Complex Loop Dynamics Underpin Activity, Specificity and Evolvability in the $(\beta\alpha)_8$ Barrel Enzymes of Histidine and Tryptophan Biosynthesis

Adrian Romero-Rivera,[1,‡,#] Marina Corbella,[1,‡] Antonietta Parracino,[1] Wayne M. Patrick[2] and Shina Caroline Lynn Kamerlin[1,*]

1. Department of Chemistry – BMC, Uppsala University, BMC Box 576, S-751 23 Uppsala, Sweden. 2. Centre for Biodiscovery, School of Biological Sciences, Victoria University of Wellington, Wellington 6012, New Zealand.

Corresponding author email addresses: lynn.kamerlin@kemi.uu.se

[‡] These authors contributed equally.

[#] Current address: Sprint Bioscience AB, Hälsovägen 7, Huddinge, Sweden

Abstract

Enzymes are conformationally dynamic, and their dynamical properties play an important role in regulating their specificity and evolvability. In this context, substantial attention has been paid to the role of ligand-gated conformational changes in enzyme catalysis; however, such studies have focused on tremendously proficient enzymes such as triosephosphate isomerase and orotidine 5'-monophosphate decarboxylase, where the rapid (μs timescale) motion of a single loop dominates the transition between catalytically inactive and active conformations. In contrast, the $(\beta\alpha)_8$-barrels of tryptophan and histidine biosynthesis, such as the specialist isomerase enzymes HisA and TrpF, and the bifunctional isomerase PriA, are decorated by multiple long loops that undergo conformational transitions on the ms (or slower) timescale. Studying the interdependent motions of multiple slow loops, and their role in catalysis, poses a significant computational challenge. This work combines conventional and enhanced molecular dynamics simulations with empirical valence bond simulations to provide rich detail of the conformational behavior of the catalytic loops in HisA, PriA and TrpF, and the role of their plasticity in facilitating bifunctionality in PriA and evolved HisA variants. In addition, we demonstrate that, similar to other enzymes activated by ligand-gated conformational changes, loops 3 and 4 of HisA and PriA act as gripper loops, facilitating the isomerization of the large bulky substrate ProFAR, albeit now on much slower timescales. This hints at convergent evolution on these different $(\beta\alpha)_8$-barrel scaffolds. Finally, our work highlights the potential of engineering loop dynamics as a powerful tool to artificially manipulate the diverse catalytic repertoire of TIM-barrel proteins.

Keywords: $(\beta\alpha)_8$-Barrel Enzymes • Protein Evolution • Catalytic Promiscuity • Computational Enzymology • Empirical Valence Bond • Enhanced Sampling

Introduction

Enzymes are dynamic systems able to explore many different conformations, and these dynamical properties are clearly connected to their biological function. Examples of this include allosteric regulation and product release,[1] as well as the role of conformational selection in enzyme catalysis,[2-7] promiscuity[4] and evolution.[8-14] In addition, such conformational dynamics can, in principle, be engineered in a targeted fashion to allow enzymes to acquire new catalytic functions and/or physiochemical properties.[8, 9, 14-17] Understanding how enzymes manipulate and modulate conformational dynamics during both natural and directed evolution is an important step in this direction. In particular, understanding the dynamical behavior of decorating loops that cover enzyme active sites is important as such loops can regulate substrate selectivity, the evolution of new activities, and potentially also turnover rates.[3, 18-32] As such, targeting the dynamics of such active site loops is attractive from an engineering perspective,[16, 33] and therefore there is substantial interest in understanding loop dynamics and its impact on selectivity and catalysis.

In this context, there have been extensive studies of a wide-range of enzymes, such as triosephosphate isomerase (TPI),[34, 35] orotidine 5'-monophosphate decarboxylase, (OMPDC)[36] glycerol 3-phosphate dehydrogenase (GPDH),[37, 38] 1-deoxy-D-xylulose-5-phosphate reductoisomerase,[39, 40] and β-phosphoglucomutase,[41] which have been demonstrated to be activated by ligand-gated conformational changes. Specifically, interactions between a key "gripper" loop decorating the active site and the non-reactive phosphodianion groups of the substrates of these enzymes trigger substantial conformational changes in the gripper loop, facilitating energetically unfavorable transitions from catalytically inactive open to catalytically active closed conformations, and these conformational transitions are central to the catalytic activities and high proficiencies of these enzymes.[22, 31]

It is noteworthy that several of the aforementioned enzymes have TIM-barrel folds. This fold comprises eight repeated (βα)-units, and most if not all TIM-barrel proteins possess decorating loops,[18, 42] the conformational diversity of which likely plays an important role in regulating specificity and function.[25, 28] These flexible loops can vary in length,[18] but are typically used to bind substrate, and to sequester the active site from solvent by closing over the active site, and it has been suggested that the active site geometries of these enzymes are shaped by the residues of these loops.[43] However, these well-characterized examples of proteins activated by ligand-gated conformational changes all focus on the roles and importance of single loops, such as gripper loop 6 in TPI. Studying the ligand-gated motion of a single loop can already pose substantial challenges;[44] systems with multiple active site loops undergoing substantial conformational changes are even more complex, and therefore unsurprisingly understudied in the literature.

We have sought to address this gap in knowledge by studying active site loop dynamics in the $(\beta\alpha)_8$-barrels of tryptophan and histidine biosynthesis. The isomerase enzymes HisA, TrpF and PriA are model systems for the evolution of specificity and activity.[28, 45-48] As shown in **Figure 1**, HisA catalyzes isomerization of the aminoaldose N'-[(5'-phosphoribosyl)-formimino]-5-aminoimidazole-4-carboxamide-ribonucleotide (ProFAR) into the aminoketose N′-[(5′-phosphoribulosyl)-formimino]-5-aminoimidazole-4-carboxamide-ribonucleotide (PRFAR). TrpF catalyses the same Amadori rearrangement on N-(5′-phosphoribosyl)anthranilate (PRA), producing 1-(2-carboxy-phenylamino)-1'-deoxyribulose-5'-phosphate (CdRP). This rearrangement proceeds *via* a Schiff acid-base mechanism, that utilizes aspartate (and in the case of TrpF) cysteine residues as acid-base pairs.[49]

Interestingly, many actinobacteria lack the *trp*F gene, instead possessing a gene for a bifunctional isomerase, PriA.[50] The PriA from *Mycobacterium tuberculosis* (*Mt*PriA) has been

particularly well characterized, and has k_{cat}/K_M values of ~10^4 M^{-1} s^{-1} for HisA activity and ~10^5 – 10^6 M^{-1} s^{-1} for TrpF activity.[25, 51] Not only are there naturally occurring bifunctional enzymes, but promiscuous TrpF activity has been detected on both ancestral and extant specialist HisA enzymes, with k_{cat} values ranging from 10^{-4} to 10^{-2} s^{-1}.[52] HisA has also been converted into TrpF by directed evolution,[53] and in serial passaging experiments.[46] In the latter study, laboratory evolution of the ProFAR-specific HisA (lacking TrpF activity) from *Salmonella enterica* over 3,000 generations yielded an extensive suite of mutations in the *S. enterica* HisA (*Se*HisA) that resulted in specialist HisA enzymes, specialist TrpF enzymes and PriA-like bifunctional enzymes.

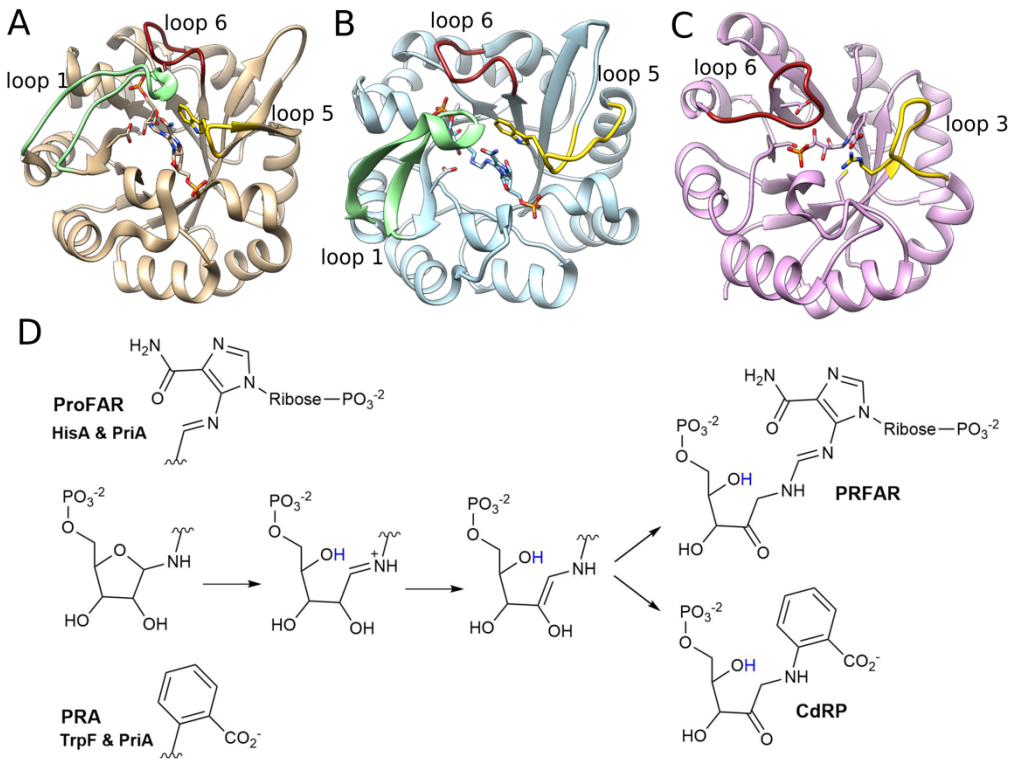

Figure 1. Tertiary structures of (**A**) *Se*HisA in complex with ProFAR (PDB ID: 5A5W[54, 55]), (**B**) *Mt*PriA in complex with PRFAR (PDB ID: 3ZS4[55]) and (**C**) *Tm*TrpF in complex with product analog rCdRP (PDB ID: 1LBM[49]). Loop 1 (residues 11-29 in HisA, 15-33 in *Mt*PriA), loop 5 (residues 142-147 in *Se*HisA, 141-151 in PriA and 30-39 in *Tm*TrpF (loop 3) and loop 6 (residues 172-182 in *Se*HisA, 172-181 in *Mt*PriA and 128-139 in *Tm*TrpF) are highlighted in light

green, yellow and dark red on each structure, respectively. Loop 1 in *Tm*TrpF is short (four residues), which is why no corresponding loop is annotated on this panel. Note that for clarity, N7D and A176D reversions were applied to the structure of *Se*HisA in complex with ProFAR (these reversions were also applied in our simulations, as described in the **Methodology** section). (**D**) The proposed mechanism for the Amadori rearrangement leading to the isomerization of substrates ProFAR and PRA by the different enzymes.[49]

As shown in **Figure 1**, HisA and PriA are decorated by three long catalytic loops, loops 1, 5, and 6 (or two analogous loops in the case of TrpF, which has lost most of loop 1).[25, 28] Of these loops, loop 5 carries key residues that are important for substrate binding, loop 6 carries the catalytically important aspartic acid side chain, and a number of mutations important for interaction with substrate PRA have been observed at position 15 of loop 1.[28, 45, 54] The mutants generated by Näsvall *et al.*[46] were the subject of detailed structural and biochemical analyses.[28] As with PriA,[25] this analysis of the evolved bifunctional *Se*HisA variants indicated that the bifunctionality is driven by competition between not just the substrates ProFAR and PRA, but also between structurally distinct conformations of loops 1 and 5, in particular[28] (**Figure S1**).

Furthermore, although the isomerization of both ProFAR and PRA proceeds through the same Amadori rearrangement (**Figure 1**), ProFAR (the native substrate of HisA) is a much larger molecule, including the presence of a second phosphate group. This group forms hydrogen bonding interactions with the side chains of R83 from loop 3 and S103 from loop 4 of *Se*HisA, and with the corresponding side chains of R85 and T105 in *Mt*PriA (**Figure 2**). Although neither of these residues are on the primary mobile loops of either enzyme (**Figure 1**), nevertheless, the interactions with the second phosphate group of ProFAR are very similar to analogous interactions in other enzymes activated by ligand-gated conformational changes,[34-41] suggesting that loops 3 and 4 in *Se*HisA and *Mt*PriA may similarly act as "gripper loops" allowing these enzymes to attain

relevant catalytically active conformations for the isomerization of the larger substrate. This effect would clearly not be present when the smaller substrate, PRA, is bound to the active site as this substrate lacks a non-reactive phosphodianion group to interact with these loops.

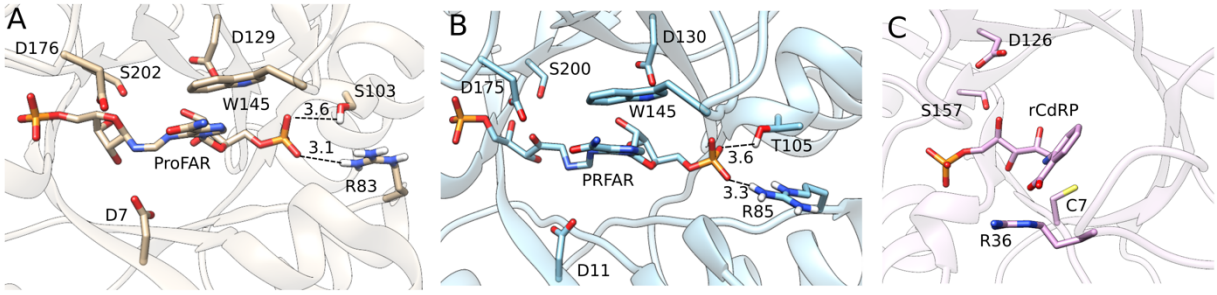

Figure 2. Comparison of the crystallographically determined active sites of (**A**) *Se*HisA in complex with substrate ProFAR (PDB ID: 5A5W[54, 55]), (**B**) *Mt*PriA in complex with product PRFAR (PDB ID: 3ZS4[55]), and (**C**) *Tm*TrpF in complex with product analog rCdRP (PDB ID: 1LBM[49, 55]). Shown here are key catalytic residues for each system, including the catalytic active site aspartic acid side chain (D176 in *Se*HisA, D175 in *Mt*PriA and D126 in *Tm*TrpF), the active site tryptophan that forms stacking interactions with the larger substrate ProFAR in *Se*HisA and *Mt*PriA (W145 in both enzymes), as well as the key gripper residues that interact with the distal phosphate group of the larger substrate/product in *Se*HisA and *Mt*PriA (R83 and S103 in *Se*HisA and R85 and T105 in *Mt*PriA). Key hydrogen bonding interactions are also highlighted, using the distances (Å) found in the corresponding crystal structures. Note that for clarity, N7D and A176D reversions were applied in *Se*HisA in complex with substrate ProFAR (and this reversion was also applied in our simulations, as outlined in the **Methodology** section).

