PNAS

www.pnas.org

Supplementary Information for

Actinobacteria challenge the paradigm:

a unique protein architecture for a well-known central metabolic complex

Eduardo M. Bruch, Pierre Vilela, Lu Yang, Alexandra Boyko, Norik Lexa-Sapart,

Bertrand Raynal, Pedro M. Alzari, Marco Bellinzoni

Correspondence to E.M. Bruch and M. Bellinzoni
Email: eduardobruch@gmail.com and marco.bellinzoni@pasteur.fr

This PDF file includes:

Figures S1 to S15
Tables S1 to S4
SI References
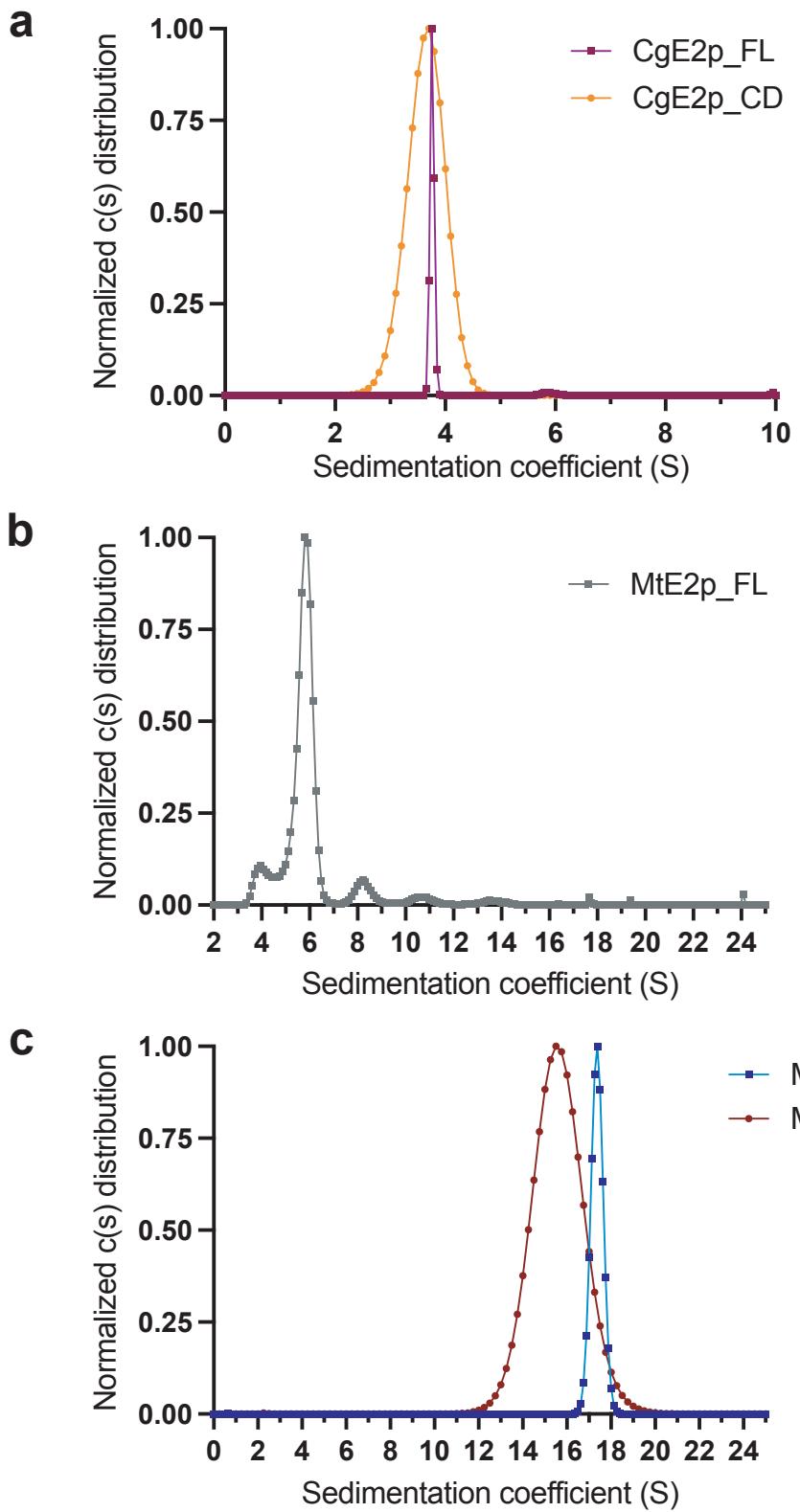
Figure S1. Analytical ultracentrifugation profiles.
Normalized sedimentation coefficient distributions c(s) are shown for each of the studied samples: **a)** *Cg*E2p_FL (violet) *vs*. *Cg*E2p_CD (orange); **b)** *Mt*E2p_FL (grey); **c)** *Mt*E2b_FL (blue) *vs*. *Mt*E2b_CD (brown). Molecular mass estimations are provided in Table 1.

a

$$\text{lipoamide + NADH} \underset{\longleftarrow}{\overset{\text{DLD}}{\longrightarrow}} \text{dihydrolipoamide + NAD}^+$$

