresponds to the computed correlation value:

$$C_{i,j} = \frac{<\Delta r_i \cdot \Delta r_j>}{\sqrt{<\Delta r_i^2><\Delta r_j^2>}}$$ **Equation 2**

where Δr_i and Δr_j are the displacement of the C_α of the *i*, *j* residue of the protein along the analysed MD trajectory with respect to its position at the most populated cluster (reference).

By applying this methodology, the enzyme and its conformational dynamics are represented and summarized with a weighted graph (see upper part of Figure 5). Further partitioning of the graph into communities by means of the Girvan-Newman algorithm[115] yields the so-called optimal community network that has been extensively applied for studying allosterically-regulated enzymes.[109, 116] Since this first formulation, further modifications have been introduced to define the edge lengths, see review [110] for a detailed overview of the available strategies. Analysis of the communities in the apo and effector-bound states provides fundamental insights on how effector binding alters the protein network, and thus its conformational landscape. More importantly, the underlying allosteric pathways are identified and key regions regulating the enzyme conformational dynamics are revealed.[109, 116] More details and examples on the application of correlation-based tools on allosterically-regulated enzymes can be found in reference [110].

6.1 Shortest Path Map (SPM) tool for enzyme design

Similar to the allosteric regulation process exerted by effector binding, in enzyme design the introduction of mutations either at the active site or at distal positions induces a shift on the conformational landscape favouring those states with the catalytically competent arrangement for the new target reaction. The high similitude between both cases, prompted us to investigate whether correlation-based tools could be developed and applied in enzyme design as well.[1, 34]

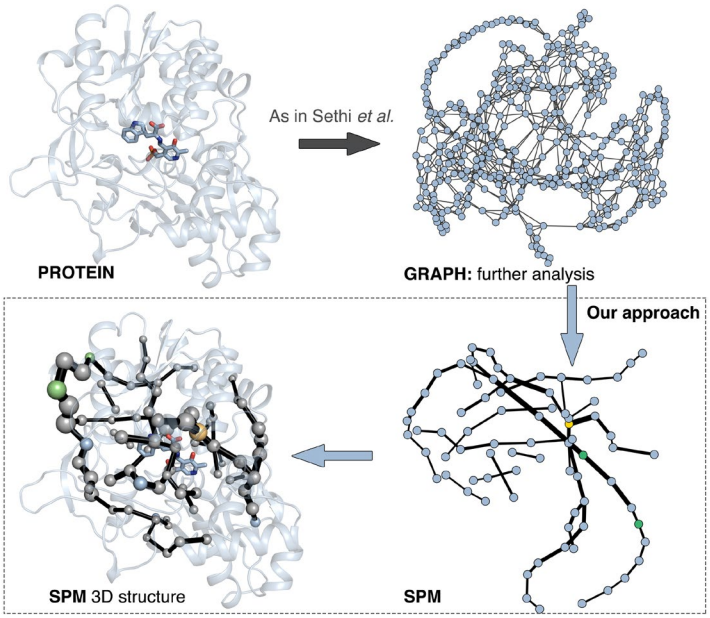


Figure 5. Developed Shortest Path Map (SPM) for enzyme design. The protein is represented with a weighted graph as done in previous allosteric studies,[114] however, the initial graph is further simplified by means of Dijkstra algorithm to characterize the shortest paths. The generated SPM is then drawn on the 3D structure.

We developed the Shortest Path Map (SPM, *DynaComm.py*) which relies on the construction of a graph based on the computed mean distances and correlation values obtained along MD simulations (see equations 1-2 above), following a similar strategy as in Sethi *et al.* protocol.[34] For each residue of the protein, a node is created and centered on the C_α. The next step is to define edges between pairs of nodes. An edge is drawn between those pairs of nodes whose C_α distances are at less than 6 Å during the whole MD simulation time. The edge distance is derived from the computed correlation values, which define the information transfer across a given edge (following equation 1). In this way, those pairs of nodes presenting larger correlation values (closer to 1 or -1) have shorter edge distances, whereas less correlated residue pairs (values closer to 0) long edge distances. At this point, a graph with nodes and edges based on proximity and correlation is created, which is further simplified as shown in Figure 5.

In contrast to previous allosteric studies that focus on identifying communities in the graph, we compute the shortest path lengths with the Dijkstra algorithm as implemented in the igraph module.[117] The algorithm goes through all nodes of the graph and identifies which is the shortest path to go from the first until the last protein residue. The method therefore identifies which are the edges of the graph that are shorter, i.e. more correlated, and that are more frequently used for going through all residues of the protein. All edges are then normalized, and only those edges having the largest contribution are represented in the shortest path map (SPM). It is useful to draw the SPM directly on the 3D structure of the protein rather than the 2D graph representation (see Figure 5). The main advantage of SPM with respect to community analysis is that the most important residues are directly identified (instead of regions), which for enzyme design is much more appealing as small libraries of hotspot positions can be directly constructed.

The application of SPM has been proven to be successful in identifying mutational spots either at the active site or at distal positions targeted by DE. SPM therefore allows, for the first time, the prediction of distal active site mutations that lead to enhanced enzymatic activity.[34] Hereafter, the cases where SPM has been applied for identifying DE mutation hotspots are presented and discussed.

6.1.1 Case example 1: Retro-Aldolase

As mentioned in the introduction, the *inside-out* protocol was successfully applied for generating the mechanistically complex retro-aldolase (RA) enzymes. The reaction pathway involves multiple intermediates, which makes the computational design quite challenging. The computational designs presented low RA activity towards the methodol substrate (10^5-fold increase in k_{cat}/k_{uncat}), and were further refined and evolved by DE in

two subsequent studies.[31, 118] DE introduced multiple mutations located all around the enzyme, which drastically enhanced the RA activity (mostly k_{cat}). The most evolved variant, RA95.5-8F, has a RA activity rivalling those of natural class I aldolases, and was achieved by applying droplet-based microfluidic screening techniques.[31] RA95.5-8F features a catalytic tetrad and has as a rate-limiting step product release rather than the C-C bond scission step (as opposed to RA95.0 and the least evolved variants).[87, 119, 120]

We reconstructed from microsecond timescale MD simulations the conformational landscapes of the different RA variants generated along the DE pathway (see Figure 6C).[34] This conformational analysis revealed that in all cases catalytically active conformational states (with short distances between the base and the Schiff-base intermediate) were sampled, although the population of these states along evolution was dramatically altered (in Figure 6C the conformational landscape of the intermediate RA95.5-5 variant is shown). The mutations introduced by DE progressively stabilised the catalytically active conformational states upon the non-productive ones, which for the case of RA95.5-8F were hardly sampled. This is in line with the change in the rate-limiting step as observed by Hilvert experimentally and Moliner computationally.[87, 119, 120]

