

Biophysical Journal, Volume 121

Supplemental information

**Distinct allosteric pathways in imidazole glycerol phosphate synthase
from yeast and bacteria**

Federica Maschietto, Aria Gheeraert, Andrea Piazzì, Victor S. Batista, and Ivan Rivalta

Supporting Information

Distinct Allosteric Pathways in Imidazole Glycerol Phosphate Synthase from Yeast and Bacteria

Federica Maschietto¹, Aria Gheeraert², Andrea Piazzini³, Victor S. Batista^{1*} and Ivan Rivalta^{2,3*}

¹Department of Chemistry, Yale University, New Haven, CT 06511, U.S.A.

²Université de Lyon, CNRS, Institut de Chimie de Lyon, École Normale Supérieure de Lyon, 46 Allée d'Italie, F-69364 Lyon Cedex 07, France

³Dipartimento di Chimica Industriale “Toso Montanari”, ALMA MATER STUDIORUM, Università di Bologna, Viale del Risorgimento 4, 40126 Bologna, Italia

Section S1. Materials and Methods

Correlation matrices for IGPS from *Thermotoga maritima* (*Tm*-IGPS) are obtained from the same trajectories and following the same protocol as in reference ¹, while yeast models (*Sc*-IGPS from *Saccharomyces cerevisiae*) are built ex-novo.

The computational structural models for apo and PRFAR bound yeast IGPS complexes are based on the crystal structure of the bienzyme complex from *S.cerevisiae* at 2.4 Å resolution (Protein Data Bank code 1OX6-B).² The HisH-HisF apo-complex having several missing residues (261-275, 301-304, and 551-552) and three extra residues at the beginning of the chain required modeling prior to simulation. To complete the structure, first, we stripped the first three residues, then we aligned and added residues 256-260 and 299-310 from 1OX4-B (removing overlapping residues from 1OX6 due to poor alignment). Finally, we added residues 550-552 from 1JVN-A, (removing residue 550 from 1OX6-B). We constructed the remaining residues (256-275) using different tools, using which we produced six different structures. One structure was generated using Modeller,³ a second one using Swiss-Model,⁴ and four suitable homology models were found on modbase. PRFAR was bound to each model by aligning each structure to the effector-bound crystal structure of yeast IGPS (PDB code 1OX5).

The twelve generated structures (six in the apo state, six bound to the effector) align with RMSD < 5 Å. To allow for a direct comparison between the dynamics of IGPS enzymes from *Tm*- and *Sc*-IGPS we kept the simulation conditions analogous to the one used for bacterial IGPS in reference.¹ For the sake of clarity, we report some essential details below. MD simulations of the apo and PRFAR-bound structures of yeast IGPS are based on the AMBER-ff99SB⁵ force field for the protein and Generalized Amber Force Field⁶ for the PRFAR ligand (see SI Text), as implemented in the Amber20 software package.⁷ We performed twelve independent MD simulations, one for each complex (apo and PRFAR-bound) for a total simulation time of 1.2 μs.

Structure refinements such as addition of hydrogen and explicit TIP3 water solvent molecules (reaching density values $\geq 0.9 \text{ mol} \cdot \text{Å}^{-3}$) are performed using AmberTools (2020). A constrained optimization with all atoms but solvent fixed at the crystal structure positions yields optimized solvated structures which are then slowly heated to 303 K, performing MD simulations (100 ps) in the canonical NVT ensemble using Langevin dynamics. We apply harmonic constraints to protein and PRFAR heavy atoms, with force constants set to $1 \text{ kcal} \cdot \text{mol}^{-1}$. During the heating procedure all positional constraints are gradually lifted until all atoms are set freed.

Unconstrained MD simulations are run for more than 9 ns, for total pre-equilibration simulation time of at least 10 ns. The pre-equilibrated systems are simulated in the NPT ensemble at 300 K and 1 atm using the Langevin dynamics for 100 ns. All simulations are performed using periodic boundary conditions. Van der Waals interactions are calculated using a switching distance of 10 Å and a cutoff of 12 Å and electrostatic interactions are treated using the Particle Mesh Ewald method.⁸ We employ the multiple time-stepping algorithm,⁹ where bonded, short-range nonbonded, and long-range electrostatic interactions are evaluated at every one, two, and four time steps, respectively, using a timestep of integration set to 1 fs.

Section S2. Details on the computation of correlation values and their analysis through the eigenvector centrality metrics, principal component analysis and allosteric pathways across yeast and bacterial IGPS.

Generalized correlation coefficients, eigenvector centrality and community network analysis

We quantify the extent of the dynamical correlation of fluctuations in the positions of Cα-atoms by computing the generalized correlation coefficient between each pair of residues,¹⁰

$$r_{\text{MI}}[x_i, x_j] = (1 - \exp(-\frac{2}{3}I[x_i, x_j]))^{1/2} \quad (1)$$

computed in terms of mutual information (MI),¹¹

$$I[x_i; x_j] = [Hx_i] + H[x_j] - H[x_i; x_j] \quad (2)$$

Here, $[Hx_i]$, $H[x_j]$, $H[x_i; x_j]$ are the marginal and joint (Shannon) entropies for atomic vector displacements $(x_i; x_j)$, computed along twelve independent 100 ns MD simulations for both apo and PRFAR-bound yeast IGPS complexes. The resulting generalized correlation coefficient values r_{MI} values fall in between 0 and +1, representing respectively uncorrelated and fully correlated variables. r_{MI} alone can be hard to decipher and require some post-processing to interpret protein behavior. Network analysis tools,^{12,13} including different centrality metrics¹⁴ can be applied for the interpretation of correlated protein motions and their allosteric behavior. Here, the Cα-atoms of the proteins' amino-acid residues constitute the nodes of a dynamical network graph, connected by edges (residue pair connection in terms of $r_{\text{MI}}[x_i, x_j]$). An adjacency matrix A is then constructed such that it can be used to identify the key amino acid residues of IGPS with high susceptibility to effector binding. A simple yet effective metric extract "central" nodes in A is the eigenvector centrality EC. The basic idea behind this measure is the assumption that the centrality index of a node is not only determined by its position in the network but also by the neighboring nodes, hence it measures how well connected a node is to other well-connected nodes in the network. The EC of a node is defined as the weighted sum of the centralities of all nodes that are connected to it by an edge, A_{ij} :

$$c_i = \epsilon^{-1} \sum_{j=1}^n A_{ij}c_j, \quad (3)$$

where c is an eigenvector associated to the largest eigenvalue of A . Being any eigenvector defined only minus a multiplicative constant we orient the eigenvector in the positive quadrant (whatever the sign obtained from the diagonalization). Additionally, an exponential damping factor with a length parameter λ can be introduced to Eq. 3, by defining A as:

$$A_{ij} = \begin{cases} 0 & \text{if } i = j \\ r_{\text{MI}}[x_i, x_j] \exp(-d_{ij}/\lambda) & \text{if } i \neq j. \end{cases} \quad (4)$$

λ controls the locality of the correlations under consideration based on the average distance between residues d_{ij} .

Hence, using short enough values of λ will result in neglecting the correlation between residues that are far away from one another, revealing the effect of the locality in the allosteric pathway. On the other hand, by setting λ to a very large value, all correlations, including those between residues separated by long distances, will be retained and $A_{ij} = r_{\text{MI}}[x_i, x_j] \quad \forall i \neq j$. In the main text the results presented correspond to a value of $\lambda = 5$.

Because we are interested in analyzing how the information transmission is affected by the allosteric stimulator, we focus on the difference centrality values computed by as $\Delta c = c_{\text{PRFAR}} - c_{\text{APO}}$. The nodes with higher eigenvector difference centrality are those acting as the principal “channels” for momentum transmission across the protein.

We visualize the c_i coefficient relative to each amino-acid in the protein structure, coloring each node from blue (zero centrality) to red (maximum centrality). In all of the cases, we apply a renormalization of the centrality values such that each falls in the -1, +1 range, as:

$$c'_i = 2 \frac{c_i - \min(c)}{\max(c) - \min(c)} - 1 \quad i = 1, \text{number of nodes} \quad (5)$$

In the present study, we calculated generalized correlation coefficients based on mutual information and EC values independently on 100 ns apo and PRFAR-bound trajectories of yeast and bacterial IGPS and averaged over six and four replicas, respectively. As mentioned before, the trajectories used for bacterial IGPS are the same as in reference,¹ hence the EC values reported both in the main text (Figure 3A) and below (Figure S6) are the same as in reference¹⁵ whereas those relative to yeast IGPS are computed ex novo, following the same procedure, as described Section S1.

The protein-network can be used to determine the optimal pathways for the information transfer between two nodes, defined as the shortest paths connecting a specific pair of nodes. In this context, edge lengths, i.e. the internode distances in the graph, are defined using the $r_{\text{MI}}[x_i, x_j]$ coefficients according to $-\log(r_{\text{MI}}[x_i, x_j])$, implying that highly correlated pairs (featuring good communication) are close in distance in the graph.

In particular we applied the Dijkstra algorithm to calculate the shortest pathways between residues *f*A233-*f*A234-A523/G524-R528 and *h*C84-C83, where each set of residues belongs to a different domain of bacterial and yeast IGPS, respectively. Hence, the computed pathways are composed of residue-to-residue steps that optimize the overall correlation (i.e., the momentum transport) between residues *f*A223-*f*A224 (at the effector site) and *h*C84 (in the glutaminase active site) in *Tm*-IGPS, and similarly residues K334, A523, G524 and C83 in *His7*.