acyl-CoA

E2

CoA-SH

S-acyldihydrolipoamide

b

acetyl-CoA added
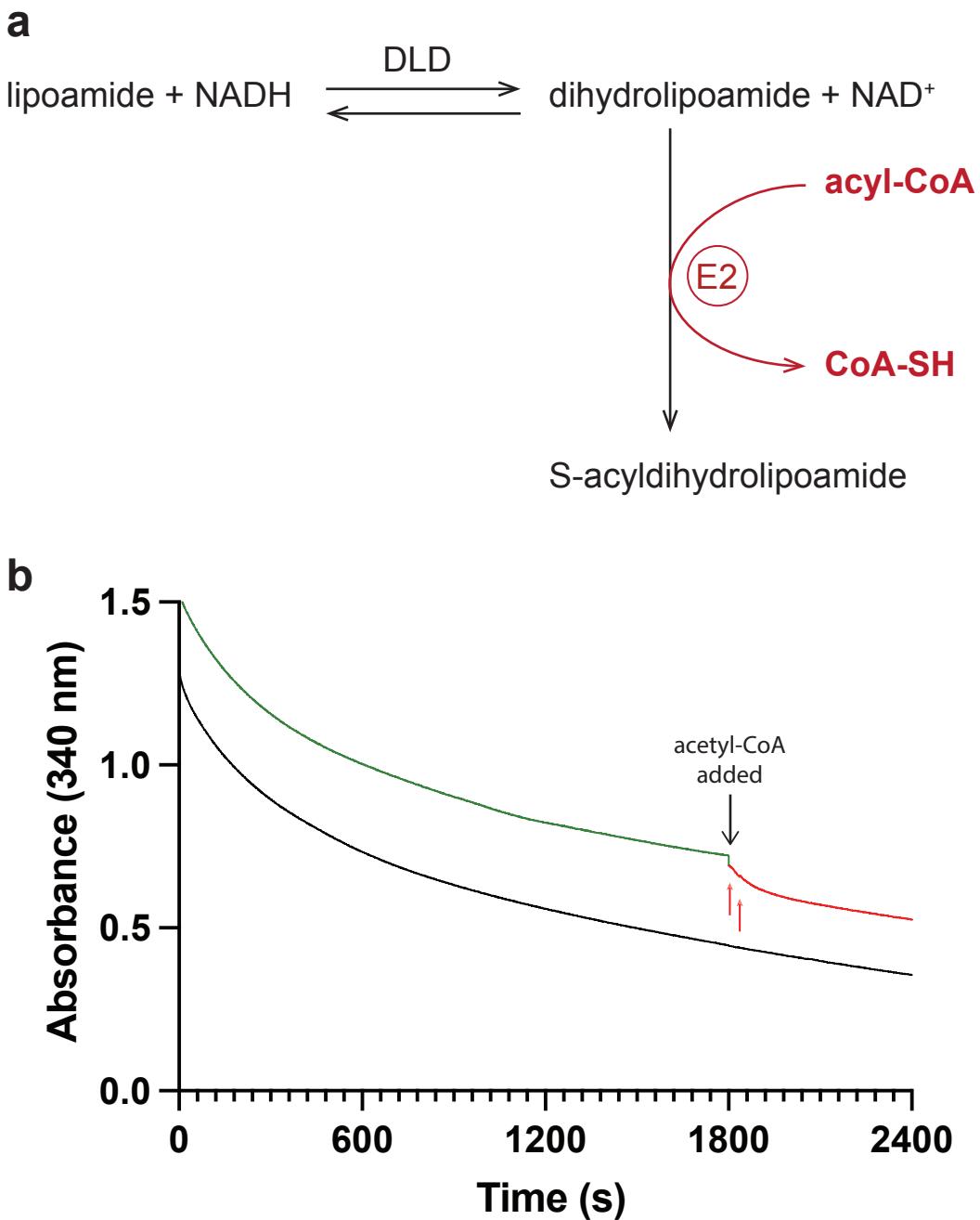
Figure S2. Transacylase assay.
a) Schema of the transacylase activity assay from the work by Hall and Weitzman (1). The assay measures the backward conversion of acyl-CoA to CoA-SH in the presence of dihydrolipoamide, in turn generated by commercial bovine dihyrolipoamide dehydrogenase (DLD) through a reservible reaction in the presence of oxidized lipoamide and NADH. Acetyl-CoA was provided as substrate in the case of *Cg*E2p and *Mt*E2p, while isobutyryl-CoA was employed for *Mt*E2b (see Materials and Methods). **b)** Representative trace of an absorbance measurement performed on *Cg*E2p_FL. After 30 min preincubation to reach a linear NADH consumpion rate (green curve), the addition of 0.15 mM acetyl-CoA triggered the E2p catatalyzed transacylase reaction (red). The red arrows beneath indicate the initial, linear range in absorbance decrease that was considered for further calculations, typically in the 1 min range. All assays were perfomed in triplicates; the background NADH consumption rate was then subtratcted to infer substrate conversion rate and E2p specific activities. The black curve represents a negative control measurement where no acetyl-CoA was provided.
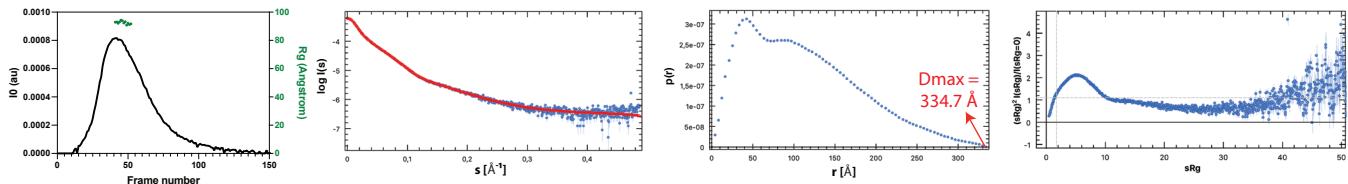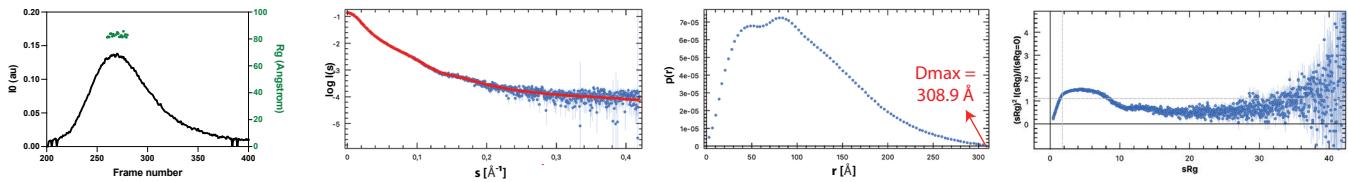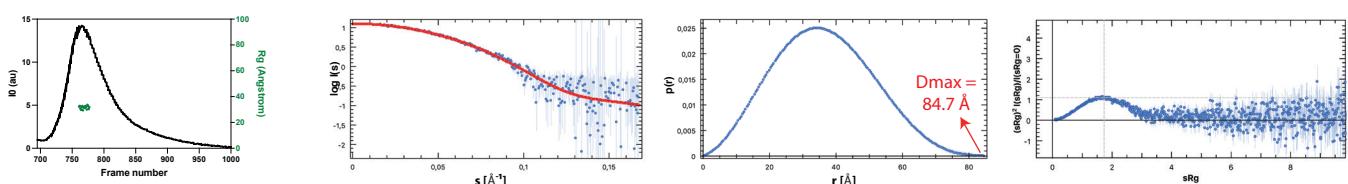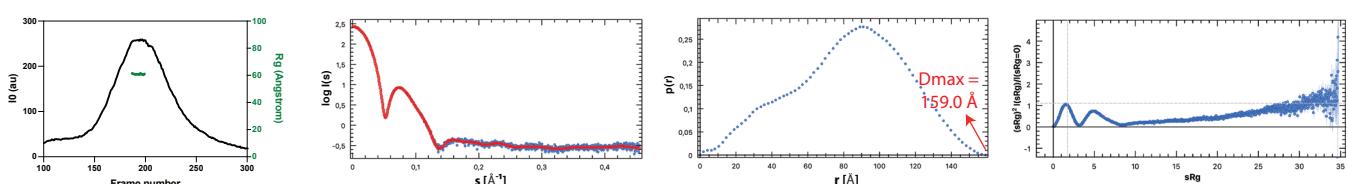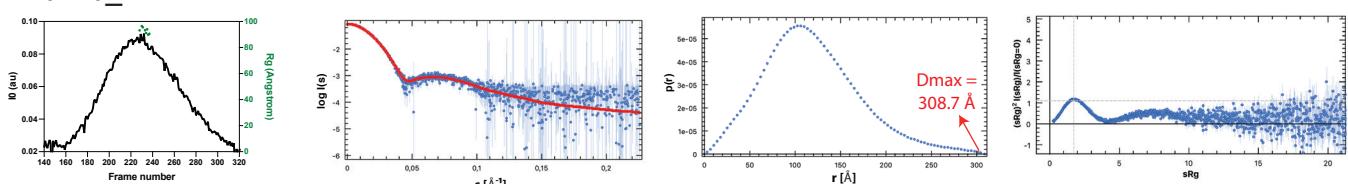
Figure S3. SAXS primary data analysis.
From left to right**:** scattering intensity per recorded frame (black) and extrapolated Rg (green); experimental scattering intensity (blue) and corresponding fitting curve (red); pair distance distribution function showing the estimated Dmax values; dimensionless Kratky plot. In the latter, dashed lines indicate the position were a globular protein maximum is predicted to be located ($sRg=\sqrt{3}$ and $(sRg)^2 I(s)/I(0) = 1.104$) (2,3).
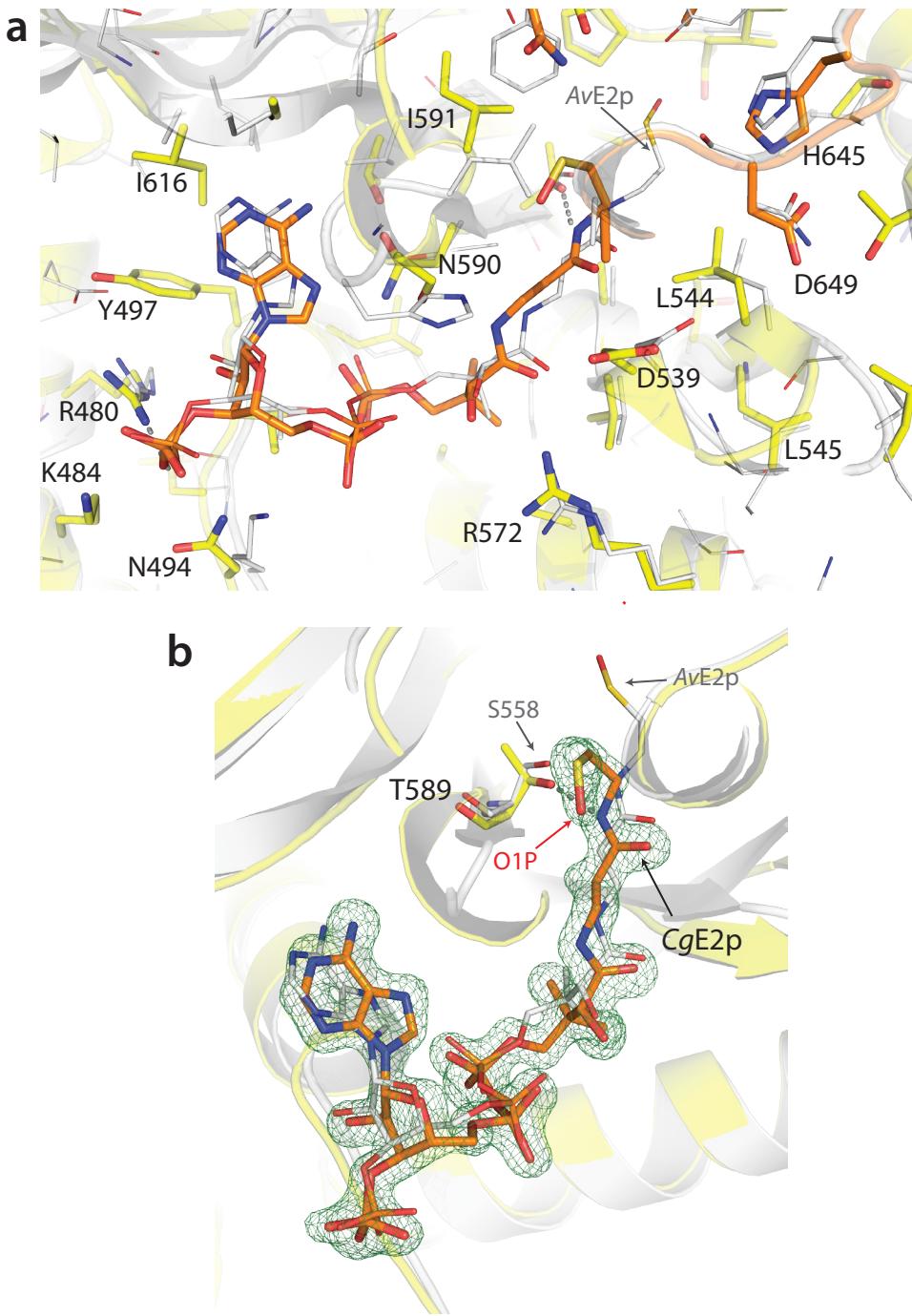
Figure S4. Comparison of the bound CoA conformation within the active site of *Cg*E2p_CD and *Av*E2p_CD.

