Supporting Information

Structure elucidation of the elusive Enzyme I monomer reveals the molecular mechanisms linking oligomerization and enzymatic activity

Trang T. Nguyen[1], Rodolfo Ghirlando[2], Julien Roche[3]*, and Vincenzo Venditti[1,3]*

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

[2]Laboratory of Molecular Biology, NIDDK, National Institutes of Health, Bethesda, Maryland 20892

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

*Corresponding authors: Julien Roche and Vincenzo Venditti

Email: jroche@iastate.edu (J. Roche), venditti@iastate.edu (V. Venditti)

Supporting Materials and Methods

NMR Spectroscopy

NMR samples were prepared in 20 mM Tris-HCl (pH 7.4), 100 mM NaCl, 4 mM $MgCl_2$, 1 mM EDTA, 2 mM DTT, and 90% H2O/10% D2O (v/v). Protein concentration was 0.4–0.8 mM (unless stated otherwise). Spectra were processed in NMRPipe (1) and analyzed using the program and SPARKY (http://www.cgl.ucsf.edu/home/sparky), respectively.

Assignments of the ^{1}H-^{15}N TROSY spectrum of wt-EIC at 1 bar 40 °C was obtained previously (2). The ^{1}H-^{15}N TROSY spectrum of wt-EIC at 2 kbar and 40 °C was assigned by tracking the peaks over a series of 2D spectra collected from 1 bar to 2 kbar. Sequential ^{1}H,^{15}N,^{13}C backbone resonance assignment of the 3m-EIC at 1 bar and 2 kbar, was carried out 40 °C using a TROSY versions of conventional 3D triple resonance correlation experiments (HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCB and HN(COCA)CB) (3). The chemical shifts assigned for 3m-EIC at 1 bar and 2 kbar have been deposited in the BioMagResBank (accession nos. 50695 and 50694, respectively).

The thermodynamics of the EIC dimer-to-monomer equilibrium were investigated by running ^{1}H-^{15}N TROSY spectra of ^{15}N-labelled wt- and 3m-EIC at increasing hydrostatic pressure from 1 bar to 2.5 kbar. Data were measured in the absence and in the presence of 10 mM αKG or PEP. Signal intensities for the NMR peaks reporting on the monomer/dimer equilibrium (i.e. NMR resonances for which distinct peaks for the monomer and dimer species are observed in the pressure titration on 3m-EIC) were quantified and their pressure dependence was fit globally using a two-state equilibrium model to obtain the free energy difference (ΔG in cal/mol) and volume changes (ΔV in mL/mol) between the dimeric and monomeric states.

$$I(p) = \frac{\mathrm{D+M}e^{\left(\frac{\Delta G + p\Delta V}{-RT}\right)}}{e^{\left(\frac{\Delta G + p\Delta V}{-RT}\right)} + 1} \qquad (1)$$

Here, *I(p)* is the intensity at each pressure, *D* is the maximum intensity (fully dimeric state) and *M* is the minimum intensity (fully monomeric state). *T* is the temperature in K and *R* is the universal gas constant.

The dimer dissociation constant is calculated from *ΔG* using the following procedure:

$$K_d = \frac{[monomer]^2}{[dimer]} = \frac{[EIC_M]^2}{0.5[EIC_D]} \quad (2)$$

where [EIC_M] is the concentration of EIC in the monomeric state and [EIC_D] is the concentration of EIC subunits in the dimeric state (which corresponds to two times the concentration of dimer). The relationship among [EIC_M], [EIC_D] and *ΔG* is given by:

$$[EIC_D] = \frac{C}{1+e^{-\Delta G/RT}} \quad (3)$$

$$[EIC_M] = C - [EIC_D] \quad (4)$$

where *C* is the total concentration of EIC in the NMR sample.

The spin-lock field for the ^{15}N-$R_{1\rho}$ experiment was set to 1 kHz. Decay durations were set to 0, 240, 560, 880, 1280, 1600, 2080, and 2400 ms for ^{15}N-R_1 and 0.2, 4.2, 7.2, 15, 23.4, 32.4, 42, 52.2, and 60 ms for ^{15}N-$R_{1\rho}$. R_1 and $R_{1\rho}$ values were determined by fitting time-dependent exponential decays of peak intensities at increasing relaxation delays. R_2 values were extracted from the measured R_1 and $R_{1\rho}$ values.

Enzymatic assay

The PEP hydrolase activity of wt- and 3m-EIC was assayed by monitoring the disappearance of the alkene proton signals of PEP using real-time ^{1}H NMR, as previously described (4). The reaction mixtures were prepared in 20 mM Tris buffer (pH 7.4), 100 mM NaCl, 4 mM $MgCl_2$, 1 mM EDTA, 2 mM DTT, 1 mM trimethylsilylpropanoic acid (TSP),

and 99.99% D_2O. Enzyme concentration was 50 μM for both wt- and 3m-EIC, unless stated otherwise.