Principal Component Analysis

Principal Component Analysis (PCA)^{16,17} has been employed to capture the essential motions of the simulated systems. In PCA, the covariance matrix of the protein C α atoms is calculated and diagonalized to obtain a new set of coordinates (eigenvectors) to describe the system motions. Each eigenvector – also called Principal Component (PC) – is associated with an eigenvalue, which denotes how much each eigenvector is representative of the system dynamics.

To avoid translational artifacts, we set the center of mass of each frame at the origin, and rotate each frame to its optimally aligned orientation relative to the average structure - computed over all apo trajectories - which also has its center of mass at the origin. Next, we evaluate the covariances of the positional fluctuations of each system over the apo and PRFAR-bound trajectories obtained by concatenation of the independent apo and effector-bound replicas. Because the motion of side-chains is mostly independent of the essential dynamics of IGPS, we restrict the covariance to the backbone atoms only. Projecting the original (centered) data onto the eigenvectors results in the PCs, whose associated eigenvalue (variance) is indicative of the portion of motion that the eigenvector describes.

Together, the first two principal components relative to *Tm*-IGPS incorporate 44% and 33% of the total motion of the bacterial apo and PRFAR-bound trajectories, respectively (Figure S3-A), while the percentages become 42% and 44% for His7 (Figure S3-B). The contribution added by the third PC is much smaller hence we limited our analysis to the first two.

The interest in projecting the trajectory coordinates onto the PCs is that we can visualize the essential motions induced by effector-binding in yeast and bacterial IGPS on the protein structure, along the trajectory. The procedure is described below.

First, we project the original trajectory onto the first two PRFAR-minus-apo difference principal components (ΔPC) and visualize their motion (details in SI). The weights over the i^{th} principal component relative to a given trajectory are given as

$$w_i(t) = r(t) - \bar{r} \cdot PC_i, \quad (6)$$

where $r(t)$ is a vector containing the stacked cartesian coordinates of the selected group of atoms at time (t) and \bar{r} are the mean (stacked x,y,z) coordinates along a selected (apo) trajectory. PC_i is the i^{th} principal component, having dimension $(3n)$, with n = number of atoms selected. The resulting weight vectors $w(t)$ are $(3n)$ dimensional and the dimension of w is equivalent to that of each row/column of the covariance matrix, and will coincide with the length of the PCs. Then, the projected coordinates on PRFAR-minus-apo difference principal components (ΔPC) are

$$r_i(t) = w_i(t) \times \Delta PC_i + \bar{r}, \quad (7)$$

Here, the product of the weights $w_i(t)$ - computed at each timestep of the apo trajectory - with the $i - th$ difference eigenvector ΔPC_i accounts for the fluctuations around the mean on that axis (i.e., the fluctuations induced by PRFAR binding), so the projected trajectory $r_i(t)$ simply describes the effector-induced fluctuations added onto the mean positions \bar{r} .

Additional comments on generalized correlation coefficients, EC and PCA

With regard to the analysis reported in the main text reported for yeast IGPS, it is worth discussing more in depth the outcomes of the single replicas as compared to the average. This analysis supports the finding in the text and shows the relevance of the simulations.

MD simulations are inherently chaotic, hence two simulations started from similar inputs may end up in significantly different configurations, making it hard to verify whether the process under interest is actually captured within the dynamic trajectory. This is why running a single trajectory may not mean much and replicates are almost always required. Indeed, allowing for high variance in the simulations - as we do, for instance, using different homology models to construct representative initial states from which to start the dynamics, is paramount to ensure that the simulations capture the process of interest (in our case the allosteric events in the enzyme's dynamics). We calculated generalized correlation coefficients based on mutual information and covariances of atomic displacements independently on each 100 ns apo and PRFAR-bound trajectories of yeast and bacterial IGPS and averaged over six and four replicas, respectively. A standard way to verify that a set of simulations contains a statistically relevant ensemble is to check that different simulations show similar ensemble average properties. The more unconstrained is the motion of a system of interest the more likely it will be that different dynamics sample different states of the system. The trade-off between considering a "large enough" number of independent simulations that will reliably capture a process of interest,

without averaging out important fluctuations, is system dependent and requires careful case-by-case examination. These observations apply to the simulations described in this work. Correlations, covariances (and therefore all the metrics derived from these) are subject to changes depending on the dynamics. For *Tm*-IGPS the four 100 ns apo/PRFAR-bound replicas, based on which we calculated the average properties discussed in the main text, showed similar features (as discussed in the original publication¹). We find rather larger deviations in the yeast as compared to *Tm*-IGPS. However, the average picture -obtained as the average apo-minus-holo correlation profile computed across the different models (shown in Figure 2) - is representative of the allosteric process although the individual simulations present different EC and PCA profiles (as shown in Figure S5). Among the six apo and PRFAR-bound replicas the dynamics that encompasses most of the allosteric traits is labelled as *sim₁* in the figures reported below. Figures 3 and 4 in the main text are associated to the representative dynamics of *sim₁*.

Section S3. Supplementary figures and tables

Breakdown of the secondary structural elements of His7

Secondary Structural Element	Residue numbers	Label	length	Secondary Structural Element	Residue numbers	Label	length
Beta strand	3 – 7	$h\beta 1$	5	Helix	277 – 288	$f\alpha 1$	12
Helix	15 – 23	$h\alpha 1$	9	Beta strand	292 – 299	$f\beta 2$	8
Beta strand	27 – 33	$h\beta 2$	7	Helix	307 – 309		3
Helix	34 – 36		3	Helix	311 – 319	$f\alpha 2$	9
Helix	39 – 41		3	Turn	320 – 322		3
				Beta strand	327 – 332	$f\beta 3$	6
Beta strand	45 – 49	$h\beta 3$	5				
Helix	53 – 62	$h\alpha 2$	10	Helix	346 – 356	$f\alpha 3$	11
Helix	66 – 74	$h\alpha 2'$	9	Beta strand	359 – 363	$f\beta 4$	5
Beta strand	79 – 83	$h\beta 4$	5	Helix	365 – 376	$f\alpha 4$	12
Helix	84 – 87	$h\alpha 3$	4	Helix	386 – 394	$f\alpha 4'$	9
				Helix	396 – 398	$f\alpha 4''$	3
Beta strand	90 – 93	$h\beta 5$	4				
				Beta strand	399 – 403	$f\beta 5$	5
Beta strand	104 – 111	$h\beta 6$	8	Beta strand	405 – 412	$f\beta 5'$	8
Turn	114 – 116		3	Helix	413 – 415	$f\alpha 5$	3
Beta strand	119 – 125	$h\beta 7$	7	Beta strand	433 – 440	$f\beta 6X$	8
Beta strand	143 – 150	$h\beta 8$	8	Turn	441 – 444		4
Helix	155 – 163	$h\alpha 4X$	9	Beta strand	445 – 450	$f\beta 6$	6
Beta strand	167 – 173	$h\beta 9$	7	Helix	451 – 460	$f\alpha 6$	10
Beta strand	176 – 184	$h\beta 10$	9	Beta strand	465 – 468		4
Beta strand	187 – 193	$h\beta 11$	7	Helix	471 – 473		3
Helix	194 – 196		3	Turn	474 – 476		3
Helix	198 – 209	$h\alpha 4$	12	Helix	482 – 491	$f\alpha 6$	10
Helix	221 – 227	$h\alpha 4'$	7	Beta strand	496 – 498		3
Helix	232 – 235	$h\alpha 4''$	4	Helix	505 – 514	$f\alpha 7$	10
Beta strand	240 – 248	$f\beta 1$	9	Beta strand	518 – 523	$f\beta 8$	6
Beta strand	250 – 252		3	Helix	524 – 527	$f\alpha 8'$	4
Beta strand	254 – 257		4				

Table S1. Full topography of secondary structural elements of yeast IGPS from <https://www.uniprot.org/uniprot/P33734>

Sequence alignment

Using the jFATCAT rigid algorithm implemented in the RCSB PDB Comparison Tool Reference (<https://www.rcsb.org/alignment>), we aligned structurally the amino acid sequence of HisH (PDB entry: 3ZR4.C), HisF (3ZR4.D) and His7 (1OX5.A). The sequence alignments are reported in Table S2. The structures of HisF and His7 are aligned for residues hM1–hS197 and P5–Q215 and the sequences of HisF and His7 are aligned for residues fM1–fE251 and G238–D553. Despite a similarity of ~ 50-60% (see Alignment Summary Table), the alignments show good structural similarity with a RMSD of the C-alpha backbone atoms ~ 2 Å. The alignments of the 3D structures are also reported in Fig. 1 in the main text.

