Actinobacteria challenge the paradigm:
a unique protein architecture for a well-known central metabolic complex

Eduardo M. Bruch[1]*, Pierre Vilela[1§], Norik Lexa-Sapart[1†], Lu Yang[1], Bertrand Raynal[2], Pedro M. Alzari[1], Marco Bellinzoni[1]*

1. Unité de Microbiologie Structurale, Institut Pasteur, CNRS, Université de Paris, F-75015 Paris, France
2. Plateforme de Biophysique Moléculaire, Institut Pasteur, F-75015 Paris, France

***Correspondence:** eduardo.bruch@pasteur.fr and marco.bellinzoni@pasteur.fr

§Present address: IGBMC, 1 rue Laurent Fries, 67404 Illkirch, France

†Present address: ALTEN, 221 bis Boulevard Jean Jaurès, 92514 Boulogne-Billancourt, France

ABSTRACT

α-ketoacid dehydrogenase complexes are large, tripartite enzymatic machineries carrying out key reactions in central metabolism. Extremely conserved across the tree of life, they have so far all considered to be structured around a high molecular weight hollow core, consisting of up to 60 subunits of the acyltransferase component. We provide here evidence that Actinobacteria break the rule by possessing an acetyltranferase component reduced to its minimally active, trimeric unit, characterized by a unique C-terminal helix that affects the oligomerization and the full 3D architecture of the complex. We show that this unique feature is characterized by an insertion, which together with OdhA is found spread over Actinobacteria. This phylum includes organisms or great interest for agriculture, industrial bio-production and many human pathogens as *Mycobacterium tuberculosis.* Moreover, components of this complex are key for *M. tuberculosis* survival in the human host, and its unique core and protein-protein interactions represent potentially "druggable" targets.

INTRODUCTION

α-oxoacid dehydrogenase complexes constitute a family of three-partite, ubiquitous metabolic complexes devoted to the oxidative decarboxylation of α-ketoacids and the concomitant production of reducing equivalents in form of NADH [1]. Three such complexes are known: the pyruvate dehydrogenase (PDH), that provides the main entrance of carbon units into the TCA cycle; the 2-oxoglutarate dehydrogenase (ODH), part of the oxidative branch of the TCA cycle; and the branched-chain α-ketoacid dehydrogenase (BCKDH), involved in the catabolism of aliphatic amino acids. These large tripartite complexes share a common molecular architecture organized around a core made by the E2 component, a flexible, multidomain protein which bears the acyltransferase activity required to transfer the acyl group from the decarboxylated substrate to the CoA-SH acceptor; the number of E2 subunits and the symmetry of the core depend on the complex and the species [1-5]. First shown by the crystal structure of *Azotobacter vinelandii* E2p, the E2 C-terminal catalytic core assumes an obligate homotrimeric state much similar to chloramphenicol acetyltransferase [6], with which it also shares the catalytic mechanism [7]. The observed higher-order oligomerisation states are made possible by intermolecular trimer-trimer interactions mediated by a well conserved, C-terminal 3_{10} hydrophobic helix which makes intermolecular symmetric interactions [5], then confirmed on other E2 enzymes and sometimes described as 'knobs and sockets' [3]. These interactions make symmetric, highly oligomeric states which adopt, in most cases, either an octahedral 432 symmetry, eight E2 homotrimers being positioned at the vertexes of a cube, or an icosahedral 532 symmetry, with 20 trimers assembled as a dodecahedron; the number of subunits depends on the complex and the species [1]. More recently, the presence of an irregularly shaped 42-mer E2 assembly has been described in the archaeon *Thermoplasma acidophilum* [8], although this peculiar oligomeric state is still based on the same kind of interactions between the C-terminal helices. Thus, the oligomeric state of the core, responsible for the large size of the complex, was observed in all analyzed complexes and is a trend commonly accepted to be universally conserved in Eubacteria, Archaea and Eukarya. While the reasons for the presence (and

63 evolutionary conservation) of such huge macromolecular scaffolds remain unclear [9], active site
64 coupling (transfer of acyl groups between lipoyl domains within the core) has often been
65 proposed as the major advantage [1,10]. Also, despite the tripartite organization of these complexes
66 as separate E1/E2/E3 enzymes had always been considered as universal, recent work on
67 *Corynebacterium glutamicum* has shown that no separate E2o component (specific to the ODH
68 complex) is present, and that an E2o succinyl transferase domain is fused to E1o, in a protein
69 called OdhA [11]. The same situation has then been confirmed for the model mycobacterium *M.*
70 *smegmatis* [12]. As the lipoyl-binding domain of E2o is absent from the E2o-E1o fusion, the ODH
71 activity depends on functional lipoyl groups provided *in trans*, and proven to be supplied by
72 E2p from the PDH complex [13], therefore suggesting the presence of a mixed PDH/ODH
73 supercomplex. By using an integrative structural biology approach, we describe here how *C.*
74 *glutamicum* E2p, that was expected to serves as the core of the mixed complex, breaks the rule
75 about the oligomeric state of acyltransferase E2 enzymes reducing its size to the minimal,
76 catalytically active trimeric unit. We also provide evidence supporting these features as a
77 common trait of Actinobacteria.
78

RESULTS

M. tuberculosis and *C. glutamicum* E2p are both trimeric in solution

To get insight into the oligomeric state of E2p in *Corynebacteriales*, full-length E2p from both *M. tuberculosis* (*Mt*E2p) and *C. glutamicum* (*Cg*E2p) were successfully expressed in *E. coli* and purified to homogeneity. Analytical ultracentrifugation (AUC) was used to assess the oligomeric assembly of the proteins (Table 1). Strikingly, the ratio between the estimated molecular weight in solution and the one predicted from the sequence indicate a trimeric state in solution for both proteins, with no experimental evidence of higher oligomeric states. The same result was observed for *Cg*E2p catalytic domain alone (*Cg*E2p_CD, comprising residues Leu437-Leu675). No larger assembly, suggestive of a 24-mer or 60-mer state, was observed for any of these proteins in the conditions tested (Table 1).