Enzymatic assays were run in duplicate at 50 ˚C and 1 bar. The integrals of the signals measured for the alkene protons were converted to mM units by reference to the internal standard TSP. Integration of the NMR signals was performed using the software MNova (https://mestrelab.com/download/mnova/). Initial velocities were determined from the linear portion of the progress curves and were fit using the Michaelis–Menten equation. Enzymatic activities reported in Figure 3B (main text) were measured at 65 ˚C and hydrostatic pressures of 1 and 3 kbar using the same protocol described above. The enzyme concentration was ~150 μM.

Analytical Ultracentrifugation

Sedimentation velocity experiments were conducted at 50,000 rpm and 20 °C on a Beckman Coulter ProteomeLab XLI analytical ultracentrifuge following standard protocols (5). Samples of the ^{15}N labeled wt-EIC and unlabeled 3m-EIC were studied at concentrations ranging from ~ 3 to 25 μM in 100 mM NaCl, 20 mM Tris (pH 7.4), 2 mM DTT, and 4 mM $MgCl_2$. Samples were loaded in standard 12-mm, two-channel centerpiece cells, and data were collected using both the absorbance (280 nm) and Rayleigh interference (655 nm) optical detection systems. Time-corrected sedimentation data (6) were analyzed in SEDFIT 16.01c (7) in terms of a continuous c(s) distribution of sedimenting species with a resolution of 0.05 S and a maximum entropy regularization confidence level of 0.68. The solution density, solution viscosity, and protein partial specific volume were calculated in SEDNTERP (8) and corrected for isotopic labeling. Sedimentation coefficients s were corrected to standard conditions $s_{20,w}$.

Molecular Dynamic simulations

The X-ray structure of dimer EIC (PDB code: 2XZ7) was used as the starting structure for the simulation of wt-EIC. The 3m-EIC dimer was created by introducing the R400E, D440R and R559E mutations on both subunits of the dimer using the PyMol/Mutagenesis tool.

Three 700-ns Gaussian accelerated molecular dynamics (GaMD) simulation (9) of dimer wt-EIC, dimer 3m-EIC and monomer 3m-EIC (chain A of the 3m-EIC dimer) were performed in explicit solvent by using the Amber ff14SB force field (10) on GPU version of AMBER 16 (11).

The starting structure was centered in a truncated octahedron box and solvated with TIP3P water molecules (12) at least 10 Å away from the solute surface. Counterions were added to neutralize the system. Each system was initially minimized for 10000 steps (50000 steepest descent then 50000 conjugate gradient) with the solute atoms fixed, and then another minimization was performed with all atoms relaxed. The system was slowly heated to 310K over 5 ns. Final system equilibration was achieved by a 5 ns of NVT and 5 ns of NPT to ensure that the simulated system had reached the appropriate density at 1 bar, following by a conventional MD (cMD) for 200 ns at 310 K. The GaMD module implemented in the GPU version of AMBER 16 was then applied to each system which included a 10 ns short cMD simulation used to collect the potential statistics for calculating GaMD acceleration parameters, a 50 ns equilibration after adding the boost potentially, and finally a GaMD production simulation with randomized initial atomic velocities. All GaMD simulations were run at the “dual-boost” level in which the total potential energies and the dihedral energies were boosted.

Periodic boundary conditions were applied for all of the simulation systems. Bonds containing hydrogen atoms were restrained with the SHAKE algorithm (13), and a 2 fs time step was used. Weak coupling to an external temperature and pressure bath was used to control both temperature and pressure (14). The electrostatic interactions were calculated using the PME (particle mesh Ewald summation) with a cutoff of 10.0 Å for long-range interactions (15). Data analysis was carried our using CPPTRAJ (16) and in-house scripts.

Generation of structural ensembles

To generate a structural ensemble representation of the monomeric and dimeric EIC tertiary fold, 700 ns aMD simulations were run in Amber 16 starting from the 3D structures

of wt-EIC and 3m-EIC as described above. In the cases of dimeric wt-EIC and 3m-EIC, 1400-ns trajectories were created by appending the 700-ns trajectory of the second subunit to the 700-ns trajectory of the first subunit of the dimer. Each trajectory was clustered to produce representative structures of the aMD with a high degree of structural diversity. Each representative structure was energy minimized, and the ensemble of representative structures was used to fit the experimental RDC data as described previously (17). In brief, back-calculation of RDCs from conformational ensembles was done using the following equation:

$$RDC_i = \sum_k D_k \left[(3\,cos^2\,\theta - 1) + \frac{3}{2}(sin^2\theta \cos 2\Phi)\right] \quad (5)$$

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 is 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. D_k, θ and ϕ were optimized to reduce the discrepancy between experimental and back-calculated RDCs using the MATLAB script downloadable at http://group.chem.iastate.edu/Venditti/downloads.html.