Alignment Summary	RMSD	Sequence Identity%	Sequence Similarity%	Length
HisF-His7	2.03	46	63	241
HisH-His7	1.93	30	52	192

HisH-His7	1	MRIGIISVGP	GNIMNLYRGVKRASENFED	VSIELVESF	---RNDLYDLLF	47	
	5	PVVHVIDVES	GNLQSLTNAIEHLG	-----YEVQLVKSPKDFN	ISGTSRLI	49	
	48	IPGVGHFGE	GMRLRENDLIDFVRKHVEDERYVVG	VCLGMQLLFEESEEA		97	
	50	LPGVGNYGHF	VDNLFNRGFEKPIREYIESGK	PIMGICVGLQALFAGSVES		99	
	98	PGVKGLSL	IEGNVVKLSR	--RLPHMGWNEVIF	---KDTFPN	---GYFFV	140
	100	PKSTGLNY	IDFKLSRFDSEKPVPEIGWNSCIPSEN	LFGLDPYKRYFV		149	
	141	HTYRAVC	-----EEE	---HVLGTTEYDGEIFPSAVRKGRILG	GFQFHPEK	181	
	150	HSFAAILN	SEKKKNLENDGW	KIAKAKYGSEEFIAAVNKN	NIFATQFHPEK	199	
	182	SSKIGRKL	LEKVIC			197	
	200	SGKAGLN	VLENFLKQQ			215	
HisF-His7	1	MLAKRIIAC	LDVKD	---GRVVKG	-GDPVELGKFYSEIGIDELVFLDITA	54	
	238	GLTRRIIAC	LDVRTNDQGD	LVVTKLGKPVQLAQKYYQQGADEV	TFLNIT	303	
	55	SVEKRKTML	ELVEKVAEQIDIPFTVGGGIHD	-----FETASELI		93	
	308	CPLKDTPL	MLEVLKQAAKT	VFVPLTVGGGIKD	IVDVGTKIPALEVASLYF	357	
	94	LRGADKVS	SINTAAVENP	-----SLITQIAQTFGSQAVVVAID		130	
	358	RSGADKVS	IGTDAVYAAEKYYELGNRGDGT	SPIETISKAYGAQAVVISVD		407	
	131	AKRVD	-----GEFMVFTYSGKKNTGILLRDW		156		
	408	PKRVVNSQ	ADTKNKVFETETYPGPNGEKY	CWYQCTIKGGRESRDLGVWEL		457	
	157	VVEVEKRG	AGEILLTSIDRDGTS	GYDTEMIRFVRPLTTLPIIASGGAGK		206	
	458	TRACEALG	AGEILLNCIDKGSNSGYDLELIEHV	KDAVKIPVIASSGAGV		507	
207	MEHFLEAFL	-AGADAALAASVPHFREIDVRELKEYLKKHGVNVRLE		251			
508	PEHFEEAFLK	TRADACLGAGMFHRGEFTVNDVKEYLLEHGLKVRMD		553			

Table S2. Sequence alignment of His7 from *Saccharomyces cerevisiae* and HisH, HisF from *Thermotoga maritima*.

Correlation matrices from Elastic Network Model

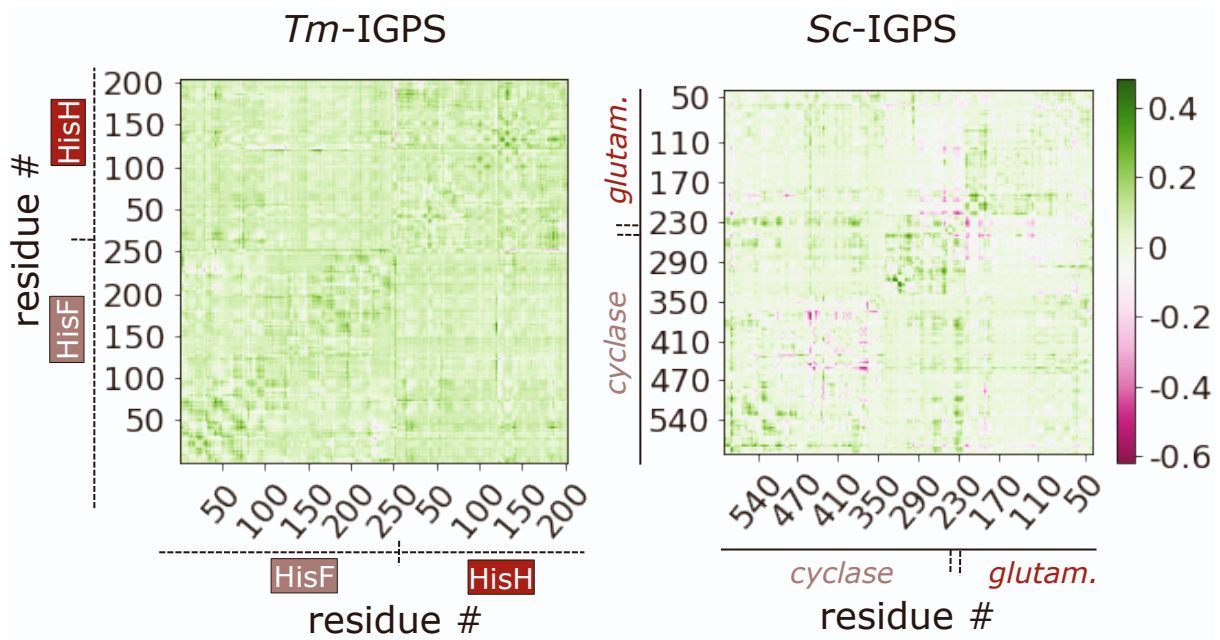


Figure S1. Difference of theoretical cross-correlation matrix between the holoenzyme and the apoenzyme in *T. maritima* (right) and *S. cerevisiae* (left). Cross-correlation matrices were computed with a Gaussian Network Model¹⁸ using the pre-equilibrated structures of model 1 for apo and PRFAR-bound of *T. maritima* and *S. cerevisiae*. Kirchoff matrices build with a cutoff of 10 Å and a spring constant of 1. Only the first 20 modes were taken into account in the computation.

Correlation matrices

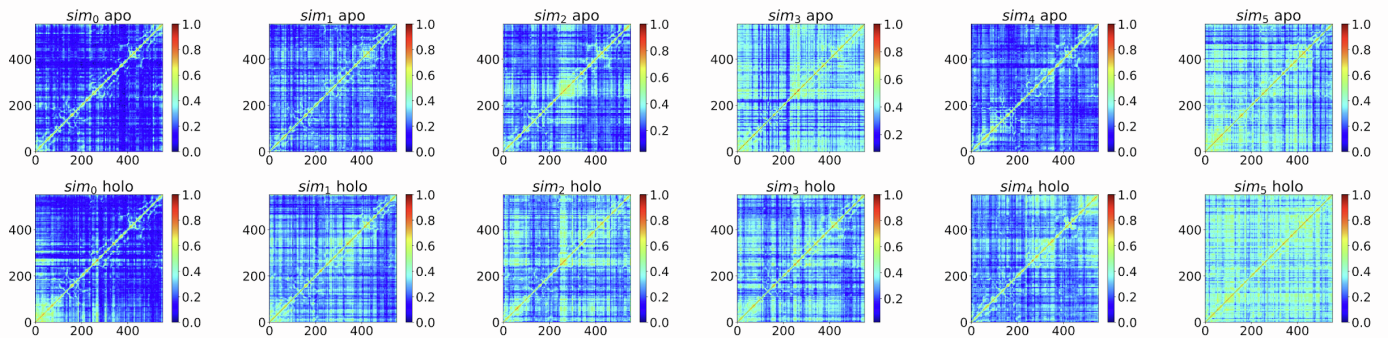


Figure S2. Generalized correlation coefficient matrices computed over six 100 ns replicas of simulated dynamics of apo and PRFAR-bound His7.

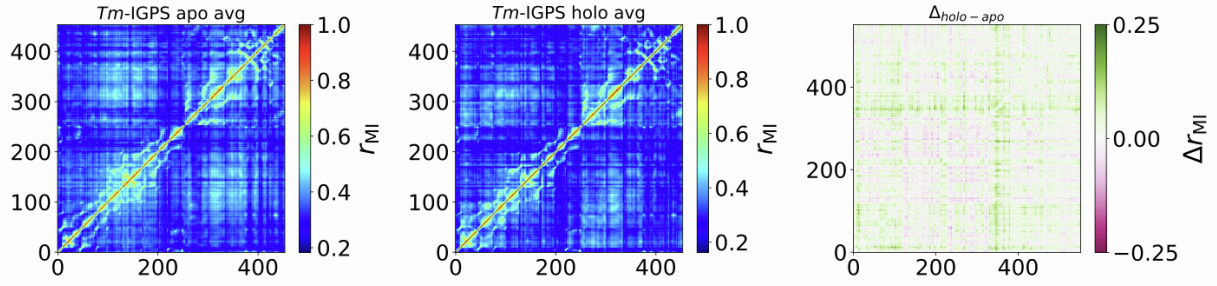


Figure S3. Left to right: average generalized correlation coefficient matrices over over apo, PRFAR-bound trajectories and difference (PRFAR-bound-minus-apo) computed over the six 100 ns replicas of simulated dynamics of *Tm*-IGPS.