The absence of the expected higher oligomeric structure was further confirmed by size exclusion chromatography coupled to SAXS analysis, were the molecular mass was estimated using MW estimation tools available in ATSAS [14,15] (Table 1 and Suppl. Table 1). For both *Cg*E2p_FL and *Cg*E2p_CD the trimeric form is predominant in solution, whereas for *Mt*E2p_FL the analysis shows a molecular weight estimation that could correspond to a dimer of trimers. However, no evidence of a cubic, dodecameric or any other higher oligomeric state was observed for any of the three proteins.

For *Cg*E2p_CD, the chromatogram showed a single peak separated from the void volume fraction. Constant R_g value through the peak indicated the presence of a unique oligomeric species. The experimental R_g and D_{max} were 2.27 nm and 8.47 nm. The dimensionless Kratky plot shows a gaussian bell shape that decreases to zero consistent with a folded, globular and compact structure, consistently with the AUC results that indicate a trimeric *Cg*E2p_CD protein with low friction coefficient. For the two full-length proteins, *Mt*E2p_FL and *Cg*E2p_FL, the dimensionless Kratky plot shows a non-gaussian bell shape followed by a shoulder that does not immediately decreases to zero, consistent with a properly folded, but not globular,

107 multidomain protein with flexible linkers (Suppl. Fig. 1). This result, together with the high
108 friction coefficient observed by AUC, is suggestive of well folded, highly elongated proteins
109 in both cases.
110

111 ***Cg*E2p_CD posseses a unique C-terminal region characterized by an insertion that blocks**
112 **the protein in its trimeric form**

113 To understand the reasons for *Cg*E2p behaving as a trimeric protein and not assuming a higher
114 oligomerization state, we decided to characterize it structurally, starting by extensive
115 crystallization trials of *Cg*E2p_CD. Crystals were obtained from the protein in the absence of
116 ligands, in the presence of the acceptor substrate CoA and in a ternary complex with CoA and
117 dihydrolipoamide (Table 2). X-ray diffraction showed that the samples belonged to three
118 different crystallographic space groups, with resolutions ranging from 2.1 to 1.35 Å. In all the
119 crystal forms observed, *Cg*E2p_CD shows the expected, chloramphenicol acetyltransferase-
120 like fold, with an obligate homotrimeric assembly (Fig. 1b), as first observed for the *A.*
121 *vinelandii* orthologue (*Av*E2p_CD) [6]. The active site environment at the interface between
122 subunits in the functional trimer is also conserved (Suppl. Fig. 2a), with His645 and Asp649 as
123 the equivalent residues to His610 and Asn614 in *Av*E2p_CD. However, the crystal structure
124 provides no evidence of any higher-order quaternary assembly. To investigate whether the
125 trimeric assembly of *Cg*E2p_CD was still competent to substrate binding, we proceeded to co-
126 crystallization trials with CoA, dihydrolipoamide ($Lip(SH)_2$), DTT (as an antioxidant) or their
127 possible combinations. Co-crystallization with CoA yielded high resolution diffracting crystals
128 (Table 2) which show a full-occupancy complex, with no significant conformational change
129 with respect to the *apo* structure. The structure is similar to the equivalent complex from *A.*
130 *vinelandii* (PDB 1EAD, [7]), with the adenosine 3'-phosphate moiety occupying basically the
131 same position in both enzymes, CoA assuming a 'IN' conformation with the pantetheine chain
132 entering the active site tunnel, and the guanidinium group of Arg480 (Arg450 in *Av*E2p_CD)
133 making a salt bridge with the CoA 3'-phosphate (Suppl. Fig. 2b).

134 A ternary complex was obtained by co-crystallization with CoA and free dihydrolipoamide,
135 similarly to *Av*E2p_CD [7], although it belonged to a different crystalline form with respect to
136 the CoA complex (Table 2). This structure shows the same, 'IN' conformation for the CoA
137 pantetheine chain and the dihydrolipoamide molecule in an orientation close, albeit not
138 identical, to the one observed for the corresponding complex in *Av*E2p_CD (PDB entry 1EAB;
139 Fig. 1c, Suppl. Fig. 3).
140
141 Overall, *Cg*E2p_CD is very similar to the available structures of homologous acyltransferase
142 domains from other E2 enzymes, with RMSDs of 1.1-1.3 Å over 220-225 aligned residues
143 (RMSD 1.16 Å over 225 residues), as well as to the other available structures of acyltransferase
144 domains from E2 enzymes, like E2p from *Bacillus stearothermophilus* [5] (pdb 1B5S, with which
145 the rms difference is 1.10 Å over 222 residues), E2o from *E. coli* [16] (pdb 1SCZ, rmsd 1.29 Å
146 over 211 residues) and E2b from *Bos taurus* [3](pdb 2II4, average rmsd 1.3 Å over 221-223
147 aligned residues), even though the sequence identity is lower than 25% in all cases (Suppl. Fig
148 5). However, all previously known enzymes share a higher order oligomeric assembly
149 displaying 432 or 532 symmetry, with the functional homotrimers acting as the 'building
150 blocks' that interact with each other through their C-terminal, 3_{10} helices [1,5]. Interestingly, this
151 helix shows the most striking differences when *Cg*E2p_CD is superimposed to *Av*E2p_CD (Fig.
152 2a/2b and Suppl. Fig 5). While in *Av*E2p_CD this trimer-trimer interacting helix (TTI-helix) is
153 roughly perpendicular to the preceding helix and protrudes outward positioning its hydrophobic
154 residues into a pocket on the facing trimer, in the case of the corynebacterial enzyme the helix
155 lays on the same monomer surface, with the C-terminal hydrophobic residues occupying the
156 same pocket in an intramolecular interaction (Fig. 2b). In other words, the position of the C-
157 terminal helix for *Cg*E2p_CD coincides exactly with the position of the same, incoming helix
158 from the facing trimer in the *A. vinelandii* cubic structure (Fig. 2b). As a consequence, the
159 'knobs and sockets' interaction between trimers is replaced by an intramolecular interaction
160 involving the same features, hindering any trimer-trimer intermolecular interaction and
161 therefore forcing the protein in an homotrimeric state (Fig. 2c).

