HHS Public Access

Author manuscript

J Am Chem Soc. Author manuscript; available in PMC 2024 January 08.

Published in final edited form as:

J Am Chem Soc. 2023 June 21; 145(24): 13347–13356. doi:10.1021/jacs.3c03425.

Solution structure ensembles of the open and closed forms of the ~130 kDa Enzyme I via AlphaFold modeling, Coarse Grained simulations and NMR

Sergey L. Sedinkin[1,‡], **Daniel Burns**[2,‡], **Divyanshu Shukla**[1], **Davit A. Potoyan**[1,2,*], **Vincenzo Venditti**[1,2,*]

[1]Department of Chemistry, Iowa State University, Ames, Iowa 50011, USA.

[2]Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, Iowa 50011, USA.

Abstract

Large scale interdomain rearrangements are essential to protein function, governing the activity of large enzymes and molecular machineries. Yet obtaining an atomic-resolution understanding of how the relative domain positioning is affected by external stimuli is a hard task in modern structural biology. Here we show that combining structural modeling by AlphaFold2 with coarse-grained molecular dynamics simulations and NMR residual dipolar coupling data is sufficient to characterize the spatial domain organization of bacterial enzyme I (EI), a ~130 kDa multidomain oligomeric protein that undergoes large scale conformational changes during its catalytic cycle. In particular, we solve conformational ensembles for EI at two different experimental temperatures and demonstrate that a lower temperature favors sampling of the catalytically competent closed state of the enzyme. These results suggest a role for conformational entropy in activation of EI and demonstrate the ability of our protocol to detect and characterize the effect of external stimuli (such as mutations, ligand binding, and post-translational modifications) on the interdomain organization of multidomain proteins. We expect the ensemble refinement protocol described here to be easily transferrable to the investigation of the structure and dynamics of other uncharted multidomain systems, and have assembled a Google Colab page (https://potoyangroup.github.io/Seq2Ensemble/) to facilitate implementation of the presented methodology elsewhere.

Graphical Abstract

*Address correspondence to: Vincenzo Venditti, Department of Chemistry, Iowa State University, Hach Hall, 2438 Pammel Drive, Ames, IA 50011, USA. venditti@iastate.edu; Tel: 515-294-1044; Fax: 515-294-7550; Davit A. Potoyan, Department of Chemistry, Iowa State University, Gilman Hall, 2415 Osborn Drive, Ames, IA 50011, USA. potoyan@iastate.edu; Tel: 515-294-9971; Fax: 515-294-7550.

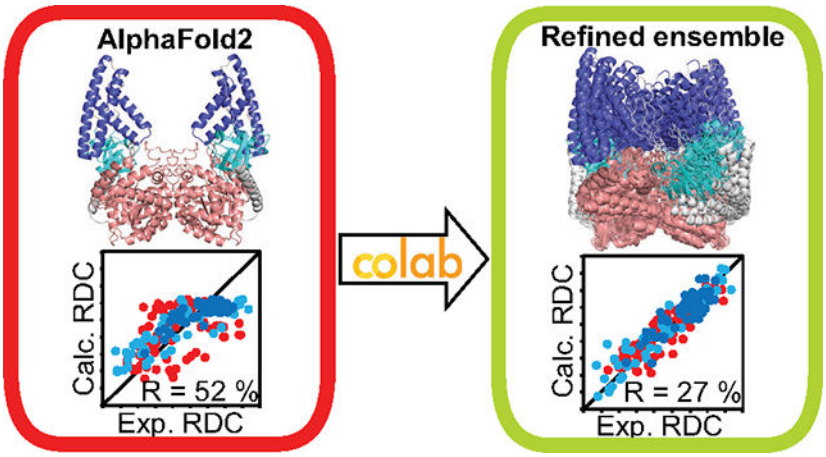
‡Equal contribution

Conflicts of Interest

The authors declare no conflict of interest.

Supporting Information

AlphaFold modeling of *T. tengcongensis* EI; comparison of the AlphaFold model with the available experimental structures; RDC datasets and distribution of the RDCs in the EI structure; 800 MHz ^{1}H-^{15}N TROSY spectra of EI; convergence of the cgMD/RDC ensemble calculation; comparison between the cgMD/RDC ensemble and the AF2 model; interdomain electrostatic interactions in the conformational ensembles; tutorial for the Seq2Ens pipeline.



Keywords

Molecular Dynamics; multidomain proteins; conformational change; AWSEM; Residual Dipolar Coupling; Colab

Introduction

Protein function depends on orchestrated rearrangements of protein domains mediated by external perturbations such as ligand binding,[1–3] post-translational modifications,[4] changes in ionic strength,[5] salt concentration, and temperature.[6] Therefore, studying how the relative positioning of protein domains is affected by such external stimuli in solution is a grand challenge in modern structural biology. Recently, structure prediction algorithms have made stunning breakthroughs, and, with the development of the DeepMind AlphaFold2 (AF2)[7] and the RoseTTAFold[8] systems, the atomic-resolution structures of almost the entire human proteome have been predicted and made available.[9] Although several studies have appeared in the literature demonstrating the ability of such predicting tools to accurately capture tertiary and quaternary folds in relatively complex proteins and assemblies,[10–13] the interdomain architecture is often misrepresented by AF2 and RoseTTAFold, especially when modeling dynamic systems that function via large-scale structural rearrangements.[13–14]

In this manuscript we explore the possibility of combining AF2 predictions with coarse grained molecular dynamics (cgMD) simulations, and sparse Residual Dipolar Coupling (RDC) NMR data to obtain an accurate description of the interdomain organization in EI from *Thermoanaerobacter tengcongensis*, a large, multidomain protein whose enzymatic cycle is finely tuned by large-scale interdomain rearrangements mediated by substrate and cofactor binding.[15–16] EI is a ~130 kDa homodimer comprising two functionally distinct domains (Figure 1). The C-terminal domain (EIC, residue 261–573) is the dimerization domain and contains the binding site for the substrate phosphoenolpyruvate (PEP).[15, 17] The N-terminal domain (EIN, residues 1–230) is divided in two subdomains, the EIN^{α} (residues 33–143) and $EIN^{\alpha\beta}$ (residues 1–20 and 148–230), that contain the phosphorylation site (His 189 on $EIN^{\alpha\beta}$) and the binding site for the phosphocarrier protein HPr (on EIN^{α}).[15, 18] EIN^{α} and $EIN^{\alpha\beta}$ are connected by two flexible loops.[15] $EIN^{\alpha\beta}$ and EIC are connected by

a 30-residue long helical linker.[15] Although the structure of the *T. tengcongensis* EI is still unsolved, structural studies on the *Escherichia coli* homologue have shown that the enzyme undergoes large interdomain rearrangements upon substrate binding.[2, 19–21] In particular, apo EI adopts an open conformation that exposes the active site to the incoming substrate. Holo EI adopts a closed conformation that presents the side chain of His 189 to PEP and shields the active site from water (Figure 1).

Here, we use AF2 to predict the structure of the full-length thermophilic EI from primary sequence. Using NMR RDC data, we show that AF2 returns accurate predictions for the structures of the individual EI domains but largely misrepresents their spatial organization within the full-length enzyme. We then combine cgMD simulations with the experimental RDCs to refine the AF2 model into a conformational ensemble that satisfies the experimental data. Our results demonstrate that apo EI does not adopt a rigid fold in solution but is rather in equilibrium among multiple conformations with diverse interdomain orientations. Interestingly, we find that the interdomain organization of EI is extremely sensitive to the experimental temperature, with low temperatures favoring the sampling of closed states similar to the catalytically competent holo structure. The fact that our protocol is capable of detecting and characterizing the temperature-induced conformational changes in EI indicates that combining AF2 predictions with cgMD simulations and RDCs is a promising approach to study the effect of external perturbations, such as mutations, ligand binding, and post-translational modifications, on the structure of large and dynamic multidomain proteins.

Results

AlphaFold modeling of full-length EI.

While the atomic-resolution structures of the isolated EIN[22] and EIC dimer[23] from *T. tengcongensis* were determined by solution NMR and crystallography, respectively, the overall fold of the full-length enzyme remains unsolved. Here, we report three structural models of the full-length EI predicted using the DeepMind AF2 algorithm. The generated models are essentially identical (heavy atom root mean square deviation, r.m.s.d., = 1.6 Å) and predicted with an average predicted local distance difference test (pLDDT) of 88 ± 12 % (Supplementary Figure S1). The top-ranked AF2 model is in very good agreement with the experimental tertiary fold of the isolated EI domains (Cα r.m.s.d. of 0.6, 1.6, and 0.4 Å for the EIN^{α}, $EIN^{\alpha\beta}$, and EIC, respectively), as well as with the quaternary structure of the EIC dimer (Cα r.m.s.d. of 0.4 Å) (Supplementary Figure S2). On the other hand, the predicted structure does not recapitulate the interdomain organization displayed by the experimental structures of the open,[19] closed,[20] and partially close[21] full-length EI from *E. coli* (Supplementary Figure S2), nor does it agree with any of the atomic resolution structures of EI from different bacterial sources.