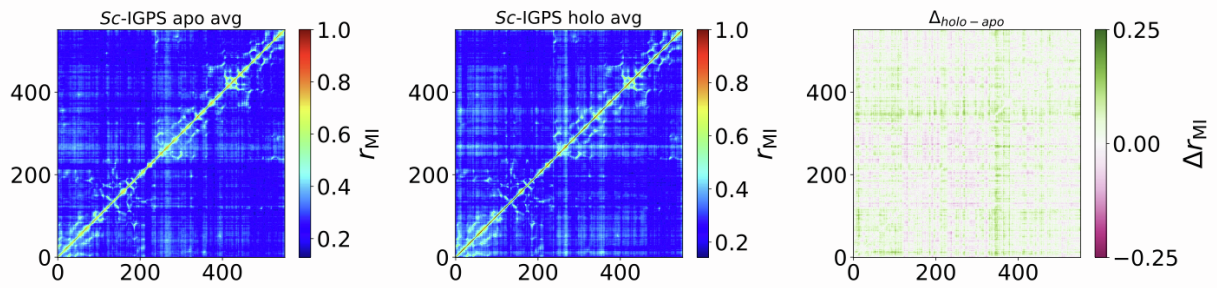


Figure S4. Left to right: average generalized correlation coefficient matrices over over apo, PRFAR-bound trajectories and difference (PRFAR-bound-minus-apo) computed over the six 100 ns replicas of simulated dynamics of His7.

Eigenvector centrality

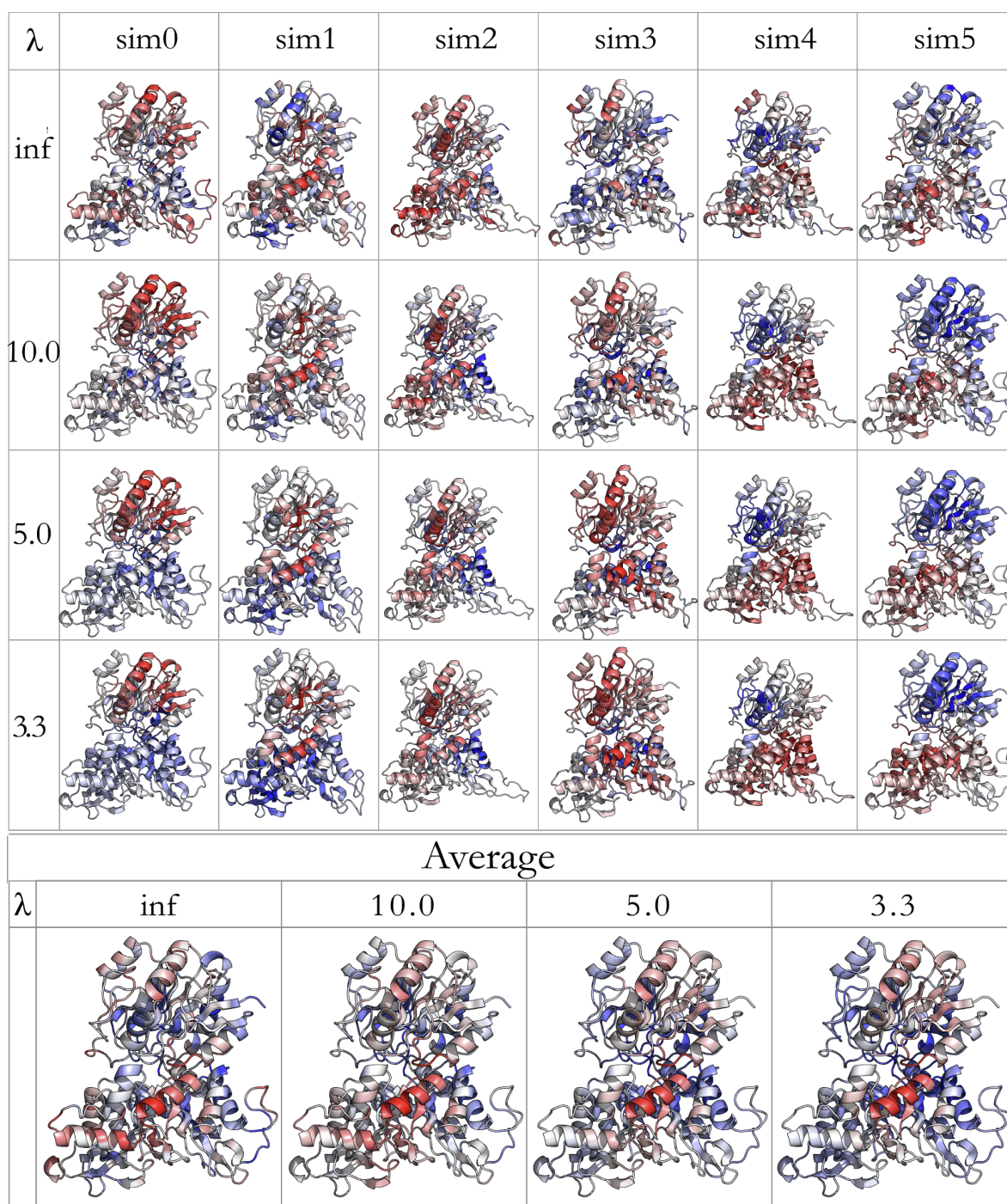


Figure S5. Centrality differences (PRFAR-bound-minus-apo) projected onto the apo structure of *Sc*-IGPS, computed different values of λ . Regions in red and blue correspond to gains and loss of centrality. To note, the EC values relative to sim_1 recover most of the allosteric traits as it can be inferred by the similarity of the centrality pictures showing the averaged values over the six independent replicas (last row).

Analysis of first and second principal components in apo and PRFAR-bound yeast and bacterial IGPS

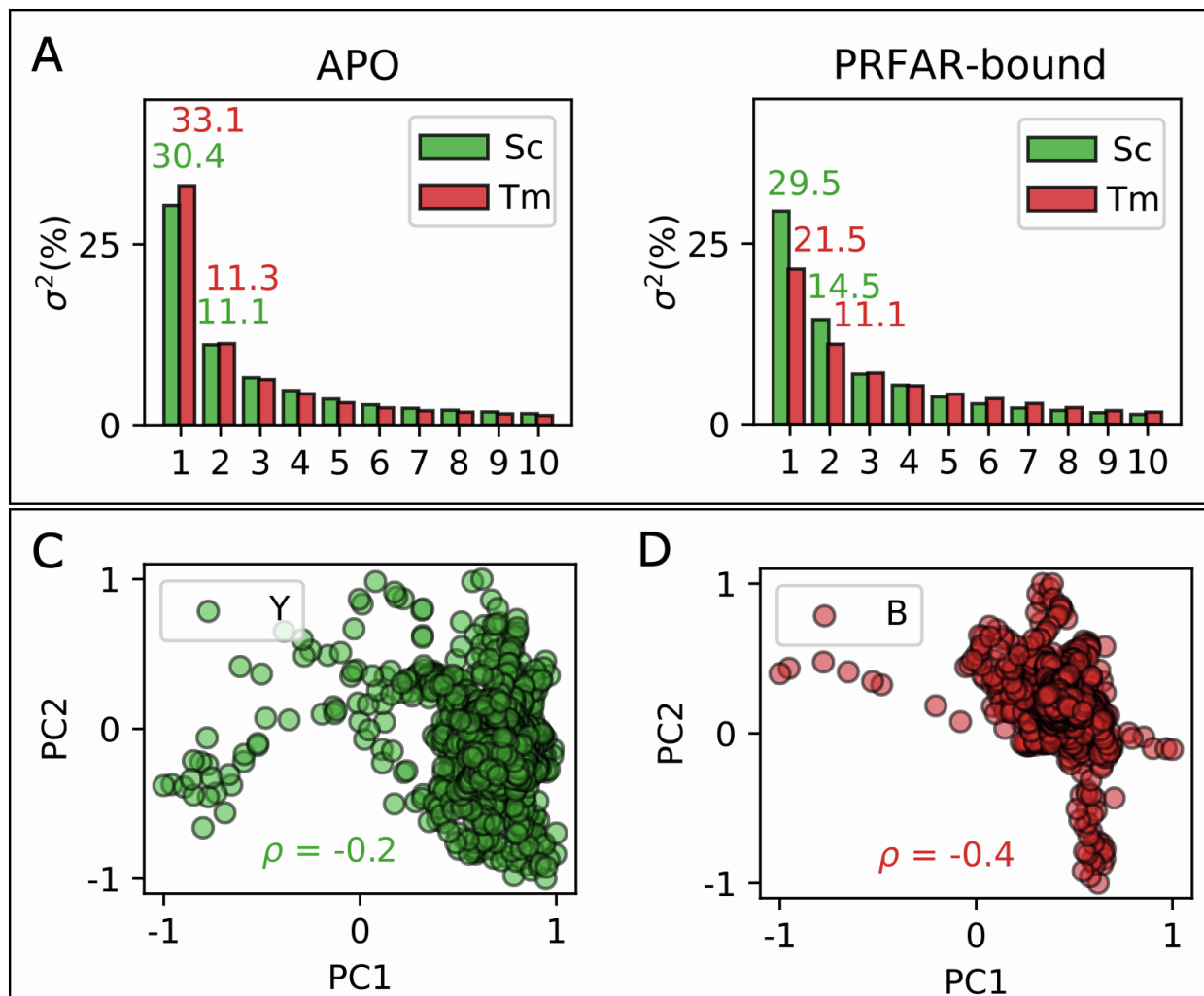


Figure S6. (A) and (B) Cumulative variance of the first and second PCs computed for the apo (A) and PRFAR-bound (B) trajectories, showing the comparison between *Tm*-IGPS (red) and *Sc*-IGPS (green). (C) and (D) show the correlation between first and second principal components computed along the trajectories of yeast and bacterial IGPS. (C) In *Sc*-IGPS PC1 and PC2 are poorly correlated, confirming that they account for distinct motions, while the higher correlation shown in (D) suggests that PC1 and PC2 in *Tm*-IGPS have some degree of overlap.