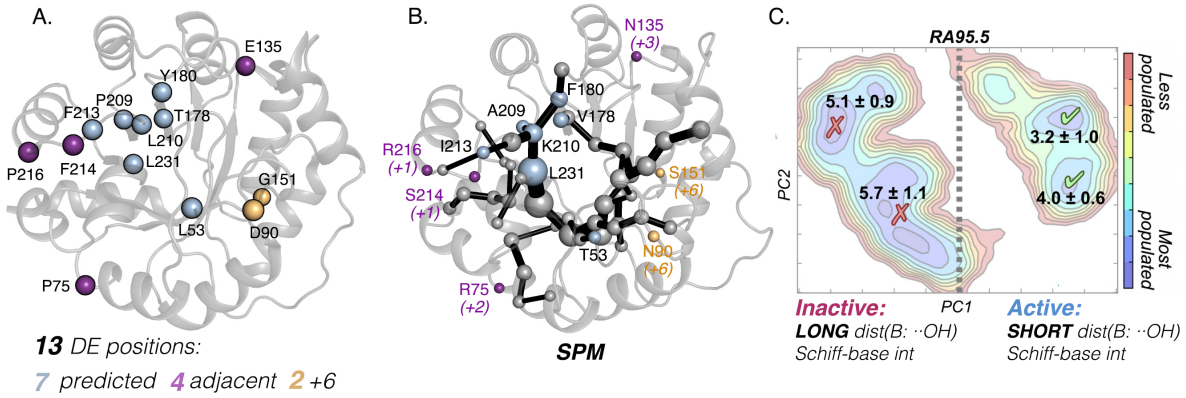


Figure 6. Retro-aldolase SPM comparison to Directed Evolution (DE) mutation hotspots. A. 13 mutations were introduced via DE, among which 7 are included in SPM (shown in blue spheres), 4 are located at an adjacent position (purple spheres), and only 2 are deviated more than 6 positions in sequence (yellow). B. SPM of RA95.5-8. C. Conformational landscape of one of the intermediate RA variants that presents moderate RA activity. Active states presenting short distances between the base and the alcohol of the Schiff-base intermediate are sampled (highlighted with a green tick), and inactive states (long distances, marked with a red cross) are displayed.

The application of SPM taking the information from these microsecond MD simulations indicated that the DE mutational hotspots could be *a priori* predicted (compare Figure 6A and 6B). For each variant along the DE pathway, we applied the SPM methodology and checked whether the positions identified with SPM were mutated in the laboratory evolution. Indeed, we found that all positions were identified, being the missing ones located at adjacent positions with a few exceptions (see Figure 6A-B, 7/13 included in SPM, 4/13 in adjacent positions, 2/13 located more than 6 positions in sequence away from SPM). Interestingly, the SPM of the most evolved RA95.5-8F highlighted that the conformational dynamics of the catalytic tetrad residues are highly correlated as they are all contained in the path. DE mutations, although located at remote positions, are also connected to the sophisticated catalytic tetrad residues, thus further showing their connection with the active site pocket for functional regulation. We speculated that the success of SPM in the particular case of RA could be related to the original scaffold of RA95.0, which is the allosterically-regulated indole 3-glycerol phosphate synthase enzyme. Of interest is also the observation that RA are $(\beta\alpha)_8$ barrel enzymes. This fold is used by many enzymes, [116] which suggests a rather general application of the SPM tool.

6.1.2 Case example 2: Monoamine oxidase (MAO-N):

Monoamine Oxidase from *Aspergillus niger* (MAO-N) is a Flavin Adenine Dinucleotide (FAD) cofactor dependent enzyme that oxidizes amines into the corresponding imines by means of cofactor reduction. The interest in MAO-N arises from its application for obtaining chiral amines especially if combined with a non-selective chemical reductant.[121] MAO-N has a complex homodimeric structure and presents the active site located at an inner

hydrophobic cavity, which lies 15 Å away from the enzyme surface (see Figure 7). This buried active site located at the end of a long tunnel is responsible for the enzyme narrow substrate scope.

Given the limited amine scope accepted by wild-type MAO-N, multiple DE studies were performed to increase its catalytic activity and broaden its substrate scope.[39, 122, 123] This was achieved by introducing mutations located at ca. 18-19 Å from the active site (see Figure 7A). One of the DE variants, MAO-N-D5, presents one active site mutation (N336S), one mutation at the substrate entrance tunnel (I246M), and three additional distal mutations (M348K, T384N, D385S). We found that these remote mutations introduced by DE have a profound effect on the enzyme conformational dynamics, which impacts substrate binding and also catalysis.[35, 124] The reconstruction of the conformational landscape of MAO-N and D5 variant revealed some pre-existing conformational states that present a β-hairpin located in close proximity to the substrate binding tunnel in either a closed (i.e. X-ray like), partially-closed or open conformation. The relative stabilities of these states as well as the kinetics were shifted along the course of DE evolution: in MAO-N open conformations of the β-hairpin were easily accessed, whereas in MAO-N D5 closed and partially closed states were substantially more stabilised. This difference in the conformational landscape of MAO-N has a great impact on substrate scope, binding and catalysis. Interestingly, the conformation of the β-hairpin in one of the MAO-N subunits profoundly alters the catalytic activity of the other subunit, thus highlighting a delicate communication between monomers that regulates the enzyme functional activity. In MAO-N, open conformations of the β-hairpin are key for properly positioning the hydrophobic active site of the other subunit for productive hexylamine oxidation. This is altered along DE evolution as D5 variant requires closed and partially-closed states for efficient alpha-methylbenzylamine oxidation, thus being open states less functionally relevant. This also explains why open conformational states were slightly destabilised in the conformational landscape of the evolved D5 variant, as they were not selected for optimisation.