a) Superposition of *Cg*E2p (yellow) and *Av*E2p (gray) active sites. Residue numbering corresponds to *Cg*E2p. The CoA ligand is visible in the center (orange for *Cg*E2p and gray for *Av*E2p). **b**) CoA binding site, showing the ligand density for CoA bound to *Cg*E2p ('polder' omit map (4) contoured at the 4σ level, in blue). The terminal thiol group of CoA is found in a non-equivalent position to *Av*E2p_CD (5), and rather points to the opposite direction due to a reorientation of the cysteamine moiety close to 180° around N4P. Furthermore, the hydroxyl group of Thr589 (equivalent to Ser558 in *Av*E2p_CD) can form a H-bond with the N4P atom from CoA to stabilize the negatively charged tetrahedral intermediate. Fourier difference electron density maps are compatible with the presence of an oxygen atom at covalent binding distance (1.7 Å) from the CoA sulfur (O1P; red arrow), suggesting the presence of, at least, a partial oxidation of the CoA thiol group to sulfenic acid (PDB 6ZZJ, this work), as observed in the *Av*E2p_CD complex (PDB 1EAD) (5).
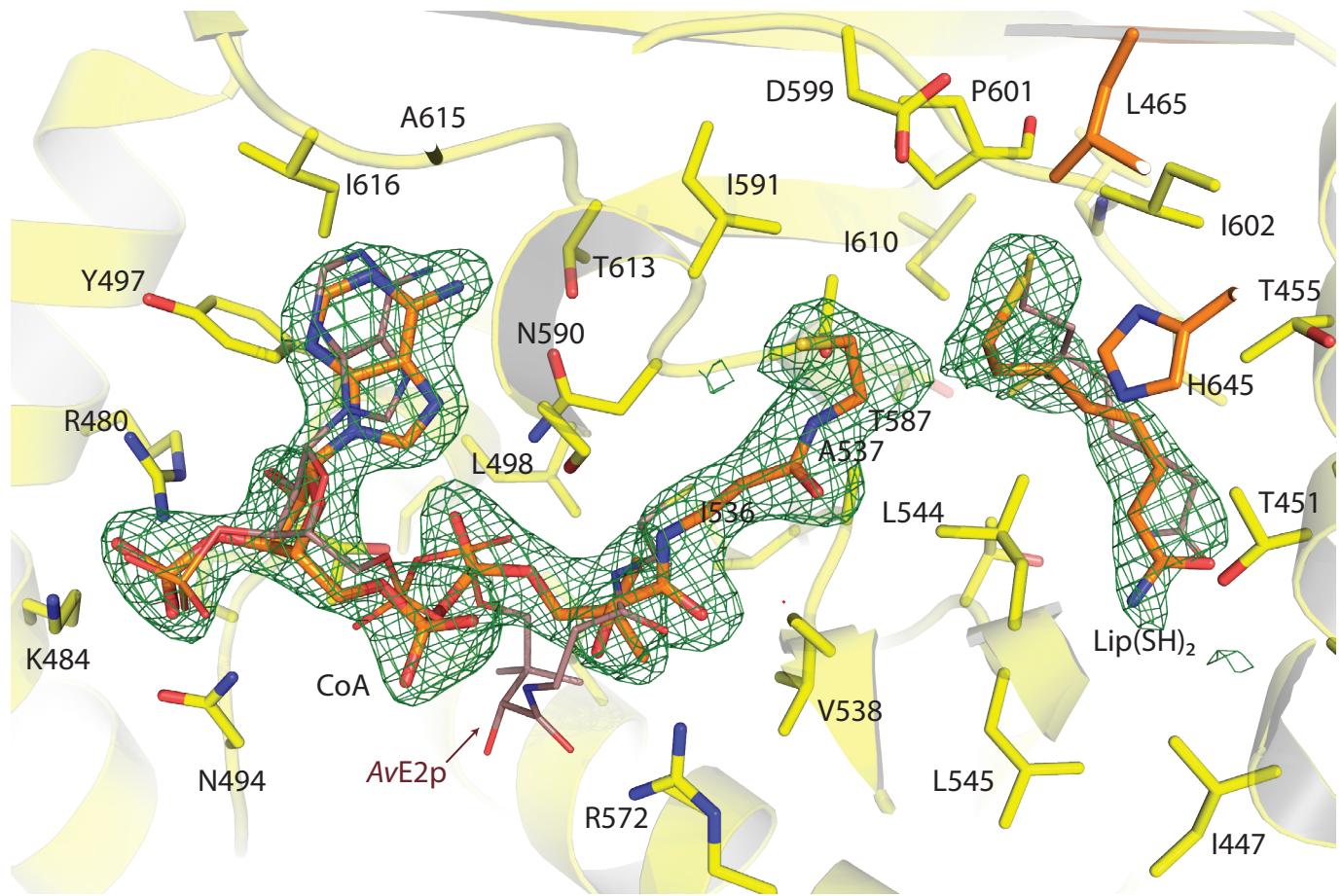
Figure S5. Comparison of the ternary complex from *Cg*E2p_CD and *Av*E2p_CD.
Superposition of the ternary complex obtained for *Cg*E2p (PDB 6ZZK) and *Av*E2p with CoA and lipoamide (PDB 1EAB) (5). *Cg*E2p is depicted in yellow, side chains of active site residues are shown as sticks (and numbered), while ligands are depicted with orange carbons. The green mesh represents the 'polder' omit map (4) contoured at the 4σ level. The *Av*E2p ligands are displayed as thinner, brown sticks (protein atoms from the *Av*E2p enzyme are not shown).

For *Av*E2p_CD, the ternary complex with CoA and Lip(SH)$_2$ presented CoA with an abortive, 'OUT' conformation in which the pantetheine chain did not reach the active site cleft but rather formed a left-handed helix with a series of intramolecular H-bonds (PDB 1EAB) (5). In the corynebacterial enzyme, despite a not strictly equivalent pose, Lip(SH)$_2$ makes active site interactions analogous to the ones reported in the case of *Av*E2p_CD, most notably the H-bond between the reactive sulphur (S8) and His645' provided by the opposite monomer, while the other sulphur atom (S6) is engaged in an H-bond with the carbonyl oxygen of Ile602, another conserved residue. The amide moiety is stabilized by H-bonds involving Thr451, the carbonyl group of Leu544 and the hydroxyl of a Ser residue preceding Leu437, left over from TEV digestion of the N-terminal His$_6$ affinity tag and forming crystal packing interactions. The other *Cg*E2p_CD monomer in the asymmetric unit, which forms a separate homotrimer via a crystallographic 3-fold axis, does not show any dihydrolipoamide bound, but a CoA molecule in the same orientation as the other protomer, and forming exactly the same interactions described before.
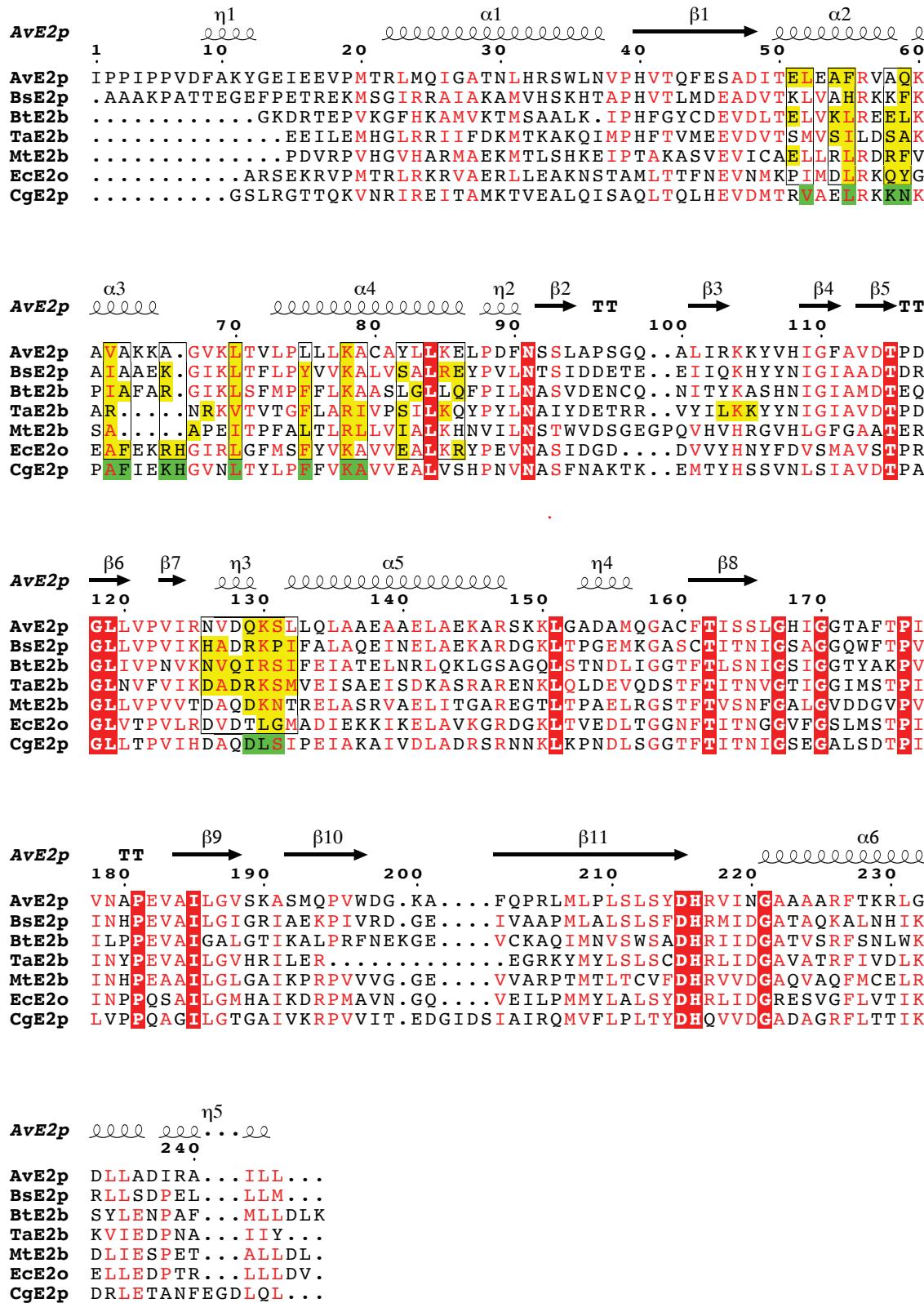
Figure S6. Sequence alignment of E2 catalytic domains of known structure.