In the present work, therefore, we combine long-timescale conventional molecular dynamics, enhanced sampling and empirical valence bond (EVB) simulations to present a comprehensive computational study of a number of wild-type and variant forms of *Se*HisA, *Mt*PriA and TrpF from *Thermotoga maritima* (*Tm*TrpF). All variants studied in this work, and the corresponding structures used, are summarized in **Table S1**. We chose these systems because, in all three cases, there are high-quality structural data in unliganded and ligand-bound forms. For *Se*HisA, we chose

to study the unliganded and substrate (ProFAR) bound forms,[54] as well as key *Se*HisA variants from ref. [28] that were selected based on their specificity patterns (specialists *vs.* generalists, **Table S2**). *Mt*PriA and *Tm*TrpF were similarly selected on the basis of high-quality structural data of each enzyme in both unliganded and product (PRFAR) or product analogue (rCdRP) bound forms respectively, as summarized in **Table S1**.

Prior simulation studies of TPI by both us and others have indicated that the large ligand-gated conformational change of the gripper loop 6 is correlated with smaller conformational motions in other decorating loops on the active site.[44, 56] We made similar observations when studying loop motions in PTP1B.[32] However, as these simulations indicate, even studying the motion of one large dominating conformational change is computationally non-trivial, and the current systems involve the interdependent conformational rearrangements of multiple loops simultaneously. Our current simulations of HisA, PriA and TrpF (1) provide rich detail of the conformational behavior of the catalytic loops in the different systems, and (2) provide insight into the link between conformational dynamics, catalytic activity and functional evolution in the different enzymes, in particular the role of loops 1 and 5 in regulating PriA and HisA's activity and selectivity, as well as the gripper loops 3 and 4 in driving ligand-gated conformational changes in these enzymes.[28, 45-48] This, in turn, is significant, because in recent years, there has been substantial (and increasing) interest in exploiting techniques such as loop grafting and related approaches in order to engineer flexible loops in enzymes as a means of controlling their catalytic activity.[16, 33] Our data provide clear evidence that this is likely to be a powerful strategy for artificially manipulating the diverse catalytic repertoire[47] of TIM-barrel proteins.

Methodology

Methodological details are presented here in brief. Full details of all simulations and any non-conventional parameters used in our simulations are provided in the **Supporting Information**.

System Preparation for Conventional and Enhanced Sampling Molecular Dynamics Simulations.

Simulations were performed on wild-type *Se*HisA, *Mt*PriA and *Tm*TrpF, as well as relevant enzyme variants, in both their unliganded forms and in complex with various ligands (substrates ProFAR and PRA and, in the case of the enhanced sampling simulations, products PRFAR and CdRP). A total of fourteen crystal structures were used to generate starting points for these simulations, and all simulations performed as well as associated structures used are summarized in **Table S1**. Where present, the D7N, D11N and D176A substitutions were reverted to wild-type using the Dunbrack 2010 Rotamer Library,[57] as implemented in UCSF Chimera, v. 1.14.[58] Missing regions in the catalytic loops were reconstructed using Modeller v. 9.23.[59] The catalytic aspartic acid side chain in the active site of each enzyme (D176 in HisA, D175 in PriA and D126 in TrpF) was kept protonated in line with the mechanism shown in **Figure 1**. All other residues (except H50 in PriA, which was doubly protonated) were kept in their default protonation states at physiological pH determined by use of PROPKA 3.1,[60] and visual inspection. The substrates ProFAR and PRA were manually placed into the relevant active sites in the same conformation as found in the structure of the HisA wild-type enzyme in complex with ProFAR and in the case of PRA (**Figure 1D**), was placed by manual overlay of the reactive part of PRA with the reactive part of ProFAR, and with the carboxylate group of PRA keeping key interactions with active site residues. Partial charges for ligands ProFAR, PRA, PRFAR and CdRP, were calculated using the standard restrained electrostatic potential (RESP) protocol using Antechamber v. 17.3,[61] and based on the

vacuum electrostatic potential calculated at the HF/6-31G(d) level of theory, using Gaussian 09 Rev. E.01.[62] All other simulation parameters were described using the general Amber force field 2 (GAFF2)[63] (see **Tables S3** to **S6**). Finally, to keep the substrate stably bound in the enzyme active sites, weak distance restraints were applied to protein-substrate distances, as described in **Table S7**.

Conventional Molecular Dynamics Simulations

Conventional MD simulations were performed using the CUDA version of the PMEMD module of the AMBER 16 simulation package.[64] The protein, ligands and solvent were described using the ff14SB force field,[65] the General AMBER Force Field 2 (GAFF2),[63] and the TIP3P water model,[66] respectively. Following initial minimization and equilibration, each system (summarized in **Table S1**) was subjected to 10 x 500 ns of molecular dynamics simulations controlled by the Langevin thermostat with a collision frequency of 2 ps^{-1},[67] and the Berendsen barostat with a 1 ps coupling constant.[68] This led to a cumulative 5 μs of production simulations per system, and a cumulative total of 70 μs of conventional MD simulations over all systems studied (**Table S1**).

Enhanced Sampling Molecular Dynamics Simulations

Steered molecular dynamics simulation (sMD) were performed using GROMACS 2018.4 in order to pull products PRFAR and CdRP out of the active site of the *Se*HisA(dup13-15/D10G/G102A/Q24L) variant, as described in the **Supporting Information**. The system preparation was performed as for the conventional MD simulations, and using the same force fields and water models as the conventional MD simulations. Following initial minimization and equilibration, 10 x 50 ns production MD simulations were performed on each system. The first 5 ns of production MD were unrestrained, after which an external force with a force constant of 10

kcal mol^{-1} $Å^{-2}$ was applied to pull the product out of the active site. This external force was then released for the last 5 ns of the MD simulation run.

Empirical Valence Bond Simulations

Following our prior success in using the empirical valence bond (EVB) approach[69] to study a wide range of analogous ring opening reactions, such as lactone[70, 71] and epoxide[72, 73] hydrolysis, we extended this approach to study the enzyme-catalyzed opening of the ribose ring of substrates ProFAR and PRA (**Figure 1**), as catalyzed by wild-type and variant forms of HisA, PriA, and TrpF. Our focus for our EVB simulations was specifically on the ribose ring-opening reaction (the first step of the mechanism shown in **Figure 1D**), as motivated in the **Results and Discussion**, and described using the valence bond states shown in **Figure 3**. Simulations were performed on wild-type *Se*HisA, *Mt*PriA and *Tm*TrpF as well as selected variants, as described in the **Supporting Information**. All simulations were performed using the *Q6* simulation package,[74, 75] using the OPLA-AA force field.[76] All EVB parameters necessary to reproduce our work, as well as a detailed description of the computational methodology and subsequent simulation analysis can be found in the **Supporting Information**, with the full parameters used in our simulations updated to Zenodo (DOI: 10.5281/zenodo.5893598). Each system was simulated using 30 individual replicas, with each replica first equilibrated for 20 ns and the endpoint of that equilibration being used as the starting point for propagating an EVB trajectory. Each EVB free energy perturbation/umbrella sampling (EVB-FEP/US)[69] simulation was simulated using 51 individual mapping windows of 200 ps of simulation time each, leading to a total of 10.2 ns of simulation time per individual EVB trajectory. Leading to a cumulative total of 600 ns equilibration and 306 ns EVB simulation time per individual system, and a cumulative 12 μs of equilibration and 6.12 μs of EVB simulation time over all 20 systems studied in this work.

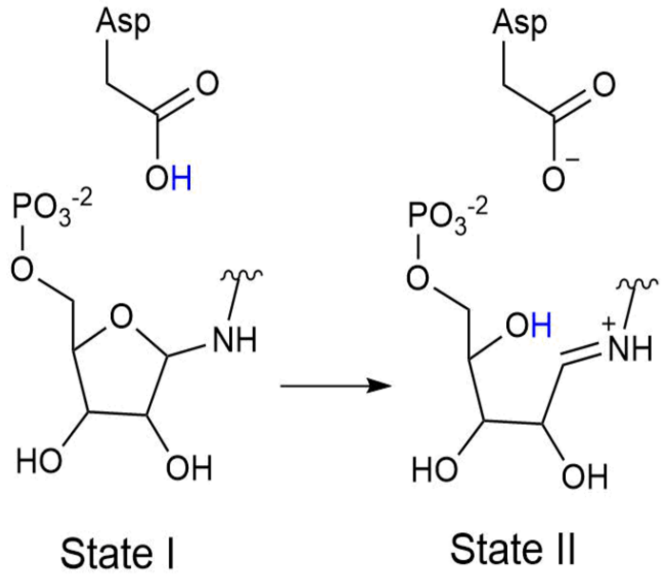

Figure 3. Valence bond states used to describe the ribose ring opening catalyzed by the different enzyme variants studied in this work. Only the ribose moiety of the substrate was placed in the EVB region, with the remainder of the substrate treated purely classically. Note that both EVB and non-EVB atoms are described using the same classical force field, and the main differences between the treatment of the EVB and surrounding regions are: (1) the use of Morse potentials and an alternative van der Waals potential in order to describe the atom pairs that are directly involved in the reaction, as well as (2) different cutoffs for evaluating nonbonded interactions in the EVB and surrounding regions (parameters provided in the **Supporting Information** and on Zenodo, DOI: 10.5281/zenodo.5893598).[38]

Analysis of Conventional and Enhanced Sampling Molecular Dynamics Simulations

Unless stated otherwise, all analysis of all conventional and enhanced sampling molecular dynamics simulations was performed using CPPTRAJ.[77] Hydrogen bonds were defined as formed if the donor−acceptor distance was ≤3.0 Å and the donor-hydrogen-acceptor angle was within 180 ± 45°. Principal component analysis (PCA) was performed by first RMS fitting to a whole protein C_α carbon atoms and then performing PCA on the C_α carbon atoms of loops 1, 5 and 6, as well as loop 1 for HisA loop 1 elongated systems analysis. Other analyses were performed as described in the **Supporting Information**.

Results and Discussion

Active Site Plasticity and Substrate Binding in the Different Enzymes

The TIM-barrel structures of *Se*HisA and *Mt*PriA are similar, with an RMSD of 1.09 Å between them (comparing the C_α atoms in PDB IDs 3ZS4[55] and 5A5W[54, 55]). *Tm*TrpF, in contrast, is a smaller enzyme with 40 fewer residues in the sequence than *Mt*PriA and *Se*HisA. We performed 10 x 500 ns conventional molecular dynamics simulations of unliganded HisA, TrpF and PriA, and calculated the average and standard deviations of the active site volumes of each enzyme using the MDpocket[78] tool, which is provided as part of the *fpocket*[79] suite of pocket detection programs, as described in the **Supporting Information**. The resulting calculated volumes are shown in **Table S8**. We obtained average volumes of 745.4 ± 129.6 Å^3, 1033.8 ± 217.0 Å^3 and 1173.5 ± 158.0 Å^3, for the active sites of TrpF, PriA and HisA, respectively during our simulations. From this, it can be seen that the TrpF active site is more compact than that of PriA and HisA, which have successively larger active site volumes, with more "flexible" pockets than TrpF (using the standard deviation on the volume as a proxy for this flexibility). For comparison, the substrates ProFAR and PRA have volumes of 829 and 559 Å^3, respectively, calculated using Alexander Balaeff's Mol_Volume program Version 1.0, with default radii of 1.7Å and a probe sphere of 0.5Å. This confirms the structural data indicating that the active site pocket of TrpF is too compact to accommodate the much larger substrate, ProFAR, leading to the selectivity of this enzyme towards PRA.[80] Furthermore, the PriA active site is the most flexible of the three, in line with structural data[25] that indicates that PriA is capable of significantly rearranging its active site (in particular loop 5 conformation) when accommodating the different substrates ProFAR and PriA (**Figure S1**).

Structures and mutagenesis experiments have identified two key active site side chains in HisA and PriA, which are important for binding of the substrate ProFAR.[25, 28, 54, 81] These are W145, which forms a stabilizing stacking interaction with the substrate, and R83 (R85), which interacts with the second phosphodianion group of the substrate (**Figure 2**). To further explore the conformational diversity of these key tryptophan and arginine residues in HisA and PriA, respectively, we examined the joint dihedral angle distribution of the side chains of these residues in simulations of unliganded HisA and PriA, as well as HisA and PriA in complex with both substrates ProFAR and PRA (W145 and W145, and R83 and R85 in HisA and PriA respectively). The corresponding dihedral data are shown in **Figures 4 and S2**, and the relevant side chain positioning is illustrated in **Figure 2**.

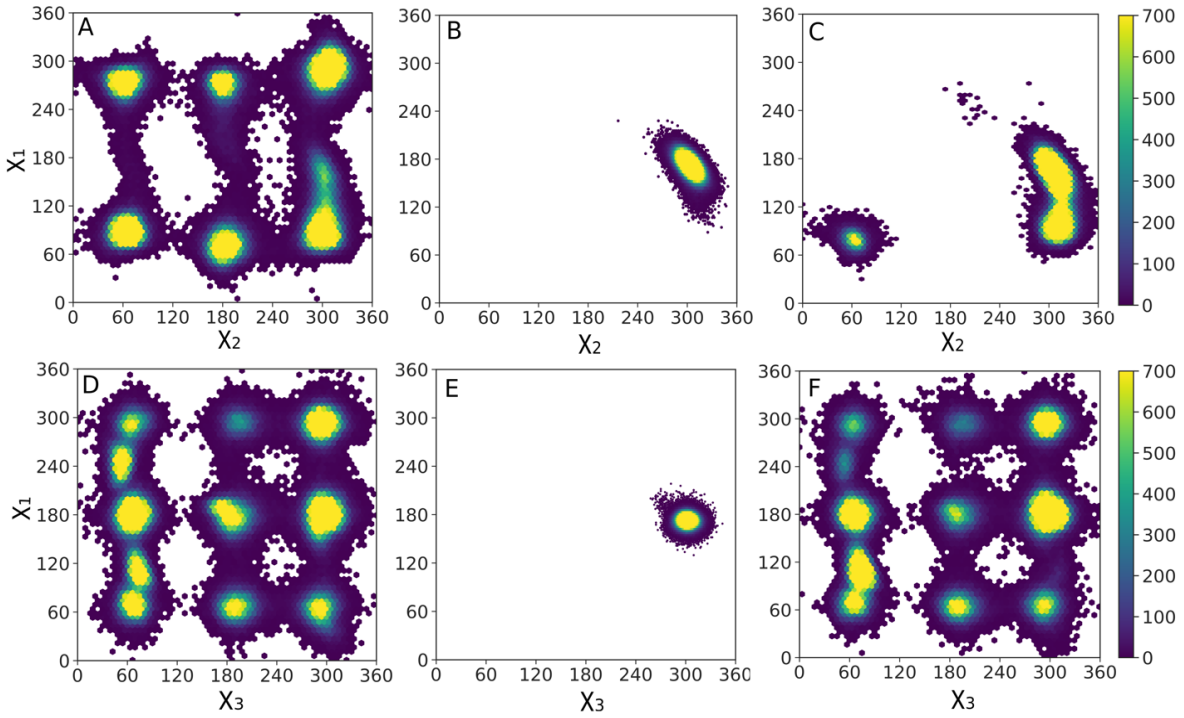

Figure 4. Joint distribution of the side chain dihedral angles of the side chains of (**A**, **B**, **C**) W145 and (**D**, **E**, **F**) R83 in (**A**, **D**) unliganded HisA, (**B**, **E**) HisA in complex with substrate ProFAR and (**C**, **F**) HisA in complex with substrate PRA. Data was extracted every 5 ps of 10 x 500 ns production simulations of each system, performed as described in the **Supporting Information**.