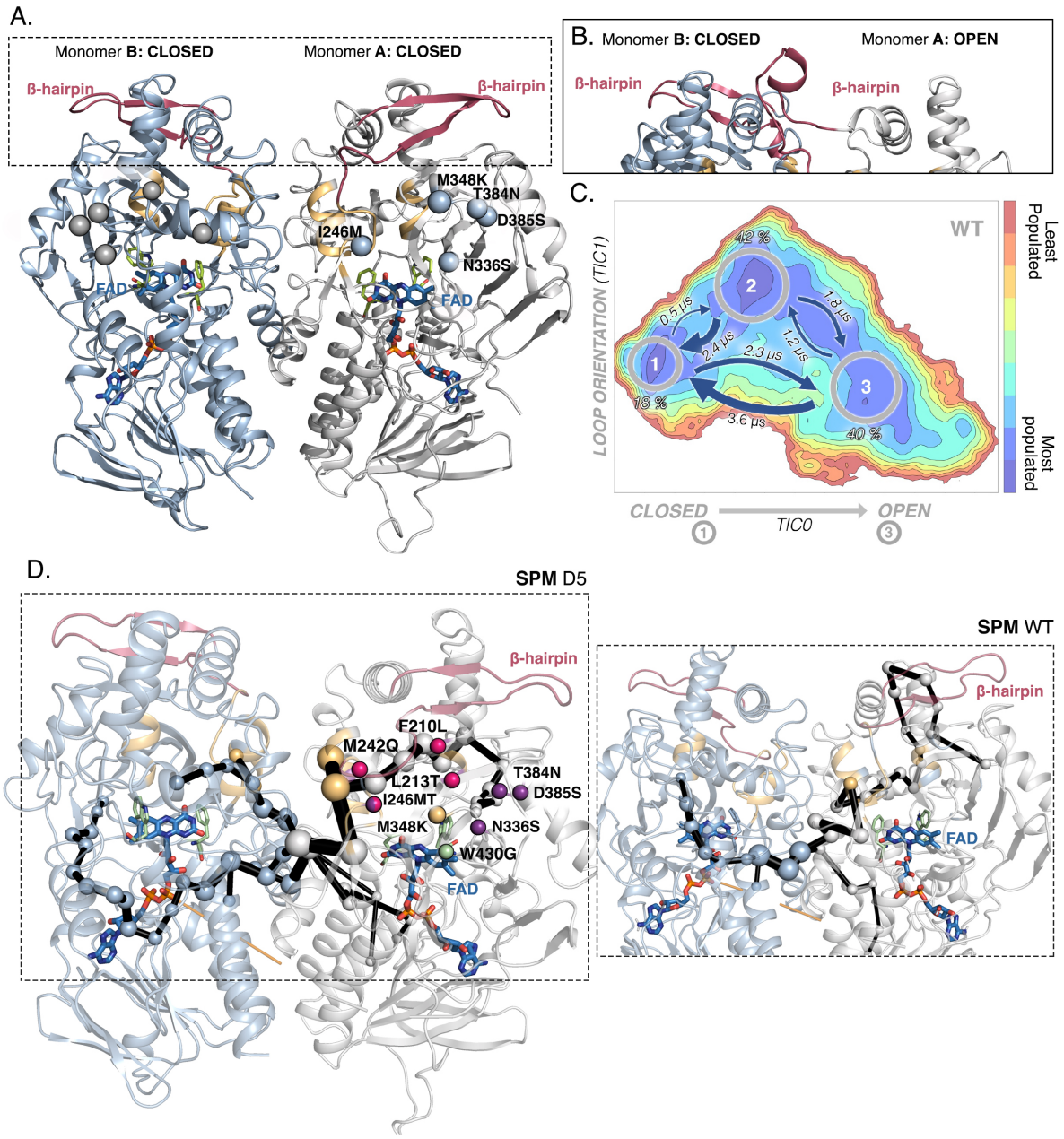


Figure 7. MAO-N Conformational landscape and SPM analysis along DE pathway. A. Representation of MAO-N homodimeric structure: the flexible β-hairpin of both monomers is in the X-ray (closed) conformation and is shown in raspberry, FAD cofactor in blue and sticks, and the positions introduced with DE for generating MAO-N D5 variant are shown in spheres. B. Zoom of the β-hairpin conformation in the open state. C. Reconstructed Free Energy Landscape with Markov State Model (MSM) for MAO-N WT. Each conformational state: open, closed, and partially closed is shown with a weighted sphere according to the relative populations, and arrows denote the conversions between states and the associated timescales. D. Computed SPM on MAO-N D5. The 5 mutations introduced with DE for generating MAO-N D5 from WT are all situated in an adjacent position to the SPM path (shown in purple) and making favourable interactions with SPM residues, except M348K (shown in yellow orange) that is far away from the path. The 4 additional mutations for generating MAO-N D9 from D5 are all of them at adjacent SPM positions (shown in pink), except the active site W430G mutation that is deviated from the path (shown in green). The SPM of MAO-N WT is also displayed, which shows how that the communication between subunits is altered along the DE pathway.

The application of the SPM in this particularly interesting system revealed two important aspects. As opposed to RA, MAO-N is not a $(\beta\alpha)_8$ barrel enzyme, still the positions identified with SPM directly contain DE positions or are located in adjacent positions (see Figure 7). Of relevance is that the mutations identified were located at distal positions, which are quite challenging to predict computationally. The SPM of D5 also contains the DE mutations for generating the most evolved D9 and D11 variants. This case example further supports that by carefully analysing the conformational dynamics of the starting enzyme through the SPM tool, a set of positions that can potentially regulate the enzyme activity can be predicted. Another interesting observation from this

study is that the SPM of MAO-N already highlights a delicate connection between subunits, as it provides a set of positions that directly link the active site of one monomer with the β-hairpin region of the other unit. Such inter-connection between subunits, is still operating in D5 although it has been partially lost due to the introduced mutations on the DE pathway. These observations suggest that SPM could also be applied for unrelated homodimeric enzymes presenting similar conformational inter-communications, but also for converting complex multimeric enzymes into stand-alone catalysts.

6.1.3 Case example 3: Tryptophan Synthase (TrpS)

Many enzyme complexes are allosterically regulated by a protein partner or by its other subunits (like in MAO-N), thus making them inefficient in the absence of their allosteric partners. From an industrial perspective, stand-alone versions of these enzymes are preferred as their use and efforts to engineer catalytic properties are simplified. Given the complexity of the allosteric communication existing between the different subunits, designing stand-alone versions of these enzymes becomes a major challenge. Intrigued by the MAO-N results that suggested the potential application of SPM for converting complex multimeric enzymes into stand-alone, we decided to focus our attention to Tryptophan synthase (TrpS) from *Pyrococcus furiosus*. TrpS is an heterodimeric enzyme complex composed by and α (TrpA) and β (TrpB) subunits, and presents a sophisticated allosteric regulation key for the multistep reaction mechanism.[125] DE was applied for converting the inefficient isolated TrpB subunit into an efficient stand-alone enzyme by the Arnold lab.[38, 126] The most evolved $TrpB^{0B2}$ variant contained six distal mutations, and was even more efficient than the original TrpS enzyme complex.