Aligned sequences include: *A. vinelandii* E2p (*Av*E2p, PDB 1EAB); *B. stearotheromophilus* E2p (*Bs*E2p, PDB 1B5S); *Bos taurus* E2b (*Bt*E2b, PDB 2I14); *T. acidophilum* E2b (*Ta*E2b, PDB 3RQC); *M. tuberculosis* E2b (*Mt*E2b, PDB 6ZZN - this work); *E. coli* SucB/E2o (EcE2o, PDB 1SCZ); *C. glutamicum* E2p (*Cg*E2p, PDB 6ZZI - this work). Highlited in yellow, residues involved in the TTI helix interactions, as determined by PISA (6) (https://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver); boxes show regions involved in these interactions in at least two sequences. In the case of *Bs*E2p only, a new model was generated by Swiss model (7) (https://swissmodel.expasy.org/interactive) starting from PDB 1B5S, in order to complete the coordinate set with the missing side chain atoms. As a reference, secondary structure features and residue number of *Av*E2p (PDB 1EAB) (5) were superimposed. Figure generated by ESPript3 (8) (http://espript.ibcp.fr).
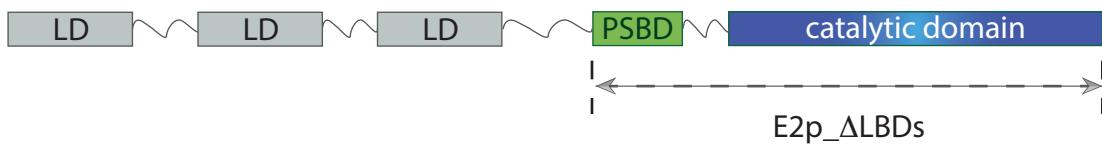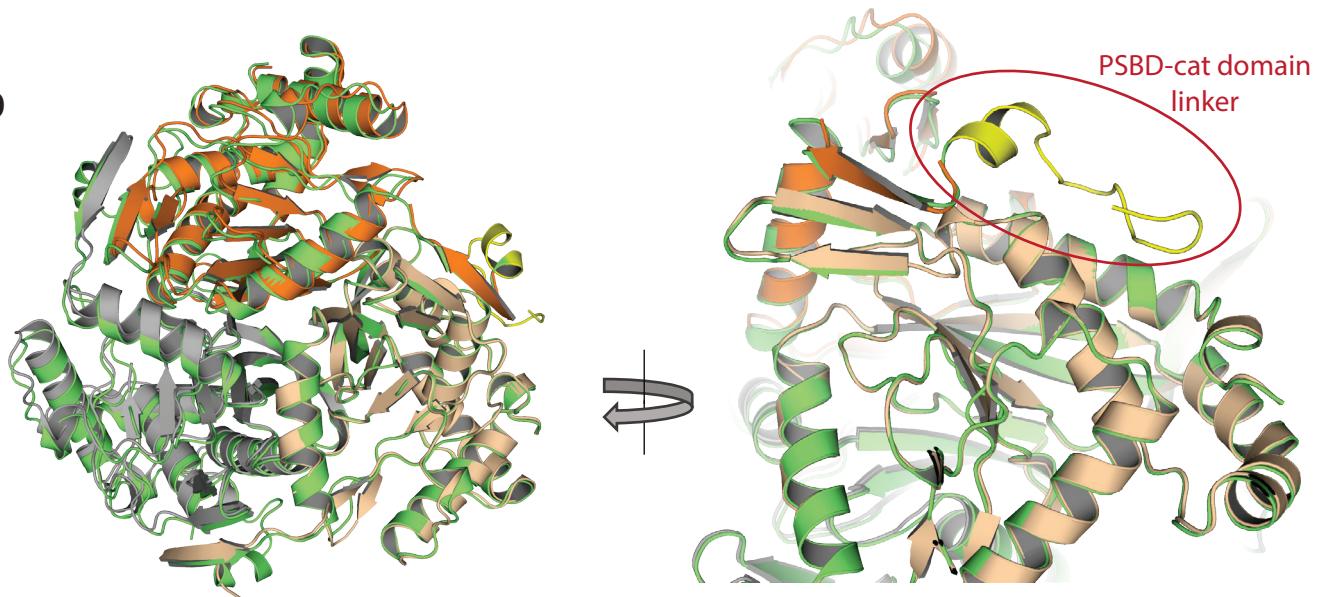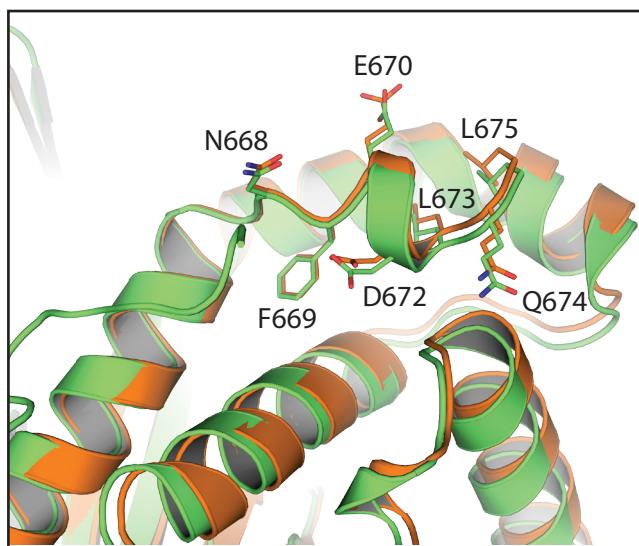
Figure S7. Crystal structure of *Cg*E2p_ΔLBDs.
a) Schematic representation of the domain architecture of *Cg*E2p_FL. The arrow indicates the region selected for *Cg*E2p_ΔLBDs construct. **b)** Front and sideview of the superposition of the structures from *Cg*E2p_CD (PDB 6ZZI) and *Cg*E2p_ΔLBDs (PDB 6ZZL), depitcted in orange and green, respectively. In yellow, the linker connecting the catalytic domain to the PSBD domain. **c)** Zoomed view on the C-terminal α-helix (TTI helix) showing the same conformation in both models.
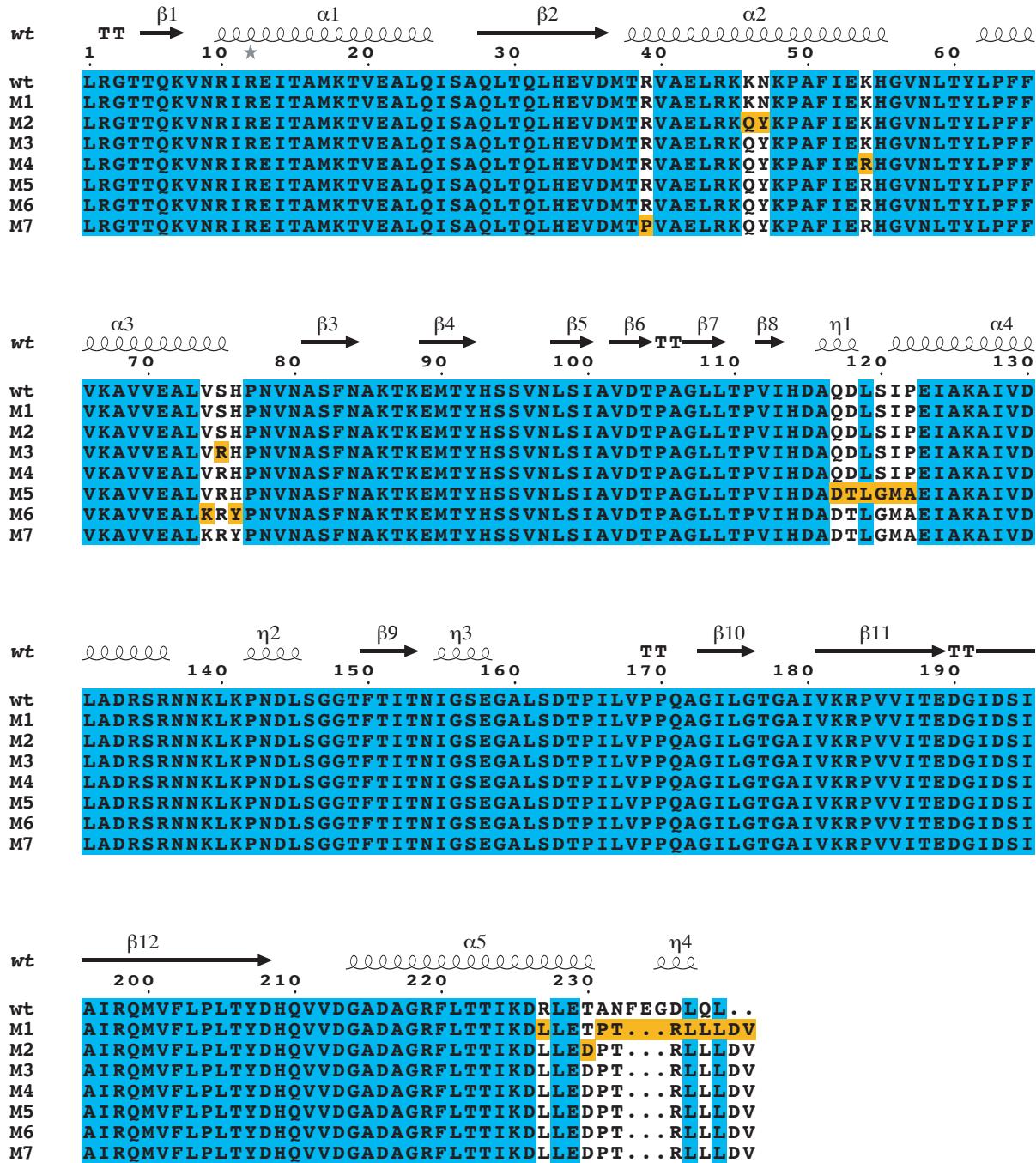
Figure S8. *Cg*E2p_CD mutant series.
Sequence alignment of the seven generated muteins of CgE2p_CD to study the effect of amino acid substitutions on protein oligomerisation. Subtitutions were introduced in an incremental way, with residues hlighlited in orange representing the first introduction with respect to the wt sequence; each substitution was then maintained in the following mutants. Cyan: identical residues. Alignment performed by MAFFT (9) and not including the N-terminal His$_6$ and TEV cleavage site, present in all constructs; figure generated with ESPript3 (8) (http://espript.ibcp.fr).
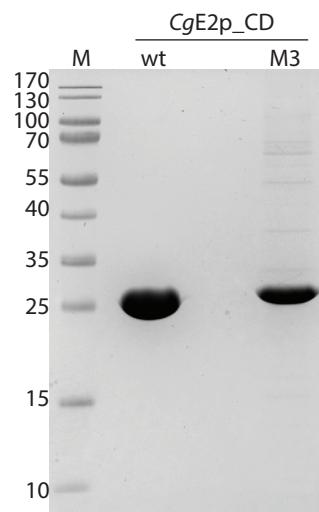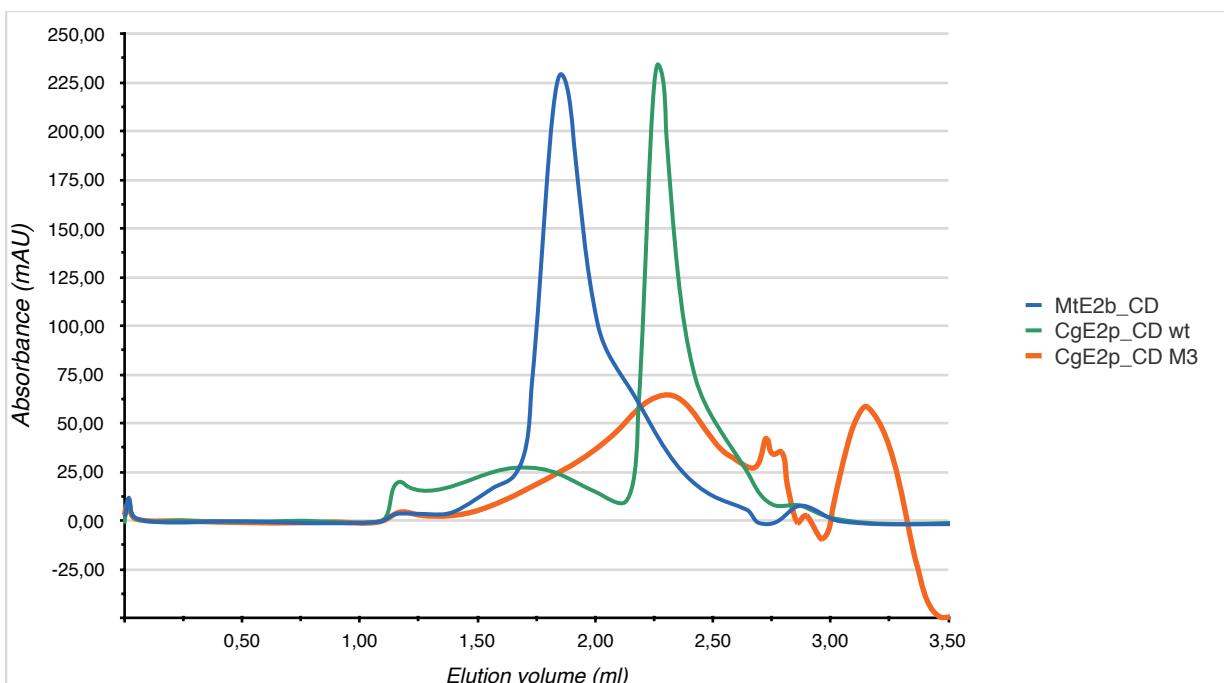
Figure S9. Purification and size-exclusion chromatography profile of _Cg_E2p_CD, wild-type _vs_. mutant M3.