The consistency between experimental and back-calculated RDC data was evaluated in terms of R-factor:

$$R - factor = \sum_i \sqrt{\left(RDC_i^{exp} - RDC_i^{calc}\right)^2 / \left(2RDC_i^{exp\,2}\right)} \quad (6)$$

where RDC_i^{exp} and RDC_i^{calc} are the experimental and back-calculated RDC for residue *i*, respectively. The protocol was iterated by increasing the number of clusters (and therefore the representative structures in the pool) until a stable R-factor was obtained.

Supporting Figures

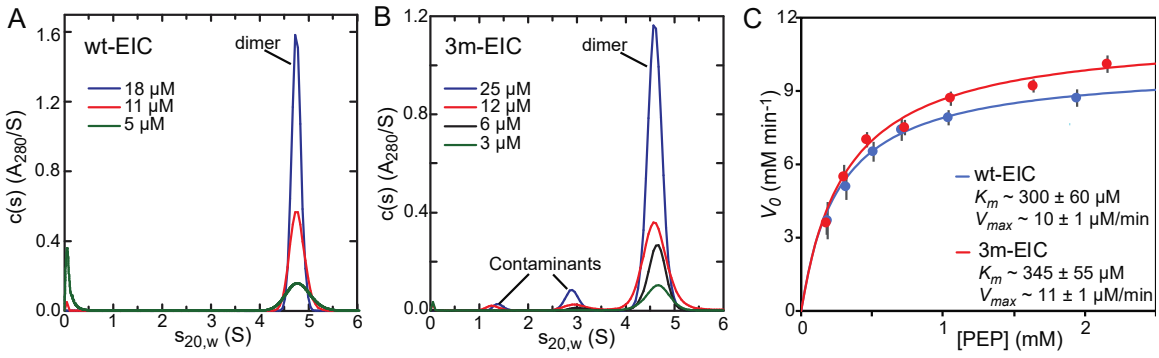

Figure S1. 3m-EIC retains the oligomeric state and activity of wt-EIC at 1 bar. c(s) distributions for (**A**) ^{15}N wt-EIC and (**B**) 3m-EIC obtained at different loading concentrations (ranging from 25 to 3 µM) based on sedimentation velocity absorbance data collected at 50,000 revolutions per minute and 20 °C (see *SI Materials and Methods*). The sedimentation experiments indicate that wt-EIC and 3m-EIC are fully dimeric within the tested concentration range. Peaks at $s_{20,w}$ < 4 S do not show concentration dependent c(s) absorbance profiles (i.e. they do not report on the monomer–dimer equilibrium) and are attributed to small amounts of contaminants in the 3m-EIC AUC sample. (**C**) Michaelis-Menten kinetics obtained for the PEP hydrolysis reaction catalyzed by 50 µM wt-EIC (blue circles) or 3m-EIC (red circles) at 50 °C and 1 bar. Modelling of the experimental data (solid curves) returns the Michaelis constant (K_m) and maximum velocity (V_{max}) of the enzymatic reaction.