The reconstruction of the conformational landscape of the allosterically regulated TrpS, and its comparison to isolated and evolved TrpB revealed interesting insights into how DE recovered the conformational ensemble in the absence of protein partner (see Figure 8).[127] For efficient catalysis, TrpB has to adopt closed, partially closed and open conformations of a lid domain called COMM (see Figure 8B). This lid prevents the substrate loss to the media and presents some catalytically relevant residues that need to be positioned near the catalytic lysine for efficient catalysis, which is only achieved after COMM closure. The distal mutations introduced were able to recover the allosterically-driven conformational ensemble of TrpS converting the isolated TrpB lacking this required conformational heterogeneity into a conformationally rich $TrpB^{0B2}$ presenting accessible closed states of the COMM. In fact, the higher activity of $TrpB^{0B2}$ with respect to TrpS is explained by its ability to access open, partially closed and closed conformational states, which are separated by rather small energy barriers. The catalytically productive closed conformation in TrpS is higher in energy (thus less accessible) than in $TrpB^{0B2}$ (see Figure 8A).

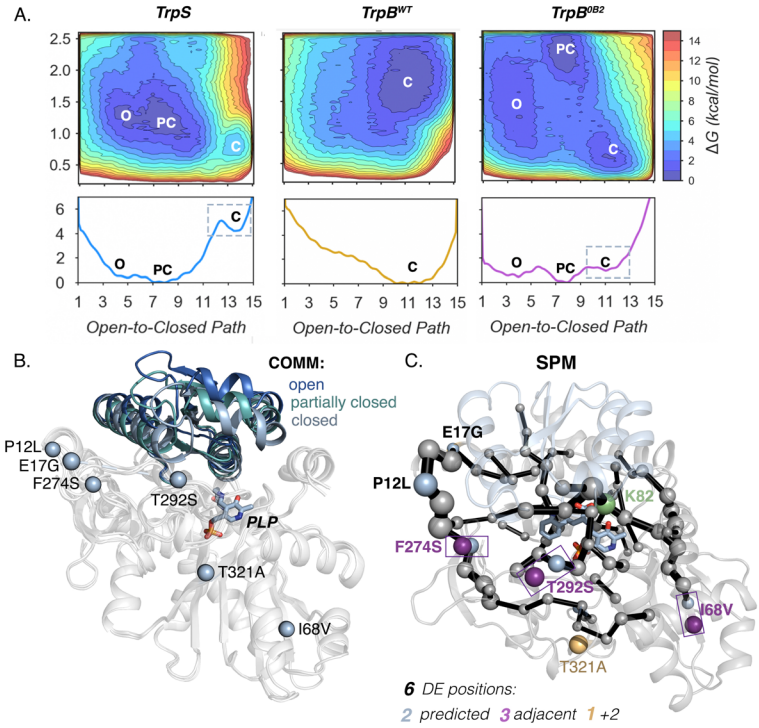


Figure 8. Tryptophan Synthase (TrpS) conformational landscapes and SPM comparison to Directed Evolution hotspots. A. The reconstructed conformational landscape of: the allosterically-regulated TrpS complex, the isolated TrpB unit, and laboratory-evolved stand-alone TrpB^{0B2}. TrpB^{0B2} recovers the allosterically-driven conformational ensemble (open (O), partially-closed (PC), and closed (C) states present). The closed state is more accessible than in TrpS, which explains the higher catalytic activity of the evolved variant. B. Representation of the open (dark blue), partially-closed (teal), and closed (light blue) conformation of the COMM domain from X-ray data. The mutations introduced with DE are shown in blue spheres. C. SPM identifies key positions for COMM domain conformational dynamics, which include 2 positions targeted with DE (P12L and E17G), 3 positions (Y301, D300 and Y69) are making persistent interactions with DE residues (F274S, T292S and I68V), and only T321A is nor in the path neither makes stable interactions.

The distal mutations introduced in TrpB^{0B2} are located either closed to the TrpA-TrpB interface (P12L, E17G), at one of the known tunnels based on X-ray data (F274S), in one of the loops situated close to the COMM domain when this adopts the closed conformation (T292S), or in other more remote regions of the enzyme (I68V, T321A). The specific effect of each mutation is not known experimentally, with the exception of T292S and P12L that provide a ca. 1- and 3-fold increase in k_{cat} with respect to isolated TrpB, respectively. Interestingly, P12L and also E17G are directly included in the computed SPM. Based on the T292-D300 interaction observed in the X-ray data, it was suggested that T292S has a key role in the open-to-closed conformational exchange of the COMM domain.[38] SPM predicts D300 as a key position for the enzyme conformational dynamics, rather than T292S that is interacting with D300 ca. 90% of the simulation time. These results suggest that by altering the interactions of the key D300 position, the COMM domain conformation can be modulated. Similarly, positions F274 and I68 make persistent non-covalent interactions with SPM residues, thus also suggesting their contribution towards TrpB conformational heterogeneity. The only mutation that is not included in SPM is T321A, which according to SPM has a minor role in COMM domain conformational dynamics (and its specific effect is not known experimentally).

The application of SPM in this particularly challenging system provides further support for its successful application in enzyme design.[127] In this particular case, SPM identifies key positions to alter the populations of the pre-existing conformational states, which in the absence of the allosteric partner can also be stabilised for stand-alone activity.

6.1.4 Conformationally driven SPM-based enzyme evolution

The application of SPM in the previously described unrelated enzymes (RA, MAO-N, TrpS) evidences that by evaluating the conformational ensemble of the starting enzyme, key positions that have a higher contribution to the enzyme conformational dynamics can be identified. This is inspiring, as it suggests that by computationally reconstructing the conformational landscape of the wild-type enzyme a reduced number of potential mutation hotspots for smart library construction can be rationally detected. In the case of RA, the application of SPM to the original scaffold used for computational enzyme design (PDB: 1LBL) already contains all the positions targeted along the DE pathways. The same trend is observed if SPM is applied to wild-type MAO-N and the allosterically-regulated TrpS complex. It is also important to highlight that the reduced set of positions provided by SPM is not restricted to the active site pocket but rather distributed to the whole protein structure, thus resolving the challenge of distal active site prediction. These observations suggest that conformationally-driven enzyme design protocols could be developed based on: i) reconstructing the conformational landscape of the starting enzyme (by means of multiple MD simulations, and/or enhanced sampling techniques), ii) identifying key positions that play a crucial role in the enzyme conformational dynamics (by means of SPM or related correlation-based approaches), and iii) selecting the best amino acid at the SPM identified position either experimentally or computationally. In this direction, the combination of SPM with other bioinformatic techniques, such as multiple sequence-alignment tools or phylogenetic approaches for ancestral enzyme reconstruction could further reduce the number of positions identified with SPM and, most importantly, suggest specific amino acid changes for each identified position.

Conclusion

The high complexity of enzyme catalysis has led to the development of a variety of computational strategies and protocols that put the focus on a reduced set of the key features of enzymatic catalysis. While initial attempts were mostly focused on the chemical step(s) of the process and overlooked the enzyme conformational dynamics often key for substrate binding and product release, recent computational approaches that based the design efforts into the conformational dynamics have been developed. The importance of considering the enzyme conformational dynamics has been highlighted by rationalizing many laboratory evolution pathways, together with ancestral enzyme reconstruction. The consideration of the enzyme as an ensemble of conformational states converts computational enzyme design into a population shift problem, i.e. predicted mutations should alter the conformational landscape of the enzyme and favour those conformational states key for the novel functionality. However, the identification of which positions might induce the desired conformational shift is not trivial.