162

163 The above differences are also reflected by the distribution of the hydrophobic residues (Fig.
164 3a), which in oligomeric *Av*E2p are clustered at the helix C-terminus whereas in *Cg*E2p are
165 arranged at one side of the helix, thus preserving the mostly hydrophobic nature of the helix-
166 pocket interaction. This property is exemplified by the replacement of Leu637 in *Av*E2p for a
167 polar residue (Gln674) in *Cg*E2p, which reflects the new amphipathic nature of the helix. Most
168 importantly, the different orientation of the terminal helix in *Cg*E2p_CD is due to the presence
169 of a three-residue insertion (Phe669-Glu670-Gly671) in *Cg*E2p (Fig. 3a). This insertion allows
170 the C-terminal helix to change its orientation and lay on the surface to reach the accommodating
171 pocket *in cis*, thus replacing an intermolecular interaction by an intramolecular one (Fig. 2c).
172 To confirm that this structural arrangement was not due to the truncated *Cg*E2p_CD protein, a
173 new *Cg*E2p construct including the peripheral subunit binding domain (PSBD), was also
174 overexpressed in *E. coli*, purified and crystallized, yielding a new, 2.5 Å resolution structure
175 (Table 2) showing the catalytic domain in addition to an N-terminal, 13-residue extension that
176 represent part of the flexible linker connecting to the PSBD (Suppl. Fig. 4). Noteworthy, no
177 difference in the oligomeric state, nor in the position of the C-terminal, 3_{10} helix was observed
178 compared to the previous structures. In addition, the E2p ortholog from *M. tuberculosis*, *Mt*E2p,
179 also possesses a three-residue insertion (Phe547-Glu548-Ala549) in an equivalent position just
180 preceding the predicted 3_{10} terminal helix (Fig. 3a), suggesting that the Phenylalanine-
181 containing insertion (PCI) could be responsible of the lack of higher order oligomerization
182 states.

183

184 **The Phe containing insertion (PCI) is a common feature in actinobacteria**

185 To analyze the distribution of the PCI, we performed sequence analysis of the E2-coding genes
186 in representative genomes distributed within the phylum Actinobacteria (Suppl. Fig. 6). Our
187 results show that the PCI is not only present in *Corynebacteriales*, including *M. tuberculosis*
188 and *C. glutamicum*, but more generally in sequences from two classes, namely *Actinobacteria*
189 (also called Actinomycetales that represent the biggest class inside the Phylum Actinobacteria)

190 and *Acidimicrobiia* (Suppl. Fig. 6). The high conservation of the Phe residue present in the
191 insertion emphasizes the importance of its interaction within the context of the hydrophobic
192 pocket. Other residues involved in H-bond interactions with the pocket (mainly Glu670,
193 Leu675 and Leu677) are also conserved (Suppl. Fig. 6). Interestingly, although the PCI is
194 conserved, the TTI-helix region has lost some of the C-terminal residues and even the full C-
195 terminal helix is absent in some of the organisms (data not shown), and is in agreement with
196 the absence of the inter-trimer interaction in which this helix is involved.
197
198 **PCI and E2 oligomerization**
199 To confirm the relationship between the presence of the three-residue, actinobacterial-specific
200 C-terminal insertion (PCI) and the oligomerization order of E2 proteins from the Actinobacteria
201 class, we followed two parallel paths. First, we structurally analyzed the orthologous protein
202 from *Corynebacterium mustelae,* a recently described species from the same genus [17], as an
203 example of orthologue bearing the same insertion (PCI). The catalytic domain of the predicted
204 E2p enzyme from this species (*Cm*E2p_CD, residues 428-666) was crystallized in complex
205 with CoA, and its structure solved by molecular replacement to 2.5 Å resolution (Table 2). The
206 crystal structure shows an homotrimeric arrangement of *Cm*E2p_CD with no significant
207 structural changes with respect to the *C. glutamicum* counterpart (RMSD of 0.79 Å over the
208 whole trimer), and without evidence of any relevant higher order oligomeric state (Fig. 3b). In
209 particular, the C-terminal 3_{10} helix has the same relative orientation as in *Cg*E2p_CD and shows
210 the same intramolecular interaction over the trimer surface (Fig. 3b), with the three-residue
211 insertion (including the well-conserved Phe) occupying an equivalent position in both
212 structures (Fig. 3c and 3a).
213 Likewise, we analyzed *Mt*E2b, the predicted E2 component of the BCKDH complex (Branched
214 Chain KetoAcid Dehydrogenase) [18]. The catalytic domain, or *Mt*E2b_CD (residues 165-393),
215 was overexpressed in *E. coli* with an N-terminal, cleavable His_6 tag and purified following the
216 same experimental strategy. The rationale for the choice of this protein resides in the absence
217 of the PCI, as shown by the sequence alignment in Fig. 3a, and confirmed by the unpublished