To test the accuracy of the AF2 prediction, we measured backbone amide RDC ($^1D_{NH}$) data for well resolved ^{1}H-^{15}N NMR cross-peaks on samples of weakly aligned EI in a dilute liquid crystalline medium of phage *pf1*[24] at 40 and 70 °C. The $^1D_{NH}$ data inform on the orientation of the N-H bond vectors with respect to an external alignment tensor.[25] Therefore, RDCs are fine reporters of protein tertiary and quaternary fold, and have been

often employed to validate and refine structural models.[26] In total, we have measured 145 and 223 $^{1}D_{NH}$ values for EI at 40 and 70 °C, respectively. However, to avoid structural noise from flexible regions, only the backbone amides from secondary structure elements were included in the following analysis (43, 25, and 47 RDCs at 40 °C, and 60, 44, and 76 RDCs at 70 °C for EIN$^{\alpha}$, EIN$^{\alpha\beta}$, and EIC, respectively) (Supplementary Figure S3 and Supplementary Table S1). Singular value decomposition (SVD) fitting of the RDCs measured for EIN$^{\alpha}$, EIN$^{\alpha\beta}$, and EIC against the coordinates of the isolated domains extracted from the AF2 model returns R-factors $\lesssim$ 36% (Figure 2). The agreement between experimental and back-calculated data confirms that the AF2 model recapitulates the tertiary fold of the EIN$^{\alpha}$, EIN$^{\alpha\beta}$, and EIC domains in solution, as well as the quaternary fold of the EIC dimer. By way of contrast, SVD fitting of the full $^{1}D_{NH}$ dataset to the structural model of full-length EI results in R-factors > 50% (Figure 2), indicating that EI must adopt a conformation or ensemble of conformations that differ substantially from the AF2 model.

Finally, it is also worth noticing that the SVD fits of the 40 °C data yield larger R-factors than the corresponding fits of the 70 °C RDCs (Figure 2). This finding is likely the result of the slower molecular tumbling of EI at low temperature that negatively impacts the signal-to-noise ratio of the NMR spectra (Supplementary Table S4) and translates into an increased experimental error.

^{15}N NMR relaxation.

To investigate the cause of the inconsistency of the RDC data with the AF2 structural model we have measured the ^{15}N longitudinal (R_1) and transverse (R_2) relaxation rates on samples of ^{2}H,^{15}N-labeled full-length EI. The backbone amide ^{15}N-R_2/R_1 values are directly proportional to the rotational tumbling of the amide bond vectors.[27] Therefore, if EI has a rigid solution structure in which the individual domains are held in position by strong interdomain interactions, a plot of the relaxation data versus the residue index will return a constant distribution around a ^{15}N-R_2/R_1 value proportional to the rotational correlation time (τ_c) of the protein. On the other hand, if the EIN$^{\alpha}$, EIN$^{\alpha\beta}$, and EIC domains tumble independently or semi-independently (i.e., the domain-domain contacts are not strong enough to keep the domains in a fixed relative orientation) the relaxation data of the backbone amides belonging to different structural domains will cluster around different ^{15}N-R_2/R_1 values, each proportional to the τ_c of a specific domain. Analysis of the ^{15}N-R_2/R_1 data measured at 40 and 70 °C indicates that EI undergoes substantial interdomain dynamics at both experimental temperatures (Figure 3). Indeed, average ^{15}N-R_2/R_1 values of 51, 103, and 144 at 40 °C and of 44, 63, and 130 at 70 °C are obtained for EIN$^{\alpha}$, EIN$^{\alpha\beta}$, and EIC, respectively, indicating that the EI domains tumble with different τ_c's (Table 1). Interestingly, the τ_c's calculated for the EI domains from the relaxation data at 40 °C (17, 24, and 29 ns for EIN$^{\alpha}$, EIN$^{\alpha\beta}$, and EIC, respectively) and 70 °C (16, 19, and 28 ns for EIN$^{\alpha}$, EIN$^{\alpha\beta}$, and EIC, respectively) are substantially larger than the τ_c values predicted for globular proteins of the size of EIN$^{\alpha}$ (5 and 3 ns at 40 and 70 °C, respectively), EIN$^{\alpha\beta}$ (5 and 3 ns at 40 and 70 °C, respectively), and the EIC dimer (29 and 27 ns at 40 and 70 °C, respectively), but substantially lower than the τ_c predicted for the 127 kDa full-length EI dimer (53 and 30 ns at 40 and 70 °C, respectively) (Table 1). This observation indicates that, although the presence of both peptide linkers and non-covalent interdomain interactions

restrict the relative positioning of the EI domains, EIN^{α}, $EIN^{\alpha\beta}$, and EIC do not tumble as a rigid, single-body object in solution and, therefore, their RDC data cannot be simultaneously accounted for by a single 3D structure.

As a final note, the experimental τ_c of EIC at 40 °C matches the τ_c predicted for a 70 kDa protein (Table 1), seemingly suggesting that at low temperature EIC tumbles in solution without feeling the presence of the linker and of the EIN domain. This counterintuitive result originates from the fact that truthful measurements of the relaxation rates were only possible for the subset of EIC signals with ^{15}N-$R_2 \lesssim 100\ s^{-1}$ (i.e. the NMR signals originating from amides that experience local motions that reduce their transverse relaxation). Therefore, the average ^{15}N-R_2/R_1 and corresponding experimental τ_c reported for EIC at 40 °C need to be considered as lower bound estimates of the actual values.

Conformational ensembles of EI at low and high temperature.

The ^{15}N relaxation data presented above demonstrate that EI is a dynamic protein in which the structural domains undergo semi-independent tumbling in solution (Figure 3). Therefore, to determine a structural model that satisfies the experimental RDC data, we have turned to a dynamic structural ensemble representation for EI.[28–31] In particular, preliminary ensembles of physically feasible conformations were generated by running fifteen 10^8-step (equivalent to ~100 microseconds of real time sampling) cgMD simulations[32] at temperatures ranging from 250 to 320 K starting from the AF2 model of the full-length EI dimer. It is important to notice that, due to the large size of the investigated enzyme, we employed a coarse-grained force field to ensure that the conformational space is exhaustively sampled by the MD trajectory. However, when investigating lower molecular weight systems, other enhanced sampling techniques coupled to all-atom force fields (i.e., accelerated MD, replica exchange, metadynamics, etc.)[33–34] might be a better option. The cgMD simulations were run using the Associative Memory, Water Mediated, Structure and Energy Model (AWSEM), a medium resolution coarse-grained force field designed for de novo prediction of protein folding and allosteric motions in native states.[35] The resulting trajectories were clustered to produce ensembles of representative structures with a high degree of structural diversity at each simulated temperature. The generated structure ensembles were used to fit the experimental RDC data by using a previously reported protocol that optimizes the alignment tensor and population of each ensemble member.[36] The agreement between experimental and back-calculated RDCs was evaluated in terms of R-factor. To evaluate the effect of the ensemble size on the fit, the protocol was iterated by increasing the number of clusters (i.e., the number of representative structures at each simulated temperature), and the best-fit conformational ensemble was determined as the smallest-sized ensemble that returns an R-factor equal to a target value (R-target) predetermined to avoid overfitting of the experimental data.[37] From the SVD analysis of the RDCs measured for full-length EI at 40 and 70 °C against the structure of the isolated EI domains, R-target values of 32 and 27 % were determined for the low and high temperature conformational ensembles, respectively (Equation 3).[37] Our refinement protocol indicates that the 6-member ensemble obtained from the 270 K cgMD and the12-member ensemble obtained from the 280 K cgMD are the best-fit solutions for the 40 and 70 °C RDCs, respectively (Figure 4A–D and Supplementary Figure S5A,B). It is also interesting to highlight that using a structure

based cgMD model that relies on the AF2 initial structure for defining native contacts[38] does not produce conformational ensembles consistent with the experimental RDC data (Supplementary Figure S5C–F). This finding further corroborates that (i) using a force field that accounts for non-native interactions is crucial for capturing large-scale dynamic excursions beyond the native structure produced by AF2, and (ii) experimental RDC data are well suited to assess the accuracy of simulated protein conformational ensembles.