Essential motions of the trajectory through principal component analysis

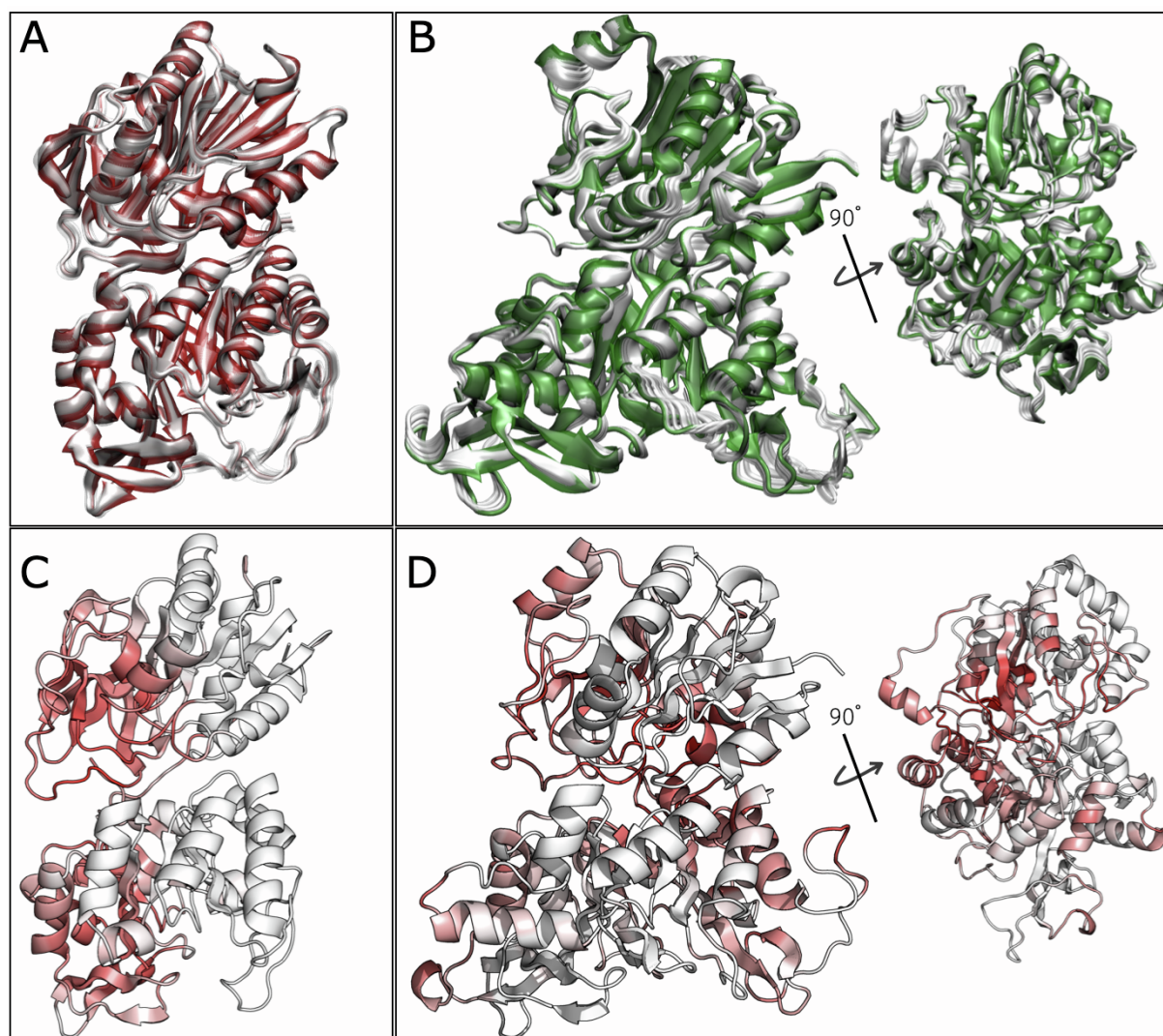


Figure S7. Projection of the original apo trajectory of *Tm*-IGPS (A) and *Sc*-IGPS onto the difference (PRFAR-minus-apo) second principal components computed along the yeast (ΔPC_1^Y) and bacterial (ΔPC_1^B) IGPS trajectories, as discussed in Section S2. This figure provides a zoom in of Figure 3E and 3F in the main text for better visualization of the dynamic low-vibrational motions of the two enzymes. Panels C and D show positive variations in the EC coefficients due to the long-range component of correlations in *Tm*-IGPS and His7 respectively. The largest increase in the long-range centrality coefficients upon PRFAR binding interests different regions in *Tm* and *Sc*. The values in *Tm* are consistent with the presence of an interdomain “breathing” motion shown with black dashed black lines and forming an angle ϕ). In *Sc*, the largest structural (long-range) rearrangements are associated with the motion of the connector and of the secondary structure elements *fa8*, *fa1*, *ha4*, *hβ9*, marked in the figure. Long-range EC centralities match description of low vibrational motions provided by the analysis of first principal components.

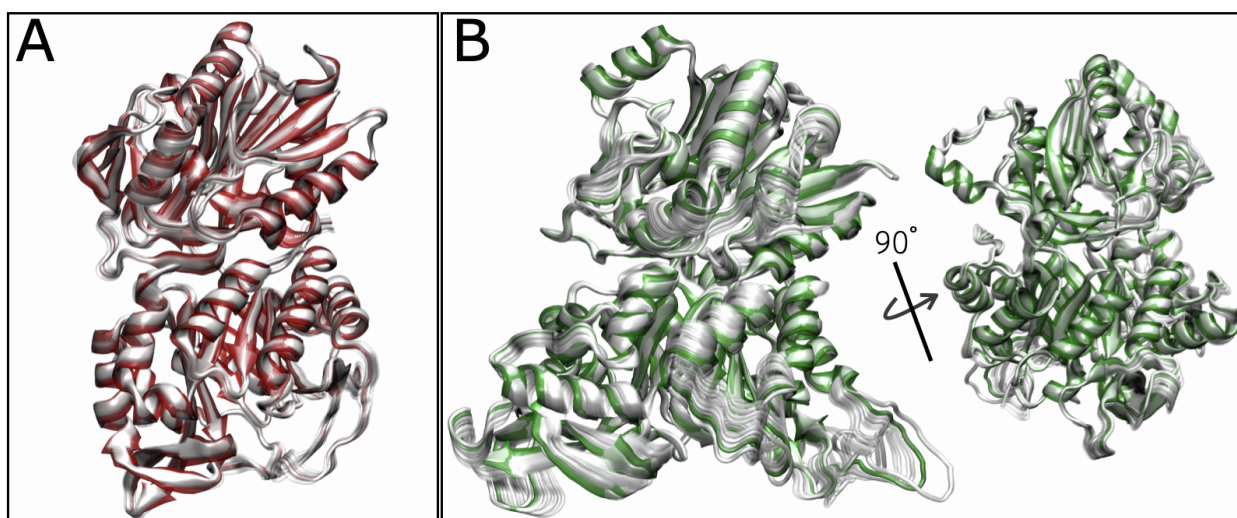


Figure 8. Projection of the apo trajectory of *Tm*-IGPS (A) and *Sc*-IGPS onto the difference (PRFAR-minus-apo) second principal components computed along the yeast (ΔPC_2^Y) and bacterial (ΔPC_2^B) IGPS trajectories, as discussed in Section S2.

Role of loop1 in *Tm*-IGPS and *Sc*-IGPS

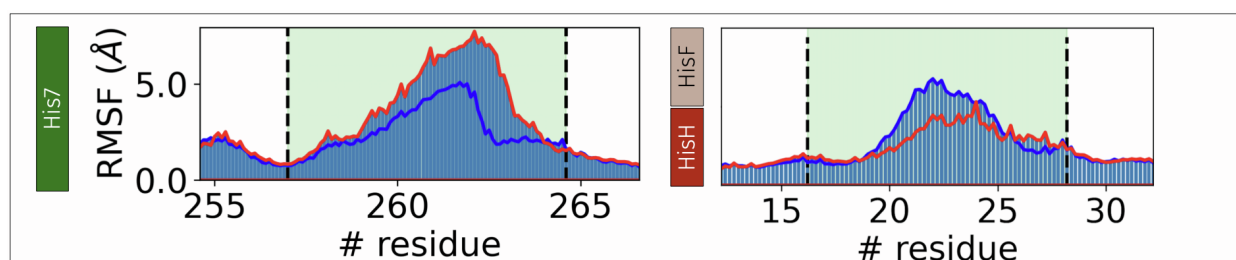


Figure S9. The motion of loop1 has a behavior in His7 and *Tm*-IGPS, upon binding of the effector. While in His7 the binding induces an increased mobility of loop1, in *Tm*-IGPS, binding of PRFAR constrains the motion of loop1. This behavior is consistent with the different role of the loop in the two systems, as suggested in the main text.

Distance profiles K334-D335 profile across six 100 ns replicas of simulated dynamics of His7.