218 crystal structure of a similar construct (pdb entry 3L60), issued from a structural genomics
219 initiative. Crystallization experiments carried out after TEV cleavage of the N-terminal His_6
220 tag yielded bipyramid shaped crystals belonging to the cubic space group F432 and diffracting
221 X-rays up to 1.5 Å resolution (Table 2). The crystal structure, solved by molecular replacement
222 using the coordinates available from the pdb, shows the 24-mer cubic arrangement observed
223 for *Av*E2p_CD or bovine E2b [3] (Suppl. Fig. 8). The terminal 3_{10} helix assumes the same
224 canonical conformation and is involved in the 'knobs and sockets' trimer-trimer interaction.
225 However, this is in striking contrast with the pdb model 3L60, which rather shows an
226 homotrimeric arrangement for an analogous construct (residues Pro165-Leu393 as deduced
227 from the information available in the pdb). To verify whether this discrepancy is due to the
228 presence of a C-terminal purification tag leftover in the previously crystallized sample, we
229 carried out further solution studies on both *Mt*E2b_FL and *Mt*E2b_CD. Both AUC data and *ab-*
230 *initio* models obtained from SAXS data are consistent with a large, cubic assembly for both
231 constructs (Table 1), indicative of the canonical 24-mer assembly (Suppl. Fig. 1 and 9).
232
233 Finally, both *Mt*E2b constructs were analyzed by single-particle cryo-EM. Isolated particles
234 indeed show a cubic shape structure for both *Mt*E2b_FL and *Mt*E2b_CD proteins, with a
235 particle side of approximately 13 nm (Fig. 4a and Suppl. Fig. 10). Several rounds of 2D-
236 classification were performed to select a subset of good particles where the class average
237 reflected the expected cube form and high-resolution features such as defined alpha helices
238 (Fig. 4b and Suppl. Fig. 11). Several *ab-initio* models were calculated and, after 3D
239 classification and the exclusion of "junk" particles showing partial dissociation of the complex,
240 a good dataset of particle was selected. After homogeneous refinement, local CTF refinement
241 and density modification the structure was solved and the corresponding cubic assembly was
242 obtained at a resolution of 3.82 Å (Fig. 4c and Suppl. Fig. 11 a and c). Real space refinement
243 was performed to fit the 24-mer, cubic crystallographic assembly into the EM map, showing a
244 very good fitting overall (Fig. 4d) and good density for the TTI helix (Fig. 4e), which, in turn,
245 confirmed its canonical orientation including the 'knobs and sockets' interactions, in striking

246 contrast with the E2p situation. The ensemble of our structural and biophysical data therefore
247 supports a strong link between the three-residue C-terminal insertion (PCI), the orientation of
248 the terminal 3_{10} helix (TTI), and the E2 oligomeric state.
249
250 **The PCI in actinobacterial E2p correlates with the presence of OdhA**
251 Unlike other canonical E1 proteins, *C. glutamicum* OdhA has a unique domain architecture
252 characterized by the fusion of an E2-like domain to E1o [13,19] (Fig. 5d). Taking advantage of this
253 unique E2-E1 fusion feature, an architecture conservation strategy was used to retrieve the
254 sequences of OdhA homologs from an extended list of available genomes. This strategy was
255 based on the search of the pfams associated to both the E2 like domain and E1o (Fig. 5a). This
256 analysis revealed that, except for a few isolated cases, OdhA homologs are a unique feature of
257 Actinobacteria (Fig. 5a). Inside this Phylum, OdhA homologs are widely present in the
258 Actinomycetales class, except for members of the order Bifidobacteriales, and the
259 Acidimicrobiia class (Fig. 5b), but not in the classes Coriobacteriia or Rubrobacteria. In the
260 three members from the class Thermoleophilia a unique architecture is present. In this
261 architecture the first domain ('2-oxoglutarate_dh_Nterminus') is replaced by two Lipoyl_biotin
262 domain and a E3-binding domain (Fig. 5d) thus bearing a full fusion of E2 and E1o (called LB-
263 PS-E2-E1). Although speculative at this stage, this protein (never reported before) could
264 perform its activity without the need of a lipoyl domain provided *in trans*, and thus avoiding
265 the interaction with E2. This is coherent with the absence of the N-terminal '2-
266 oxoglutarate_dh_Nterminus', present in OdhA, which is important for the interaction with E2
267 (Fig. 5d).
268
269 The correlation between the PCI in E2 and the presence of an OdhA homolog (devoid of a
270 lipoyl-binding domain) could be observed for all the representative genomes selected (Fig. 5c).
271 Both features are present in classes Actinobacteria and Acidimicrobiia and absent in the classes
272 Thermoleophilia, Rubrobacteria or Coriobacteria. Since OdhA requires E2p to perform its
273 activity in the context of the PDH/ODH supercomplex, it is tempting to speculate that loss of