Analysis of the best-fit conformational ensembles confirms that the EI domains adopt rigid local folds that agree well with the AF2 prediction. Indeed, a comparison of the conformations displayed by EIN^{α}, $EIN^{\alpha\beta}$, and EIC in the 40 and 70 °C ensembles reveals near-identical folds (Cα r.m.s.d. of 1.9, 1.0, and 2.3 Å for EIN^{α}, $EIN^{\alpha\beta}$, and EIC, respectively) that align well with the structures of the individual domains in the predicted structural model (Supplementary Figure S6). On the other hand, the refined ensembles indicate that EI undergoes extensive interdomain dynamics at both 40 and 70 °C that are not captured by AF2 modeling. Indeed, a high degree of interdomain fluctuations is observed within the cgMD/RDC ensembles, but not within the ensemble of (3) structures predicted by AlphaFold (Figure 4E–H). Of note, a plot of the root mean square fluctuations (r.m.s.f.) values versus the residue index shows that EIN^{α}, $EIN^{\alpha\beta}$, and EIC fluctuate with different average r.m.s.f. in the cgMD/RDC conformational ensembles (average r.m.s.f. values of 16, 8, and 2 Å at 40 °C and of 21, 15, and 2 Å at 70 °C are obtained for EIN^{α}, $EIN^{\alpha\beta}$, and EIC, respectively) (Figure 4G,H), which is consistent with the ^{15}N-R_2/R_1 data indicating that the EI domains undergo semi-independent tumbling in solution (Figure 3 and Table 1).

An open-to-closed transition induced by temperature.

Inspection of the conformational ensembles calculated at 40 and 70 °C indicates that EI samples a conformational space that goes from the open state observed in the AF2 model to the closed state observed in the crystal structure of the activated EI from *E. coli* (Figure 4A,C). Of note, all conformations included in our ensembles are more closed than the solution structure of the apo *E. coli* enzyme (Figure 4A,C). Such result is the consequence of extended electrostatic interactions established between $EIN^{\alpha\beta}$ and EIC in the refined ensembles (Supplementary Figures S7 and S8).

In order to discriminate between open and closed conformations, we have determined the 3D solvent exposure of the active site pocket in each member of the refined ensembles. The calculation was done using the software SADIC[39] by calculating the exposed volume (eV) of a sphere of 15 Å radius centered on the phosphorus atom of an hypothetical PEP molecule bound in the active site (Figure 5A). The eV parameter is a better definition of solvent accessibility than the exposed surface area as it allows better discrimination between a surface atom located in a deep pocket and a surface atom localized at the tip of an exposed loop.[39] As expected, the eV values calculated from the ensemble members are much smaller than the one calculated from the NMR structure of the apo *E.coli* EI, and cluster between the eV values calculated from the AF2 structure and the crystal structure of closed EI (Figure 5C). From a visual inspection of the EI structures, an eV value of 5 % was taken as the threshold to discriminate between closed ($eV < 5$ %) and open ($eV > 5$ %) states, since structures with $eV < 5$ % have a binding pocket that is completely occluded to solvent

(Figure 5A). Interestingly, while the entire 70 °C ensemble is in the open form, three out of six members of the 40 °C ensemble sample closed conformations (Figure 5C). Of note, the $EIN^{\alpha\beta}$/EIC orientation in these closed conformations closely resembles the X-ray structure of activated EI (Figure 5B,D), suggesting that lowering the temperature from 70 to 40 °C induces an open-to-closed conformational change in the thermophilic EI that is similar to the open-to-closed conformational change induced by PEP binding on the *E. coli* homologue at physiological temperature.

Seq2Ensmeble: An open-source computational pipeline for generating conformational ensembles consistent with RDC NMR data from protein sequence alone.

The Seq2Ensemble pipeline used for generating the EI conformational ensembles presented above consists of computational tools for turning the primary sequence of proteins into conformational ensembles consistent with the NMR data. The pipeline starts by generating a 3D structure from sequence via AF2 and following up by running high-throughput multi-replica cgMD simulations at different temperatures and initial states for each temperature. Simulation data is then subjected to iterative clustering and fitting to obtain conformational ensembles consistent with the RDC data. Seq2Ensemble is open source and implements all the tools inside Google Colab, allowing users to run the entire pipeline on research-grade GPUs without installing any software. Our pipeline is partially based on ColabFold, an implementation of AlphaFold on the Google Colab environment.[40] Detailed instructions and examples on how to run the Seq2Ensemble pipeline are provided in Methods and Supplementary Files.

Discussion

Recent advancements in the algorithms for prediction of protein structures have made possible to obtain atomic-resolution models of virtually any protein sequence in a matter of few minutes.[7–9] This ability to generate large, inexpensive databases of structural models introduces the need for standardized methods for assessing the accuracy of the predicted models and for their eventual refinement. To this end, independent works have appeared in the literature form the Bax, Montelione, and Zweckstetter groups demonstrating that NMR RDC data can be conveniently used to test the quality of AF2 predictions.[11–13] More recently, Adams and coworkers have shown that density maps from X-ray crystallography and electron cryo-microscopy data can be used to improve the regions of the model that are not accurately predicted by the neural network.[41] However, even with the latest methodological advancements, predicting the spatial relationship among protein domains remains a challenge, which is exacerbated in dynamic systems that lack extended surfaces of conserved interdomain contacts and/or well-defined experimental density maps.[13–14]

In principle, the interdomain accuracy problem could be alleviated by running extended MD simulations starting from the modelled structure. During an MD simulation, the time-dependence of each interdomain interaction is calculated and, where present, the existence of multiple domain orientations revealed and characterized. When investigating large, multidomain systems, obtaining sufficient sampling of the conformational space with all-atom force fields can become too computationally expensive, and it is nowadays common

to employ enhanced sampling techniques and coarse-grained models.[33–34] Such, advanced computational methods can speed up the calculation by several orders of magnitude and generate exhaustive conformational ensembles that can be validated against experimental data. Here, we have combined the ability of cgMD to produce a large pool of physically feasible protein conformations with the ability of RDCs to provide an experimental assessment of the generated structures to refine an AF2 model into a structure ensemble that is consistent with the experimental data. The protocol was used to investigate the temperature-induced conformational transitions in bacterial EI, a multidomain dynamic protein that undergoes large scale interdomain rearrangements during its catalytic cycle (Figure 1).

In particular, SVD fitting of the EI AF2 model against the backbone amide RDC data showed that, while the structure of the individual protein domains is accurately reproduced by AF2, the predicted spatial relationship among EIN^{α}, $EIN^{\alpha\beta}$, and EIC is inconsistent with the experimental data (Figure 2). Coherent with this observation, refinement of the AF2 model with our cgMD/RDC protocol revealed the presence of extensive interdomain fluctuations in EI that are not accounted for by AF2 modeling (Figure 4). Of note, the interdomain motions captured by the cgMD/RDC ensemble are consistent with ^{15}N-R_2/R_1 data indicating that the EI domains undergo semi-independent tumbling in solution (Figure 3). Overall these results highlight once again (i) that AF2 predictions are insufficient atomic-resolution descriptions of multidomain and dynamic proteins and (ii) that RDC data are fine experimental reporters of structure quality. In addition, the presented results demonstrate the ability of our cgMD/RDC approach to refine an AF2 prediction into a conformational ensemble that is consistent with the experimental data and that returns a more reliable structural model for multidomain dynamic systems, such as EI.

Analysis of the cgMD/RDC ensembles calculated for the apo EI at 40 and 70 °C showed that the enzyme adopts conformations that are more closed than the one predicted by AF2, and that the thermodynamic balance between the open and closed states is dramatically affected by the experimental temperature (Figures 4 and 5). Indeed, while the 70 °C ensemble is entirely composed by open EI conformations, 50 % of the 40 °C ensemble samples closed conformations that closely resemble the crystallographic structure of the catalytically competent state of the *E. coli* EI (Figure 5). One possible explanation for this change in balance is that the transition to the closed form of EI reduces the entropy of the system (i.e. it is coupled to a negative ΔS term and, therefore, it is more likely to occur at low temperature), which is consistent with the hypothesis that substrate binding facilitates transition to the closed form by reducing conformational disorder within EIC.[2] Indeed, at physiological conditions the entropic penalty required to transition to the closed form of EI cannot be offset by reducing the external temperature. On the other hand, the local reduction in conformational disorder induced by PEP binding to the C-terminal domain would result in a less negative ΔS associated with the transition from the (entropically favored) open to the (entropically disfavored) closed form of holo EI compared to the apo enzyme, and shift the conformational equilibrium toward the catalytically competent closed state (Figure 6).