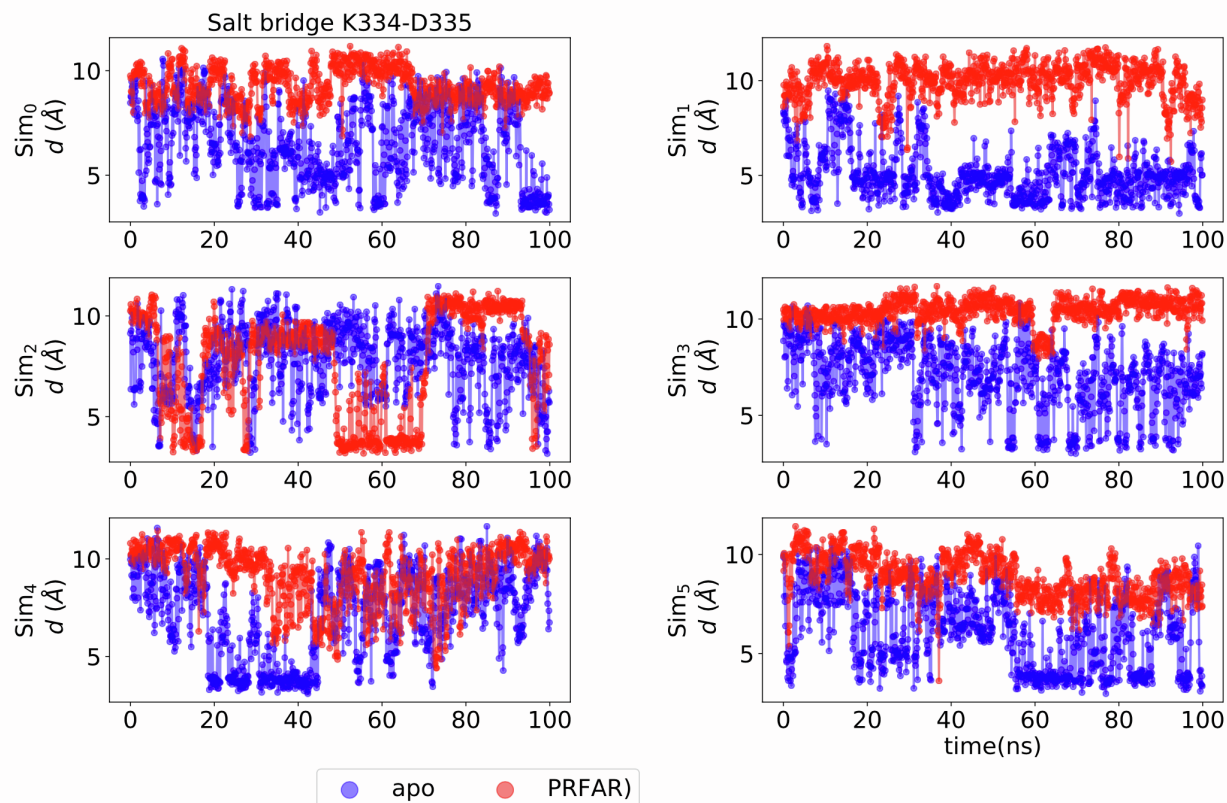


Figure S10. The K334-D335 salt bridge is mostly present in the APO simulation and breaks upon PRFAR binding as the effector interacts with residue K334. The dissolution of the K334-D335 is particularly evident in Sim₁, in accordance with our observation of Sim₁ best capturing the allosteric process.

Distance profiles of the hydrophobic cluster across six 100 ns replicas of simulated dynamics of His7.

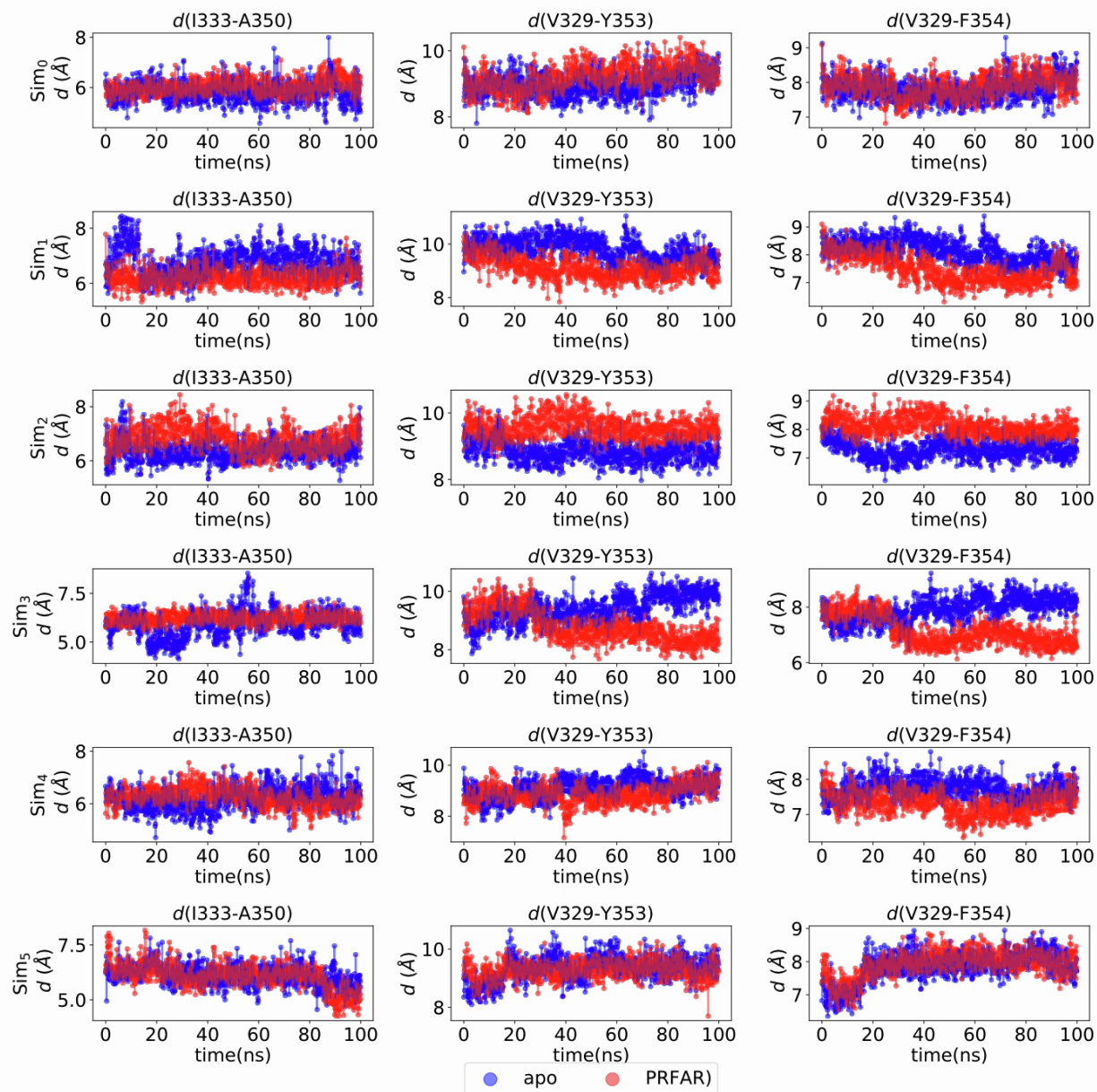


Figure S11. Distance profiles of I333-A350, V329-Y353, V329-F354 computed across six apo (blue) and PRFAR-bound (red) 100 ns replicas of simulated dynamics of His7.

Hinge Motion

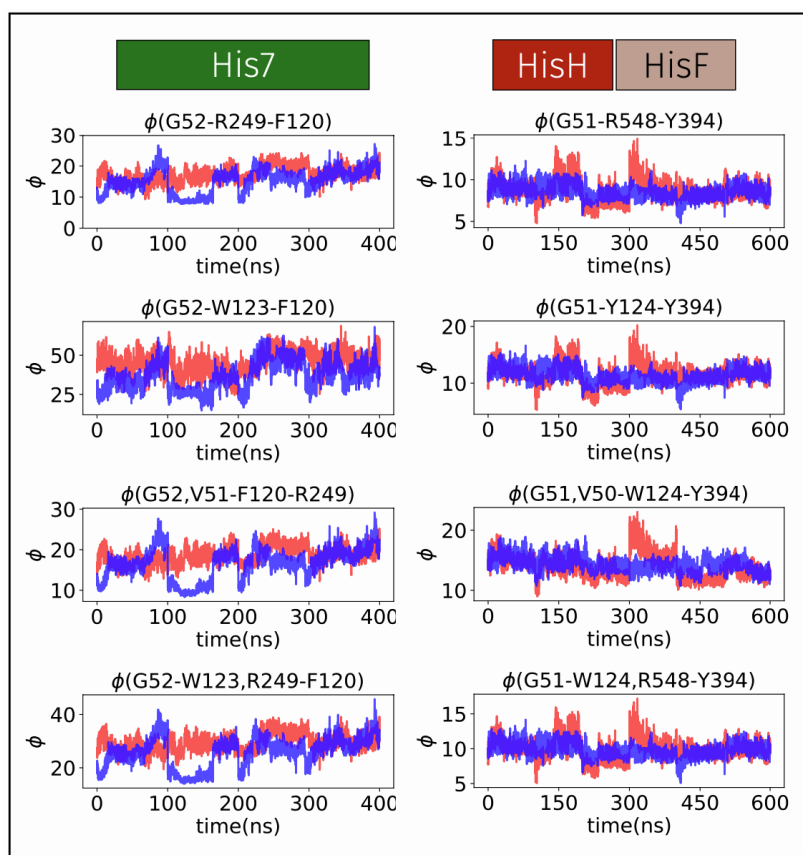


Figure S12. Hinge motion profile six apo (blue) and PRFAR-bound (red) over the concatenated dynamics of *Tm*-IGPS(left) and Hisy (right), measured through the angle ϕ defined using different residues. The standard definition of ϕ (G51-W124-Y394) used in other publications is included. The oscillation mostly ranges between 10 and 20 degrees with no significant changes to the PRFAR bound profiles as compared to the apo.