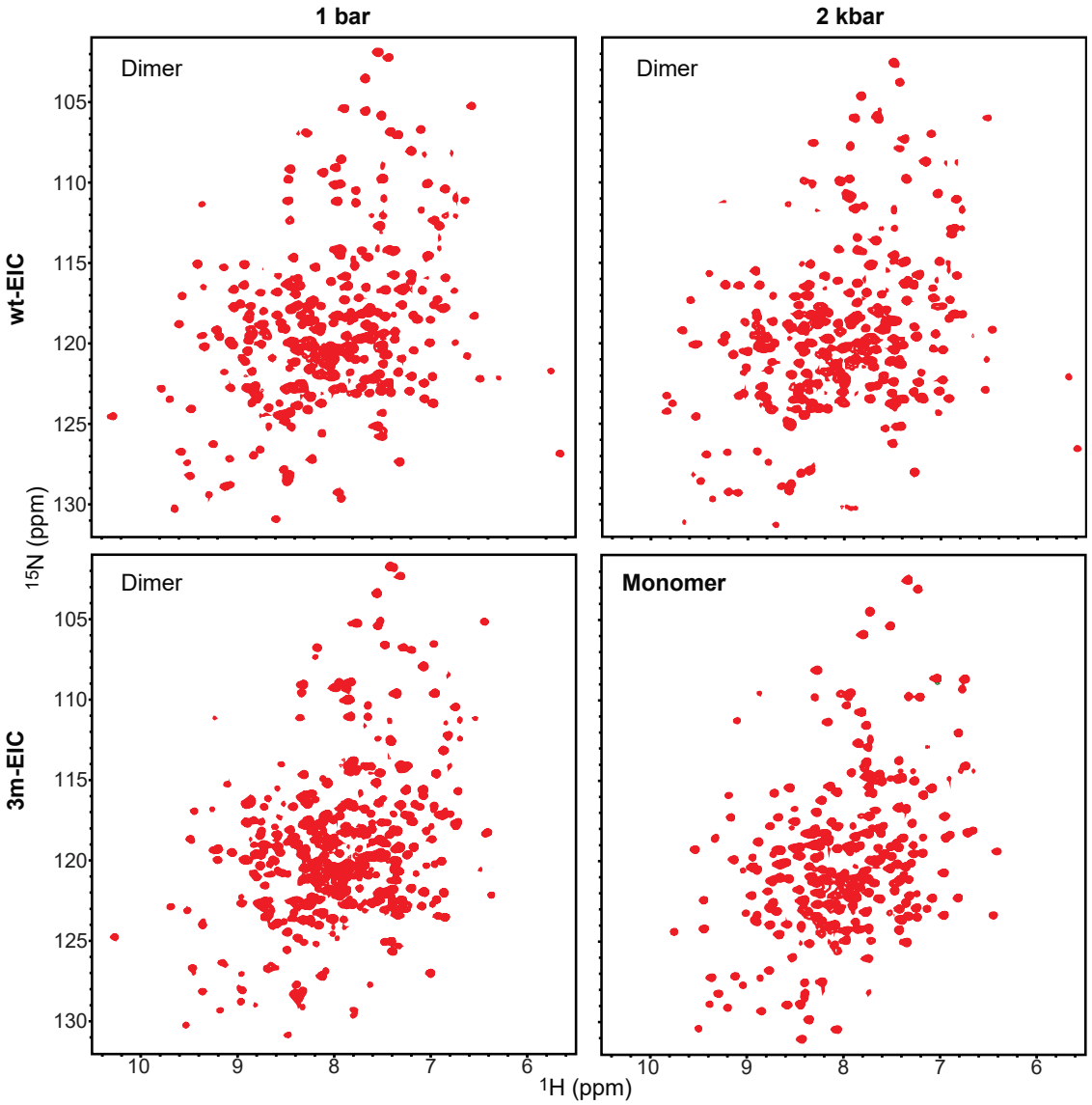

Figure S2. A folded 3m-EIC monomer at 2 kbar. 700 MHz ^{1}H-^{15}N TROSY spectra of ^{15}N-labeled wt-EIC (top) and 3m-EIC (bottom) at 1bar (left) and 2 kbar (right).