Finally, the results presented in this work demonstrated the effectiveness of coupling AF2 modeling with cgMD calculations and experimental RDC data to detect and structurally

characterize the effect of a change in the experimental temperature on the spatial interdomain organization of EI. We expect the approach presented here to be easily transferrable to investigating the effect of other external perturbations (i.e., ligand binding, mutations, post-translational modifications, etc.) on the structure and dynamics of other multidomain and dynamic proteins. To facilitate implementation of our protocol in other laboratories and research settings we have developed a Google Colab page that allows users to (i) model any primary sequence using AF2, (ii) run cgMD simulations at the desired temperatures, (iii) cluster the resulting trajectories, and (iv) check the consistency of the generated ensembles with experimental RDC data within their local browser.

As a final note, it is important to highlight that our method differs from previously described approaches that use the RDCs as structural restraints to directly bias the MD trajectory and obtain conformational ensembles consistent with the experimental data.[42–43] However, similarly with what reported for the latter approaches, the cgMD/RDC protocol described here would benefit from the use of multiple orthogonal sets of RDC data measured with different alignment media and/or for additional pairs of coupled spins. In the absence of these additional data sets, part of the conformational dynamics could be absorbed by the alignment tensor, and the calculated conformational ensemble could be an underestimation on the conformational space sampled by the system.[44]

Materials and Methods

Preparation of NMR samples.

Uniformly ^{2}H, N *T. tengcongensis* EI was expressed and purified as previously described.[45] NMR samples were prepared in 20 mM Tris-HCl (pH 7.4), 100 mM NaCl, 4 mM $MgCl_2$, 1 mM ethylenediaminetetraacetic acid (EDTA), 2 mM dithiothreitol (DTT), 0.02 % (w/v) NaN_3, and 90 % H_2O / 10 % D_2O (v/v). Protein concentration was 0.5–1.0 mM. All NMR spectra were acquired on a Bruker 800 MHz spectrometer equipped with Z-shielded gradient triple resonance TCI cryoprobe. Spectra were processed and analyzed using the programs NMRPipe[46] and POKY,[47] respectively. ^{1}H-^{15}N correlated spectra were assigned based on previously deposited assignment (BioMagResBank entry 27762).[45]

Backbone amide $^{1}D_{NH}$ RDCs were measured at 40 and 70 °C by taking the difference in $^{1}J_{NH}$ scalar couplings in aligned and isotropic media. The alignment media employed was phage *pf1* (16 mg/ml; ASLA Biotech),[24] and $^{1}J_{NH}$ couplings were measured using the ARTSY pulse scheme.[48] SVD analysis of RDCs was carried out using Xplor-NIH.[49]

^{15}N-R_1 and $R_{1\rho}$ experiments were carried out at 40 and 70 °C using heat-compensated pulse schemes with a TROSY readout.[50] The spin-lock field for the $R_{1\rho}$ experiment was set to 1 kHz. Decay durations were 0, 80, 200, 320, 440, 560, 720, and 840 ms for R_1, and 0.2, 4.2, 7.2, 15.0, 23.4, 32.4, 42.0, 52.2, and 60.0 ms for $R_{1\rho}$. R_1 and $R_{1\rho}$ values were determined by fitting time-dependent exponential restoration of peak intensities at increasing relaxation delays. R_2 values were extracted from the measured R_1 and $R_{1\rho}$ values.

Ensemble calculation.

To generate structural ensembles of EI, 100,000,000-step cgMD simulations were run starting from the AF2 model of the EI dimer using the AWSEM coarse-grained force field implementation in the OpenMM molecular dynamics library.[51] The energy function of AWSEM combines knowledge-based and physics-based terms designed for de novo prediction of 3D protein fold and dynamics. In AWSEM, each residue is coarse-grained into three beads which allows for resolving detailed backbone stereochemistry, hydrogen bonding, and water-mediated tertiary interactions using physics-based potentials.

As reported for other symmetric dimer systems,[29] the sampling of the tertiary fold is better evaluated from combined trajectories in which the trajectory of the one subunit is appended at the end of the trajectory of the second subunit. Therefore, a 200,000,000-step trajectory was generated by appending the 100,000,000-step trajectory of the second subunit to the 100,000,000-step trajectory of the first subunit of the dimer. The combined trajectory was then clustered using the r.m.s.d. based K-Means algorithm implemented in the MDAnalysis python library.[52] After specifying the maximum ensemble size "N" to generate, the clustering algorithm was iterated over the trajectory, producing from 1 to N clusters with 10 initializations and 300 iterations each (see https://docs.mdanalysis.org/stable/documentation_pages/analysis/encore/clustering.html for details). At each iteration the centroids of each cluster were saved as PDB files. These centroids served as representative structures of the cgMD (i.e. the members of the conformational ensembles) with a high degree of structural diversity.

Back-calculation of RDCs from conformational ensembles was done using the following equation:

$$RDC_i = \sum_k D_k\left[\left(3cos^2\theta - 1\right) + \frac{3}{2}\left(sin^2\theta\cos 2\phi\right)\right] \quad (1)$$

where θ is the angle formed between the internuclear bond vector of the amide group of residue i and the z axis of the alignment tensor, ϕ the angle between the xy plane projection of the internuclear bond vector and the x axis, and D_k is the magnitude of the alignment tensor for ensemble member k multiplied by its fractional population in the ensemble.[37] A Matlab script to fit experimental RDC data to an ensemble is available for download at http://group.chem.iastate.edu/Venditti/downloads.html.[36] Alternatively, RDC fitting can be performed using the Colab page https://potoyangroup.github.io/Seq2Ensemble/.

The consistency between experimental and back calculated was evaluated in terms of R-factor:[37,53]

$$R = \sqrt{\frac{\langle D_{obs} - D_{calc}\rangle^2}{2\langle D_{obs}^2\rangle}} \quad (2)$$

where D_{obs} and D_{calc} are the observed and back calculated RDCs, respectively. A target R-factor (R-target) for the calculation of conformational ensemble was determined using the equation: [6]

$$R-target=\sum_i \frac{R_i N_i}{N} \quad (3)$$

where R_i is the R-factor determined from SVD fitting against the AF2 model of the i^{th} structural domain (EIN$^{\alpha}$, EIN$^{\alpha\beta}$, or EIC), and N_i is the number of RDCs measured for the i^{th} structural domain, and N is the total number of experimental RDCs measured for the multidomain protein.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

We thank Dr. Julien Roche for critical reading of the manuscript. This work was supported by funds from the National Institute of General Medical Sciences with grant no. R35GM133488 (to V.V.) and grant no. R35GM138243 (to D.A.P).

References

1. Tang C; Schwieters CD; Clore GM, Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. Nature 2007, 449 (7165), 1078–82. [PubMed: 17960247]
2. Venditti V; Tugarinov V; Schwieters CD; Grishaev A; Clore GM, Large interdomain rearrangement triggered by suppression of micro- to millisecond dynamics in bacterial Enzyme I. Nat Commun 2015, 6, 5960. [PubMed: 25581904]
3. Rademacher N; Kuropka B; Kunde SA; Wahl MC; Freund C; Shoichet SA, Intramolecular domain dynamics regulate synaptic MAGUK protein interactions. Elife 2019, 8.
4. Tsai CJ; Del Sol A; Nussinov R, Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. Mol Biosyst 2009, 5 (3), 207–16. [PubMed: 19225609]
5. Suh JY; Tang C; Clore GM, Role of electrostatic interactions in transient encounter complexes in protein-protein association investigated by paramagnetic relaxation enhancement. J Am Chem Soc 2007, 129 (43), 12954–5. [PubMed: 17918946]
6. Medvedev KE; Kolchanov NA; Afonnikov DA, High temperature and pressure influence the interdomain orientation of Nip7 proteins from P. abyssi and P. furiosus: MD simulations. J Biomol Struct Dyn 2018, 36 (1), 68–82. [PubMed: 27924686]
7. Jumper J; Evans R; Pritzel A; Green T; Figurnov M; Ronneberger O; Tunyasuvunakool K; Bates R; Zidek A; Potapenko A; Bridgland A; Meyer C; Kohl SAA; Ballard AJ; Cowie A; Romera-Paredes B; Nikolov S; Jain R; Adler J; Back T; Petersen S; Reiman D; Clancy E; Zielinski M; Steinegger M; Pacholska M; Berghammer T; Bodenstein S; Silver D; Vinyals O; Senior AW; Kavukcuoglu K; Kohli P; Hassabis D, Highly accurate protein structure prediction with AlphaFold. Nature 2021, 596 (7873), 583–589. [PubMed: 34265844]
8. Baek M; DiMaio F; Anishchenko I; Dauparas J; Ovchinnikov S; Lee GR; Wang J; Cong Q; Kinch LN; Schaeffer RD; Millan C; Park H; Adams C; Glassman CR; DeGiovanni A; Pereira JH; Rodrigues AV; van Dijk AA; Ebrecht AC; Opperman DJ; Sagmeister T; Buhlheller C; Pavkov-Keller T; Rathinaswamy MK; Dalwadi U; Yip CK; Burke JE; Garcia KC; Grishin NV; Adams PD; Read RJ; Baker D, Accurate prediction of protein structures and interactions using a three-track neural network. Science 2021, 373 (6557), 871–876. [PubMed: 34282049]
9. Tunyasuvunakool K; Adler J; Wu Z; Green T; Zielinski M; Zidek A; Bridgland A; Cowie A; Meyer C; Laydon A; Velankar S; Kleywegt GJ; Bateman A; Evans R; Pritzel A; Figurnov M; Ronneberger O; Bates R; Kohl SAA; Potapenko A; Ballard AJ; Romera-Paredes B; Nikolov S; Jain R; Clancy E; Reiman D; Petersen S; Senior AW; Kavukcuoglu K; Birney E; Kohli P; Jumper J; Hassabis