274 E2 higher oligomerization could be a feature evolved to optimize the assembly of such a mixed
275 complex, where a E2o-E1o fusion enzyme has to compete with a canonical E1p for the lipoyl
276 groups provided by the same E2p. The latest could explain why this small size E2p was only
277 detected in the actinobacteria phylum and not in any other organism in the tree of life.
278
279 **DISCUSSION**
280 The ensemble of enzymatic reactions that make up central metabolism and concur to the
281 production of reducing equivalents in heterotrophic organisms is commonly considered to be
282 extremely conserved. The same has longtime been believed for the respective enzymes and
283 their assemblies, for which most of our current knowledge comes from studies made decades
284 ago on model organisms, which, for prokaryotes, were in most cases represented by *E. coli* or
285 *B. subtilis*. 2-oxoacid dehydrogenase complexes, which are key players in central metabolism,
286 were also object of intensive study during this 'golden age' of enzymology, with most of our
287 knowledge coming from seminal work carried out, over several decades, by the groups led by
288 renowned biochemists like Richard N. Perham [1], Lester Reed [20], and John Guest [21]. Their
289 extensive biochemical work was then largely confirmed by structural evidence that came out
290 during the '90s and the following years, initiated by the work performed in Wim Hol's lab on
291 the diazotroph, Gram-negative model *A. vinelandii*, and the first determination of a crystal
292 structure of AvE2p_CD [6], which in turn confirmed observations dating back to the '60s on *E.*
293 *coli* E2p [22] and E2o [23]. Most of the following structural studies have focused on the PDH
294 complexes, especially the ones from *E. coli*, *B. stearothermophilus* or their counterparts from
295 human or yeast [1,24,25], while ODH and BCKDH complexes, despite a number of biochemical
296 studies [26], were largely assumed to share PDH-like molecular structures. In particular, 2-
297 oxoacid dehydrogenase complexes have all been considered to conform to the 'rule' of a
298 hollow, large E2 core showing either cubic (432) or icosahedral symmetry (532), assembled
299 according to an ensemble of equivalent, and quasi-equivalent interactions involving the C-
300 terminal end of each E2 monomer [5]. The more unusual 42-mer E2 assembly described in *T.*
301 *acidophilum* [8], although deviating from the standard symmetry, still possess a hollow core

302 whose assembly is based on the same kind of C-terminal helices interactions, and thus does not
303 represent a real exception to the rule. Questions about the necessity of such large oligomeric
304 assemblies for 2-oxoacid dehydrogenase activity were already raised in the past, when the same
305 team that described the *T. acidophilum* E2 showed that precluding the formation of higher E2
306 oligomers did not lead to loss of catalytic activity [9], in contrast with previous reports. Our work
307 here shows how *Corynebacteriales*, and most *Actinomycetales*, break the 'dogma' regarding
308 the three-dimensional architecture of 2-oxoacid dehydrogenase complexes, and raise new
309 questions about the origin and evolution of these enzymatic machineries.
310
311 Actinobacteria represent one of the largest and ubiquitous prokaryotic phylum, with more than
312 50 families. Some of its members are of great economic importance to humans, because
313 agriculture and forests depend on their contributions to the soil system, as *Streptomyces*, one of
314 the largest bacterial genera and a major source of antibiotics. This phylum also includes
315 important human pathogens such as *M. tuberculosis*, as well as major cell factories as *C.*
316 *glutamicum*. Despite the wide applications of *C. glutamicum* in the biotech industry, which
317 range from the production of amino acids to that of biofuels [27], relatively little is known about
318 how *Corynebacteriales* regulate their central metabolism. Since the discovery that the FHA
319 protein OdhI is a phospho-dependent regulator of OdhA, the *C. glutamicum* E1o subunit [11], a
320 bunch of evidence indicate that the ODH activity in *Corynebacteriales* – including *M.*
321 *tuberculosis* and *M. smegmatis* – is regulated by a signal transduction pathway triggered by the
322 Ser/Thr kinase PknG in response to nutrient availability [28-30].
323
324 The evidence reported here supports a strong correlation between the presence of E2o and E1o
325 activities on a single OdhA-like polypeptide and the loss of highly oligomeric structure of the
326 PDH core, two properties that are most likely related to the dependence of both the PDH and
327 ODH activities on the lipoyl domains provided by a single E2p protein, and, therefore, their
328 coexistence in a mixed multienzymatic machinery. The presence of such kind of supercomplex,
329 first hypothesized in 2006 for *C. glutamicum* [11], is in turn suggestive of the presence of

330 regulation mechanisms that might well be unique to Actinobacteria. In this respect, the
331 PknG/GarA (OdhI) signaling pathway, that acts directly on the E1o activity through the OdhI
332 regulator that exerts an allosteric inhibition [31], has only be described in Corynebacteriales and
333 might indeed be related to the unusual architecture of the PDH/ODH supercomplex. An
334 additional ODH allosteric regulator is acetyl-CoA, the product of the PDH reaction, identified
335 as an allosteric activator of *M. smegmatis* KGD [12] but also, much earlier, as an activator of
336 corynebacterial ODH [32], suggesting, in turn, positive feedback mechanisms by product
337 channeling from PDH to ODH. On the other hand, little is known about the regulation of the
338 PDH reaction in Actinobacteria, apart from the known involvement of the RamB transcriptional
339 regulator in the expression of the *aceE* (E1p) gene in *C. glutamicum* [33].
340
341 Recent developments in metabolomics have started to unravel the mechanisms of metabolic
342 regulation in *M. tuberculosis* [34,35], a particularly important facet as metabolic plasticity is a key
343 capability of the pathogen to switch from active replication to dormant states. The strategy of
344 targeting central metabolism for new drugs has been suggested by several groups, mainly due
345 to the presence of unique *M. tuberculosis* pathways that are crucial for pathogen survival and
346 latent state persistence within the host [34,36,37]. Unlike other central metabolic enzymes, the low
347 similarity with its eukaryotic counterparts makes of the PDH/ODH supercomplex a good drug
348 target candidate. Indeed, the enzymatic components of the complex were shown to participate
349 in the pathogen response to oxidative stress [38] or virulence [18]. In particular, *M. tuberculosis*
350 DlaT (*Mt*E2p), that we show here as being structured as an homotrimer, has been shown to be
351 required for TB infection in the Guinea pig model, and was also involved in the pathogen's
352 response to reactive nitrogen intermediates [36]. Because of its unique quaternary arrangement,
353 the components of the PDH/ODH supercomplex are good candidates to reveal new 'druggable'
354 surfaces, allowing to target essential, but previously unknown, protein-protein interactions. By
355 revealing the structural principles of the atypical molecular assembly of these enzymes, our
356 work thus opens the way to the development of novel therapeutic approaches.
357