From these data it can be seen that both the tryptophan and arginine side chains are highly conformationally flexible in the unliganded enzymes. However, while the binding of the substrate ProFAR to HisA restricts the conformational space of the arginine side chain on the "gripper" loop 3 to a catalytically competent position that helps stabilize the bound substrate, in PriA, the R143 side chain is only 4.8Å from the "gripper" residue R85 (distance between the two side chain carbon atoms, based on PDB ID: 3ZS4[55]). This in turn creates electrostatic repulsion between the two arginine side chains, thus destabilizing loop 5 as well as the interaction between the substrate ProFAR and the R85 side chain (**Figure S3**). In contrast, when PRA, that lacks the second phosphodianion group, is bound to the PriA active site, the R83/R85 side chains increase their conformational flexibility again, sampling more or less the same conformational space as in the case of the unliganded enzyme (**Figure S2**). Therefore, the interaction of these residues with the larger substrate ProFAR is likely playing an important role in the ability of these enzymes to bind and isomerize this compound.

In the case of W145, this side chain slightly increases its conformation flexibility when the smaller substrate PRA is bound to HisA (note that PRA was placed manually in the active site by overlay with substrate ProFAR, as described in the **Supporting Information**). However, in the case of PriA, both W145 the and R143 side chains are conformationally restricted to a catalytically competent position due to a rearrangement of loop 5 upon PRA binding, that swaps the position between these two residues compared to when ProFAR is bound to PriA, preventing the electrostatic repulsion between the R85 and R143 side chains that is observed when ProFAR (PRFAR) is bound to the active site (**Figure S1**).[25] As noted previously, the R83/R85 side chain is one of a number of key residues on the "gripper" loop that interact with the distal phosphodianion group of the larger substrate ProFAR, contributing to the stabilization of the

substrate in HisA active site. Hence, in PriA, the loop 5 rearrangement required for ProFAR substrate binding[25] prevents R85 from gripping the second phosphodianion group and thus showing clear preference for the isomerization of PRA substrate.

We note the similarity of these gripper interactions to corresponding interactions in enzymes such as TIM, OMPDC and GPDH,[34-38] where interactions with the non-reactive phosphodianion group of the substrate drives a ligand-gated conformational change. This in turn stabilizes otherwise energetically unfavorable but catalytically important closed conformations of key catalytic loops over the respective active sites of these enzymes. In the case of the current systems, the interaction between the HisA gripper residues and the remote phosphodianion group of ProFAR appears to be similarly important for maintaining the closed conformation of loop 5, and when this interaction is lost, as in PriA, we see corresponding opening of loop 5 (**Figure S3)**. This supports the likelihood that HisA and PriA are also activated by ligand-gated conformational changes, albeit with more complex loop dynamics (due to the involvement of not one but three highly mobile and long catalytic loops) than in other previously characterized systems.

Conformational Dynamics of Key Catalytic Loops of HisA, PriA, and TrpF

To further explore the impact of the binding of the two substrates on loop dynamics, we extended our simulations to also included simulations of TrpF in both its unliganded form, and in complex with PRA (**Table S1**). We then performed Principal Component Analysis (PCA) to characterize the motion of the key catalytic loops during our simulations in each of the individual systems, similarly to prior analysis we have performed on triosephosphate isomerase,[44] except in our prior work substantial conformational changes take place in only one and not two (*Tm*TrpF) or three (*Se*HisA; *Mt*PriA) distinct loops. The PCA analysis was performed on the mass-weighted

Cartesian coordinates of each enzyme compared to the coordinates of the corresponding closed state, allowing us to explore the variation of the conformations of these loops in coordinate space.

Figure 5 shows an overview of the structural changes in the catalytic loops along the first two principal components, PC1 and PC2, as well as the minimum and maximum deviations of each of the key catalytic loops from the corresponding closed reference state along each principal component for each enzyme. Significant conformational motion is observed along both PCs; however, PC1 primarily described loop transitions between closed and open conformational states, whereas PC2 described conformational variation in the loops during these loop transitions, including transitions between different open conformations of these loops, with contributions to the overall variance of 41.2%, 70.9% and 51.0% for PC1, and 16.4%, 6.0% and 13.9% for PC2, for each of HisA, PriA and TrpF, respectively.

We subsequently projected the free energies for each enzyme along the most dominant motions, PC1 and PC2, from simulations of each of HisA, PriA and TrpF in their unliganded forms as well as in complex with substrates ProFAR and PRA, respectively (here, the smaller substrate PRA was artificially placed in the HisA active site by manual overlay with the reactive part of ProFAR, as outlined in the **Supporting Information**). This allowed us to compare the free energy surfaces defined by these two principal components both between the different enzymes, and the effect of ligand binding on these surfaces. The resulting data are shown in **Figure 6**. Note that, as shown in **Figure 5**, these projected free energy surfaces show the combined motion of all key catalytic loops along each principal component.

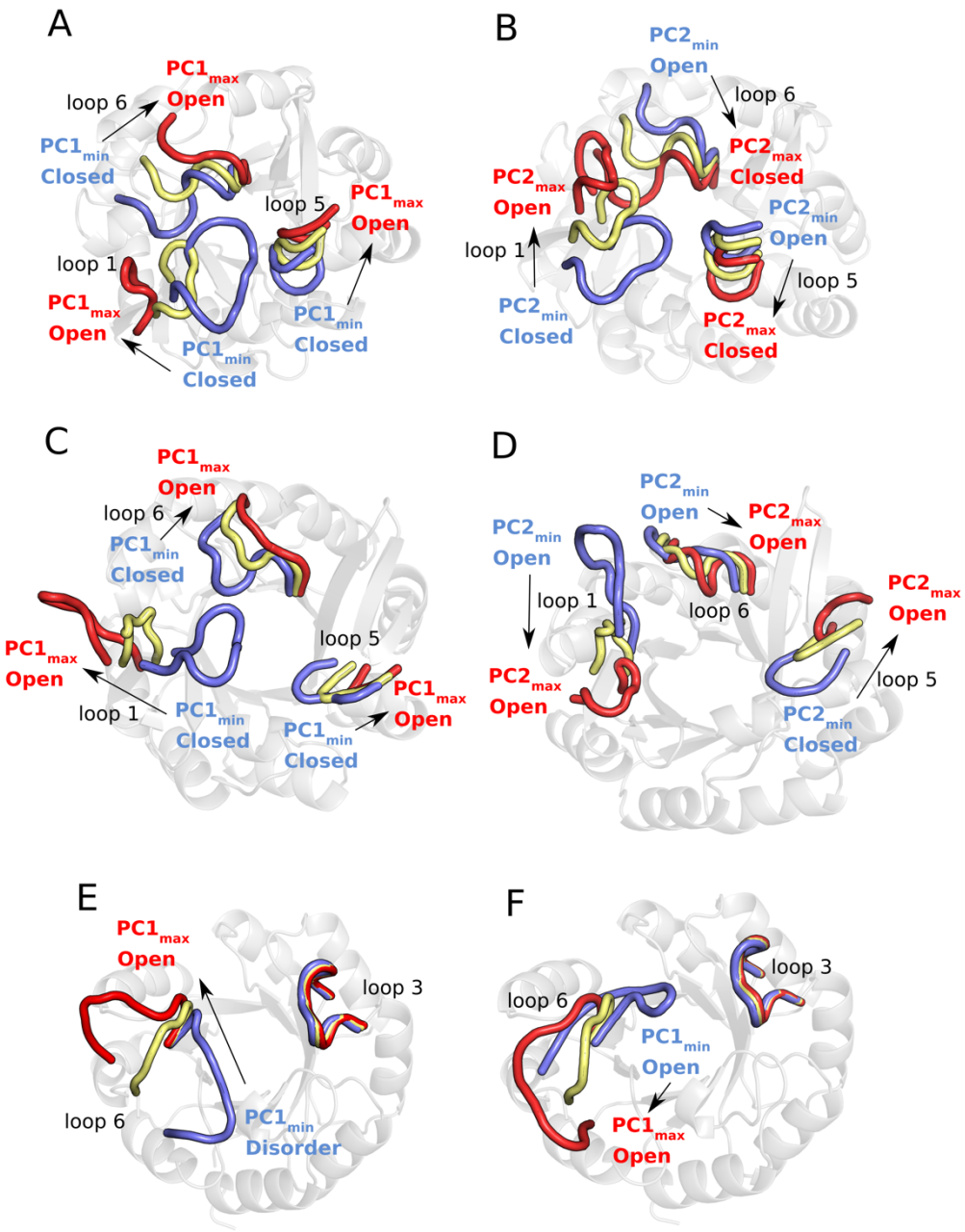

Figure 5. Overview of the variations in the conformational states of key catalytic loops of (**A**, **B**) HisA, (**C**, **D**) PriA and (**E**, **F**) TrpF in coordinate space, along the first two principal components (**A**, **C**, **E**) PC1 and (**B**, **D**, **F**) PC2 obtained from PCA analysis performed on our conventional MD simulations of each system. PCA analysis was performed relative to the corresponding loop closed states for each enzyme, and the minimum and maximum deviations of each loop from this closed state along each PC are highlighted in blue and red respectively, with an intermediate state observed in our analysis highlighted in yellow.

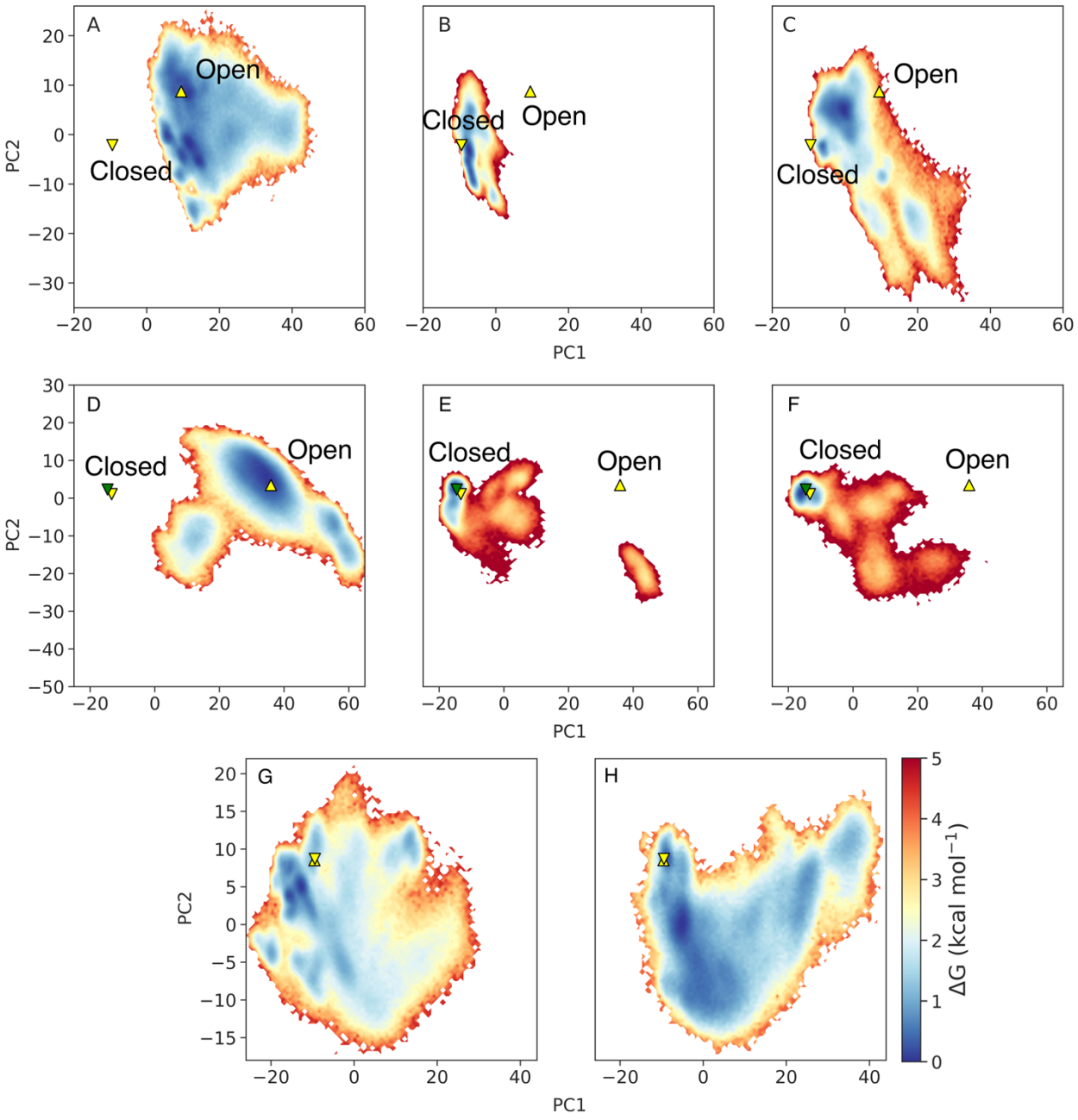

Figure 6. Projected free energy surfaces (kcal mol^{-1}) along the first two principal components (PC1 and PC2) obtained from applying Cartesian principal component analysis (PCA) to our conventional MD simulations of: (**A, B, C**) HisA in its (**A**) unliganded form, and in complex with (**B**) ProFAR and (**C**) PRA, (**D, E, F**) PriA in its (**D**) unliganded form and in complex with (**E**) ProFAR and (**F**) PRA, and (**G, H**) TrpF in its (**G**) unliganded form and in complex with (**H**) PRA. The crystallographic loop open and loop closed states of the enzyme are indicated on this surface using the symbols ▲ and ▼. In the case of HisA, these states are defined based on the loop conformations found in PDB IDs:

5AHE[54] and 5A5W,[54] for the unliganded open, ProFAR-bound closed and PRA-bound closed conformations, where the latter two systems have the same crystallographic closed conformation. In the case of PriA, these states are defined based on PDB IDs: 2Y89,[25] 3ZS4[55] and 2Y85,[25, 55] for the unliganded open, ProFAR-bound closed (yellow triangle) and PRA-bound closed (blue triangle) conformations. For TrpF, the corresponding conformations of the catalytic loops are defined based on PDB IDs: 1NSJ[55, 82] and 1LBM[49, 55]. The overlap of these triangles indicate that the loops are found in similar positions in the crystal structure, irrespective of whether the starting structure is liganded or unliganded. Note that this figure considers the motion of all relevant catalytic loops along these two PC, as illustrated in **Figure 5**.