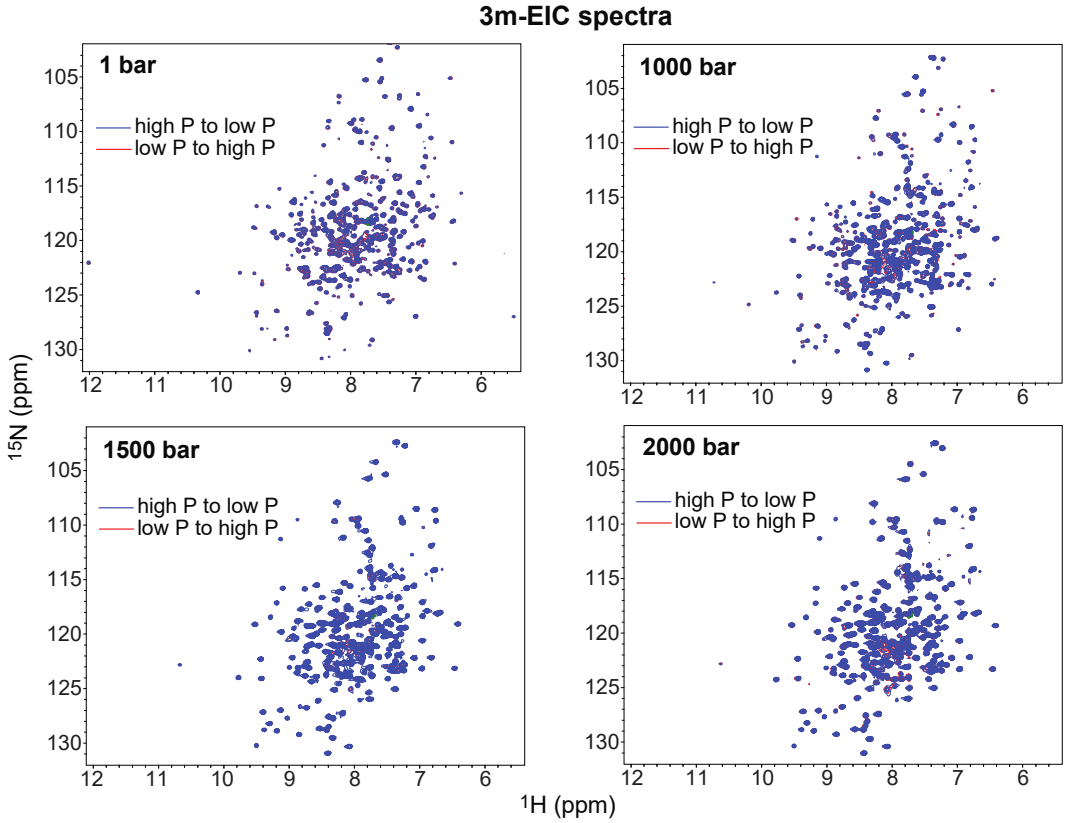

Figure S3. Reversibility of the NMR chemical shifts in the pressure titration experiment. NMR spectra of ^{15}N-labeled 3m-EIC were acquired at 1 (top left), 1000 (top right), 1500 (bottom left), and 2000 (bottom right) bar. Red spectra were measured while increasing the external pressure from 1 to 2500 bar. The blue spectra were measured while decreasing the pressure from 2500 to 1 bar. The perfect overlap between the red and blue spectra at every tested pressure indicates that the chemical shifts of the monomeric and dimeric 3m-EIC are fully reversible during the pressure titration experiment.

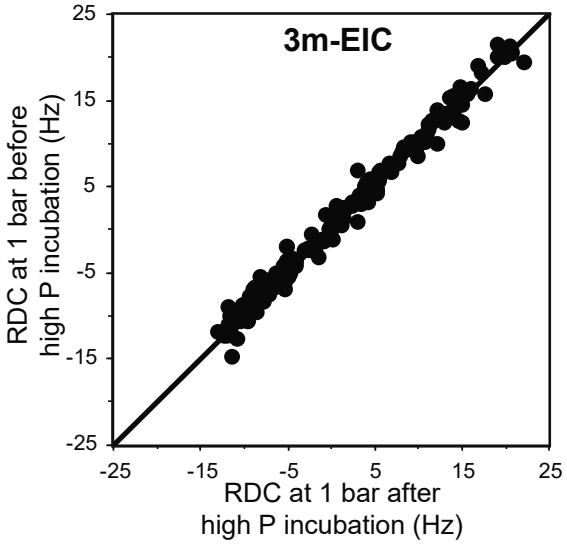

Figure S4. Reversibility of the backbone amide RDCs in the pressure titration experiment. The backbone amide RDC measured at 1 bar on a fresh sample of 3m-EIC are plotted versus the RDC measured at 1 bar after dissociation of the dimer at 2500 bar and reconstitution of the oligomeric species at atmospheric pressure. The two datasets are highly correlated indicatingthe 3m-EIC dimer adopts the same tertiary and quaternary folds in the two samples.