a) SDS-PAGE of samples of purified wild-type _Cg_E2p_CD and its mutant version M3 (sequences provided in Suppl. Fig. 7). **b)** Representative size-exclusion chromatography profiles of _Cg_Ep_CD wild-type (green curve) and mutant M3 (orange), superimposed to _Mt_E2b_CD (blue). Samples run on a Superose 6 increase 5/150 GL column (GE Healthcare).
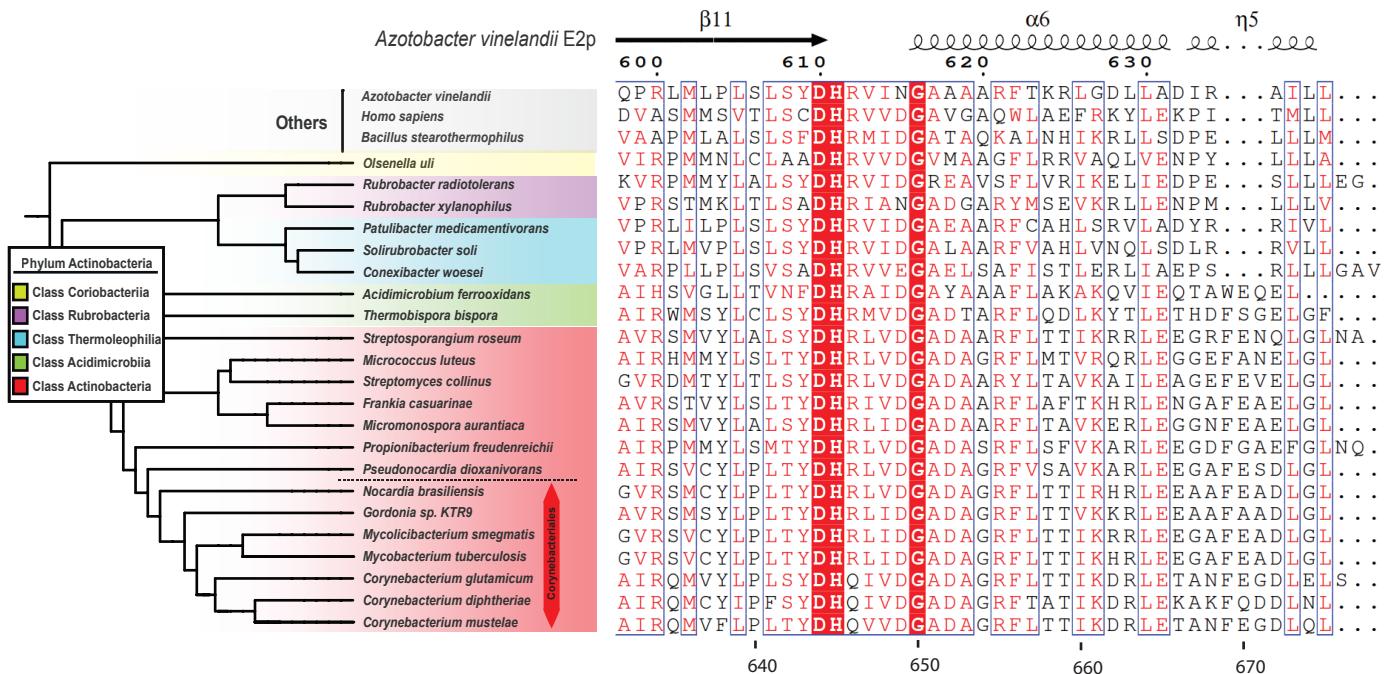
Figure S10. Analysis of the PCI distribution within Actinobacteria.