In the unliganded forms of all three enzymes, the catalytic loops can explore a range of "wide-open" conformations and are overall highly conformationally diverse (**Figure 5**). This observation is consistent with our prior simulation studies on both triosephosphate isomerase,[44] and the protein tyrosine phosphatases PTP1B and YopH,[32] as well as chimeric forms of these enzymes.[83] However, and consistently with structural data, the binding of ProFAR to HisA fully restricts the conformational sampling of all three loops (**Figure 6B**). In the case of PriA, the binding of both ProFAR and PRA also stabilizes the closed conformation of the three active site loops, but still allows for some conformational flexibility in these loops (**Figure 6**, based on both the topologies of the projected free energy surfaces, and the corresponding energies). In sharp contrast, in the liganded form of TrpF, our MD simulations show that PRA is not stable in the active site due to the flexibility of loop 6, which explores transitions towards open conformations, similarly to the unliganded system (**Figures 5E, F** and **6G, H**). In the crystal structure of the unliganded enzyme (PDB ID: 1NSJ[55, 82]), this loop is present in a closed position but with missing density, whereas in our simulations of both the liganded and unliganded form of the enzyme, loop 6 samples open conformations, suggesting that it is not the correct closed state to stabilize the substrate PRA in

the active site. We note that the structure used for these simulations (PDB ID 1LBM[49, 55]) was solved in complex with the product analog rCdRP. Our simulations suggest that the loop conformations observed in this structure are a conformational state on the trajectory to product release, rather than an ideal conformational state for stabilizing the Michaelis complex.

In addition, in HisA and PriA, we observe the formation of a stacking interaction between the substrate ProFAR and the side chain of W145 in our conventional MD simulations (**Figure S4**), with an average distance of 3.9 ± 0.3 Å and an angle γ = 11.7 ± 5.5° between the center of mass of the imidazole ring of ProFAR and the indole ring of the W145 side chain during our simulations. Our results confirm the role for W145 that was proposed previously, based on experimentally determined structures.[54] This is furthermore consistent with prior structural analysis that indicates that HisA activity is abolished in *Se*HisA(dup13-15), because the extended conformation of loop 1 blocks this side chain from interacting with ProFAR.[28]

In contrast, in the case of PriA, and again in agreement with prior structural analysis,[25] we observe two possible conformations of loop 5, depending on which substrate is bound to the active site. That is, when ProFAR is bound to the active site, we sample a conformation similar to that observed in wild-type HisA (**Figure S1A**) with a similar stacking interaction between ProFAR and W145 (**Figure S4**), however, the loop 5 rearrangement required to optimize the stacking position of W145 with ProFAR creates transient electrostatic repulsion between loop 5 and the rest of the enzyme, making this loop more conformationally dynamic, which we observe in our analysis in the form of an increased standard deviation in the distance and angle of the corresponding stacking interaction (d = 4.1 ± 0.6 Å, γ = 19.2 ± 13.9°) (**Figures S3** and **S4**). The greater plasticity of this interaction, in turn, decreases substrate stability in the active site (the ProFAR RMSF increases

from 18.4 Å in our simulations of wild-type HisA to 22.2 Å in our simulations of wild-type PriA), and thus the corresponding ProFAR isomerization activity of PriA.

For comparison, in our simulations of PriA in complex with substrate PRA, we sample an active site conformation in which the R143 side chain (loop 5 residue) forms salt bridges with the side chains of with D130 and D175, and the arginine acts as a "shield" dampening the electrostatic repulsion between the D130 side chain and the anthranilate carboxylate group of PRA. This interaction also stabilizes the catalytic aspartic (D175), placing it in an optimal position for catalysis (**Figure S5** and **Table S9**), as shown in previous studies[25]. This PriA conformation is similar to the "TrpF-active" conformation observed in the *Se*HisA(dup13-15/D10G/Q24L/G102A) crystal structure[28] (PDB ID: 5AB3[28, 55], **Figure S6,** with manual placement of PRA in the active site), where the arginine is close to residue D129. While we observe this conformation in our PriA simulations, we do not observe the formation of a corresponding interaction in our simulations of wild-type HisA in complex with PRA, the negative charge repulsion between PRA and the D129 side chain destabilizes the position of the substrate in the active site (**Figure S7**), as well as the stability of the loop 6 carrying the key catalytic aspartic acid side chain (**Figure S7**). We do, however, observe a similar interaction with the R169 side chain in the *Se*HisA(L169R) variant, with interactions with D129 and, in this case, a salt bridge interaction with the anthranilate carboxylate group of PRA (**Figure S7** and **Table S9**), consistent with experimental work that demonstrated that the introduction of the L169R substitution in HisA induces TrpF activity (**Table S2**). [46]

Finally, in the case of TrpF in complex with PRA, we observe a salt-bridge interaction between E184 and the side chain of R36 on loop 3 (**Figure S6** and **Table S9**), and we can see that as for HisA and PriA, that R36 is again acting as a "shield" avoiding possible negative repulsion

interactions between the substrate and the negatively charged side chain. However, we do not observe clear interactions between the R36 side chain and the substrate PRA (**Figure S6** and **Table S9**, with the fraction of simulation time in which this interaction is observed being <0.1).

Overall, we observed that this arginine plays an essential role in the introduction of TrpF activity, by shielding electrostatic repulsion between negatively charged side chains and the anthranilate carboxylate group of PRA. This is in agreement with experiments where introducing an arginine or removing the negative residue introduce TrpF activity in HisA systems.[46]

Conformational Dynamics of Loop 1 in HisA and PriA and Its Impact on Selectivity

While loops 5 and 6 of HisA and PriA have been clearly identified as being important for binding and catalysis (loop 6 carries the catalytic aspartic acid side chain, **Figure 1**),[25, 28, 54] the precise catalytic role of loop 1 remains unclear, although extending the conformation of loop 1 through duplication of residues 13-15 (HisA(dup13-15)) plays an important role in the acquisition of bifunctionality[28, 84] in a real-time evolution experiment on HisA,[46] and substitutions at position 15 on this loop appear to be important for facilitating the TrpF activity of this enzyme.[28, 54] Therefore, we also performed simulations of the *Se*HisA(dup13-15/D10G) variant, as described in the **Methodology** section. *Se*HisA(dup13-15/D10G) is a bifunctional enzyme that can catalyze the isomerization of both ProFAR and PRA with modest catalytic efficiencies,[28] and the corresponding crystal structure (PDB ID: 5AC7[28, 55]) shows the enzyme in a 'PRA-active' conformation with loops 1 and 6 in a closed state, and loop 5 in an open state.

When initiating simulations of the unliganded *Se*HisA(dup13-15/D10G) variant starting from these loop 1 closed conformations, we did not observe any opening of loop 1. This is in contrast to our simulations of the wild-type enzyme, where we sampled open conformations of this loop when we started from the unliganded closed conformation observed in PDB ID: 5A5W,[54]

removing the ProFAR substrate. This provides evidence, in addition to the discussion in ref. [28], that the elongation of loop 1 heavily stabilizes this PRA-active conformation. This is further supported by examining the root mean square fluctuations of all C_α atoms in our simulations of these two enzymes (**Figure 7**), where we observe that loop 1 is more flexible than either of loops 5 or 6 in simulations of the wild-type enzyme, but has reduced flexibility in simulations of the *Se*HisA(dup13-15/D10G) variant.

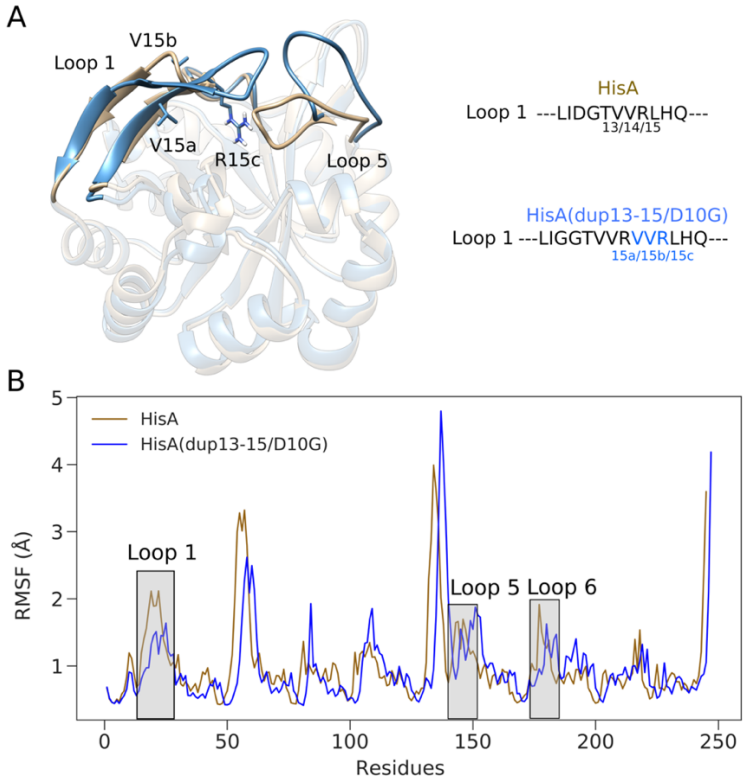

Figure 7. Root mean square fluctuation (RMSF, Å) of all backbone C_α atoms in wild-type HisA, as well as in the *Se*HisA(dup13-15/D10G) variant (based on crystallographic coordinates from PDB IDs: 5A5W[54, 55] and 5AC7,[28, 55] respectively). (**A**) The position of the key catalytic loops 1 and 5 is highlighted in this figure (loop 6 is not clearly visible in this figure as it is behind loops 1 and 5, but has the same conformation in both systems). A comparison between the loop 1 sequences in wild-type HisA and in the *Se*HisA(dup13-15/D10G) variant is provided as an inset. (**B**) Root mean square fluctuation (RMSF, Å) of all backbone C_α atoms in wild-type HisA and the *Se*HisA(dup13-15/D10G) variant.

In addition, as it is possible that product release would trigger a change from a closed to a wide-open conformation of loop 1, we also performed steered molecular dynamics (sMD) simulations of the only crystal structure of a loop-elongated HisA variant with product bound to the active site (*Se*HisA(dup13-15/D10G/G102A/Q24L), PDB ID: 5AB3[28]). Here, we performed pulling simulations of products PRFAR and CdRP out of the enzyme from the bottom of the TIM barrel, again as described in the **Methodology** section. In doing so, we observe that the larger molecule, PRFAR, always induces a conformational change in loops 1 and 6 from a closed to an open and/or wide-open conformation upon pulling it out of the active site (**Figure S8A** and **C**, backbone RMSD of loop residues of up to ~4.0 Å in loop 1 and ~3.5 Å in loop 6, compared to the starting closed conformation). In contrast, the small substrate is able to leave the active site without inducing large conformational changes in loop 1 (loop 1 opens in only one out of ten replicates), but pulling it out from the active site still does induce a conformational change in catalytic loop 6 (**Figure S8B** and **D**, backbone RMSD of loop residues of up to ~2.0 Å in loop 1 and ~3.5 Å in loop 6, compared to the starting closed conformation). This suggests substantial loop rearrangement is required for efficient product release, making it possible that a slow (rate-limiting) product release step is the reason for the low turnover numbers observed for the catalysis what is otherwise a very intrinsically facile reaction[85] (**Table S2**).

Interestingly, while the elongation of loop 1 through the duplication of residues 13-15 (VVR) appears to be essential for the change of specificity towards the isomerization of PRA (facilitated by the presence of a new stabilizing arginine side chain close to the active sites, **Figure S1**,[28] simply the loop duplication by itself is not enough to induce bifunctionality. That is, while the duplication elongates loop 1, it also rigidifies it, such that *Se*HisA(dup13-15) does show some ability to isomerize PRA ($k_{cat} > 0.15$ s^{-1}), but at the expense of losing all ability to isomerize the

larger substrate ProFAR (no detectable activity).[28] Therefore, the duplication by itself simply leads to a switch in activity from a modestly efficient isomerase towards ProFAR (k_{cat} 7.8 s^{-1}) towards a less efficient isomerase with activity towards PRA. Critical to the bifunctionality is the inclusion of an additional substitution, present in both variants studied above, namely D10G. This substitution increases the flexibility of the elongated loop 1 (**Figure 8**), allowing for the loop to take on wide-open configurations which in turn facilitate the entry and binding of ProFAR to the active site (**Figure 8B,** wide-open conformation).