358 **MATERIAL AND METHODS**

359

360 ***Plasmid construction***

361 *C. glutamicum* open reading frames were amplified from genomic DNA by PCR and cloned in
362 pET-28a vector by restriction free cloning [39], while expression constructs for the other proteins
363 (*Cm*E2p, *Mt*E2p, *Mt*E2b) were provided by Genscript (Leiden, the Netherlands). In all cases,
364 a sequence coding for a His_6-tag followed by the TEV protease cleavage site (ENLYFQG) were
365 fused to the N-terminus of the protein of interest. Constructs were made to express: for *Cg*E2p
366 (Uniprot accession no. Q8NNJ2), the full-length protein (*Cg*E2p_FL, residues 1-675), the
367 catalytic domain (*Cg*E2p_CD, corresponding to residues 437-675), and the catalytic core and
368 the peripheral subunit binding domain (PSBD; *Cg*E2p-ΔLBDs, corresponding to residues 372
369 to 675). For *Cm*E2p (Uniprot accession no. A0A0G3H170), the catalytic core was expressed
370 (*Cm*E2p_CD, residues 428-666), while for *Mt*E2p (Uniprot accession no. P9WIS7), the full-
371 length protein was expressed (*Mt*E2p_FL, residues 1-553).
372 For *Mt*E2b (Uniprot accession no. O06159), two constructs were built: the first one
373 corresponding to the full-length protein (*Mt*E2b_FL, residues 1-393), and the second one
374 corresponding to the catalytic core region (*Mt*E2b_CD, defined as residue 165-393). All
375 plasmids were verified by DNA sequencing.

376

377 ***Protein purification***

378 Each plasmid was transformed into *E. coli* BL21DE3 using standard protocols. In the case of
379 *Cg*E2p_FL and *Cg*E2p_CD, transformed cells were growth in LB medium at 37 ºC until optical
380 density measure at 600 nm reach ~0.6. Afterwards, temperature was dropped to either 18 ºC or
381 30 ºC (see table below) and incubated for 18 hours after adding 0.5 mM isopropyl-β-D-
382 thiogalactopyranoside (IPTG). For *Mt*E2p_FL, *Mt*E2b_FL or *Mt*E2b_CD autoinduction buffer
383 was used so cells were growth at 37 ºC for 4 hours and then temperature was dropped to either
384 18 or 30 ºC and cells incubated overnight.

385 Cells were recovered by centrifugation and washed twice with the corresponding Lysis buffer
386 (see buffer table). Cell lysis was performed using a CF2 cell disruptor (Constant Systems Ltd.,
387 Daventry, UK). The soluble fraction was separated from debris by centrifugation at 26,800 g
388 for 1 hour. The recombinant proteins were purified by Ni^{2+}-IMAC on HisTrap columns (GE
389 Healthcare). Fractions, as confirmed by SDS-PAGE, containing the protein of interest were
390 dialyzed against GF buffer (see buffer table), supplemented with 1 mM dithiothreitol (DTT),
391 for 12 hours at 18°C, after adding His-tagged TEV protease [40] at a 1:30 ratio (w/w) to remove
392 the N-terminal His_6 tag from the recombinant protein. The cleaved, recombinant proteins were
393 then separated from the His-tagged TEV protease (and the cut N-terminal portion) by gravity-
394 flow separation on Ni-NTA agarose (Qiagen). Finally, the recovered recombinant proteins were
395 further purified by size exclusion chromatography using a Sephacryl S-400 16/60 column (GE
396 Healthcare). Fractions containing the protein were pulled, concentrated and flash frozen in
397 liquid nitrogen.
398

	Lysis buffer	**Gel filtration buffer**	**Induction temperature**
*Cg*E2p_FL	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	18 ºC
*Cg*E2p_CD	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	18 ºC
*Cg*E2p_ΔLBDs	25 mM Tris-HCl pH8.5, 300 mM NaCl, 25 mM Imidazole	50 mM Tris-HCl pH8.5, 150 mM NaCl, 5% glycerol	18 ºC
*Mt*E2p_FL	25 mM imidazole, 300 mM, NaCl, Tris-HCl pH 8.5, 5% glycerol	25 mM Tris-HCl pH 8.5, 150 mM NaCl, 5% glycerol	30 ºC
*Mt*E2b_FL	25 mM imidazole, 300 mM NaCl, 5% glycerol, Tris-HCl pH 8.5	25 mM Tris-HCl pH 8.5, 150 mM, NaCl, 5% glycerol, 1 mM DTT	30 ºC
*Mt*E2b_CD	25 mM Tris-HCl pH 8.5, 50 mM NaCl	25 mM Tris-HCl pH 8.5, 50 mM NaCl	18 ºC