Sequence alignment of several E2 of known structure together with predicted E2 from species from the phylum Actinobacteria. Representative sequences from the phylum were selected to represent the 5 classes (Coriobacteriia, yellow; Rubrobacteria, violet; Thermoleophilia, cyan; Acidimicrobiia, yellow; Actinobacteria, red). Within the biggest classes (Actinobacteria and Actinomycetales), sequences selected represent different orders including the order *Corynebacteriales*, that include *M. tuberculosis* and *C. glutamicum*. The sequences were aligned using MUSCLE (10) and the presence of the PCI was analyzed in the region corresponding to the insertion observed in *Cg*E2p and *Mt*E2p. As a guide, secondary structure features and residue number of *Av*E2p were superimposed using ESPript3 (8) (http://espript.ibcp.fr).
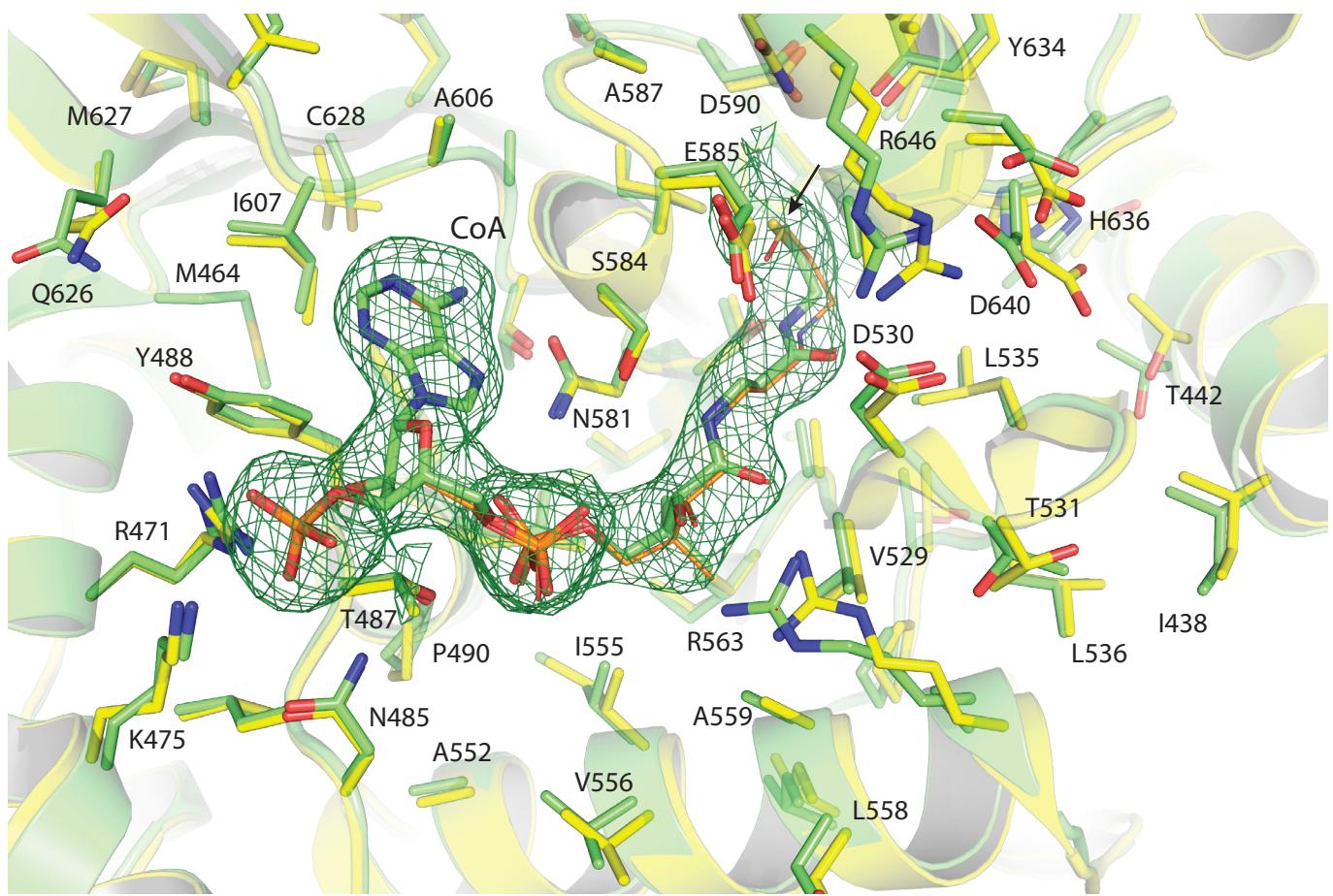
Figure S11. Comparison of the active site pockets of *Cm*E2p_CD *vs*. *Cg*E2p_CD.
Ribbon representation of *Cm*E2p_CD in complex with CoA (green, PDB 6ZZM), superimposed to *Cg*E2p_CD (yellow) in complex with oxidized CoA (thinner sticks, orange; PDB 6ZZJ). Side chains of residues in the active site environment are shown in sticks; numerotation refers to *Cm*E2p_CD. The bound CoA molecule shows the same 'IN' conformation observed in *Cg*E2p_CD, including the peculiar, bent conformation of the terminal thiol group (black arrow). The green mesh represents the 'polder' omit electron density map (4) for the ligand, contoured at the 4σ level.
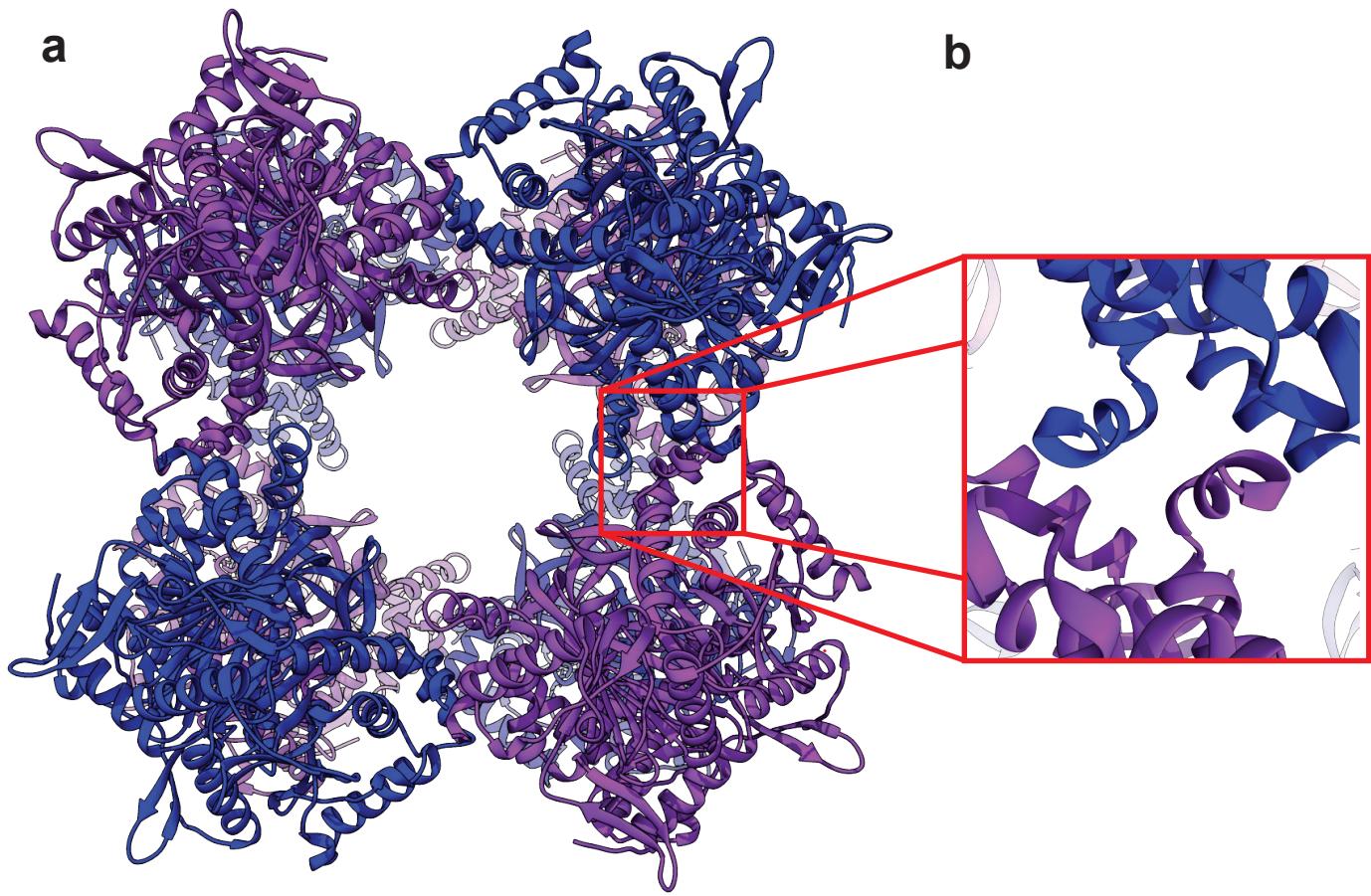
Figure S12. Crystal structure of *Mt*E2b_CD.
a) Ribbon representation of the 24-mer *Mt*E2b_CD assembly (PDB 6ZZN; this work). Protomers are related by crystallographic symmetry, the content of the asymmetric unit corresponding to a single *Mt*E2b_CD monomer (space group F 4 3 2). Eight trimers are positioned at the corners of the cube (alternating blue and violet colours). **b)** Zoomed view of the region corresponding to the inter-trimer interaction where the role of the TTI-helix, main determinant of the trimer-trimer interaction, is highligted.
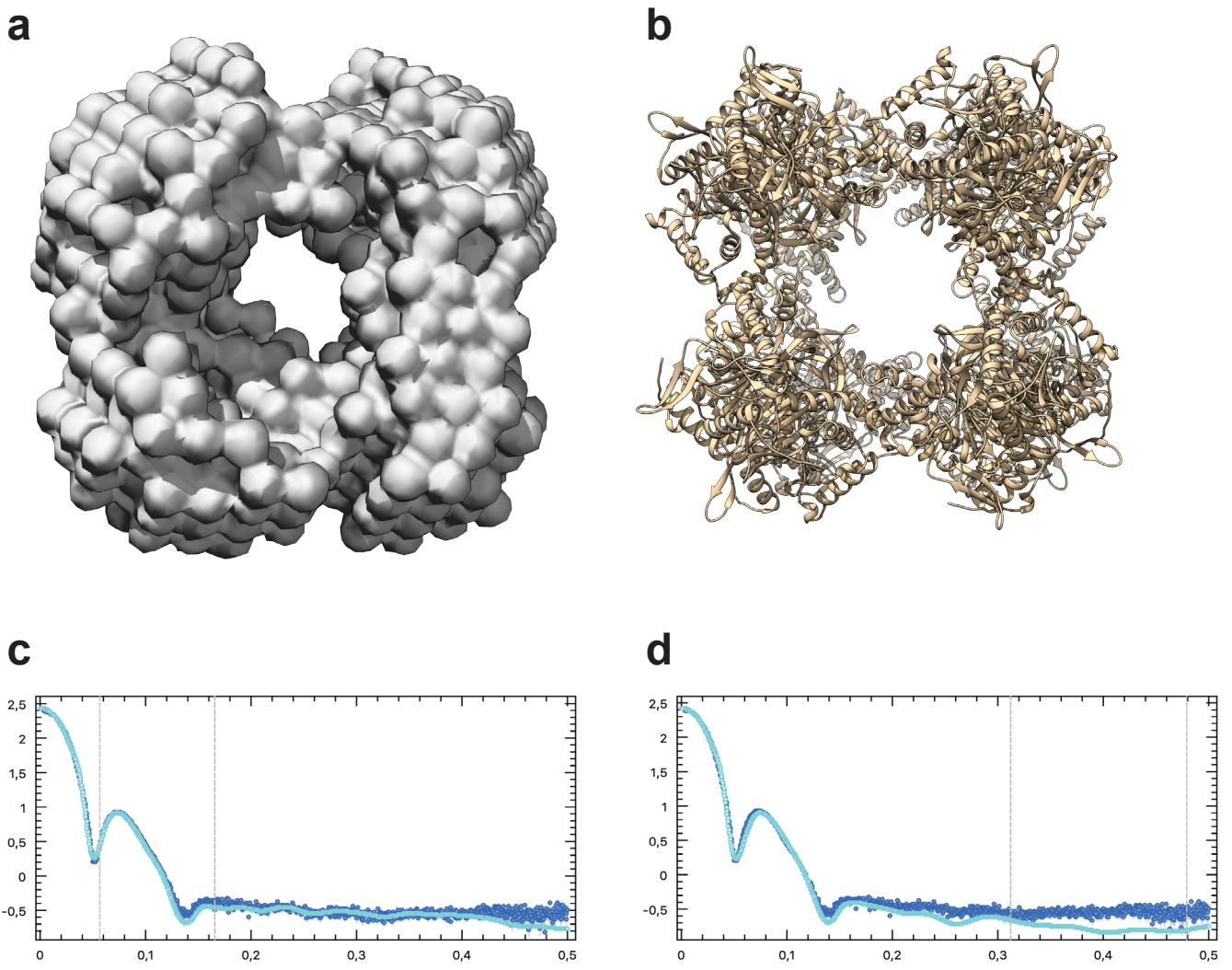
Figure S13. *Mt*E2b_CD SAXS analysis.
a) Front view from the DAMMIN (11) generated *ab-initio* model of *Mt*E2b_CD from the SAXS experimental data. **b**) Crystal structure of the same protein (PDB 6ZZN, this work; Suppl. Fig. 12). **c)** Experimental scattering intensity (blue dots) plotted and compared with the theoretical fit obtained for the DAMMIN model (cyan curve). **d)** Experimental (blue dots) and calculated scattering intensity (cyan) from the crystallographic model, obtained using CRYSOL (12).
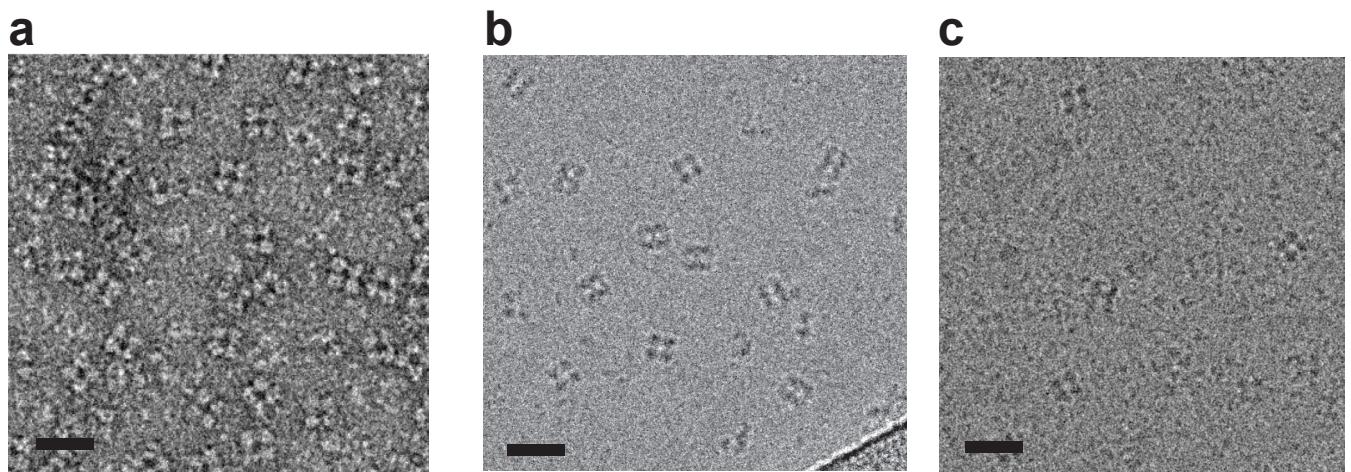
Figure S14. Electron microscopy visualization of individual _Mt_E2b particles.