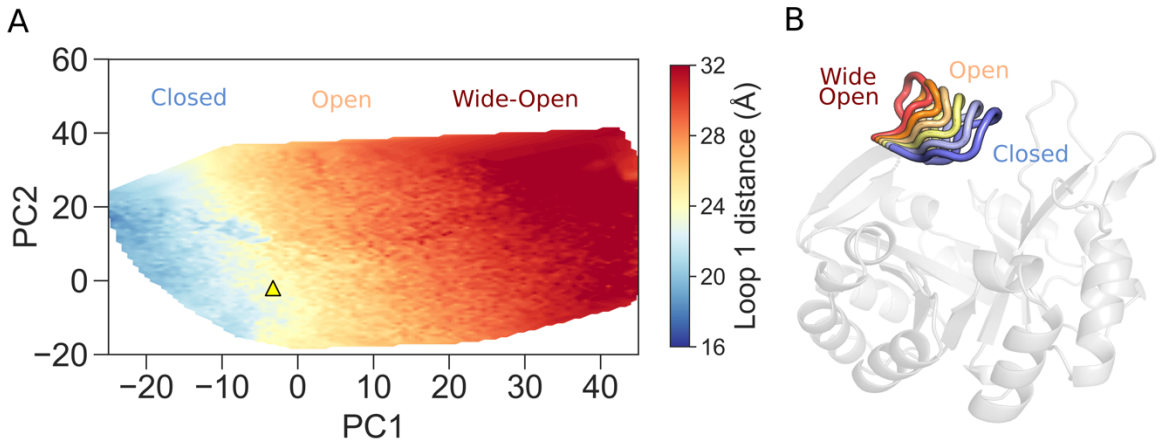

Figure 8. (**A**) Global projected loop 1 distance surfaces (Å) along the first two principal components (PC1 and PC2) obtained from applying Cartesian PCA to combined analysis of our conventional MD simulations of unliganded *Se*HisA(dup13-15) (PDB ID: 5G2I[28]), *Se*HisA (dup13-15/D10G) (PDB ID: 5AC7,[28, 55]), *Se*HisA(dup13-15/D10G/G102A) (PDB ID: 5AC8[28]), and *Se*HisA(dup13-15/D10G/G102A/Q24L/V15[b]M) (PDB ID: 5G1Y[28]). The starting *Se*HisA(dup13-15) loop 1 open state is indicated on this surface by a yellow triangle, ▲, and the same starting loop 1 conformation is used for simulations of all four systems. Note that this figure considers the motion of loop 1 along the PCs, as shown for PC1, projections along which include transitions from closed to wide-open states of loop 1 (**B**). The PCA analysis was performed on the mass-weighted Cartesian coordinates of the HisA variants compared to the coordinates of the corresponding open state, allowing us to explore the variation of the conformations of this loop in coordinate space. The loop 1 distance is defined as the center of mass of residues 15-25 of loop 1 and residue 129 from the barrel scaffold. Here, we considered distances <22.0 Å closed state (blue), between 23.0 - 28.0

Å open states (yellow), and >29.0 Å wide-open conformations (red), based on a combination of visual inspection and comparison of the closed state crystal structures (PDB IDs: 5AB3[28, 55] and 5AC7[28, 55]). A 1 Å difference between states was applied to avoid fuzzy states between transition from closed to open state and from open to wide-open state.

Our conventional MD simulations (10 x 500 ns per system) are overall short, taking in particular into account the slow turnover numbers of these enzymes (that suggest loop motions on the ms to s timescale).[28] However, our observations that the D10G substitution leads to increased flexibility of loop 1 are in good agreement with prior NMR relaxation dispersion experiments.[28] These detected μs to ms motions at 14 backbone ^{15}N positions in the *Se*HisA(dup13-15/D10G) variant, compared to only three positions for the *Se*HisA(dup13-15) variant, and with two resonances that are unique to *Se*HisA(dup13-15/D10G). As a result, adding this substitution is sufficient to convert *Se*HisA(dup13-15/D10G) back to a bifunctional enzyme, through exploitation of conformational dynamics, with k_{cat} of 0.09 s^{-1} for the isomerization of PRA, and 0.05 s^{-1} for the isomerization of ProFAR (**Table S2**).[28]

To explore this further, we therefore examined simulations of four loop elongated variants, specifically: dup13-15 (PDB ID: 5G2I[28]), dup13-15/D10G (PDB ID: 5AC7,[28, 55]), dup13-15/D10G/G102A (PDB ID: 5AC8[28]), and dup13-15/D10G/G102A/Q24L/V15[b]M (PDB ID: 5G1Y[28]). The first and last of these variants are only active towards the isomerization of PRA, whereas the middle two variants are bifunctional towards both PRA and ProFAR (**Table S2**). In all cases, when performing conventional MD simulations starting from the closed conformation of loop 1, this loop is very stable, and remains closed over our simulation timescales (**Table S1**). We therefore also initiated trajectories starting from the loop 1 open conformations of these variants to see if the loop prefers to remain open, transition to wide-open conformations, or to transition back to a closed conformation. PCA analysis was then performed on the mass-weighted Cartesian

coordinates of each enzyme compared to the coordinates of the corresponding loop 1. Using the *Se*HisA(dup13-15) variants as a reference, we defined the conformations of the loop as closed, open or wide-open, based on the distance between the center of mass of residues 15-25 of loop 1, and residue L126, on the far side of the protein scaffold from this loop, with distances <22 Å corresponding to closed conformations, distances between 23 and 28 Å corresponding to open conformations, and distances higher than 29 Å corresponding to wide-open conformations (see **Figures 8** and **S9** for an illustration of the different conformations).

Table 1. Relative populations of the closed, open and wide-open conformations of loop 1 sampled in 10 x 500 ns MD simulations of wild-type *Se*HisA and variants.[a]

System	Function	Closed	Open	Wide-Open
Wild-Type	HisA	13	60	27
HisA(D10G)	HisA	33	49	18
HisA(dup13-15)	TrpF	35	59	6
HisA(dup13-15/D10G)	HisA+TrpF	29	53	18
HisA(dup13-15/D10G/G102A)	HisA+TrpF	13	65	22
HisA(dup13-15/D10G/G102A/Q24L/V15bM)	TrpF	64	31	5

[a] The relative populations (%) of the different loop 1 conformational states sampled during our simulations with HisA dup13-15 variants were determined by the number of frames with distances <22.0 Å for closed state, between 23.0 - 28.0 Å for open states, and >29.0 Å wide-open conformations, shown schematically in **Figure 8**, divided by the sum of all states considered for this analysis. A 1 Å difference between states was applied to avoid fuzzy states between transition from closed to open state and from open to wide-open state for the *Se*HisA(dup13-15) variants.

From analysis of our MD simulations (**Table 1**), we clearly see how all variants carrying the D10G substitution are able to populate all three of closed, open and wide-open conformations. However, the relative populations of these states depends strongly on enzyme variant: already, the D10G substitution by itself appears to be sufficient to cause a conformational shift towards a closed conformation, and the *Se*HisA(dup13-15/D10G/Q102A/Q24L/V15[b]M) variant, which shows the

highest TrpF activity of all variants studied in ref. [28] (**Table S2**), also shows the most significant population shift towards sampling a closed conformation of loop 1. This is consistent with structural analysis,[28] which indicated that the Q24L substitution is important because it introduces a new stabilizing interaction with V15b, as well as an even better interaction with V15[b]M, such that the Q24L interaction is just as important as the VVR duplication for the adaptive benefit of the V15[b]M substitution to be realized. Finally, the wide-open conformation is also only rarely sampled in the *Se*HisA(dup13-15) variant, which does not carry the D10G substitution. We hypothesize that in the case of these variants, this is due to the presence of a Gly-Gly dyad in the hinge of loop 1, which provides the loop with enough flexibility to explore these wide-open conformations. Clearly, dup(13-15), as well as the inclusion of additional substitutions, is significantly impacting the conformational space sampled by this loop, shifting towards a loop-closed conformation of loop 1 that is favorable for TrpF activity.

Empirical Valence Bond Simulations of the Enzyme-Catalyzed Ribose Ring Opening Step in the Isomerization of Substrates ProFAR and PRA

To further probe the role of loop 1 in HisA and PriA, we have complemented our conventional and enhanced sampling molecular dynamics simulations with empirical valence bond (EVB)[69] simulations of the initial ribose ring opening step in the Amadori reaction of substrates ProFAR and PRA (**Figure 1**), as catalyzed by wild-type and variant forms of the two enzymes (see the **Methodology** section for details of the simulation and parameterization procedures).

We note that the calculated activation free energies for this step are not trivial to directly compare with the experimental turnover numbers: no kinetic data exists on the rates of the individual chemical steps. Furthermore, the observed k_{cat} values for these systems are extremely low – on the order of 1 s^{-1} (or lower) for both the HisA and TrpF reactions catalyzed by HisA and

its variants.[28] However, *E. coli* TrpF has a k_{cat} value of 30-40 s^{-1}, with the rate-limiting step being a spontaneous keto-enol tautomerization step that occurs off the enzyme, after the ring-opening step.[86] When also taking into account the potential involvement of loop dynamics in determining the turnover rates, as is the case in other enzymes with catalytically important conformational changes such as protein tyrosine phosphatases,[26, 29, 32, 83] this means that the experimental turnover numbers for the isomerization of ProFAR and PRA by the enzymes of interest do not correspond to a chemical step occurring in the enzyme active site. However, as the Amadori reaction that occurs between the enzyme-catalyzed ring-opening reaction and the non-enzymatic tautomerization is likely to be fast (even the uncatalyzed reaction occurs spontaneously at 25 °C[85]), the ring-opening reaction is likely the slowest enzyme-catalyzed chemical step in the catalytic cycle. This is supported by QM/MM studies of the mechanism of the HisA-catalyzed reaction,[84] although this work does not take into account that chemistry is not rate-limiting here. However, even if the experimentally measured turnover numbers (k_{cat}) do not directly correspond to this step, they *do* produce a lower limit for the rate of this step (and thus an upper limit for the corresponding activation free energy for the rate-limiting step of the enzymatic reaction). Thus, comparing the calculated activation free energies for the ring-opening in different variants, and in different conformational states of loop 1, can still provide insight into the impact of loop dynamics and key amino acid substitutions on the rate of the slowest enzyme-catalyzed step.

Taking these limitations into account, the resulting experimental and calculated activation energies are shown in **Tables S10** and **S11** and **Figure 9**, with representative Michaelis complexes, transition states and ring-open intermediates for the ribose-ring opening reaction catalyzed by each wild-type enzyme shown in **Figures 10** (ProFAR) and **S10** (PRA). We note that the smaller substrate PRA has much more conformational freedom in the HisA and PriA active sites, which

are both evolved to (also) accommodate the larger substrate ProFAR, and is thus capable of sampling a broad number of different conformations at the Michaelis complex, leading to an increased standard error of the mean in comparison to systems where ProFAR is bound to the active site of the these enzymes.

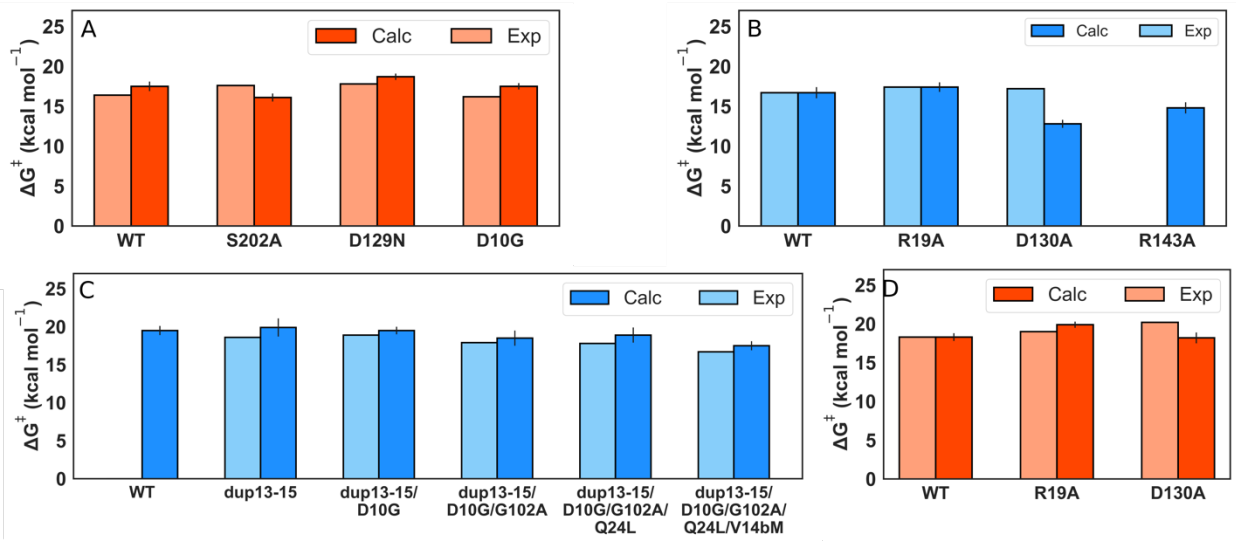

Figure 9. Calculated and experimentally derived activation energies for the isomerization of substrates (**A**,**D**) ProFAR (orange bars) and (**B,C)** PRA (blue bars), as catalyzed by (**A,C**) wild-type *Se*HisA and variants, and (**B,D**) *Mt*PriA and variants. The calculated free energies ($\Delta G^{\ddagger}_{calc}$**)** correspond to the ribose ring-opening step (**Tables S10** and **S11**). The experimental activation free energies ($\Delta G^{\ddagger}_{exp}$**)** were derived from the kinetic data presented in refs. [25, 28] .

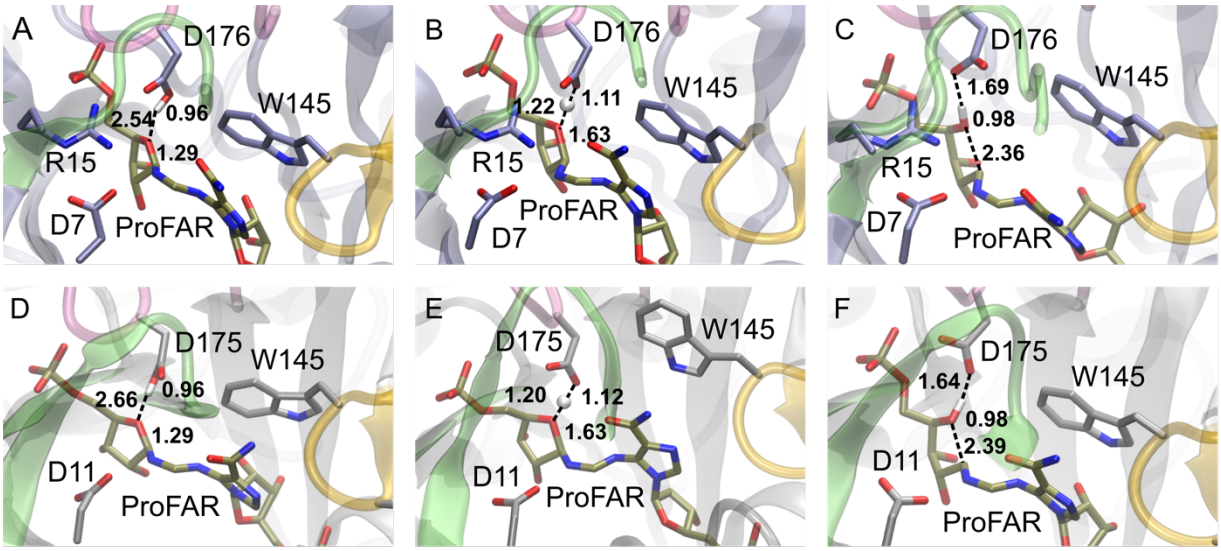

Figure 10. Representative structures of stationary points at the Michaelis complexes (MC), transition states (TS) and intermediate states (IS), respectively, for the ribose-ring opening step of the (**A, B, C**) HisA- and (**D, E, F**) PriA-catalyzed isomerization of ProFAR (by the wild-type enzymes). For EVB simulation details, see the **Methodology** section. Structures were selected based on clustering analysis using the hierarchical agglomerative algorithm, as implemented in CPPTRAJ.[77] Note that the annotated catalytic distances are average values over 6000 snapshots extracted for each state from our EVB trajectories (from 30 x individual 200 ps EVB mapping windows per stationary point/system). For a full list of reacting distances across all variants, see **Tables S12** and **13**.