D, Highly accurate protein structure prediction for the human proteome. Nature 2021, 596 (7873), 590–596. [PubMed: 34293799]

10. Fowler NJ; Williamson MP, The accuracy of protein structures in solution determined by AlphaFold and NMR. Structure 2022, 30 (7), 925–933 e2. [PubMed: 35537451]
11. Robertson AJ; Courtney JM; Shen Y; Ying J; Bax A, Concordance of X-ray and AlphaFold2 Models of SARS-CoV-2 Main Protease with Residual Dipolar Couplings Measured in Solution. J Am Chem Soc 2021, 143 (46), 19306–19310. [PubMed: 34757725]
12. Tejero R; Huang YJ; Ramelot TA; Montelione GT, AlphaFold Models of Small Proteins Rival the Accuracy of Solution NMR Structures. Front Mol Biosci 2022, 9, 877000. [PubMed: 35769913]
13. Zweckstetter M, NMR hawk-eyed view of AlphaFold2 structures. Protein Sci 2021, 30 (11), 2333–2337. [PubMed: 34469019]
14. David A; Islam S; Tankhilevich E; Sternberg MJE, The AlphaFold Database of Protein Structures: A Biologist's Guide. J Mol Biol 2022, 434 (2), 167336. [PubMed: 34757056]
15. Chauvin F; Brand L; Roseman S, Enzyme I: the first protein and potential regulator of the bacterial phosphoenolpyruvate: glycose phosphotransferase system. Res Microbiol 1996, 147 (6–7), 471–9. [PubMed: 9084757]
16. Nguyen TT; Ghirlando R; Venditti V, The oligomerization state of bacterial enzyme I (EI) determines EI's allosteric stimulation or competitive inhibition by alpha-ketoglutarate. J Biol Chem 2018, 293 (7), 2631–2639. [PubMed: 29317499]
17. Dotas RR; Nguyen TT; Stewart CE Jr.; Ghirlando R; Potoyan DA; Venditti V, Hybrid Thermophilic/Mesophilic Enzymes Reveal a Role for Conformational Disorder in Regulation of Bacterial Enzyme I. J Mol Biol 2020, 432 (16), 4481–4498. [PubMed: 32504625]
18. Garrett DS; Seok YJ; Liao DI; Peterkofsky A; Gronenborn AM; Clore GM, Solution structure of the 30 kDa N-terminal domain of enzyme I of the Escherichia coli phosphoenolpyruvate:sugar phosphotransferase system by multidimensional NMR. Biochemistry 1997, 36 (9), 2517–30. [PubMed: 9054557]
19. Schwieters CD; Suh JY; Grishaev A; Ghirlando R; Takayama Y; Clore GM, Solution structure of the 128 kDa enzyme I dimer from Escherichia coli and its 146 kDa complex with HPr using residual dipolar couplings and small- and wide-angle X-ray scattering. J Am Chem Soc 2010, 132 (37), 13026–45. [PubMed: 20731394]
20. Teplyakov A; Lim K; Zhu PP; Kapadia G; Chen CC; Schwartz J; Howard A; Reddy PT; Peterkofsky A; Herzberg O, Structure of phosphorylated enzyme I, the phosphoenolpyruvate:sugar phosphotransferase system sugar translocation signal protein. Proceedings of the National Academy of Sciences of the United States of America 2006, 103 (44), 16218–16223. [PubMed: 17053069]
21. Venditti V; Schwieters CD; Grishaev A; Clore GM, Dynamic equilibrium between closed and partially closed states of the bacterial Enzyme I unveiled by solution NMR and X-ray scattering. Proc Natl Acad Sci U S A 2015, 112 (37), 11565–70. [PubMed: 26305976]
22. Evangelidis T; Nerli S; Novacek J; Brereton AE; Karplus PA; Dotas RR; Venditti V; Sgourakis NG; Tripsianes K, Automated NMR resonance assignments and structure determination using a minimal set of 4D spectra. Nat Commun 2018, 9 (1), 384. [PubMed: 29374165]
23. Oberholzer AE; Bumann M; Schneider P; Bachler C; Siebold C; Baumann U; Erni B, Crystal structure of the phosphoenolpyruvate-binding enzyme I-domain from the Thermoanaerobacter tengcongensis PEP: sugar phosphotransferase system (PTS). J Mol Biol 2005, 346 (2), 521–32. [PubMed: 15670601]
24. Clore GM; Starich MR; Gronenborn AM, Measurement of Residual Dipolar Couplings of Macromolecules Aligned in the Nematic Phase of a Colloidal Suspension of Rod-Shaped Viruses. J Am Chem Soc 1998, 120, 10571–10572.
25. Tjandra N; Bax A, Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. Science 1997, 278 (5340), 1111–4. [PubMed: 9353189]
26. Bax A; Grishaev A, Weak alignment NMR: a hawk-eyed view of biomolecular structure. Curr Opin Struct Biol 2005, 15 (5), 563–70. [PubMed: 16140525]