The detailed study of many laboratory-engineered enzymes through Directed Evolution has also shown that this change in the conformational landscape is often achieved by distal mutations. The key role exerted by remote mutations on the active site of the enzyme suggests that allostery (*i.e.* regulation of enzyme function by distal positions) might be an intrinsic characteristic of enzymes, which can be exploited for enzyme evolution. Indeed, recently developed computational strategies that focus on the reconstruction of the enzyme conformational landscape have shown that if combined with correlation-based tools, such as the Shortest Path Map (SPM), potential hotspots for altering the conformational landscape and inducing the desired change in function can be identified. This new tool opens the door to new unprecedented computational protocols based on the introduction of active site and also distal mutations, which given the broad sequence space of enzymes are extremely challenging to predict. The evaluation of conformational landscapes and the elucidation of the complex enzymatic network of interactions could potentially bring the computational enzyme design one step forward.

Funding Information

I thank the Generalitat de Catalunya for the emerging group CompBioLab (2017 SGR-1707), Spanish MINECO for project PGC2018-102192-B-I00, and the funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ERC-2015-StG-679001).

Research Resources

I am grateful for the computer resources, technical expertise, and assistance provided by the Barcelona Supercomputing Center - Centro Nacional de Supercomputación.

Acknowledgments

I thank Miguel Angel Maria-Solano, Eila Serrano-Hervás, and Christian Curado for the helpful discussions and comments on the manuscript.

Notes

References

1. 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.
2. Romero-Rivera A, Garcia-Borràs M, Osuna S. Computational tools for the evaluation of laboratory-engineered biocatalysts. *Chem. Commun.* 2017, 53:284-297.
3. Damborsky J, Brezovsky J. Computational tools for designing and engineering enzymes. *Curr. Opin. Chem. Biol.* 2014, 19:8-16.
4. Arnold FH. The nature of chemical innovation: new enzymes by evolution. *Q. Rev. Biophys.* 2015, 48:404-410.
5. Currin A, Swainston N, Day PJ, Kell DB. Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chem. Soc. Rev.* 2015, 44:1172-1239.
6. Bornscheuer UT, Huisman GW, Kazlauskas RJ, Lutz S, Moore JC, Robins K. Engineering the third wave of biocatalysis. *Nature* 2012, 485:185-194.
7. Francis JC, Hansche PE. DE FIRST. *Genetics* 1972, 70:59-73.
8. Lutz S, Bornscheuer UT. *Protein Engineering Handbook*. Vol. Volume 1 & Volume 2: Wiley-VCH Verlag GmbH & Co. KGaA; 2008.
9. Packer MS, Liu DR. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* 2015, 16:379-394.
10. Jiang L, Althoff EA, Clemente FR, Doyle L, Rothlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF, et al. De novo computational design of retro-aldol enzymes. *Science* 2008, 319:1387-1391.
11. Rothlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, et al. Kemp elimination catalysts by computational enzyme design. *Nature* 2008, 453:190-U194.
12. Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, St.Clair JL, Gallaher JL, Hilvert D, Gelb MH, Stoddard BL, et al. Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science* 2010, 329:309-313.
13. Kourist R, Jochens H, Bartsch S, Kuipers R, Padhi SK, Gall M, Böttcher D, Joosten H-J, Bornscheuer UT. The α/β-Hydrolase Fold 3DM Database (ABHDB) as a Tool for Protein Engineering. *ChemBioChem* 2010, 11:1635-1643.
14. Kuipers RK, Joosten H-J, van Berkel WJH, Leferink NGH, Rooijen E, Ittmann E, van Zimmeren F, Jochens H, Bornscheuer U, Vriend G, et al. 3DM: Systematic analysis of heterogeneous superfamily data to discover protein functionalities. *Proteins* 2010, 78:2101-2113.

15. Kuipers RKP, Joosten H-J, Verwiel E, Paans S, Akerboom J, van der Oost J, Leferink NGH, van Berkel WJH, Vriend G, Schaap PJ. Correlated mutation analyses on super-family alignments reveal functionally important residues. *Proteins* 2009, 76:608-616.
16. Addington TA, Mertz RW, Siegel JB, Thompson JM, Fisher AJ, Filkov V, Fleischman NM, Suen AA, Zhang C, Toney MD. Janus: Prediction and Ranking of Mutations Required for Functional Interconversion of Enzymes. *J. Mol. Biol.* 2013, 425:1378-1389.
17. Pavelka A, Chovancova E, Damborsky J. HotSpot Wizard: a web server for identification of hot spots in protein engineering. *Nucleic Acids Res.* 2009, 37:W376-W383.
18. Kazlauskas RJ, Bornscheuer UT. Finding better protein engineering strategies. *Nat. Chem. Biol.* 2009, 5:526-529.
19. Turner NJ. Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* 2009, 5:567-573.
20. Currin A, Swainston N, Day PJ, Kell DB. SpeedyGenes: an improved gene synthesis method for the efficient production of error-corrected, synthetic protein libraries for directed evolution. *Protein Eng. Des. Sel.* 2014, 27:273-280.
21. Xiao H, Bao Z, Zhao H. High Throughput Screening and Selection Methods for Directed Enzyme Evolution. *Ind. Eng. Chem. Res.* 2015, 54:4011-4020.
22. Yang KK, Wu Z, Arnold FH. Machine-learning-guided directed evolution for protein engineering. *Nat. Methods* 2019, 16:687-694.
23. Mazurenko S, Prokop Z, Damborsky J. Machine Learning in Enzyme Engineering. *ACS Catal.* 2020, 10:1210-1223.
24. Romero PA, Arnold FH. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* 2009, 10:866-876.
25. Jaeckel C, Kast P, Hilvert D. Protein design by directed evolution. In: *Annu. Rev. Biophys.* Vol. 37; 2008, 153-173.
26. Renata H, Wang ZJ, Arnold FH. Expanding the Enzyme Universe: Accessing Non-Natural Reactions by Mechanism-Guided Directed Evolution. *Angew. Chem. Int. Ed.* 2015, 54:3351-3367.
27. Campbell E, Kaltenbach M, Correy GJ, Carr PD, Porebski BT, Livingstone EK, Afriat-Jurnou L, Buckle AM, Weik M, Hollfelder F, et al. The role of protein dynamics in the evolution of new enzyme function. *Nat. Chem. Biol.* 2016, 12:944-950.
28. Huffman MA, Fryszkowska A, Alvizo O, Borra-Garske M, Campos KR, Canada KA, Devine PN, Duan D, Forstater JH, Grosser ST, et al. Design of an in vitro biocatalytic cascade for the manufacture of islatravir. *Science* 2019, 366:1255.
29. Pavlidis IV, Weiß MS, Genz M, Spurr P, Hanlon SP, Wirz B, Iding H, Bornscheuer UT. Identification of (S)-selective transaminases for the asymmetric synthesis of bulky chiral amines. *Nat. Chem.* 2016, 8:1076-1082.
30. Savile CK, Janey JM, Mundorff EC, Moore JC, Tam S, Jarvis WR, Colbeck JC, Krebber A, Fleitz FJ, Brands J, et al. Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. *Science* 2010, 329:305-309.
31. Obexer R, Godina A, Garrabou X, Mittl PRE, Baker D, Griffiths AD, Hilvert D. Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat. Chem.* 2017, 9:50-56.
32. Jiménez-Osés G, Osuna S, Gao X, Sawaya MR, Gilson L, Collier SJ, Huisman GW, Yeates TO, Tang Y, Houk KN. The role of distant mutations and allosteric regulation on LovD active site dynamics. *Nat. Chem. Biol.* 2014, 10:431-436.
33. Gunasekaran K, Ma B, Nussinov R. Is allostery an intrinsic property of all dynamic proteins? *Proteins* 2004, 57:433-443.
34. 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.