As can be seen from these data, our calculations reproduce the experimental trend in activation free energies derived from the turnover numbers relatively well for both substrates ProFAR (**Table S11**) and PRA (**Table S12**), reproducing these values to within 2.0 kcal mol^{-1} for either substrate. This was unexpected, considering the experimental turnover number does not correspond to a chemical step in the enzyme, as discussed above, and indicates that the observed changes in activity nevertheless do have a chemical component.

While both wild-type PriA and HisA are capable of catalyzing isomerization of the substrate ProFAR, only PriA can catalyze the isomerization of the smaller substrate PRA. As already seen

from the PriA crystal structures with both products bound, loop 5 can be rearranged either to accommodate one substrate or the other, displaying two slightly different conformations of the loop (**Figure S1**).[25] These are a "knot-like" pro-ProFAR conformation of loop 5, with W145 pointing "in" towards the substrate ProFAR (PDB ID: 3ZS4[55]), and a pro-PRA β-hairpin conformation of loop 5, with R143 pointing "in" towards the substrate PRA (PDB ID: 2Y85[25, 55]), extrapolating the substrate positioning from the position of the analogous product PRFAR and product analog rCdRP in the respective conformations. Here, we used PriA in its pro-ProFAR conformation as a reference state to calibrate our EVB simulations of the initial ring-opening of ProFAR. The resulting EVB parameters were then used unchanged in all relevant systems.

Based on these parameters, we obtain an activation free energy of 17.5 ± 0.6 kcal mol^{-1} for the analogous reaction catalyzed by wild-type HisA, which is only 1.1 kcal mol^{-1} higher than the experimental value (derived from k_{cat}) of 16.4 kcal mol^{-1}. To further validate our PriA/HisA-ProFAR results, we performed single amino acid substitutions in each enzyme (R19A and D130A in PriA and S202A, D129N and D10G in HisA), and performed EVB simulations on these variants. In most of the systems we obtain activation free energies ~1 kcal mol^{-1} higher than the corresponding experimental values, as for wild-type HisA. However, for D130A in PriA and S202A in HisA, we underestimate the activation free energies by 1.5 - 2.0 kcal mol^{-1} in comparison with the experimental value, suggesting that the experimental effect is due to either a change in substrate positioning or loop conformation or dynamics, which we are unable to capture in our simulations when simply starting from the wild-type crystal structure and manually truncating these residues.

In the case of the substrate PRA, we again used the reaction catalyzed by wild-type *Mt*PriA as our EVB reference state, this time with the loop 5 in its pro-PRA conformation. We note that as

shown in **Table S2**, wild-type *Se*HisA does not show TrpF activity, whereas variants in which loop 1 is extended through dup13-15 do.[28] In the *Se*HisA(dup13-15) variant, we obtained an activation free energy of 19.9 ± 1.2 kcal mol^{-1}, which is within 1.3 kcal mol^{-1} of the experimental value (derived from k_{cat}) of 18.6 kcal mol^{-1} (noting again that the rate-limiting step for the enzyme-catalyzed reaction occurs off the enzyme,[86] so this value is only a proxy for the barrier for the ring-opening). We then extended our EVB calculations to model PRA ring-opening as catalyzed by a set of variants of *Mt*PriA and *Se*HisA(dup13-15) (**Figure 9** and **Table S11**). In the case of the *Se*HisA(dup13-15) variants, all the variants yielded results within reasonable agreement (~1 kcal mol^{-1}) with experimental values. We note as an aside that we also performed simulations on wild-type HisA for PRA substrate (which is not active toward this substrate) and obtained an activation free energy of 19.5 ± 0.6 kcal mol^{-1}, very similar to the one obtained for *Se*HisA(dup13-15), suggesting that in theory the wild-type enzyme could catalyze this reaction if all loops are in the correct conformation and the substrate is optimally positioned, and that the experimental lack of activity is not due to a high barrier to the chemical reaction catalyzed by this enzyme. Tying in with this, as described in the **Supporting Information**, our PRA simulations are initated from an idealized position of this substrate in the active site, based on overlay with the position of the larger substrate ProFAR. However, the stability of this ProFAR conformation in the active site is facilitated by interaction with the gripper loop 3, whereas PRA lacks the distal phosphodianion group of ProFAR and is thus not able to make this interaction (**Figure 3**). This suggests that if PRA *could* be gripped properly (and thus optimally aligned), turnover could in turn happen.

In the case of *Mt*PriA we modeled three single amino acid substitutions (R19A, D130A and R143A) and extended our EVB simulations to model the effect of these substitutions (**Table S11**, **Figure 9**), in order to specifically capture the impact of the loss of electrostatic contribution of

each truncated side chain on the activation free energy. In the case of the R19A variants, we obtain excellent agreement with the experimental value. However, in the case of the D130A variant, our model significantly under-estimates the activation free energy difference compared to experiment, again suggesting that the experimental effect is rather related to a change in substrate positioning or loop dynamics, that is not captured in our EVB simulations. In the case of the R143A variant, we obtain an activation free energy 1.9 kcal mol^{-1} lower than the wild-type enzyme. We note that, experimentally, this substitution has been shown to significantly impair the isomerization activity of *Mt*PriA towards PRA (k_{cat}/K_M reduced from 1.7 x 10^5 M^{-1} s^{-1} to 6.0 x 10^3 M^{-1} s^{-1} on introduction of this substitution[25]). However, it is unclear if this is an effect on k_{cat}, K_M or both, and it is plausible that the loss of activity is due to structural effects that prohibit productive substrate binding, that are not captured in our simulations.[38] This latter issue would be similar to our observations from a recent study of an analogous system activated by a ligand-gated conformational change, glycerol-3-phosphate dehydrogenase (GPDH).[38] In this system a substantial loss of activity upon truncation of a key catalytic arginine to alanine could only be explained by structural rearrangements (predominantly blocking of the closure of a key catalytic loop), that were observed upon crystallization of this variant. In contrast, this loss of activity could not be captured simply by performing a truncation of this side chain on the wild-type enzyme and considering only electrostatic effects which only accounted for a smaller part of the observed change in activity.

Finally, in order to examine how important loop 1 closure is for PRA isomerization, we also performed EVB simulations of ProFAR and PRA ring-opening as catalyzed by *Se*HisA (both wild-type and the *Se*HisA(dup13-15/D10G/G102A/Q24L/V15[b]M variant)) and *Mt*PriA (wild-type) with loop 1 in the open conformation (**Tables S11** and **S12**). Due to the lack of relevant crystal structures, we used structures with loop 1 in an open conformation extracted from our conventional

MD simulations as starting points for the EVB simulations. As can be seen from these data, for the catalyzed isomerization of PRA, when the reaction was modelled with loop 1 in an open conformation, we obtain much higher activation free energies for ring-opening than when modeling the reaction from a loop-closed conformation, due to a combination of the loss of key interactions between loop 1 and the substrate, and also extra solvent-exposure of the active site when this loop opens up. However, we observe no impact on the activation free energy when modelling the catalyzed isomerization of ProFAR, staring with loop 1 in open conformation. Therefore, while the catalytic importance of loops 5 and 6 is well-established,[25, 28, 54] our EVB calculations show a clear role also for correct closure of loop 1 for PRA isomerization, with full closure of the loop into a catalytically competent conformation being essential for efficient isomerization of PRA.

Overview and Conclusions

In the present work, we use a combination of conventional and enhanced sampling molecular dynamics simulations, as well as empirical valence bond calculations, to explore the role of loop dynamics in dictating the selectivity and evolvability of the evolutionarily important model enzymes, HisA and TrpF,[28, 45-48] which selectively catalyze the isomerization of substrates ProFAR and PRA, respectively (**Figure 1**), as well as the bifunctional isomerase PriA, which catalyzes both reactions in bacteria such as *M. tuberculosis*.[87] The roles of loop dynamics and ligand-gated conformational changes in TIM-barrel proteins and proteins from other folds have been a topic of substantial research interest (*e.g.* refs. [18, 34-38, 88-97], among many others). However, what makes the current enzymes stand out from these prior studies is the importance of not one but two (TrpF) or even *three* (HisA and PriA) long, mobile loops (**Figure 1**), the specific conformations of which

have been suggested to play an important role in facilitating the selectivity of PriA and evolved HisA variants.[25, 28, 54] Thus, these enzymes undergo more complex loop dynamics than the aforementioned systems. In addition, prior enzymes that have been characterized as being activated by ligand-gated conformational changes, such as TPI and OMPDC, are extremely proficient enzymes.[34-38, 98-100] In contrast, all enzymes studied here are relatively inefficient (**Table S2**),[25, 28, 80], with turnover numbers of ~10 s^{-1} or less,[28, 52, 101] despite catalyzing a reaction that is intrinsically very fast.[85] Related to this, while loop motions in highly proficient TIM-barrel enzymes such as TPI are relatively fast (on the μs timescale[96]), motions of up to the ms timescale have been detected in the evolved HisA variants,[28] and thus loop motions are likely to be (at least partially) rate-limiting in these enzymes.

At the simplest level, our simulations show, in agreement with structural data,[25] that the enzymes TrpF, PriA and HisA have increasingly large (in terms of active site volume) and "breathable" active sites (**Table S8**), allowing for the accommodation of substrate ProFAR by HisA and PriA unlike PRA-specific TrpF, the active site of which is clearly too small to accommodate the larger substrate.[25] More importantly, HisA and PriA both possess "gripper residues" interacting with the non-reactive phosphodianion group of ProFAR (R83 and S103 in HisA, R85 and T105 in PriA, **Figure 2**) that are very similar to analogous interactions in other enzymes activated by ligand-gated conformational changes, such as TPI.[34, 35] Of note, however, is that the HisA/PriA "gripper" residues are contributed from the less mobile loops 3 and 4, unlike the primary gripper loop, loop 6 in TPI, which undergoes a substantial conformational change upon ligand binding.[22] Other TIM-barrel proteins such as OMPDC possess analogous gripper loops to TPI loop 6,[22] showing evidence for convergent evolution on these different $(\beta\alpha)_8$-barrel scaffolds.

Our simulations show that while the gripper interaction is stable in HisA throughout the simulations, in PriA, there is electrostatic repulsion between R85 and an additional active site arginine, R143, which causes instabilities in the catalytic loops (in particular loop 5, **Figure S3**), as well as in the substrate positioning in the active site, such that the larger ProFAR is bound less stably in the PriA active site than in the HisA active site. This is in effect a ligand-gated effect, where interaction with the non-reactive phosphodianion (which is not present in the smaller substrate, PRA) facilitates the stability of catalytically important loop 5. Thus, the underlying principles driving loop stability are similar to those of other enzymes that are activated by ligand-gated changes.

Following from this, PCA analysis on our simulations shows that the active site loops in these enzymes are not rigid, but can sample a range of wide-open conformations with transitions between them, with their conformational flexibility being stabilized by ligand binding (although less so in the bifunctional PriA than the ProFAR-specific *Se*HisA). In contrast, in TrpF, which binds a smaller substrate and lacks the gripper, the active site loops remain dynamic, in particular loop 6 (**Figure 1**), which samples a range of open conformations even with substrate PRA bound to the active site. Related to this, all HisA variants from the real-time evolution experiment[46] studied here also sample a range of open and wide-open conformations. In this context, however, the single D10G substitution on loop 1 appears to be sufficient by itself to increase the population of the closed conformation sampled during our simulations (**Table 1**), and the highest proportion of closed conformation is observed in simulations of the *Se*HisA(dup13-15/D10G/Q102A/Q24L/V15[b]M) variant, which has the highest TrpF activity[28] (**Table S2**).

Furthermore, pulling simulations where we pull products PRFAR and CdRP out of the *Se*HisA(dup13-15/D10G/G102A/Q24L) active site (the only variant from the real-time evolution

experiment[28] with a PRFAR-bound crystallographic structure) show significant conformational changes in both loops 1 and 6 when pulling PRFAR out of the active site, whereas when pulling CdRP out of the active site, loop 1 is much more stable and the main requirement is for loop 6 to open. This suggests that loop 1 dynamics are more important for binding of ProFAR and subsequent release of PRFAR, than for the smaller substrate PRA and its product CdRP, whereas loop 1 dynamics appears to be critical to catalysis (**Table S11**). In addition, the substantial rearrangements that we observe for both compounds suggests that a slow (potentially rate-limiting) product release step is the reason for the low turnover numbers observed for an otherwise facile reaction, further providing evidence that turnover rates are being regulated by loop dynamics.

In addition, in contrast to HisA, which undergoes conformational changes of loop 1 during the real-time evolution experiment that changes its selectivity from ProFAR-specific to PRA-specific,[28, 46] the bifunctional enzyme PriA is already able to rearrange its active site in its wild-type form, to accommodate the different substrates through alternation between Pro-ProFAR and Pro-PRA conformations of loop 5.[25] These conformational changes both reduce repulsion between the two active site arginine side chains that in turn destabilize loop 5 dynamics (**Figure S3**), as well as disrupting the stacking interaction between the W145 side chain and the substrate ProFAR (**Figure S4**). This rationalizes the preference of this enzyme towards PRA rather than ProFAR, despite its similarities with the HisA active site, and we note also that loss of the stacking interaction between W145 and the substrate ProFAR was also presented as one aspect of the gain of PRA-isomerization activity in the *Se*HisA(dup13-15) variant from the real time evolution experiment.[28, 46]

Finally, we performed EVB simulations of the first ring-opening step of ProFAR and (where relevant) PRA isomerization (**Figure 1**) by wild-type HisA, PriA and variants. As described above,

the actual rate of the chemical step in these enzymes is unknown, since the rate-limiting step is likely to occur off the enzyme.[86] However, of the steps that occur in the enzyme active site, this is the step that is likely to be the slowest, and therefore it is of interest how the substitutions affect the rate of the ring-opening reaction. Here, we see that our calculated activation free energies for the ring-opening reaction trend well with the differences in activation free energies derived using the measured turnover numbers as an upper limit for this value, suggesting there is both a chemical and a dynamical component to the observed changes in activity upon substitution and/or duplication of key residues. Furthermore, in order to reproduce the relevant PriA activation free energies, it was necessary to start from different structures of loop 5, following earlier structural analysis that demonstrate the loop can exist in either a knot-like pro-PRA or beta-hairpin pro-ProFAR conformation (**Figure S1**), depending on what product is bound to the active site.[25] Clearly, the ease with which this rearrangement can occur will also impact the selectivity of this enzyme. Also, EVB simulations of wild-type *Se*HisA, *Mt*PriA and the *Se*HisA(dup13-15/D10G/G102A/Q24L/V15[b]M) variant with loop 1 in an open conformation all yield substantially higher energies for PRA isomerization, whereas ProFAR isomerization (by wild-type HisA) seems to be unaffected (**Tables S10** and **11**). This further emphasizes the importance of the correct closure of loop 1 for isomerization of the smaller substrate PRA. This is in contrast to substrate binding, where the conformational plasticity of loop 1 appears to be far more important for facilitating correct binding of ProFAR than PRA (**Figure S8**).