27. Kay LE; Torchia DA; Bax A, Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry 1989, 28 (23), 8972–9. [PubMed: 2690953]
28. Khatiwada B; Nguyen TT; Purslow JA; Venditti V, Solution structure ensemble of human obesity-associated protein FTO reveals druggable surface pockets at the interface between the N- and C-terminal domain. J Biol Chem 2022, 298 (5), 101907. [PubMed: 35398093]
29. Nguyen TT; Ghirlando R; Roche J; Venditti V, Structure elucidation of the elusive Enzyme I monomer reveals the molecular mechanisms linking oligomerization and enzymatic activity. Proc Natl Acad Sci U S A 2021, 118 (20).
30. Purslow JA; Nguyen TT; Khatiwada B; Singh A; Venditti V, N (6)-methyladenosine binding induces a metal-centered rearrangement that activates the human RNA demethylase Alkbh5. Sci Adv 2021, 7 (34).
31. Angyan AF; Gaspari Z, Ensemble-based interpretations of NMR structural data to describe protein internal dynamics. Molecules 2013, 18 (9), 10548–67. [PubMed: 23999727]
32. Tozzini V, Coarse-grained models for proteins. Curr Opin Struct Biol 2005, 15 (2), 144–50. [PubMed: 15837171]
33. Bernardi RC; Melo MCR; Schulten K, Enhanced sampling techniques in molecular dynamics simulations of biological systems. Biochim Biophys Acta 2015, 1850 (5), 872–877. [PubMed: 25450171]
34. Yang YI; Shao Q; Zhang J; Yang L; Gao YQ, Enhanced sampling in molecular dynamics. J Chem Phys 2019, 151 (7), 070902. [PubMed: 31438687]
35. Davtyan A; Schafer NP; Zheng W; Clementi C; Wolynes PG; Papoian GA, AWSEM-MD: protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. J Phys Chem B 2012, 116 (29), 8494–503. [PubMed: 22545654]
36. Purslow JA; Nguyen TT; Egner TK; Dotas RR; Khatiwada B; Venditti V, Active Site Breathing of Human Alkbh5 Revealed by Solution NMR and Accelerated Molecular Dynamics. Biophys J 2018, 115 (10), 1895–1905. [PubMed: 30352661]
37. Venditti V; Egner TK; Clore GM, Hybrid Approaches to Structural Characterization of Conformational Ensembles of Complex Macromolecular Systems Combining NMR Residual Dipolar Couplings and Solution X-ray Scattering. Chem Rev 2016, 116 (11), 6305–22. [PubMed: 26739383]
38. Noel JK; Levi M; Raghunathan M; Lammert H; Hayes RL; Onuchic JN; Whitford PC, SMOG 2: A Versatile Software Package for Generating Structure-Based Models. PLoS Comput Biol 2016, 12 (3), e1004794. [PubMed: 26963394]
39. Varrazzo D; Bernini A; Spiga O; Ciutti A; Chiellini S; Venditti V; Bracci L; Niccolai N, Three-dimensional computation of atom depth in complex molecular structures. Bioinformatics 2005, 21 (12), 2856–60. [PubMed: 15827080]
40. Mirdita M; Schutze K; Moriwaki Y; Heo L; Ovchinnikov S; Steinegger M, ColabFold: making protein folding accessible to all. Nat Methods 2022, 19 (6), 679–682. [PubMed: 35637307]
41. Terwilliger TC; Poon BK; Afonine PV; Schlicksup CJ; Croll TI; Millan C; Richardson JS; Read RJ; Adams PD, Improved AlphaFold modeling with implicit experimental information. Nat Methods 2022, 19 (11), 1376–1382. [PubMed: 36266465]
42. De Simone A; Aprile FA; Dhulesia A; Dobson CM; Vendruscolo M, Structure of a low-population intermediate state in the release of an enzyme product. Elife 2015, 4.
43. Lange OF; Lakomek NA; Fares C; Schroder GF; Walter KF; Becker S; Meiler J; Grubmuller H; Griesinger C; de Groot BL, Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. Science 2008, 320 (5882), 1471–5. [PubMed: 18556554]
44. Salmon L; Blackledge M, Investigating protein conformational energy landscapes and atomic resolution dynamics from NMR dipolar couplings: a review. Rep Prog Phys 2015, 78 (12), 126601. [PubMed: 26517337]
45. Dotas RR; Venditti V, Resonance assignment of the 128 kDa enzyme I dimer from Thermoanaerobacter tengcongensis. Biomol NMR Assign 2019, 13 (2), 287–293. [PubMed: 31025174]

46. Delaglio F; Grzesiek S; Vuister GW; Zhu G; Pfeifer J; Bax A, NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 1995, 6 (3), 277–93. [PubMed: 8520220]

47. Lee W; Rahimi M; Lee Y; Chiu A, POKY: a software suite for multidimensional NMR and 3D structure calculation of biomolecules. Bioinformatics 2021, 37 (18), 3041–3042. [PubMed: 33715003]

48. Fitzkee NC; Bax A, Facile measurement of (1)H-(1)5N residual dipolar couplings in larger perdeuterated proteins. J Biomol NMR 2010, 48 (2), 65–70. [PubMed: 20694505]

49. Schwieters CD; Kuszewski JJ; Tjandra N; Clore GM, The Xplor-NIH NMR molecular structure determination package. J Magn Reson 2003, 160 (1), 65–73. [PubMed: 12565051]

50. Lakomek NA; Ying J; Bax A, Measurement of (1)(5)N relaxation rates in perdeuterated proteins by TROSY-based methods. J Biomol NMR 2012, 53 (3), 209–21. [PubMed: 22689066]

51. Lu W; Bueno C; Schafer NP; Moller J; Jin S; Chen X; Chen M; Gu X; Davtyan A; de Pablo JJ; Wolynes PG, OpenAWSEM with Open3SPN2: A fast, flexible, and accessible framework for large-scale coarse-grained biomolecular simulations. PLoS Comput Biol 2021, 17 (2), e1008308. [PubMed: 33577557]

52. Michaud-Agrawal N; Denning EJ; Woolf TB; Beckstein O, MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. J Comput Chem 2011, 32 (10), 2319–27. [PubMed: 21500218]

53. Clore GM; Garrett DS, R-factor, free R, and complete crossvalidation for dipolar coupling refinement of NMR structures. J Am Chem Soc 1999, 121, 9008–9012.

54. Schwieters CD; Clore GM, Reweighted atomic densities to represent ensembles of NMR structures. J Biomol NMR 2002, 23 (3), 221–5. [PubMed: 12238594]

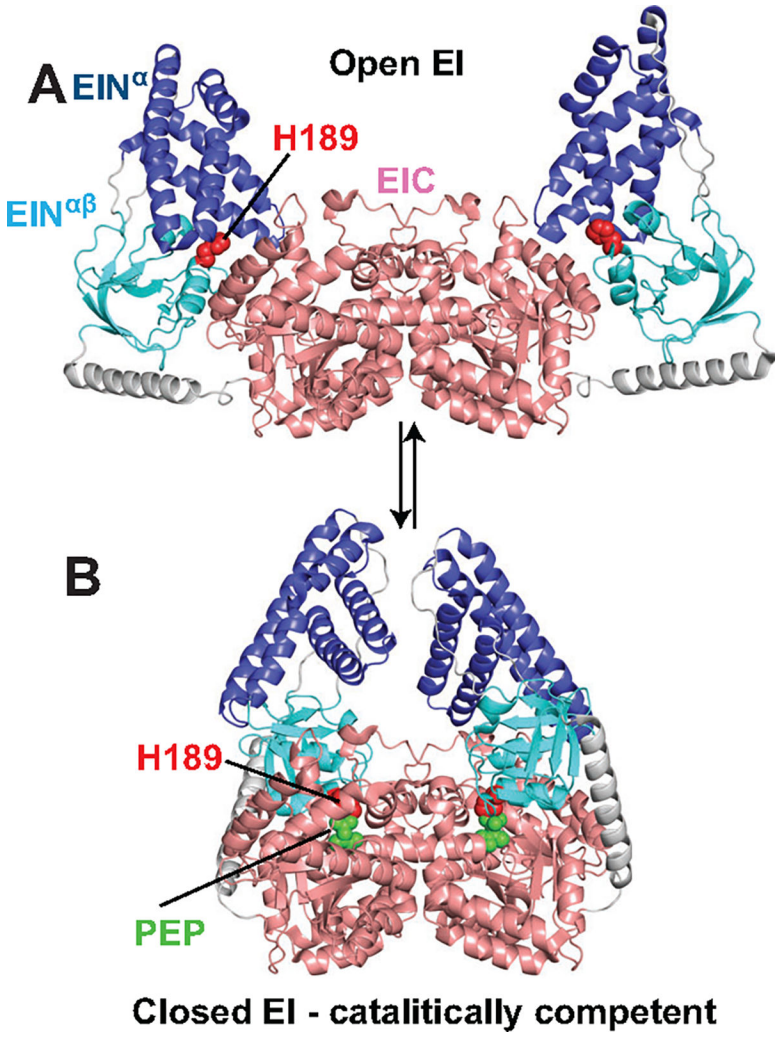


Figure 1. Structures of *E. coli* EI.

(**A**) Solution structure of open EI. (**B**) Crystal structure of phosphorylated EI bound to oxalate. The crystal structure corresponds to the activated closed form of EI bound to PEP. Therefore, the phosphoryl group (covalently attached to His 189) and the oxalate molecule are replaced by a PEP molecule in the figure. EIN$^{\alpha}$, EIN$^{\alpha\beta}$, and EIC are shown as blue, cyan, and salmon ribbons, respectively. The His 189 phosphorylation site is shown as red spheres. PEP is shown as green spheres.