a_)_ Negatively stained EM grid prepared using uranyl acetate as the staining agent, showing cubic particles of _Mt_E2b_CD. (**b,c)** - EM grid prepared in cryo condition showing cubic particles of _Mt_E2b_CD (**b**) and _Mt_E2b_FL (**c**), respectively. The scale bar represents 20 nm in all cases.
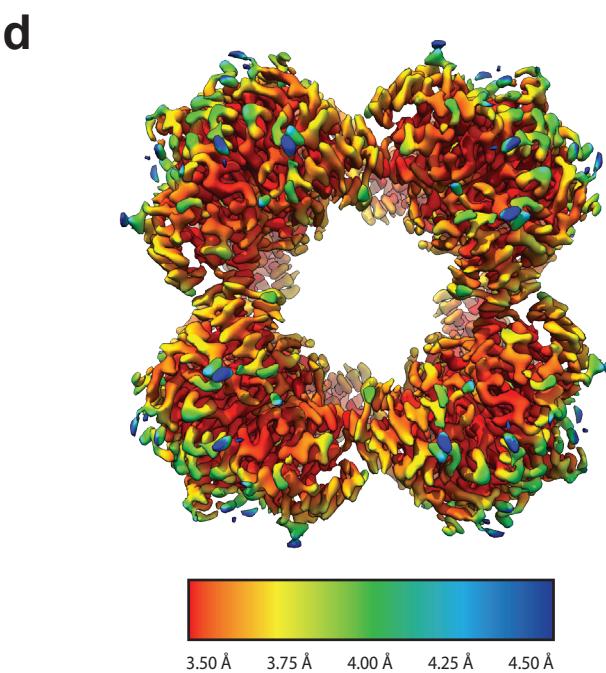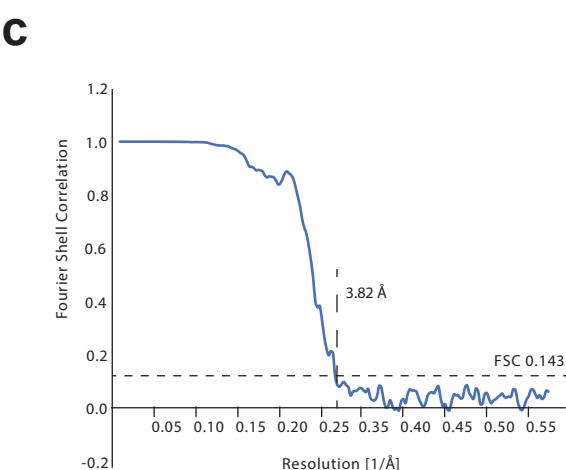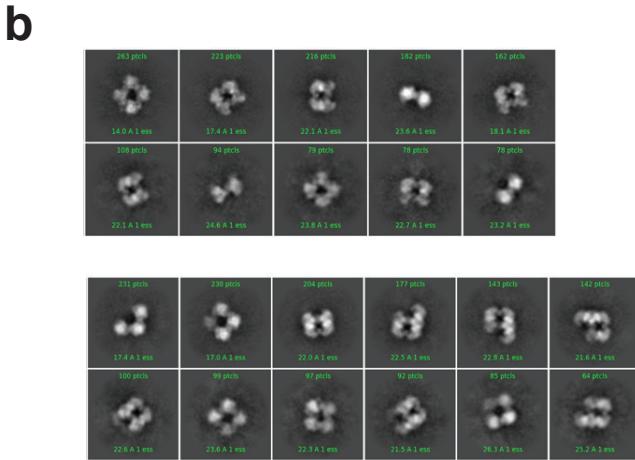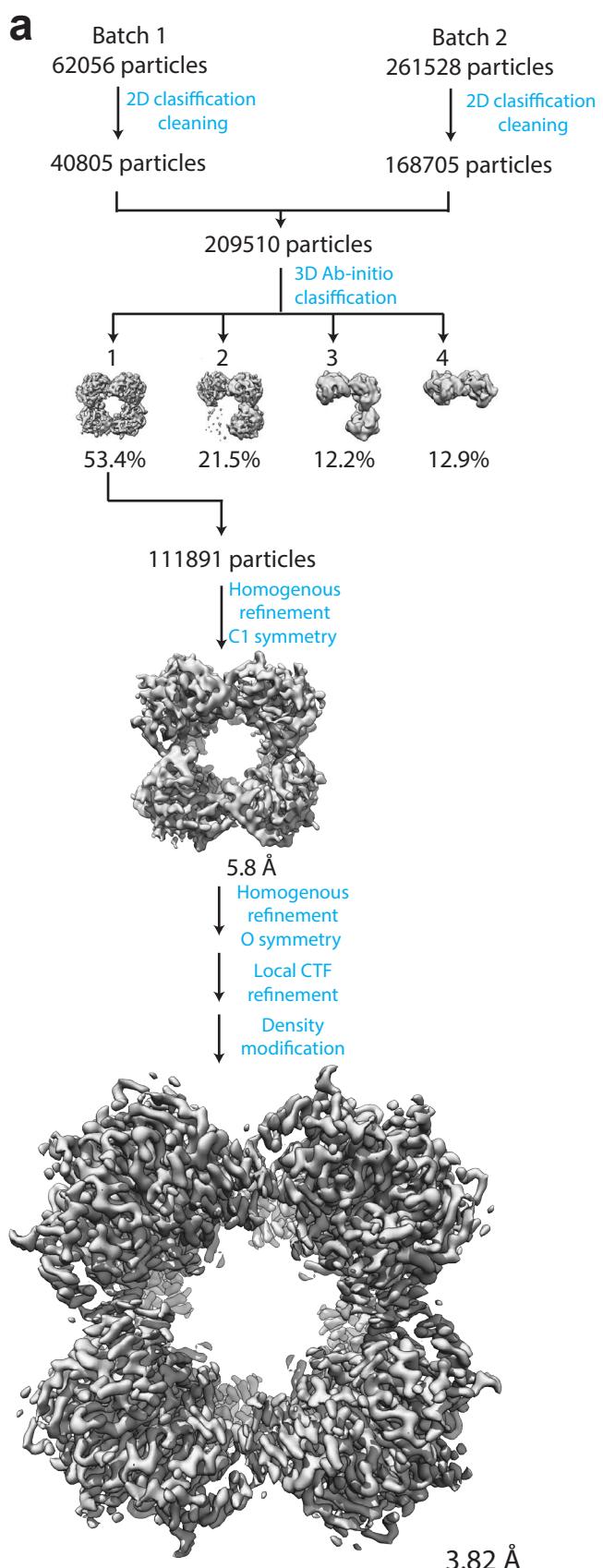
Figure S15. Cryo-EM data analysis pipeline (*Mt*E2b_CD).
a) Flow chart of the CryoEM data processing from 2D classification to the final model, performed with the Cryosparc software (13); final density modification was performed with the phenix suite (14). **b)** Representative 2D classes correponding to model 4 (top) and models 2 and 3 (bottom). **c)** FSC plots and resolution estimation using the gold-standard 0.143 criterion. **d)** Final refined map colored according to local resolution calculated with the Blocres tool in Cryosparc (15).