The distribution of ϕ supports our hypothesis of a reduced importance of the hinge motion in the allosteric process of His7, as compared to that of *Tm*-IGPS.

Relevant salt-bridge interactions in *Tm*- and *Sc*-IGPS

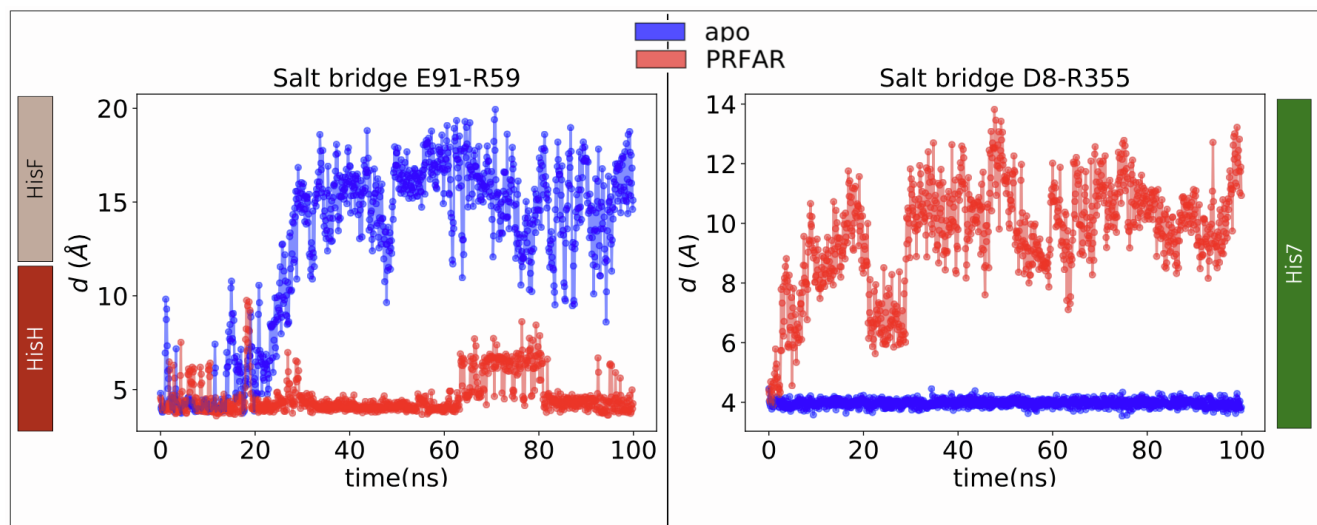


Figure S13. Profile of the E91-R59 (*Tm*-IGPS) and D8-R355 (His7) salt-bridge interactions along 100 ns of apo and PRFAR bound states of dynamics. As suggested by the large modification of the profiles upon effector binding these interactions are crucial in the signaling mechanism of bacterial and yeast IGPS, respectively.

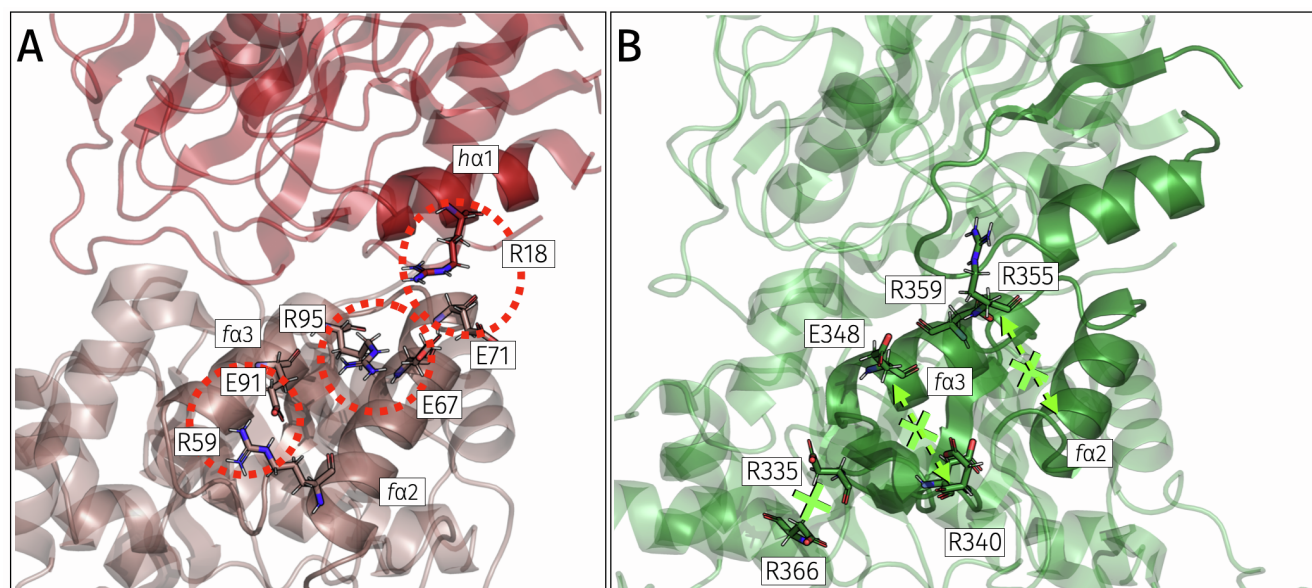


Figure S14. (A) Residues that participate in the network of salt-bridges at sideR of *Tm*-IGPS, induced by PRFAR binding. In *Sc*-IGPS (B), there are no corresponding surface-charged residues that can allow communication between the two active sites through the coupling of fα3-fα2 similar to that of *Tm*-IGPS. Instead, in *Sc*-IGPS the signal travels across fα3 and fβ3.

Closeup view of the glutaminase active site and PGVG sequences in yeast and bacterial IGPS

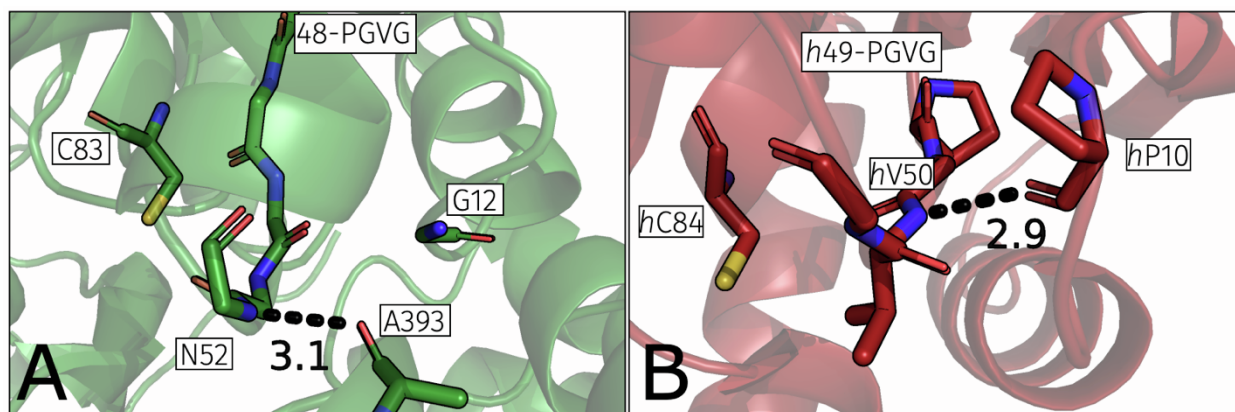


Figure S15. (A) The hydrogen bond between A393-N52 is mostly present in the apo structure and loosens upon PRFAR binding. Unlike in *Sc* apo, the *h*48-PGVG sequence *Tm*-IGPS is not within hydrogen-bonding distance to the glutaminase domain. (B) In *Tm* apo, residue *h*V50 is tightly bound in a hydrogen bond with *h*P10, while the corresponding distance varies significantly across the dynamics of *Sc*-apo, suggesting a different cross communication between the Ω -loop and PGVG in the two enzymes. The hydrogen bond between *h*V50 and *h*P10 in *Tm*-IGPS dissolves in the presence of PRFAR. This bond rupture marks the transition between the inactive state (apo) and the pro-active state.