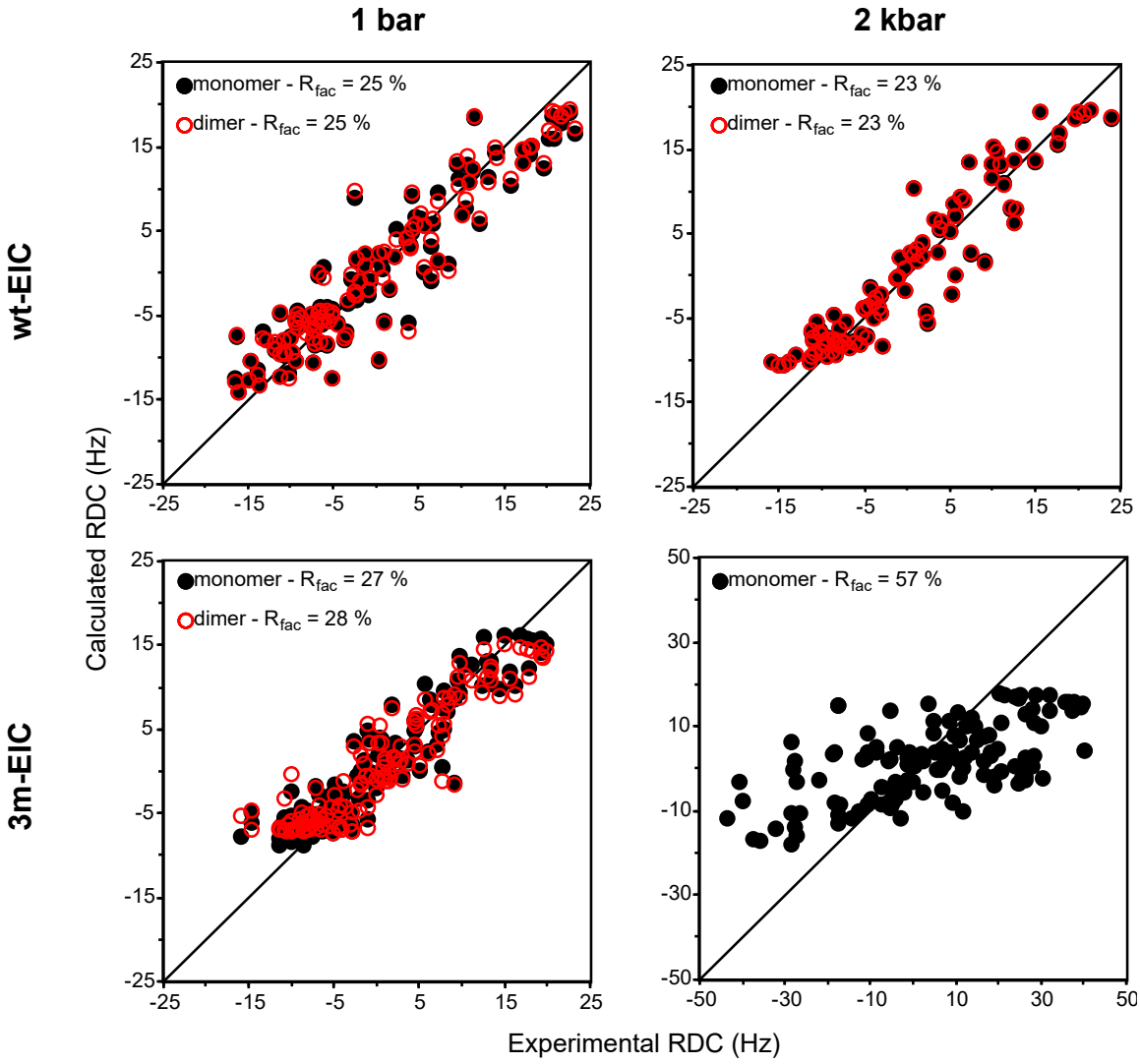

Figure S5. The tertiary fold of 3m-EIC is affected by monomerization. SVD fit of the experimental $^{1}D_{NH}$ RDC data acquired for well-defined secondary structures of 3m-EIC (top) and wt-EIC (bottom) at 1 bar (left) and 2 kbar (right) to the coordinates of the X-ray structure of wt-EIC (PDB code: 2XZ7). Fits to a single subunit and to the full dimer are shown as filled black and open red circles, respectively. More details are reported in Table S2. Note that an ensemble representation of monomeric 3m-EIC that satisfies the experimental RDC data measured at 2 kbar (bottom right panel) shows that monomerization of the enzyme induces structural and dynamic changes in the active site

loops and, at a minor extent, at the C-terminal end, but does not perturb the secondary structure elements of EIC (Fig. 2C). Therefore, the poor correlation between experimental and back-calculated data observed in the bottom right panel is likely due to the inability to fit the experimental RDCs using a single alignment tensor.