Taken together, these observations highlight the critical role of multiple decorating loops in HisA, PriA and TrpF in facilitating catalysis. These enzymes stand out from prior systems that have been demonstrated to be activated by ligand-gated conformational changes[34-41] due to a number of factors. First, we have shown the inter-dependent motion of three long loops (or two in

TrpF), none of which dominates and each of which is capable of undergoing substantial conformational changes to facilitate the turnover of different substrates. Second, unlike prior systems which show substantial rate accelerations compared to the uncatalyzed reactions with comparatively rapid loop motions, in these enzymes, the catalyzed reaction is already intrinsically fast[85] whereas loop-motion is slow and appears to be controlling the reaction rate.

It has been argued that a PriA-like gene product could have been the common evolutionary ancestor for both HisA and TrpF.[53, 87, 102] Ancestral sequence reconstruction has also been used to suggest that ancient HisA precursors were likely bifunctional, and that this bifunctionality persisted over at least a two-billion-year time span.[52] However, as shown in **Figure 5**, HisA and PriA exploit loops 1, 5 and 6 to facilitate activity, whereas TrpF lacks an analog for loop 1, and isomerizes PRA harnessing just two catalytic loops, 3 and 6. Our results suggest that an evolutionary trajectory from a PriA-like ancestor to an extant TrpF would be surprisingly complex. Loop 1 would need to be truncated (and not extended, as when *Se*HisA was artificially evolved into a TrpF[28]) and the inter-dependency of this third loop would need to be lost, raising the question of what evolutionary path would take a PriA-like precursor to TrpF, while completely abolishing this loop.

Despite the novel aspects of the systems studied here, a key similarity with prior systems is the generalizability of ligand-gated conformational changes across a wide range of systems, in particular TIM-barrel proteins,[34] which tend to possess flexible loops decorating their active sites. The conservation of such ligand-gated conformational changes—albeit triggered in different loops—suggests that these decorating loops evolve independently of the barrel providing a starting point for the emergence and divergence of new enzyme activities.[31, 47] In addition, TrpF, for example, has been shown to be highly tolerant of variations in loop 6 sequence, such that grafting

sequences from related enzymes such as TrpA, HisA and PriA onto the TrpF scaffold did not abolish activity.[103] This is significant considering the high evolvability of this scaffold,[18] and the wide range of chemistry it supports,[47] which makes it very desirable as a starting point for protein engineering efforts. In addition, it could be argued that the real-time evolution experiment that bestowed PRA-isomerization activity to HisA[46] effectively performed "natural" loop-engineering by altering the conformations of key active site loops.[28] Our work suggests therefore that, more broadly, loop grafting and engineering is a powerful tool for generating novel enzymes with tailored activities and specificities, even in complex systems with multiple highly mobile and inter-dependent catalytic loops.

Supporting Information

Full details of computational methodology, additional structural and simulation analysis. Representative EVB input files, starting structures, parameter files, and *Q*6 topology files have been uploaded to Zenodo, DOI: 10.5281/zenodo.5893598.

Acknowledgments

This work was supported by the Knut and Alice Wallenberg Foundation (grant numbers 2018.0140 and 2019.0431), the Human Frontier Science Program (grant RGP0041/2017), and the Swedish Research Council (grant number 2019-03499). This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 890562. The simulations were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at multiple supercomputing centers (NSC, HPC2N, UPPMAX), partially funded by the Swedish Research

Council through grant agreement no. 2016-07213. Further simulations were performed at the Barcelona Supercomputing Center (QSB-2019-3-0003). We also thank Dan Andersson, Maria Selmer, José Manuel Sanchéz Ruiz, Annika Söderholm, Fernanda Duarte, Kshatresh Dutta Dubey and Klaudia Szeler for helpful and insightful discussions, and in particular Dan Andersson, Maria Selmer and José Manuel Sanchéz Ruiz for inspiring us to undertake this study.

References

1. Gora, A.; Brezovsky, J.; Damborsky, J., Gates of Enzymes. *Chem. Rev.* **2013,** *113*, 5871-5923.

2. McGeagh, J. D.; Ranaghan, K. E.; Mulholland, A. J., Protein Dynamics and Enzyme Catalysis: Insights from Simulations. *Biochim. Biophys. Acta, Proteins Proteomics* **2011,** *1814*, 1077-1092.

3. Villali, J.; Kern, D., Choreographing an Enzyme's Dance. *Curr. Opin. Chem. Biol.* **2010,** *14*, 636-643.

4. Khersonsky, O.; Tawfik, D. S., Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective. *Annu. Rev. Biochem.* **2010,** *79*, 471-505.

5. Doshi, U.; Hamelberg, D., The Dilemma of Conformational Dynamics in Enzyme Catalysis: Perspectives from Theory and Experiment. *Adv. Exp. Med. Biol.* **2014,** *805*, 221-243.

6. Kohen, A., Role of Dynamics in Eznyme Catalysis: Substantial versus Semantic Controversies. *Acc. Chem. Res.* **2015,** *48*, 466-473.

7. Klinman, J. P., Dynamically Achieved Active Site Precision in Enzyme Catalysis. *Acc. Chem. Res.* **2015,** *48*, 449-456.

8. James, L. C.; Tawfik, D. S., Conformational Diversity and Protein Evolution - A 60 Year Old Hypothesis Revisited. *Trends Biochem. Sci.* **2003,** *28*, P361-P368.

9. Tokuriki, N.; Tawfik, D. S., Protein Dynamism and Evolvability. *Science* **2009,** *324*, 203-207.

10. Petrović, D.; Risso, V. A.; Kamerlin, S. C. L.; Sanchez-Ruiz, J. M., Conformational Dynamics and Enzyme Evolution. *J. R. Soc. Interface* **2018,** *15*, 20180330.

11. Johansson, K. E.; Lindorff-Larsen, K., Structural Heterogeneity in Protein Evolution and Design. *Curr. Opin. Struct. Biol.* **2018,** *48*, 157-163.

12. Maria-Solano, M. A.; Serrano-Hervás, E.; Romero-Rivera, A.; Iglesias-Fernández, J.; Osuna, S., Role of Conformational Dynamics in the Evolution of Novel Enzyme Functions. *Chem. Commun.* **2018,** *54*, 6622-6634.

13. Campitelli, P.; Modi, T.; Kumar, S.; Ozkan, S. B., The Role of Conformational Dynamics and Allostery in Modulating Protein Evolution. *Annu. Rev. Biochem.* **2020,** *49*, 267-288.

14. Damry, A. M.; Jackson, C. J., The Evolution and Engineering of Enzyme Activity Through Tuning Conformational Landscapes. *Prot. Eng. Des. Sel.* **2021,** *34*, gzab009.

15. Campbell, E. C.; Correy, G. J.; Mabbitt, P. D.; Buckle, A. M.; Tokuriki, N.; Jackson, C. J., Laboratory Evolution of Protein Conformational Dynamics. *Curr. Opin. Struct. Biol.* **2018,** *50*, 49-57.

16. Crean, R. M.; Gardner, J. M.; Kamerlin, S. C. L., Harnessing Conformational Plasticity to Generate Designer Enzymes. *J. Am. Chem. Soc.* **2020,** *142*, 11324-11342.

17. Osuna, S., The Challenge of Predicting Distal Active Site Mutations in Computational Enzyme Design. *WIREs Comp. Mol. Sci.* **2020,** *11*, e1502.

18. Wierenga, R. K., The TIM-Barrel Fold: A Versatile Framework for Efficient Enzymes *FEBS Lett.* **2001,** *492*, 193-198.

19. Rajagopalan, P. T. R.; Benkovic, S. J., Preorganization and Protein Dynamics in Enzyme Catalysis. *Chem. Rec.* **2002,** *2*, 24-36.

20. McElheny, D.; Schnell, J. R.; Lansing, J. C.; Dyson, H. J.; Wright, P. E., Defining the Role of Active-Site Loop Fluctuations in Dihydrofolate Reductase Catalysis. *Proc. Natl. Acad. Sci. USA* **2005,** *102*, 5032-5037.

21. Tawfik, D. S., Loop Grafting and the Origins of Enzyme Species. *Science* **2006,** *311*, 475-476.

22. Malabanan, M. M.; Amyes, T. L.; Richard, J. P., A Role for Flexible Loops in Enzyme Catalysis. *Curr. Opin. Struct. Biol.* **2010,** *20*, 702-710.

23. Wierenga, R. K., Triosephosphate Isomerase: A Highly Evolved Biocatalyst. *Cell. Mol. Life Sci.* **2010,** *67*, 3961-3982.

24. Bhabha, G.; Lee, J.; Ekiert, D. C.; Gam, J.; Wilson, I. A.; Dyson, H. J.; Benkovic, S. J.; Wright, P. E., A Dynamic Knockout Reveals That Conformational Fluctuations Influence the Chemical Step of Enzyme Catalysis. *Science* **2011,** *332*, 234-238.

25. Due, A. V.; Kuper, J.; Geerlof, A.; von Kries, J. P.; Wilmanns, M., Bisubstrate Specificity in Histidine/Tryptophan Biosynthesis Isomerase from *Mycobacterium Tuberculosis* by Active Site Metamorphosis. *Proc. Natl. Acad. Sci. USA* **2011,** *108*, 3554-3559.

26. Whittier, S. K.; Hengge, A. C.; Loria, J. P., Conformational Motions Regulate Phosphoryl Transfer in Related Protein Tyrosine Phosphatases. *Science* **2013,** *341*, 899-903.

27. Blaha-Nelson, D.; Krüger, D. M.; Szeler, K.; Ben-David, M.; Kamerlin, S. C. L., Active Site Hydrophobicity and the Convergent Evolution of Paraoxonase Activity in Structurally Divergent Enzymes: The Case of Serum Paraoxonase 1. *J. Am. Chem Soc.* **2017,** *139*, 1155-1167.

28. Newton, M. S.; Guo, X.; Söderholm, A.; Näsvall, J.; Lundström, P.; Andersson, D. I.; Selmer, M.; Patrick, W. M., Structural and Functional Innovations in the Real-Time Evolution of New $(\beta\alpha)_8$ Barrel Enzymes. *Proc. Natl. Acad. Sci. USA* **2017,** *114*, 4727-4732.

29. Moise, G.; Morales, Y.; Beaumont, V.; Caradonna, T.; Loria, J. P.; Johnson, S. J.; Hengge, A. C., A YopH PTP1B Chimera Shows the Importance of the WPD-Loop Sequence to the Activity, Structure, and Dynamics of Protein Tyrosine Phosphatases. *Biochemistry* **2018,** *57*, 5315-5326.

30. Moreira, C.; Calixto, A. R.; Richard, J. P.; Kamerlin, S. C. L., The Role of Ligand-Gated Conformational Changes in Enzyme Catalysis. *Biochem. Soc. Trans.* **2019,** *47*, 1449-1460.

31. Richard, J. P., Protein Flexibility and Stiffness Enable Efficient Enzymatic Catalysis. *J. Am. Chem. Soc.* **2019,** *141*, 3320-3331.

32. Crean, R. M.; Biler, M.; W., v. d. K. M.; Hengge, A. C.; Kamerlin, S. C. L., Loop Dynamics and Enzyme Catalysis in Protein Tyrosine Phosphatases. *J. Am. Chem. Soc.* **2021,** *143*, 3830-3845.

33. Nestl, B. M., Engineering of Flexible Loops in Enzymes. *ACS Catal.* **2014,** *4*, 3201-3211.

34. Richard, J. P.; Zhai, X.; Malabanan, M. M., Reflections on the Catalytic Power of a TIM-barrel. *Bioorg. Chem.* **2014,** *57*, 206-212.

35. Richard, J. P., A Paradigm for Enzyme-Catalyzed Proton Transfer at Carbon: Triosephosphate Isomerase. *Biochemistry* **2012,** *51*, 2652-2661.

36. Richard, J. P.; Amyes, T. L.; Reyes, A. C., Orotidine 5′-Monophosphate Decarboxylase: Probing the Limits of the Possible for Enzyme Catalysis. *Acc. Chem. Res.* **2018,** *51*, 960-969.

37. He, R.; Reyes, A. C.; Amyes, T. L.; Richard, J. P., Enzyme Architecture: The Role of a Flexible Loop in Activation of Glycerol-3-Phosphate Dehydrogenase for Catalysis of Hydride Transfer. *Biochemistry* **2018,** *57*, 3227-3236.

38. Mhashal, A. R.; Romero-Rivera, A.; Mydy, L. S.; Cristobal, J. R.; Gulick, A. M.; Richard, J. P.; Kamerlin, S. C. L., Modeling the Role of a Flexible Loop and Active Site Side Chains in Hydride Transfer Catalyzed by Glycerol-3-phosphate Dehydrogenase. *ACS Catal.* **2020,** *10*, 11253-11267.

39. Kholodar, S. A.; Murkin, A. S., DXP Reductoisomerase: Reaction of the Substrate in Pieces Reveals a Catalytic Role for the Non-Reacting Phosphodianion Group. *Biochemistry* **2013,** *52*, 2302-2308.

40. Kholodar, S. A.; Allen, C. L.; Gulick, A. M.; Murkin, A. S., The Role of Phosphate in a Multistep Enzymatic Reaction: Reactions of the Substrate and Intermediate in Pieces. *J. Am. Chem Soc.* **2015,** *137*, 2748-2756.

41. Ray, W. J.; Long, J. W.; Owens, J. D., An Analysis of the Substrate-Induced Rate Effect in the Phosphoglucomutase System. *Biochemistry* **1976,** *15*, 4006-4017.

42. Kadumuri, R. V.; Vadrevu, R., Diversity in αβ and βα Loop Connections in TIM Barrel Proteins: Implications for Stability and Design of the Fold. *Interdiscip. Sci. Comput. Life Sci.* **2018,** *10*, 805-812.

43. Saraste, M., The TIM-Barrel Fold: A Versatile Framework for Efficient Enzymes. *FEBS Lett.* **2001,** *492*, 193-198.