35. 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.
36. Petrović D, Risso VA, Kamerlin SCL, Sanchez-Ruiz JM. Conformational dynamics and enzyme evolution. *J. R. Soc. Interface* 2018, 15.
37. Campitelli P, Modi T, Kumar S, Ozkan SB. The Role of Conformational Dynamics and Allostery in Modulating Protein Evolution. *Annu. Rev. Biophys.* 2020, 49:267-288.
38. 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. USA* 2015, 112:14599-14604.
39. Ghislieri D, Green AP, Pontini M, Willies SC, Rowles I, Frank A, Grogan G, Turner NJ. Engineering an Enantioselective Amine Oxidase for the Synthesis of Pharmaceutical Building Blocks and Alkaloid Natural Products. *J. Am. Chem. Soc.* 2013, 135:10863-10869.
40. Benkovic SJ, Hammes-Schiffer S. A Perspective on Enzyme Catalysis. *Science* 2003, 301:1196-1202.
41. Garcia-Viloca M, Gao J, Karplus M, Truhlar DG. How Enzymes Work: Analysis by Modern Rate Theory and Computer Simulations. *Science* 2004, 303:186-195.
42. Marti S, Roca M, Andres J, Moliner V, Silla E, Tunon I, Bertran J. Theoretical insights in enzyme catalysis. *Chem. Soc. Rev.* 2004, 33:98-107.
43. Nagel ZD, Klinman JP. A 21st century revisionist's view at a turning point in enzymology. *Nat. Chem. Biol.* 2009, 5:543-550.
44. Warshel A, Sharma PK, Kato M, Xiang Y, Liu H, Olsson MHM. Electrostatic Basis for Enzyme Catalysis. *Chem. Rev.* 2006, 106:3210-3235.
45. Boehr DD, Nussinov R, Wright PE. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* 2009, 5:789-796.
46. Hammes GG, Benkovic SJ, Hammes-Schiffer S. Flexibility, Diversity, and Cooperativity: Pillars of Enzyme Catalysis. *Biochemistry* 2011, 50:10422-10430.
47. Bhabha G, Lee J, Ekiert DC, Gam J, Wilson IA, Dyson HJ, Benkovic SJ, Wright PE. A Dynamic Knockout Reveals That Conformational Fluctuations Influence the Chemical Step of Enzyme Catalysis. *Science* 2011, 332:234-238.
48. Silva RG, Murkin AS, Schramm VL. Femtosecond dynamics coupled to chemical barrier crossing in a Born-Oppenheimer enzyme. *Proc. Natl. Acad. Sci. USA* 2011, 108:18661-18665.
49. Glowacki DR, Harvey JN, Mulholland AJ. Taking Ockham's razor to enzyme dynamics and catalysis. *Nat. Chem.* 2012, 4:169-176.
50. Kamerlin SCL, Warshel A. At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins* 2010, 78:1339-1375.
51. Henzler-Wildman KA, Thai V, Lei M, Ott M, Wolf-Watz M, Fenn T, Pozharski E, Wilson MA, Petsko GA, Karplus M, et al. Intrinsic motions along an enzymatic reaction trajectory. *Nature* 2007, 450:838-844.
52. Neu A, Neu U, Fuchs A-L, Schlager B, Sprangers R. An excess of catalytically required motions inhibits the scavenger decapping enzyme. *Nat. Chem. Biol.* 2015, 11:697-704.
53. Tokuriki N, Tawfik DS. Protein Dynamism and Evolvability. *Science* 2009, 324:203-207.
54. Khersonsky O, Tawfik DS. Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* 2010, 79:471-505.
55. Campbell EC, Correy GJ, Mabbitt PD, Buckle AM, Tokuriki N, Jackson CJ. Laboratory evolution of protein conformational dynamics. *Curr. Opin. Struc. Biol.* 2018, 50:49-57.
56. Crean RM, Gardner JM, Kamerlin SCL. Harnessing Conformational Plasticity to Generate Designer Enzymes. *J. Am. Chem. Soc.* 2020.