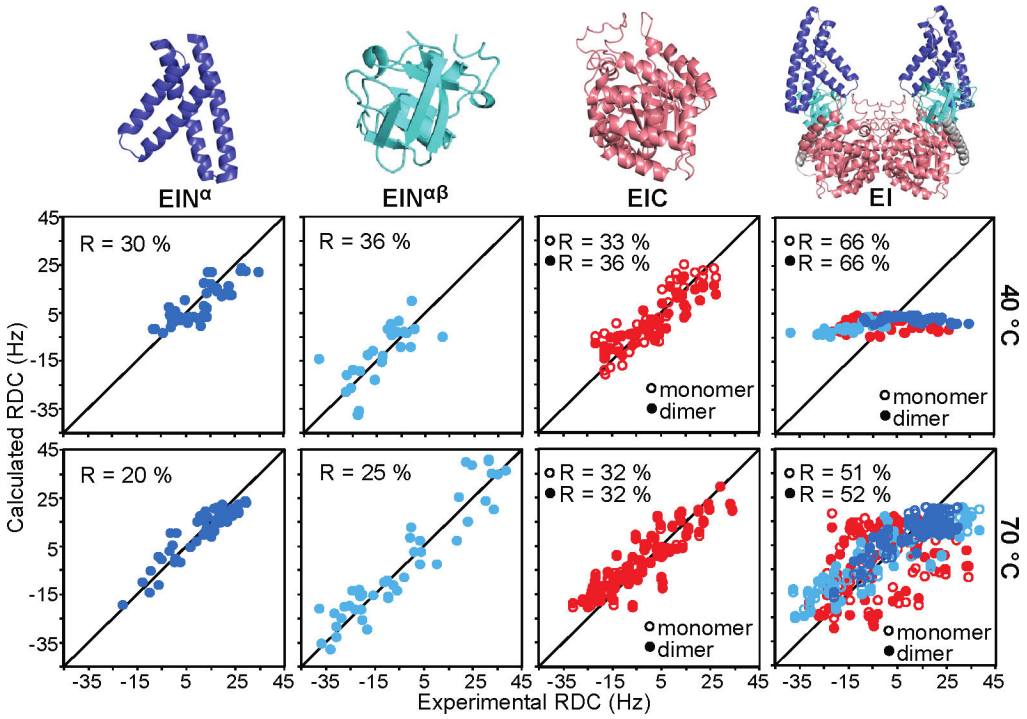


Figure 2. RDC analysis of the AF2 model.

Singular value decomposition fitting (SVD) of the RDC data against the AF2 model. The first, second, and third panels from left to right show fits against the data of the EIN$^{\alpha}$, EIN$^{\alpha\beta}$, and EIC, respectively. The forth panel (left to right) reports SVD against the full RDC dataset. The middle and bottom rows correspond to fits against the 40 and 70 °C RDC data, respectively. The data from EIN$^{\alpha}$, EIN$^{\alpha\beta}$, and EIC are colored blue, cyan, and red, respectively. In the third and fourth panels (left to right), open and filled circles correspond to fits using a single subunit or both subunits of the AF2 model of the EI dimer, respectively. The top row shows the structure of EIN$^{\alpha}$, EIN$^{\alpha\beta}$, EIC, and the full-length EI in the AF2 model. EIN$^{\alpha}$, EIN$^{\alpha\beta}$, and EIC are shown as blue, cyan, and salmon ribbons.

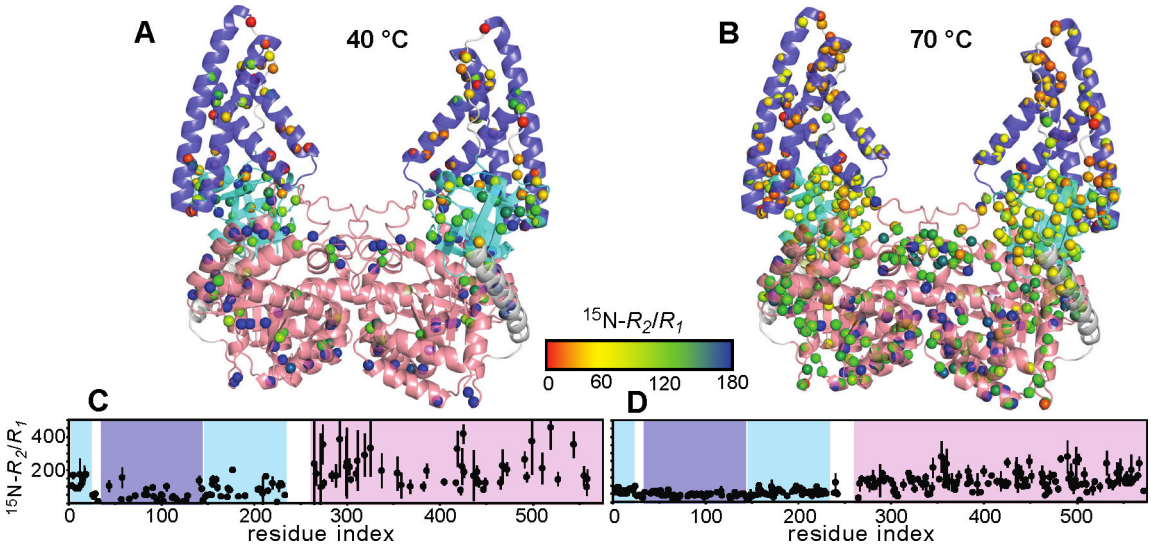


Figure 3. ^{15}N-R_2/R_1 analysis of the full-length EI.

The 800MHz ^{15}N-R_2/R_1 data of EI measured at 40 °C (**A**) and 70 °C (**B**) are plotted on the AF2 model of EI. Analyzed amides are shown as spheres. The relationship between the sphere color and ^{15}N-R_2/R_1 value is depicted by the color bar. EIN$^{\alpha}$, EIN$^{\alpha\beta}$, and EIC are shown as blue, cyan, and salmon ribbons. In (**C**) and (**D**) the ^{15}N-R_2/R_1 data measured at 40 and 70 °C, respectively, are plotted versus the residue index. The blue, cyan, and salmon boxes show the boundaries of EIN$^{\alpha}$, EIN$^{\alpha\beta}$, and EIC, respectively.

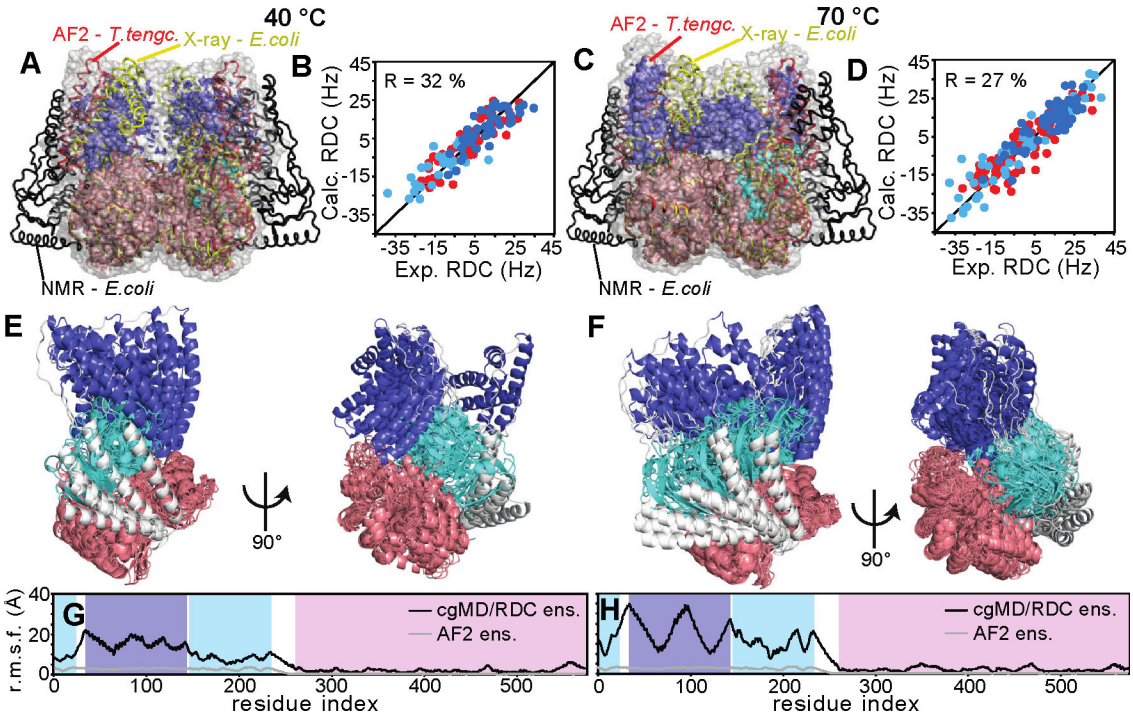


Figure 4. Conformational ensembles of *T. tengcongensis* EI.