44. Liao, Q.; Kulkarni, Y.; Sengupta, U.; Petrović, D.; Mulholland, A. J.; van der Kamp, M. W.; Strodel, B.; Kamerlin, S. C. L., Loop Motion in Triosephosphate Isomerase is not a Simple Open and Shut Case. *J. Am. Chem. Soc.* **2018,** *140*, 15889-15903.

45. Lundin, E.; Näsvall, J.; Andersson, D. I., Mutational Pathways and Trade-Offs Between HisA and TrpF Functions: Implications for Evolution via Gene Duplication and Divergence. *Front. Microbiol.* **2020,** *11*, 588235.

46. Näsvall, J.; Sun, L.; Roth, J. R.; Andersson, D. I., Real-Time Evolution of New Genes by Innovation, Amplification and Divergence. *Science* **2012,** *338*, 384-387.

47. Romero-Romero, S.; Kordes, S.; Michel, F.; Höcker, B., Evolution, Folding, and Design of TIM Barrels and Related PRoteins. *Curr. Opin. Struct. Biol.* **2021,** *68*, 94-104.

48. Copley, S. D., Evolution of New Enzymes By Gene Duplication and Divergence. *FEBS J.* **2020,** *287*, 1262-1283.

49. Henn-Sax, M.; Thoma, R.; Schmidt, S.; Hennig, M.; Kirschner, K.; Sterner, R., Two $(\beta\alpha)_8$-Barrel Enzymes of Histidine and Tryptophan Biosynthesis Have Similar Reaction Mechanisms and Common Strategies for Protecting Their Labile Substrates. *Biochemistry* **2002,** *41*, 12032-12042.

50. Juárez-Vázquez, A. L.; Edirisinghe, J. N.; Verduzco-Castro, E. A.; Michalska, K.; Wu, C.; Noda-Garcia, L.; Endres, M.; Medina-Ruíz, S.; Santoyo-Flores, J.; Carrillo-Tripp, M.; Ton-That, H.; Joachimiak, A.; Henry, C. S.; Barona-Gómez, F., Evolution of Substrate Specificity in a Retained Enzyme Driven by Gene Loss. *eLife* **2017,** *6*, e22679.

51. Kuper, J.; Dönges, C.; Wilmanns, M., Two-fold Repeated $(\beta\alpha)_4$ Half-Barrels May Provide a Molecular Tool for Dual Substrate Specificity. *EMBO Rep.* **2005,** *6*, 134-139.

52. Plach, M. G.; Reisinger, B.; Sterner, R.; Merkl, R., Long-Term Persistence of Bi-functionality Contributes to the Robustness of Microbial Life to Exaptation. *PLoS Genet.* **2016,** *12*, e1005836.

53. Jürgens, C.; Strom, A.; Wegener, D.; Hettwer, S.; Wilmans, S.; Sterner, R., Directed Evolution of a $(\beta\alpha)_8$-Barrel Enzyme to Catalyze Related Reactions in Two Different Metabolic Pathways. *Proc Natl Acad Sci USA* **2000,** *97*, 9925-9930.

54. Söderholm, A.; Guo, X.; Newton, M. S.; Evans, G. B.; Näsvall, J.; Patrick, W. M.; Selmer, M., Two-step Ligand Binding in a $(\beta\alpha)_8$ Barrel Enzyme: Substrate-Bound Structures Shed New Light on the Catalytic Cycle of HisA. *J. Biol. Chem.* **2015,** *290*, 24657-24668.

55. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res.* **2000,** *28*, 235-242.

56. Katebi, A. R.; Jernigan, R. L., The Critical Role of the Loops of Triosephosphate Isomerase for its Oligomerization, Dynamics and Functionality. *Protein Sci.* **2014,** *23*, 213-228.

57. Shapovalov, M. X.; Dunbrack Jr., R. L., A Smoothed Backbone-Dependent Rotamer Library for Proteins Derived from Adaptive Kernel Density Estimates and Regressions. *Structure* **2011,** *19*, 844-858.

58. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comp. Chem.* **2004,** *25*, 1605-1612.

59. Šali, A.; Blundell, T. L., Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993,** *234*, 779-815.

60. Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H., PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK_a Predictions. *J. Chem. Theory Comput.* **2011,** *7Th* 525-537.

61. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006,** *25*, 247-260.

62. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranansinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssel, l., K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J., *Gaussian 09, Revision E.01*. Gaussian, Inc.: Wallingford, CT, 2016.

63. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollmann, P. A.; Case, D. A., Development and Testing of a General AMBER Force Field. *J. Comput. Chem.* **2004,** *25*, 1157-1174.

64. Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev,

A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; Kollman, P. A., *AMBER 2016*. University of California: San Francisco, 2016.

65. Maier, J. A.; Martinez, C.; Kasavaghala, K.; Wickström, L.; Hauser, K. E.; Simmerling, C. L., FF14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from FF99SB. *J. Chem. Theory Comput.* **2015,** *11*, 3696-3713.

66. Jorgensen, W. L.; Chandresekhar, J.; Madura, J. D., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79*, 926.

67. Schneider, T.; Stoll, E., Molecular-Dynamics Study of a Three-Dimensional One-Component Model for Distortive Phase Transitions. *Phys. Rev. B* **1978,** *17*, 1302-1322.

68. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984,** *81*, 3684-3690.

69. Warshel, A.; Weiss, R. M., An Empirical Valence Bond Approach for Comparing Reactions in Solutions and in Enzymes. *J. Am. Chem. Soc.* **1980,** *102*, 6218-6226.

70. Blaha-Nelson, D.; Krüger, D. M.; Szeler, K.; Ben-David, M.; Kamerlin, S. C. L., Active Site Hydrophobicity and the Convergent Evolution of Paraoxonase Activity in Structurally Divergent Enzymes: The Case of Serum Paraoxonase 1. *J. Am. Chem. Soc.* **2017,** *139*, 1155-1167.

71. Ben-David, M.; Soskine, M.; Dubovetskyi, A.; Cherukuri, K.-P.; Dym, O.; Sussman, J. L.; Liao, Q.; Szeler, K.; Kamerlin, S. C. L.; Tawfik, D. S., Enzyme Evolution: An Epistatic Ratchet versus a Smooth Reversible Transition. *Mol. Biol. Evol.* **2020,** *37*, 1133-1147.

72. Amrein, B. A.; Bauer, P.; Duarte, F.; Janfalk Carlsson, Å.; Naworyta, A.; Mowbray, S. L.; Widersten, M.; Kamerlin, S. C. L., Expanding the Catalytic Triad in Epoxide Hydrolases and Related Enzymes. **2015**, *5,* 5702-5713.

73. Bauer, P.; Janfalk Carlsson, Å.; Amrein, B. A.; Dobritzsch, D.; Widersten, M.; Kamerlin, S. C. L., Conformational Diversity and Enantioconvergence in Potato Epoxide Hydrolase 1. *Org. Biomol. Chem.* **2016,** *14*, 5639-5651.

74. Marelius, J.; Kolmodin, K.; Feierberg, I.; Åqvist, J., Q: A Molecular Dynamics Program for Free Energy Calculations and Empirical Valence Bond Simulations in Biomolecular Systems. *J. Mol. Graph. Model.* **1998,** *16*, 213-225.

75. Bauer, P.; Barrozo, A.; Purg, M.; Amrein, B. A.; Esguerra, M.; Wilson, P. B.; Major, D. T.; Åqvist, J.; Kamerlin, S. C. L., *Q6*: A Comprehensive Toolkit for Empirical Valence Bond and Related Free Energy Calculations. *SoftwareX* **2018,** *7*, 388-395.

76. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996,** *118*, 11225-11236.

77. Roe, D. R.; Cheatham III, T. E., PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013,** *9*, 3084-3095.

78. Schmidtke, P.; Bidon-Chanal, A.; Luque, F. J.; Barril, X., MDpocket: Open-Source Cavity Detection and Characterization on Molecular Dynamics Trajectories. *Bioinformatics* **2011,** *27*, 3276-3285.

79. Le Guilloux, V.; Schmidtke, P.; Tuffery, P., Fpocket: An Open Source Platform for Ligand Pocket Detection. *BMC Bioinformat.* **2009,** *10*, 168.

80. Sterner, R.; Kleemann, G. R.; Szadkowski, H.; Lustig, A.; Hennig, M.; Kirschner, K., Phosphoribosyl Anthranilate Isomeraes from *Thermotoga Maritima* is an Extremely Stable and Active Homodimer. *Prot. Sci.* **1996,** *5*, 2000-2008.

81. Noda-Garcia, L.; Juárez-Vázquez, A. L.; Ávila-Arcos, M. C.; Verduzco-Castro, E. A.; Montero-Morán, G.; Gaytán, P.; Carrillo-Tripp, M.; Barona-Gómez, F., Insights Into the Evolution of Enzyme Substrate Promiscuity After the Discovery of $(\beta\alpha)_8$ Isomerase Evolutionary Intermediates from a Diverse Metagenome. *BMC Evol. Biol.* **2015,** *15*, 107.

82. Hennig, M.; Sterner, R.; Kirschner, K.; Jansonius, J. N., Crystal Structure at 2.0 Å Resolution of Phosphoribosyl Anthranilate Isomerase from the Hyperthermophile *Thermotoga Maritima*: Possible Determinants of Protein Stability. *Biochemistry* **1997,** *36*, 6009-6016.

83. Shen, R.; Crean, R. M.; Olsen, K. J.; Richan, T.; Brandão, T. A. S.; Berry, R. D.; Tolman, A.; Loria, J. P.; Johnson, S. J.; Kamerlin, S. C. L.; Hengge, A. C., Insights into the Importance of WPD-Loop Sequence for Activity and Structure in Protein Tyrosine Phosphatases. *ChemRxiv* **2021**, 10.33774/chemrxiv-2021-8f6mc.

84. Dutta Dubey, K.; Singh, W., Simulations Reveal the Key Role of Arg15 in the Promiscuous Activity in the HisA Enzyme. *Org. Biomol. Chem.* **2021,** *19*, 10652-10661.

85. Nursten, H., *The Maillard Reaction: Chemistry, Biochemistry and Implications*. Royal Society of Chemistry: Great Britain, 2005.

86. Hommel, U.; Eberhard, M.; Kirschner, K., Phosphoribosyl Anthranilate Isomerase Catalyzes a Reversible Amadori Reaction. *Biochemistry* **1995,** *34*, 5249-5439.

87. Barona-Gómez, F.; Hodgson, D. A., Occurence of a Putative Ancient-Like Isomerase Involved in Histidine and Tryptophan Biosynthesis. *EMBO Rep.* **2003,** *4*, 296-300.

88. Brändén, C.-I., The TIM Barrel - The Most Frequently Occuring Folding Motif in Proteins. *Curr. Opin. Struct. Biol.* **1991,** *1*, 978-983.

89. Katebi, A. R.; Jernigan, R. L., The Critical Role of the Loops of Triosephosphate Isomerase for Its Oligomerization, Dynamics and Functionality. *Prot. Sci.* **2014,** *23*, 213-228.

90. Joseph, D.; Petsko, G.; Karplus, M., Anatomy of a Conformational Change: Hinged "Lid" Motion of the Triosephosphate Isomerase Loop. *Science* **1990,** *249*, 1425-1428.

91. Desamero, R.; Rozovsky, S.; Zhadin, N.; McDermott, A.; Callender, R., Active Site Loop Motion in Triosephosphate Isomerase: T-Jump Relaxation Spectroscopy of Thermal Activation. *Biochemistry* **2003,** *42*, 2941-2951.

92. Kadumuri, R. V.; Vadrevu, R., Diversity in αβ and βα Loop Connections in TIM Barrel Proteins: Implications for Stability and Design of the Fold. *Interdiscip. Sci.* **2018,** *10*, 805-812.

93. Liao, Q.; Kulkarni, Y.; Sengupta, U.; Petrovic, D.; Mulholland, A. J.; van der Kamp, M.; Strodel, B.; Kamerlin, S. C. L., Loop Motion in Triosephosphate Isomerase Is Not a Simple Open and Shut Case. *J. Am. Chem. Soc.* **2018,** *140*, 15889-15903.

94. Mhashal, A. R.; Romero-Rivera, A.; Mydy, L. S.; Cristobal, J. R.; Gulick, A. M.; Richard, J. P.; Kamerlin, S. C. L., Modeling the Role of a Flexible Loop and Active Site Side Chains in Hydride Transfer Catalyzed by Glycerol-3-phosphate Dehydrogenase. *ACS Catal.* **2020,** *19*, 11253-11267.

95. Rozovsky, S.; Jogl, G.; Tong, L.; McDermott, A. E., Solution-State NMR Investigations of Triosephosphate Isomerase Active Site Loop Motion: Ligand Release in Relation to Active Site Loop Dynamics. *J. Mol. Biol.* **2001,** *310*, 271-280.

96. Rozovsky, S.; McDermott, A. E., The Time Scale of the Catalytic Loop Motion in Triosephosphate Isomerase. *J. Mol. Biol.* **2001,** *310*, 259-270.

97. Wang, Y.; Berlow, R.; Loria, J. P., The Role of Loop-Loop Interactions in Coordinating Motions and Enzymatic Function in Triosephosphate Isomerase. *Biochemistry* **2009,** *48*, 4548-4556.

98. Radzicka, A.; Wolfenden, R., A Proficient Enzyme. *Science* **1995,** *267*, 90-93.

99. Knowles, J. R.; Albery, W. J., Perfection in Enzyme Catalysis: The Energetics of Triosephosphate Isomerase. *Acc. Chem. Res.* **1977,** *10*, 105-111.

100. Knowles, J. R., Enzyme Catalysis: Not Different, Just Better. *Nature* **1991,** *350*, 121-124.

101. Claren, J.; Malisi, C.; Höcker, B.; Sterner, R., Establishing Wild-Type Levels of Catalytic Activity on Natural and Artificial (βα)8-Barrel Protein Scaffolds. *Proc. Natl. Acad. Sci. USA* **2009,** *106*, 3704-3709.

102. Lang, D.; Thoma, R.; Henn-Sax, M.; Sterner, R.; Wilmanns, M., Structural Evidence for Evolution of the β/α Barrel Scaffold by Gene Duplication and Fusion. *Science* **2000,** *289*, 1546-1550.

103. Ochoa-Leyva, A.; Montero-Morán, G.; Saab-Rincón, G.; Brieba, L. G.; Soberón, X., Alternative Splice Variants in TIM Barrel Proteins from Human Genome Correlate with the Structural and Evolutionary Modularity of this Versatile Protein Fold. *PLoS One* **2013,** *8*, e70582.

Table of Contents Graphic