399
400
401
402 ***SEC-SAXS and ab initio modelling***

403 SAXS experiments were performed at SOLEIL synchrotron on the SWING beamline. The
404 beam wavelength was set to 1.033 Å. The detector was located at a distance of 2 meters to
405 collect data at q-range between 0.005 and 0.55 $Å^{-1}$ ($q=4\pi \sin \theta/\lambda$ and 2θ is the scattering angle).
406 Protein sample were injected into a size exclusion column (Superose™ 6 Increase 5/150 GL,
407 GE Healthcare Bio-Sciences) previously equilibrated with corresponding buffer and eluted
408 directly into the SAXS flow-through capillary cell at a flow rate of 0.2 ml/min. Frames were
409 collected continuously during the whole protein elution time.
410
411 Frames were analyzed using FOXTROT software. Buffer corresponding frames were averaged
412 and subtracted from sample frames. The forward scattering I(0) and the Rg for each sample
413 subtracted frame were derived from Guinier approximation. The I(0) vs frame curve was used
414 to select the region corresponding to the protein's elution peak. Frames corresponding to the
415 elution peak and with constant Rg were averaged. All subsequent data processing was
416 performed using ATSAS suite [14].
417 PRIMUS software, from the ATSAS suite, was used to recalculate the I(0) and Rg, while
418 GNOM was used to calculate the pair-distance distribution function P(r) and Dmax [15]. Finally,
419 the POROD volume and the tools available in ATSAS were used to estimate the molecular
420 weights of the proteins [41]. The dimensionless Kratky plot was calculated according to previous
421 work [42,43], using the value determined by Guinier analysis.
422 *Ab initio* modelling was also performed using software from the ATSAS suite. For *Mt*E2b_CD,
423 10 independent models were generated using DAMMIF, with no imposed symmetry. In both
424 cases best models were obtained assuming oblate anisometry. The models were then compared
425 and align with DAMSEL and DAMSUP and output files were generated using DAMAVER,
426 DAMFILT and DAMSTART, all within the ATSAS suite. Finally, the resulting damstart file
427 was used in DAMMIN to obtain a dummy atom model that best fit the data.
428
429 ***AUC***

430 Sedimentation velocity (SV) analytical ultracentrifugation assays were performed using a
431 Beckman Coulter ProteomeLab XL-I analyticalultracentrifuge equipped with UV-Vis
432 absorbance and Raleigh interference detection systems, using the 8-hole Beckman An-50Ti
433 rotor at 20°C. The speeds chosen are 10,000 rpm for *Mt*E2b_CD and *Mt*E2b_FL and 36,000
434 rpm for *Mt*E2p_FL, *Cg*E2p_FL and *Cg*E2p_CD. Three concentrations of each protein were
435 prepared for this experiment in their corresponding buffer and loaded into analytical
436 ultracentrifugation cells. During the run SV was followed using by measuring absorbance at
437 290 nm.
438 SEDFIT 15.01 (available at http://analyticalultracentrifugation.com; [44]) was used to calculate
439 the sedimentation coefficient distribution C(s), then corrected to standard conditions to get the
440 final standard values. These coefficients were plotted as a function of the different
441 concentrations and an extrapolation to zero concentration was made to obtain the standard value
442 for the main oligomer. From these values, molecular mass and friction ratio were obtained.
443
444 ***Crystallization***
445 The crystallization experiments were performed at either 18°C or 4 °C by the sitting drop vapor
446 diffusion technique in 96-well plates, according to established protocols at the Crystallography
447 Core Facility of the Institut Pasteur [45]. Crystallization conditions were: for *Cg*E2p_CD in apo
448 form, 0.1 M Hepes-Na pH 7.5, 5% PEG4000, 30% 2-methyl-2,4-penthane diol (MPD), at 18°C;
449 for *Cg*E2p_CD in complex with CoA, 0.1 M Hepes-Na pH 7.5, 1.56 M tri-sodium citrate, at
450 4°C; for *Cg*E2p_CD in ternary complex with CoA and dihydrolipoamide, 0.1 M Tris-HCl pH
451 7.3, 1.62 M tri-sodium citrate, at 18°C; for *Cg*E2p_ΔLBDs, 30% PEG1500 at 18°C; for
452 *Cm*E2p_CD in complex with CoA, 0.1 M Hepes-Na pH 7.5, 150 mM NaCl, 30% PEG4000, at
453 18°C; for *Mt*E2b_CD, see 0.1M Imidazole pH 6.5, 4 M NH_4^+ acetate at 18°C.
454
455 ***Data collection and structure solution***
456 Diffraction data were acquired from crystals maintained at 100 K, either at the SOLEIL
457 Synchrotron (Saint-Aubin, France), on the beamline Proxima-1 or Proxima-2A, or on the

458 automated beamline MASSIF-1 at the ESRF (Grenoble France) [46]. The data were processed
459 with XDS [47], run through either XDSME (available at https://github.com/legrandp/xdsme) or
460 autoPROC [48], and scaled with Aimless, available from the CCP4 suite [49], or STARANISO as
461 provided within autoPROC. The structures were solved by molecular replacement through the
462 program PHASER [50]. Coordinates of the *E. coli* E2p catalytic domain (PDB entry 4N72; [51])
463 were used as the search model to first solve the structure of *Cg*E2p_CD in apo form, which, in
464 turn, served as molecular replacement search model for the following datasets, including
465 *Cm*E2p_CD, but not MtE2b_CD whose structure was solved by using the previously released
466 coordinates of the same protein (pdb entry 3L60). All rebuilding and adjustments of the models
467 were performed with COOT [52]. The refinement was carried out with BUSTER, applying local
468 structure similarity restraints for non-crystallography symmetry (NCS) [53] and a Translation-
469 Libration-Screw (TLS) model. Chemical dictionaries for ligands were generated with the Grade
470 server (http://grade.globalphasing.org). Validation of models was performed with
471 MOLPROBITY [54] and the validation tools in the PHENIX suite [55]. Data collection, refinement,
472 model statistics and pdb accession codes for coordinates and structure factors are indicated in
473 Table 2. Graphical representations were rendered with Pymol [56].
474
475 ***CryoEM***
476
477 For grid preparation 3.5 µl of 0,2 mg/ml *Mt*E2b_CD were applied into a Lacey Carbon Film on
478 200 Mesh Copper Grids. Grid vitrification was performed in a Vitrobot (FEI) with 95%
479 humidity, ashless filter paper (Standard Vitrobot Filter Paper, Ø55/20mm, Grade 595) and
480 using a blot time of 3.5 seconds and a blotting force of 0. Grids were then storage and collected
481 in a 200 keV Thermo Scientific Glacios® Cryo-Transmission Electron Microscope. Images
482 were recorded using a Falcon IIIEC direct electron detector in linear mode. The images were
483 collected using EPU software 2.2.0.65REL, with an under focus range from 0.8 µm to 2.6 µm.
484 The pixel size was set to 0.96 Å. Per movie 37 frames were collected in total, with an overall
485 dose of 42.5 e-/Å^2.