Atomic probability density maps[54] showing the conformational space sampled by EI in the cgMD/RDC ensemble at (**A**) 40 °C and (**C**) 70 °C. The atomic probability maps are plotted at a threshold ranging from 1 % (transparent gray) to 40 % (opaque blue, cyan, and salmon for EIN^{α}, $EIN^{\alpha\beta}$, and EIC, respectively) of maximum. The solution structure of open *E. coli* EI, the X-ray structure of closed *E. coli* EI, and the AF2 model of *T. tengcongensis* EI are superimposed to the maps as black, yellow, and red cartoons, respectively. The ensembles were calculated for a single subunit of the EI dimer (see Methods). The density maps for the dimer were reconstructed by superimposing the EIC coordinates of each ensemble member into the EIC coordinates of the AF2 model (note that this procedure is justified by the fact that EI behaves as a symmetric dimer in solution and that the EIC RDC data fit well to the AF2 structure, Figure 2). Panels (**B**) and (**D**) show the correlation between the experimental RDCs and the RDCs back calculated from the 40 and 70 °C ensembles, respectively. Filled blue, cyan, and red circles are the data originating from EIN^{α}, $EIN^{\alpha\beta}$, and EIC, respectively. Panels (**E**) and (**F**) show the overlay of the 6 and 12 members of the 40 and 70 °C ensembles, respectively. Structures were aligned by superimposing the coordinates of the EIC domain. EIN^{α}, $EIN^{\alpha\beta}$, and EIC are shown as blue, cyan, and salmon cartoons, respectively. Panels (**G**) and (**H**) show the r.m.s.f. versus residue index calculated from the 40 and 70 °C ensembles, respectively (black line). The ensemble members were superimposed using the coordinate of the EIC domain to emphasized the semi-independent motions of the EI domains. The r.m.s.f. calculated from the AF2 ensemble (Supplementary Figure S1) is shown as a gray line for comparison. The blue, cyan, and salmon boxes show the boundaries of EIN^{α}, $EIN^{\alpha\beta}$, and EIC, respectively.

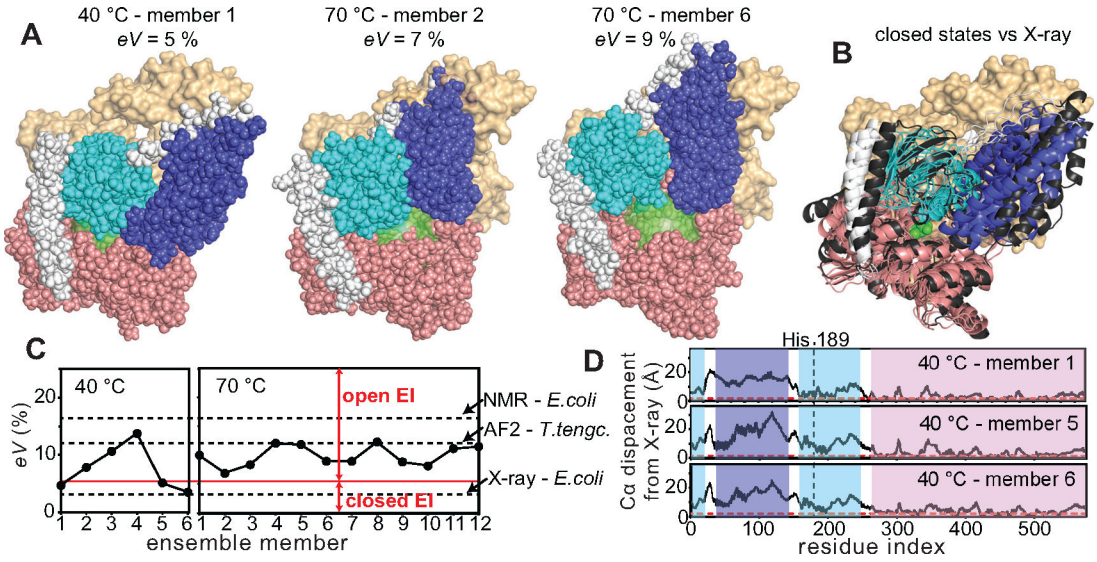


Figure 5. A temperature-induced open-to-closed switch.

(**A**) Three members of the calculated ensembles with different degrees of active site solvent accessibility are shown. The sphere (15 Å radius) used to calculate the exposed volume (*eV*) of the active site is colored green. The EI structure is shown as spheres (EIN^{α}, $EIN^{\alpha\beta}$, and EIC are blue, cyan, and salmon, respectively). The second subunit of the dimer from the AF2 model is shown as orange surface. (**B**) The three ensemble members adopting closed conformations (member 1, 5, and 6 of the 40 °C ensemble; EIN^{α}, $EIN^{\alpha\beta}$, and EIC are blue, cyan, and salmon, respectively) are superimposed on the crystal structure of closed EI (black ribbons). The second subunit of the EI dimer is shown as orange surface. The PEP molecule is shown as green spheres. (**C**) *eV* value versus ensemble member from the 40 (left) and 70 (right) °C ensembles. The *eV* values calculated from the solution structure of *E. coli* EI, the crystal structure of closed *E. coli* EI, and the AF2 model of *T. tengcongensis* EI are shown as dashed lines. The boundary between open and closed conformations is shown as a solid red line. (**D**) Cα displacement of the members 1 (top), 5 (middle), and 6 (bottom) of the 40 °C ensemble from the X-ray structure of closed EI is plotted versus the residue index. The blue, cyan, and salmon boxes show the boundaries of EIN^{α}, $EIN^{\alpha\beta}$, and EIC, respectively. The position of the catalytic His 189 residue is highlighted by a vertical dashed line. The horizontal dashed red line is drawn at a Cα displacement of 2 Å.

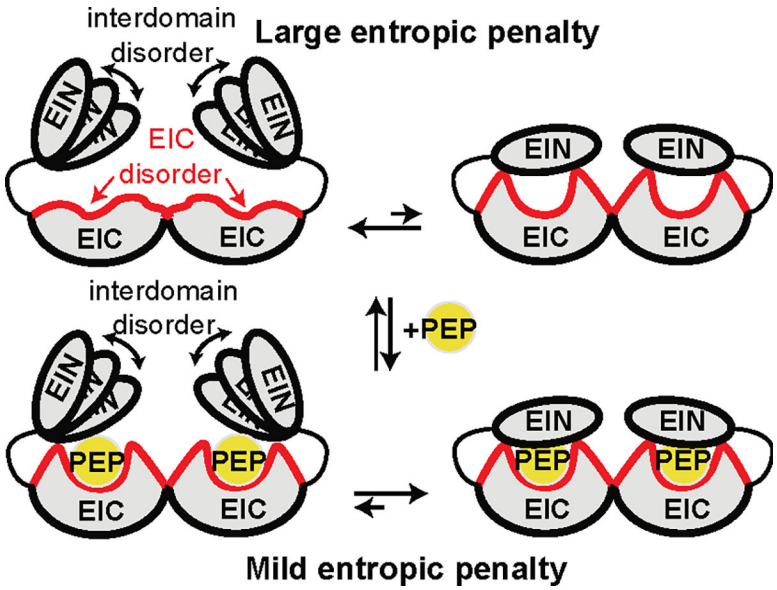


Figure 6. Model for conformational disorder in apo and holo EI.

In the absence of PEP transition to the catalytically competent closed state is coupled to a large entropic penalty because it requires both ordering of EIC[2] and a quenching of interdomain dynamics (top). Local EIC disorder is quenched in the enzyme-substrate complex.[2] Therefore, PEP binding reduces the entropic cost to transition from the open to the closed state of EI (bottom).

Table 1.

Average ^{15}N-R_2/R_1 values and rotational correlation time

		^{15}N-R_2/R_1 [a]		Exp. τ_c (ns) [b]		Pred. τ_c (ns) [c]	
	MW (kDa)	**40 °C**	**70 °C**	**40 °C**	**70 °C**	**40 °C**	**70 °C**
EIN$^{\alpha}$	12.6	51	44	17	16	5	3
EIN$^{\alpha\beta}$	11.5	103	63	25	19	5	3
EIC dimer	69.7	144	130	29	28	29	17
EI dimer	126.9	-	-	-	-	53	30

[a] Average ^{15}N-R_2/R_1 values calculated over residues from EIN$^{\alpha}$ (first row), EIN$^{\alpha\beta}$ (second row), EIC (third row), and full-length EI (forth row).

[b] Experimental τ_c'sare calculated from the average ^{15}N-R_2/R_1 values using equation $\tau_c = \frac{1}{4\pi\upsilon_N}\sqrt{6\frac{R_2}{R_1} - 7}$, where ν_N is the ^{15}N resonance frequency in Hz

[c] Predicted τ_c's are calculated from the molar weight (MW) by using http://nickanthis.com/tools/tau (empirical formula)