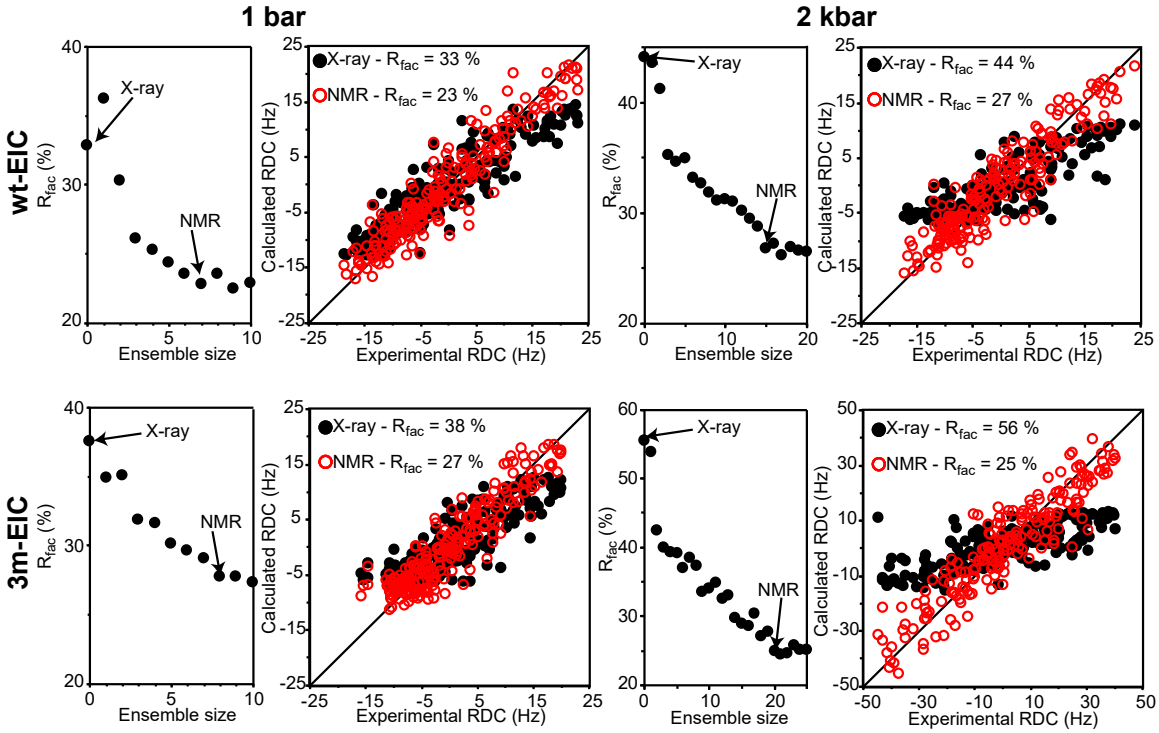

Figure S6. Ensemble refinement results. The results of the ensemble refinement are shown for wt-EIC (top) and 3m-EIC (bottom) at 1 bar (left) and 2 kbar (right). The plot of the R-factor versus ensemble size was used to determine the smallest ensemble size needed to obtain a stable R-factor. Ensemble sizes of 7, 15, 8 and 20 were needed to satisfy the solution NMR data for wt-EIC at 1bar, wt-EIC at 2 kbar, 3m-EIC at 1 bar and 3m-EIC at 2 kbar, respectively. The R-factor to the X-ray structure of wt-EIC is displayed at ensemble size = 0. The correlation between the experimental RDCs and the RDCs back-calculated from the best conformational ensemble (open red circles) or from the reference X-ray structure (filled black circles) is also shown. Note that all RDC data (including the ones from flexible loops) are included in these calculations. More details are reported in Table S2.

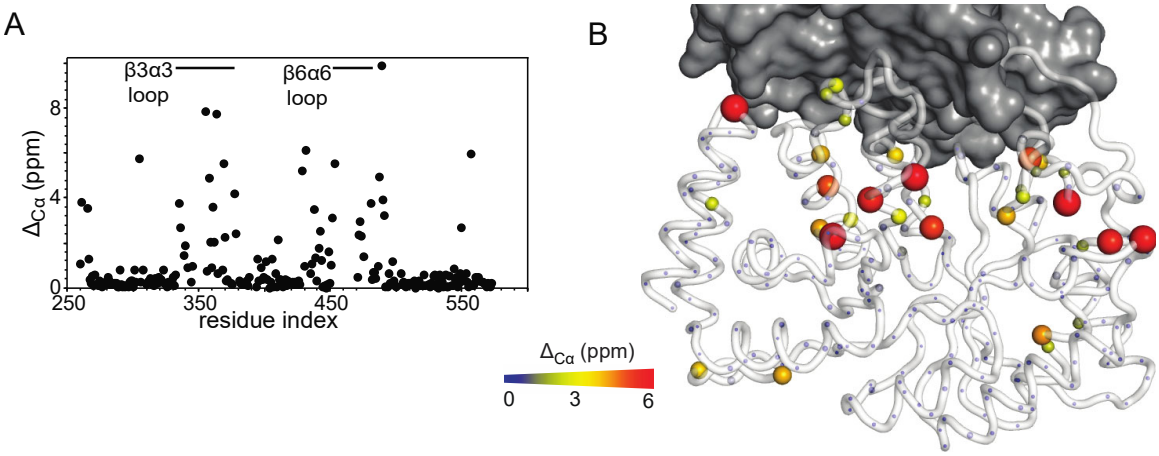

Figure S7. Chemical shift changes upon monomerization. (**A**) Changes in Cα chemical shift induced in 3m-EIC by increasing pressure from 1 to 2000 bar. (B) The changes in Cα chemical shift are plotted on the crystal structure of the EIC dimer according to the color bar. The second subunit of the EIC dimer is shown as a gray surface.