Time-evolution of hydrogen bond at the interface of His7

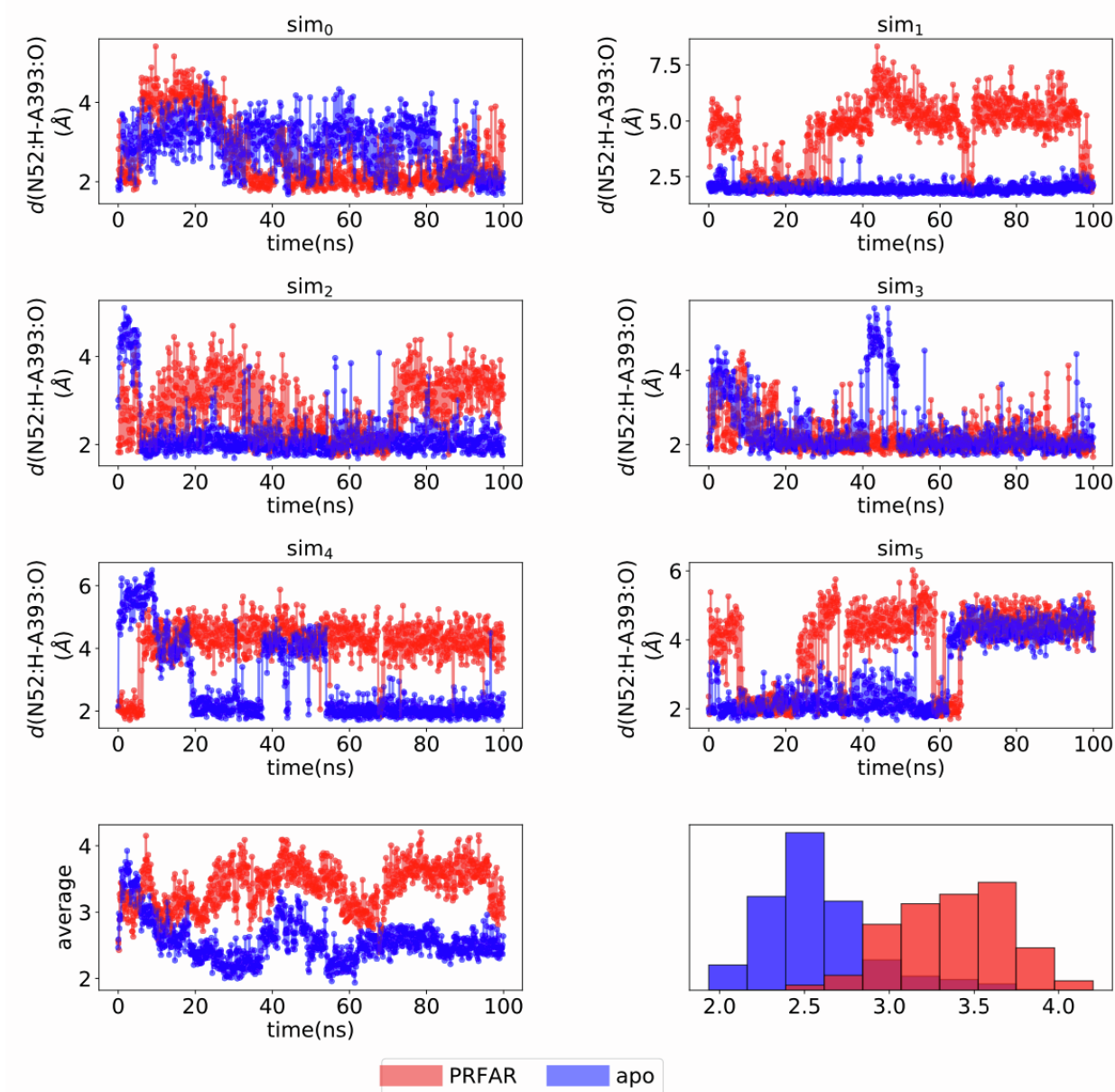


Figure S16. Distance profile of the N52:H-A393:O bond computed along the six replicas of the apo (blue) and PRFAR-bound (red) simulated dynamics of His7. The last row shows the mean values averaged on the different replicas, as well as a histogram representation of the same distribution. At the interface, the hydrogen bond between the backbone atoms of A393 and N52 elongates in the presence of the effector.

Movie legends

A way to investigate the essential motions of the trajectory is to project the original trajectory onto each of the principal components, to visualize the motion of the principal component. The resulting trajectories computed by projecting the original coordinates onto the first difference (ΔPC_1) and second (ΔPC_2) principal components are shown in the enclosed jupyter-notebook:

A way to investigate the essential motions of the trajectory is to project the original trajectory onto each of the principal components, to visualize the motion of the principal component. Instead of including all atoms of the trajectories one can focus on selected atom groups, for instance the backbone atoms.

The principal component analysis presented in this work is performed by selecting the backbone atoms of the apo and PRFAR-bound trajectories of either *Tm*-IGPS or *Sc*-IGPS.

The product of the weights $w_i(t)$ for the i^{th} principal component relative to the apo trajectory with the difference eigenvector $\Delta PC_i = PC_i^{PRFAR-bound} - PC_i^{APO}$ describes the fluctuations around the mean on that axis, induced by PRFAR binding,

$$r_i(t) = PC_i(t) \times u_i + r$$

The projected trajectory $r_i(t)$ is simply the fluctuations added onto the mean positions. (See description in section *Principal Component Analysis* at page 2).

The resulting trajectories computed by projecting the original coordinates onto the difference first (ΔPC_1) and second (ΔPC_2) principal components are shown in the three videos enclosed to the Supplementary material.

- 1- Video named DELTA_PC1.mov shows projected trajectories of *Tm*-IGPS and *Sc*-IGPS along the **first difference principal component** ΔPC_1 (PRFAR-bound-minus-apo).
- 2- Video named DELTA_PC1rot.mov shows projected trajectories of *Tm*-IGPS and *Sc*-IGPS along the **first difference principal component** ΔPC_1 (PRFAR-bound-minus-apo), where the *Sc*-IGPS is shown in a rotated view with respect to the Video 1, to highlight the motion of the connector.
- 3- Video named DELTA_PC2.mov shows projected trajectories of *Tm*-IGPS and *Sc*-IGPS along the **second difference principal component** ΔPC_2 (PRFAR-bound-minus-apo).

The videos altogether show the difference in the dynamics of the low-vibrational motions of the two enzymes. While *Tm*-IGPS adopts a hinge-like breathing motion that modifies the opening of the interface between the two subunits, *Sc*-IGPS displays a rather different spring-like motion located at the core of the enzyme, coupled with large variations at the connector site and loop1.

Supporting References

- (1) Rivalta, I.; Sultan, M. M.; Lee, N.-S.; Manley, G. A.; Loria, J. P.; Batista, V. S. Allosteric Pathways in Imidazole Glycerol Phosphate Synthase. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (22), E1428–E1436.
- (2) Chaudhuri, B. N.; Lange, S. C.; Myers, R. S.; Davisson, V. J.; Smith, J. L. Toward Understanding the Mechanism of the Complex Cyclization Reaction Catalyzed by Imidazole Glycerolphosphate Synthase: Crystal Structures of a Ternary Complex and the Free Enzyme. *Biochemistry* **2003**, *42* (23), 7003–7012.
- (3) Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Protein Sci.* **2016**, *86*, 2.9.1–2.9.37.
- (4) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* **2018**, *46* (W1), W296–W303.
- (5) Wang, J.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.* **2000**, *21* (12), 1049–1074.
- (6) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *Journal of Computational Chemistry*. 2004, pp 1157–1174.
- (7) Citations for Amber <https://ambermd.org/CiteAmber.php> (accessed Jan 7, 2021).
- (8) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092.
- (9) Grubmüller, H.; Heller, H.; Windemuth, A.; Schulten, K. Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-Range Interactions. *Mol. Simul.* **1991**, *6* (1-3), 121–142.
- (10) Lange, O. F.; Grubmüller, H. Generalized Correlation for Biomolecular Dynamics. *Proteins* **2006**, *62* (4), 1053–1061.
- (11) Cover, T. M.; Thomas, J. A. *Elements of Information Theory*; John Wiley & Sons, 2012.
- (12) Yang, Z.; Algesheimer, R.; Tessone, C. J. A Comparative Analysis of Community Detection Algorithms on Artificial Networks. *Sci. Rep.* **2016**, *6*, 30750.
- (13) Rivalta, I.; Batista, V. S. Community Network Analysis of Allosteric Proteins. *Methods Mol. Biol.* **2021**, *2253*, 137–151.
- (14) Oldham, S.; Fulcher, B.; Parkes, L.; Arnatkevic Iūtė, A.; Suo, C.; Fornito, A. Consistency and Differences between Centrality Measures across Distinct Classes of Networks. *PLoS One* **2019**, *14* (7), e0220061.
- (15) Negre, C. F. A.; Morzan, U. N.; Hendrickson, H. P.; Pal, R.; Lisi, G. P.; Loria, J. P.; Rivalta, I.; Ho, J.; Batista, V. S. Eigenvector Centrality for Characterization of Protein Allosteric Pathways. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (52), E12201–E12208.
- (16) Lange, O. F.; Grubmüller, H. Full Correlation Analysis of Conformational Protein Dynamics. *Proteins* **2008**, *70* (4), 1294–1312.
- (17) David, C. C.; Jacobs, D. J. Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins. *Methods Mol. Biol.* **2014**, *1084*, 193–226.
- (18) Atilgan, A. R. *et al.* Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophys J* **80**, 505–515 (2001).