Table S1. SAXS data collection parameters.

Experimental settings

Beamline	**SOLEIL SWING**	**ESRF BM29** (*Cg*E2p_CD sample only)
Detector	CCD-based AVIEX	Pilatus3 1M
Detector distance (m)	2.000 / 4.000	2.867
Beam size (μm^2)	~ 800 x 150	~ 700 x 700
Wavelength (Å)	1.033	0.992
s range ($Å^{-1}$) [a]	0.0064 < s < 0.50	0.034 < s < 0.50
Exposure time/frame (s)	1.5	1.0
Temperature (K)	288	293

Structural parameters

Sample	*Cg*E2p_CD	*Cg*E2p_FL	*Mt*E2p_FL	*Mt*E2b_CD	*Mt*E2b_FL
R_g Guinier (Å)	27.7 ± 3.8	88.9 ± 2.53	80.97±1.06	62.11±1.85	92.79 ± 1.16
$I_{(o)}$ Guinier (arbitrary units)	12.38 ± 0.58	59 ± 2.1	0.13±0.006	260.94±7.2	0.086± 0.0068
R_g *P*(r) (Å)	27.7	92.95	84.37	62.81	91.58
$I_{(o)}$ *P*(r) (arbitrary units)	12.38	59	0.13	267	0.08
D_{max} (Å)	84.7	334.74	308.91	159,04	308,72
Vol POROD ($Å^3$)	117250	578000	577000	1170000	3930000
MW Bayesian (kDa)	76.3	242.6	318.4	479.1	1178.4
MW credibility interval Bayesian	71.4-81.9	221.0-372.7	221.0-372.7	264.2-556.0	916.6-inf

Abbreviations: MW: Molecular weight; R_g : Radius of gyration; D_{max} : maximal particle dimension.

[a] Momentum transfer $|s| = 4\pi\sin(\theta)/\lambda$

Table S2. Crystallographic data collection and refinement statistics (highest resolution shell in parenthesis). In the case of the *Cg*E2p_ΔLBDs dataset only, resolution limits were determined by applying an anisotropic cut-off by STARANISO, as implemented in the autoPROC data processing software (16).

Data collection	*Cg*E2p_CD	*Cg*E2p_CD Ox-CoA complex	*Cg*E2p_CD CoA/Lip(SH)$_2$ ternary complex	*Cg*E2p_ΔLBDs	*Cm*E2p_CD CoA complex	*Mt*E2b_CD
Synchrotron beamline	SOLEIL Proxima 2A	SOLEIL Proxima 1	SOLEIL Proxima 1	SOLEIL Proxima 1	ESRF MASSIF-1	ESRF MASSIF-1
Wavelength (Å)	0.9801	0.9801	1.5498	0.9786	0.9660	0.9660
Space group	P3	P6$_3$22	P2$_1$3	P6$_4$	P6$_3$22	F432
Cell dimensions *a*, *b*, *c* (Å)	158.28, 158.28, 63.02	73.74, 73.74, 192.19	132.13, 132.13, 132.13	170.39, 170.39, 67.56	100.62, 100.62, 281.57	214.07, 214.07, 214.07
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 90	90, 90, 120	90, 90, 120	90, 90, 90
Resolution (Å)	79.14 – 1.93 (1.96 – 1.93)	48.05 – 1.35 (1.37 – 1.35)	93.43 – 2.09 (2.13 – 2.09)	85.20 – 2.23 (2.37 – 2.23)	47.38 – 2.50 (2.60 -2.50)	49.11 – 1.50 (1.53 -1.50)
R_{pim}	0.043 (0.369)	0.020 (0.545)	0.028 (0.334)	0.027 (0.484)	0.028 (0.487)	0.023 (0.391)
$I / \sigma(I)$	14.0 (2.2)	16.1 (1.1)	16.6 (2.2)	16.0 (1.5)	18.7 (1.7)	14.9 (1.9)
Completeness (%)	99.3 (100.0)	99.6 (92.8)	100.0 (100.0)	95.0 (74.8)	100.0 (100.0)	100.0 (100.0)
CC(1/2)	0.998 (0.736)	0.998 (0.844)	0.999 (0.797)	0.999 (0.629)	0.999 (0.593)	0.999 (0.748)
Multiplicity	3.8 (3.8)	25.0 (21.2)	19.2 (19.1)	20.6 (15.9)	19.0 (19.6)	17.2 (17.4)
Refinement						
Resolution (Å)	1.93	1.35	2.09	2.23	2.50	1.50
No. reflections	131357	68431	45719	40014	30177	67274
R_{work}/ R_{free} (%)	16.6 / 18.5	16.8 / 19.4	16.4 / 17.2	19.0 / 21.3	20.1 / 22.0	19.0 / 19.3
No. atoms						
Protein	10944	1830	3659	5721	3617	1668
Ligands/ions	-	64	108	17	96	13
Solvent	1257	300	320	308	131	337
Average B-factors						
Protein	29.5	23.4	42.2	63.6	69.0	26.4
Ligand/ions	-	25.3	51.5	81.4	66.1	25.3
Solvent	38.2	38.8	51.6	58.8	64.8	40.4
R.m.s deviations						
Bond lengths (Å)	0.008	0.010	0.008	0.008	0.008	0.008
Bond angles (º)	0.91	1.04	0.91	0.96	0.91	0.93
PDB entry code	6ZZI	6ZZJ	6ZZK	6ZZL	6ZZM	6ZZN

Table S3. Number of genomes with identified *odhA*-like genes and total number of available genomes analyzed. To identify OdhA-like predicted proteins, the following pfam domain accession no. were searched: PF00198, PF16078, PF00676, PF16870, PF02779 (http://pfam.xfam.org/). JGI: Joint Genome Institute, USA (genome online database available at https://gold.jgi.doe.gov/).

Superkingdom / Phylum	Number of finished genomes in JGI databases	Number of genomes with *odhA*-like genes	Percentage of genomes with *odhA*-like genes
Archaea	173	0	0
Eukaryota	37	0	0
Bacteria (except Actinobacteria)	2957	9	0,3%
Actinobacteria	356	280	78,6%

Table S4. Autoinduction temperatures and purification buffers employed for the purification of the recombinant proteins used in this study.

Sample	Induction temperature	Lysis buffer	Gel filtration buffer
*Cg*E2p_FL	18 ºC	50 mM Tris-HCl pH 8.5, 500 mM NaCl, 1 mM DTT	50 mM Tris-HCl pH 8.5, 150 mM NaCl, 5% glycerol
*Cg*E2p_CD	18 ºC	50 mM Tris-HCl pH 8.5, 500 mM NaCl, 1 mM DTT	50 mM Tris-HCl pH 8.5, 150 mM NaCl, 5% glycerol
*Cg*E2p_ΔLBDs	18 ºC	25 mM Tris-HCl pH 8.5, 300 mM NaCl, 25 mM imidazole	50 mM Tris-HCl pH 8.5, 150 mM NaCl, 5% glycerol
*Mt*E2p_FL	30 ºC	25 mM Tris-HCl pH 8.5, 300 mM NaCl, 25 mM imidazole, 5% glycerol	25 mM Tris-HCl pH 8.5, 150 mM NaCl, 5% glycerol
*Mt*E2b_FL	30 ºC	25 mM Tris-HCl pH 8.5, 300 mM NaCl, 25 mM imidazole, 5% glycerol	25 mM Tris-HCl pH 8.5, 150 mM NaCl, 5% glycerol
*Mt*E2b_CD	18 ºC	25 mM Tris-HCl pH 8.5, 50 mM NaCl	25 mM Tris-HCl pH 8.5, 50 mM NaCl

SI References

1. E. R. Hall, P. D. Weitzman, A continuous spectrophotometric assay for the transacylase (E2) component of pyruvate and alpha-oxoglutarate dehydrogenase enzyme complexes. *Anal. Biochem.* **62**, 286–290 (1974).

2. J. Trewhella, *et al.*, 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: an update. *Acta Crystallogr. D Struct. Biol* .**73**, 710–728 (2017).

3. T. Bizien, *et al.*, A Brief Survey of State-of-the-Art BioSAXS. *Protein Pept. Lett.* **23**, 217–231 (2016).

4. D. Liebschner, *et al.*, Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr. D Struct. Biol.* **73**, 148–157 (2017).

5. A. Mattevi, G. Obmolova, K. H. Kalk, A. Teplyakov, W. G. J. Hol, Crystallographic Analysis of Substrate Binding and Catalysis in Dihydrolipoyl Transacetylase (E2p). *Biochemistry* **32**, 3887–3901 (1993).

6. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).

7. A. Waterhouse, *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296-W303 (2018).

8. X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–4 (2014).

9. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol* .**30**, 772–780 (2013).

10. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).

11. D. I. Svergun, Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886 (1999).

12. D. Franke, *et al.*, ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50,** 1212–1225 (2017).

13. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

14. T. C. Terwilliger, S. J. Ludtke, R. J. Read, P. D. Adams, P. V. Afonine, Improvement of cryo-EM maps by density modification. *Nat. Methods* **17**, 923–927 (2020).

15. G. Cardone, J. B. Heymann, A. C. Steven, One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. *J. Struct. Biol.* 184, 226–236 (2013).

16. C. Vonrhein, *et al.*, Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 293–302 (2011).