57. Huang P-S, Boyken SE, Baker D. The coming of age of de novo protein design. *Nature* 2016, 537:320.
58. Kipnis Y, Baker D. Comparison of designed and randomly generated catalysts for simple chemical reactions. *Protein Sci.* 2012, 21:1388-1395.
59. Kries H, Blomberg R, Hilvert D. De novo enzymes by computational design. *Curr. Opin. Chem. Biol.* 2013, 17:221-228.
60. Jessen LE, Hoof I, Lund O, Nielsen M. SigniSite: Identification of residue-level genotype-phenotype correlations in protein multiple sequence alignments. *Nucleic Acids Res.* 2013, 41:W286-W291.
61. Ebert MC, Pelletier JN. Computational tools for enzyme improvement: why everyone can – and should – use them. *Curr. Opin. Chem. Biol.* 2017, 37:89-96.
62. Khersonsky O, Lipsh R, Avizemer Z, Ashani Y, Goldsmith M, Leader H, Dym O, Rogotner S, Trudeau DL, Prilusky J, et al. Automated Design of Efficient and Functionally Diverse Enzyme Repertoires. *Mol. Cell* 2018, 72:178-186.e175.
63. Zanghellini A, Jiang L, Wollacott AM, Cheng G, Meiler J, Althoff EA, Rothlisberger D, Baker D. New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci.* 2006, 15:2785-2794.
64. Kiss G, Çelebi-Ölçüm N, Moretti R, Baker D, Houk KN. Computational Enzyme Design. *Angew. Chem. Int. Ed.* 2013, 52:5700-5725 and references cited therein.
65. Świderek K, Tuñón I, Moliner V, Bertran J. Computational strategies for the design of new enzymatic functions. *Arch. Biochem. Biophys.* 2015, 582:68-79.
66. Privett HK, Kiss G, Lee TM, Blomberg R, Chica RA, Thomas LM, Hilvert D, Houk KN, Mayo SL. Iterative approach to computational enzyme design. *Proc. Natl. Acad. Sci. USA* 2012, 109:3790-3795.
67. Kiss G, Röthlisberger D, Baker D, Houk KN. Evaluation and ranking of enzyme designs. *Protein Sci.* 2010, 19:1760-1773.
68. Wijma HJ, Floor RJ, Bjelic S, Marrink SJ, Baker D, Janssen DB. Enantioselective Enzymes by Computational Design and In Silico Screening. *Angew. Chem.* 2015, 127:3797-3801.
69. Li R, Wijma HJ, Song L, Cui Y, Otzen M, Tian Ye, Du J, Li T, Niu D, Chen Y, et al. Computational redesign of enzymes for regio- and enantioselective hydroamination. *Nat. Chem. Biol.* 2018, 14:664-670.
70. Wijma HJ, Fürst MJLJ, Janssen DB. A Computational Library Design Protocol for Rapid Improvement of Protein Stability: FRESCO. In: Bornscheuer UT, Höhne M, eds. *Protein Engineering: Methods and Protocols*. New York, NY: Springer New York; 2018, 69-85.
71. Smith CA, Kortemme T. Backrub-Like Backbone Simulation Recapitulates Natural Protein Conformational Variability and Improves Mutant Side-Chain Prediction. *J. Mol. Biol.* 2008, 380:742-756.
72. Huang P-S, Ban Y-EA, Richter F, Andre I, Vernon R, Schief WR, Baker D. RosettaRemodel: A Generalized Framework for Flexible Backbone Protein Design. *PLoS ONE* 2011, 6:e24109.
73. Fu X, Apgar JR, Keating AE. Modeling Backbone Flexibility to Achieve Sequence Diversity: The Design of Novel α-Helical Ligands for Bcl-xL. *J. Mol. Biol.* 2007, 371:1099-1117.
74. Ding F, Dokholyan NV. Emergence of Protein Fold Families through Rational Design. *PLoS Comput. Biol.* 2006, 2:e85.
75. Larson SM, England JL, Desjarlais JR, Pande VS. Thoroughly sampling sequence space: Large-scale protein design of structural ensembles. *Protein Sci.* 2002, 11:2804-2813.
76. Davey JA, Chica RA. Multistate approaches in computational protein design. *Protein Sci.* 2012, 21:1241-1252.
77. St-Jacques AD, Eyahpaise M-ÈC, Chica RA. Computational Design of Multisubstrate Enzyme Specificity. *ACS Catal.* 2019, 9:5480-5485.

78. Lindert S, Meiler J, McCammon JA. Iterative Molecular Dynamics—Rosetta Protein Structure Refinement Protocol to Improve Model Quality. *J. Chem. Theory Comput.* 2013, 9:3843-3847.
79. Mandell DJ, Kortemme T. Backbone flexibility in computational protein design. *Curr. Opin. Biotech.* 2009, 20:420-428.
80. Friedland GD, Kortemme T. Designing ensembles in conformational and sequence space to characterize and engineer proteins. *Curr. Opin. Struc. Biol.* 2010, 20:377-384.
81. Nosrati GR, Houk KN. SABER: A computational method for identifying active sites for new reactions. *Protein Sci.* 2012, 21:697-706.
82. Malisi C, Kohlbacher O, Höcker B. Automated scaffold selection for enzyme design. *Proteins* 2009, 77:74-83.
83. Fazelinia H, Cirino PC, Maranas CD. OptGraft: A computational procedure for transferring a binding site onto an existing protein scaffold. *Protein Sci.* 2009, 18:180-195.
84. Huang X, Han K, Zhu Y. Systematic optimization model and algorithm for binding sequence selection in computational enzyme design. *Protein Sci.* 2013, 22:929-941.
85. Amrein BA, Steffen-Munsberg F, Szeler I, Purg M, Kulkarni Y, Kamerlin SCL. CADEE: Computer-Aided Directed Evolution of Enzymes. *IUCrJ* 2017, 4:50-64.
86. Mondal D, Kolev V, Warshel A. Combinatorial Approach for Exploring Conformational Space and Activation Barriers in Computer-Aided Enzyme Design. *ACS Catal.* 2020, 10:6002-6012.
87. De Raffele D, Martí S, Moliner V. Understanding the Directed Evolution of de novo Retro-Aldolases from QM/MM Studies. *ACS Catal.* 2020.
88. Świderek K, Tuñón I, Moliner V. Predicting enzymatic reactivity: from theory to design. *WIREs Comput. Mol. Sci.* 2014, 4:407-421.
89. Grisewood MJ, Gifford NP, Pantazes RJ, Li Y, Cirino PC, Janik MJ, Maranas CD. OptZyme: Computational Enzyme Redesign Using Transition State Analogues. *PLoS ONE* 2013, 8:e75358.
90. Monza E. AS, Lucas M.F., Guallar V. Molecular Modeling in Enzyme Design, Toward In Silico Guided Directed Evolution. In: *Directed Enzyme Evolution: Advances and Applications*: Springer; 2017.
91. Alonso S, Santiago G, Cea-Rama I, Fernandez-Lopez L, Coscolín C, Modregger J, Ressmann AK, Martínez-Martínez M, Marrero H, Bargiela R, et al. Genetically engineered proteins with two active sites for enhanced biocatalysis and synergistic chemo- and biocatalysis. *Nat. Catal.* 2020, 3:319-328.
92. Mateljak I, Monza E, Lucas MF, Guallar V, Aleksejeva O, Ludwig R, Leech D, Shleev S, Alcalde M. Increasing Redox Potential, Redox Mediator Activity, and Stability in a Fungal Laccase by Computer-Guided Mutagenesis and Directed Evolution. *ACS Catal.* 2019, 9:4561-4572.
93. Wagner JR, Sørensen J, Hensley N, Wong C, Zhu C, Perison T, Amaro RE. POVME 3.0: Software for Mapping Binding Pocket Flexibility. *J. Chem. Theory Comput.* 2017, 13:4584-4592.
94. Kreß N, Halder JM, Rapp LR, Hauer B. Unlocked potential of dynamic elements in protein structures: channels and loops. *Curr. Opin. Chem. Biol.* 2018, 47:109-116.
95. Kokkonen P, Bednar D, Pinto G, Prokop Z, Damborsky J. Engineering enzyme access tunnels. *Biotechnol. Adv.* 2019, 37:107386.
96. Prokop Z, Gora, A., Brezovsky, J., Chaloupkova, R., Stepankova, V., Damborsky, J. Engineering of protein tunnels: the keyhole-lock-key model for catalysis by enzymes with buried active sites. In: Wiley-VCH, ed. *Protein Engineering Handbook*. Vol. 3. Weinheim: Wiley-VCH; 2012, 421-464.