Supporting Tables

Table S1. Thermodynamics of the 3m-EIC dimer-to-monomer equilibrium

	3m-EIC	3m-EIC + 20 mM αKG	3m-EIC + 20 mM PEP
ΔG (kcal/mol)[a]	5.0 ± 0.1	6.6 ± 0.1	7.3 ± 0.2
ΔV (ml/mol)	-153 ± 2	-152 ± 2	-146 ± 3
K_d (M)	$9.10^{-11} \pm 2.10^{-13}$	$5.10^{-13} \pm 7.10^{-15}$	$6.10^{-14} \pm 3.10^{-15}$

[a] This ΔG value corresponds to the free energy required to transition one subunit of EIC from the dimeric to the monomeric species.

Table S2. Backbone amide $^1D_{NH}$ RDC analysis

Used RDC	Number of RDCs	D_a^{NH} (Hz)	Rhombicity	R-factor (%)
SVD fit to the X-ray structure (PDB code 2XZ7)				
wt-EIC 1 bar (fit to monomer)	115	9.5	0.36	25
wt-EIC 1 bar (fit to dimer)	230	9.7	0.34	25
wt-EIC 2 kbar (fit to monomer)	95	9.8	0.12	23
wt-EIC 2 kbar (fit to dimer)	190	9.8	0.12	23
3m-EIC 1 bar (fit to monomer)	137	8.0	0.13	27
3m-EIC 1 bar (fit to dimer)	274	7.5	0.02	28
3m-EIC 2 kbar (fit to monomer)	129	9.9	0.51	57
Ensemble refinement (only one subunit was considered)				
wt-EIC 1 bar (fit to X-ray)	208	7.2	0.56	33
wt-EIC 1 bar (fit to ensemble)	208	ND	ND	23
wt-EIC 2 kbar (fit to X-ray)	190	5.6	0.15	44
wt-EIC 2 kbar (fit to ensemble)	190	ND	ND	27
3m-EIC 1 bar (fit to X-ray)	239	6.3	0.15	38
3m-EIC 1 bar (fit to ensemble)	239	ND	ND	27
3m-EIC 2 kbar (fit to X-ray)	196	7.8	0.44	56
3m-EIC 2 kbar (fit to ensemble)	196	ND	ND	25

Supporting References

1. Delaglio F*, et al.* (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3):277-293.
2. Dotas RR & Venditti V (2019) Resonance assignment of the 128 kDa enzyme I dimer from Thermoanaerobacter tengcongensis. *Biomol NMR Assign* 13(2):287-293.
3. Clore GM & Gronenborn AM (1998) Determining the structures of large proteins and protein complexes by NMR. *Trends in Biotechnology* 16(1):22-34.
4. Dotas RR*, et al.* (2020) Hybrid Thermophilic/Mesophilic Enzymes Reveal a Role for Conformational Disorder in Regulation of Bacterial Enzyme I. *Journal of Molecular Biology* 432(16):4481-4498.
5. Zhao H, Brautigam CA, Ghirlando R, & Schuck P (2013) Overview of Current Methods in Sedimentation Velocity and Sedimentation Equilibrium Analytical Ultracentrifugation. *Current Protocols in Protein Science* 71(1):20.12.21-20.12.49.
6. Ghirlando R*, et al.* (2013) Improving the thermal, radial, and temporal accuracy of the analytical ultracentrifuge through external references. *Analytical Biochemistry* 440(1):81-95.
7. Schuck P (2000) Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophysical Journal* 78(3):1606-1619.
8. Cole JL, Lary JW, P Moody T, & Laue TM (2008) Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. *Methods Cell Biol* 84:143-179.
9. Miao Y, Feher VA, & McCammon JA (2015) Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *Journal of Chemical Theory and Computation* 11(8):3584-3595.
10. Maier JA*, et al.* (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* 11(8):3696-3713.

11. Case DA*, et al.* (2005) The Amber biomolecular simulation programs. *Journal of Computational Chemistry* 26(16):1668-1688.
12. Mark P & Nilsson L (2001) Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *The Journal of Physical Chemistry A* 105(43):9954-9960.
13. Ryckaert J-P, Ciccotti G, & Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* 23(3):327-341.
14. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, & Haak JR (1984) Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* 81(8):3684-3690.
15. Essmann U*, et al.* (1995) A smooth particle mesh Ewald method. *The Journal of Chemical Physics* 103(19):8577-8593.
16. Roe DR & Cheatham TE (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* 9(7):3084-3095.
17. Purslow JA*, et al.* (2018) Active Site Breathing of Human Alkbh5 Revealed by Solution NMR and Accelerated Molecular Dynamics. *Biophysical Journal* 115(10):1895-1905.