97. Chovancova E, Pavelka A, Benes P, Strnad O, Brezovsky J, Kozlikova B, Gora A, Sustr V, Klvana M, Medek P, et al. CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *PLOS Comput. Biol.* 2012, 8:e1002708.
98. Röthlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, et al. Kemp elimination catalysts by computational enzyme design. *Nature* 2008, 453:190-195.
99. Khersonsky O, Kiss G, Roethlisberger D, Dym O, Albeck S, Houk KN, Baker D, Tawfik DS. Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proc. Natl. Acad. Sci. USA* 2012, 109:10358-10363.
100. Khersonsky O, Röthlisberger D, Dym O, Albeck S, Jackson CJ, Baker D, Tawfik DS. Evolutionary Optimization of Computationally Designed Enzymes: Kemp Eliminases of the KE07 Series. *J. Mol. Biol.* 2010, 396:1025-1042.
101. Khersonsky O, Röthlisberger D, Wollacott AM, Murphy P, Dym O, Albeck S, Kiss G, Houk KN, Baker D, Tawfik DS. Optimization of the In-Silico-Designed Kemp Eliminase KE70 by Computational Design and Directed Evolution. *J. Mol. Biol.* 2011, 407:391-412.
102. Privett HK, Kiss G, Lee TM, Blomberg R, Chica RA, Thomas LM, Hilvert D, Houk KN, Mayo SL. Iterative approach to computational enzyme design. *Proceedings of the National Academy of Sciences* 2012, 109:3790-3795.
103. Blomberg R, Kries H, Pinkas DM, Mittl PR, Grutter MG, Privett HK, Mayo SL, Hilvert D. Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* 2013, 503:418-421.
104. Broom A, Rakotoharisoa RV, Thompson MC, Zarifi N, Nguyen E, Mukhametzhanov N, Liu L, Fraser JS, Chica RA. Evolution of an enzyme conformational ensemble guides design of an efficient biocatalyst. *bioRxiv* 2020:2020.2003.2019.999235.
105. Korendovych IV, Kulp DW, Wu Y, Cheng H, Roder H, DeGrado WF. Design of a switchable eliminase. *Proc. Natl. Acad. Sci. USA* 2011, 108:6823-6827.
106. Merski M, Shoichet BK. Engineering a model protein cavity to catalyze the Kemp elimination. *Proc. Natl. Acad. Sci. USA* 2012, 109:16179-16183.
107. 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.
108. Devamani T, Rauwerdink AM, Lunzer M, Jones BJ, Mooney JL, Tan MAO, Zhang Z-J, Xu J-H, Dean AM, Kazlauskas RJ. Catalytic Promiscuity of Ancestral Esterases and Hydroxynitrile Lyases. *J. Am. Chem. Soc.* 2016, 138:1046-1056.
109. 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. USA* 2020, 117:346-354.
110. Guo J, Zhou H-X. Protein Allostery and Conformational Dynamics. *Chem. Rev.* 2016, 116:6503-6515.
111. Ma B, Nussinov R. Protein dynamics: Conformational footprints. *Nat. Chem. Biol.* 2016, 12:890-891.
112. Gobeil Sophie MC, Clouthier Christopher M, Park J, Gagné D, Berghuis Albert M, Doucet N, Pelletier Joelle N. Maintenance of Native-like Protein Dynamics May Not Be Required for Engineering Functional Proteins. *Chem. Biol.* 2014, 21:1330-1340.
113. Atilgan AR, Akan P, Baysal C. Small-world communication of residues and significance for protein dynamics. *Biophys. J.* 2004, 86:85-91.
114. Sethi A, Eargle J, Black AA, Luthey-Schulten Z. Dynamical networks in tRNA:protein complexes. *Proc. Natl. Acad. Sci. USA* 2009, 106:6620-6625.
115. Girvan M, Newman MEJ. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* 2002, 99:7821-7826.

116. Rivalta I, Sultan MM, Lee N-S, Manley GA, Loria JP, Batista VS. Allosteric pathways in imidazole glycerol phosphate synthase. *Proc. Natl. Acad. Sci. USA* 2012, 109:E1428-E1436.
117. Csárdi G, Nepusz T. The igraph software package for complex network research. *InterJournal* 2006, Complex Systems:1695-1704.
118. Giger L, Caner S, Obexer R, Kast P, Baker D, Ban N, Hilvert D. Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat. Chem. Biol.* 2013, 9:494-498.
119. De Raffele D, Martí S, Moliner V. QM/MM Theoretical Studies of a de Novo Retro-Aldolase Design. *ACS Catal.* 2019, 9:2482-2492.
120. Zeymer C, Zschoche R, Hilvert D. Optimization of Enzyme Mechanism along the Evolutionary Trajectory of a Computationally Designed (Retro-)Aldolase. *J. Am. Chem. Soc.* 2017, 139:12541-12549.
121. Herter S, Medina F, Wagschal S, Benhaïm C, Leipold F, Turner NJ. Mapping the substrate scope of monoamine oxidase (MAO-N) as a synthetic tool for the enantioselective synthesis of chiral amines. *Bioorganic Med. Chem.* 2017:30391-30397.
122. Carr R, Alexeeva M, Enright A, Eve TSC, Dawson MJ, Turner NJ. Directed Evolution of an Amine Oxidase Possessing both Broad Substrate Specificity and High Enantioselectivity. *Angew. Chem. Int. Ed.* 2003, 42:4807-4810.
123. Dunsmore CJ, Carr R, Fleming T, Turner NJ. A Chemo-Enzymatic Route to Enantiomerically Pure Cyclic Tertiary Amines. *J. Am. Chem. Soc.* 2006, 128:2224-2225.
124. Curado-Carballada C, Feixas F, Osuna S. Molecular Dynamics Simulations on Aspergillus niger Monoamine Oxidase: Conformational Dynamics and Inter-monomer Communication Essential for Its Efficient Catalysis. *Adv. Synth. Catal.* 2019, 361:2718-2726.
125. Dunn MF. Allosteric regulation of substrate channeling and catalysis in the tryptophan synthase bienzyme complex. *Arch. Biochem. Biophys.* 2012, 519:154-166.
126. 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.
127. 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.