486

487 Movies frames were aligned using Motiocor2 [57,58], and CTF estimation was performed using
488 Gctf [59]. Movies were imported and analyzed in Cryosparc v2.15 [60]. Using the curate exposure
489 feature, 1235 out of 1476 movies, followed by 2483 out of 3533 movies from a second data
490 collection were selected for further analysis. A first round of manual picking was performed
491 and after two rounds of 2D classification, the three most populated classes were selected for
492 template-based particle picking against a dataset containing the selected micrographs. The
493 selected particles were cleaned using the inspect pick function of Cryosparc and three rounds
494 of 2D classification and a final dataset of 40805 and 168705 good particles, corresponding to
495 collection session 1 and 2, respectively, was selected.
496 A round of *ab initio* and 3D classification into 4 classes was performed to select a subset of
497 111891 particles that could be refined to 5.8 Å. After local CTF refinement and imposing
498 Octahedral O3 symmetry, this dataset could be refined to 3.92 Å. Finally, density modification
499 was applied using the tool resolve_cryo_em from the PHENIIX suite [61], reaching a final map
500 at 3.82 Å resolution. Model fitting into cryo-EM maps was performed using the programs
501 UCSF Chimera [62] and phenix.real_space_refine [63]. The final, sharpened map from Cryosparc
502 as well as the density modified map from resolve_cryo_em were deposited to the EMDB under
503 the accession code 11600. Figures were generated and rendered with UCSF Chimera.
504

505 ***Bioinformatic analysis***

506 ***a0) Analysis of PCI insertion***

507

508 Multiple sequence alignments were performed using MUSCLE
509 (https://www.ebi.ac.uk/Tools/msa/muscle/) [64]. Further alignment details were carried out in
510 AliView version 1.26 [65] . Sequences from *Mt*E2b, *Mt*E2p, *Cg*E2p, *Cm*E2p and E2 proteins with
511 known structure were compare to the representative sequences to account for the PCI presence.
512 The protein sequence from E2 proteins with known structure were obtained from the pdb
513 database (pdb entries 1EAB, 3B8K, 1SCZ, 3RQC, 1B5S). The final multiple sequence

514 alignment was merged with the pdb entry 1EAB to obtain final figures using ESPript 3.0
515 (http://espript.ibcp.fr) [66]. Finally, the presence of the PCI was analyzed in the region
516 corresponding to the C-terminal α-helix (residues 630 to 637 in *Av*E2p).
517
518 ***a1) OdhA like gene distribution inside selected genomes***
519
520 To identify the OdhA like genes into the genomes the gene search tool from IMG/M was used
521 (Integrated Microbial Genomes & Microbiomes system v.5.0) [67]. To simplify the analysis, only
522 genomes annotated as finished were selected. For each genome OdhA like genes were identified
523 by their conserved Pfam architecture using the associated Pfams: Pfam16078 (2-
524 oxogl_dehyd_N), Pfam00198 (2-oxoacid dH), Pfam00676 (E1-dh), Pfam02279 (Transket_pyr)
525 and Pfam16870 (OxoGdeHyase_C).
526

Domain/Phylum	Number of finished genomes in JGI	Number of genomes with *odhA*-like genes (pfam: 16078, 00198, 00676, 02779, 16870)	Percentage of genomes with *odhA*-like genes
Archae	173	0	0
Eukaryota	37	0	0
Bacteria (non Actino)	2957	9	0,3%
Actinobacteria	356	280	78,6%

527
528 **b) *odhA like gene distribution inside Phylum Actinobacteria***
529
530 All the finished genomes corresponding to the phylum actinobacteria and two non-finished
531 genomes corresponding to the class Thermoleophila were selected from IMG/M available

532 genomes (Thermoleophila counts to the date with only one finished genome). After elimination
533 of duplicated genomes OdhA like genes and E2-E1 like genes were identified using its
534 corresponding Pfam architecture.
535 For E2-E1 like genes the associated genomes correspond to: Pfam00364 (Lypoil binding
536 domain), Pfam02817 (E3 binding domain), Pfam00198 (2-oxoacid dH), Pfam00676 (E1-dh),
537 Pfam02279 (Transket_pyr) and Pfam16870 (OxoGdeHyase_C). For the selected genomes a
538 tree was created using the JGI distance tree tool. The tree is created using the alignment of 16S
539 genes based on the SILVA database and dnadist and neighbor tools from the PHYLIP package
540 (available at https://evolution.genetics.washington.edu/phylip/). Figures were built using the
541 Interactive Tree Of Life (iTOL) v4 online tool [68].
542
543 **c) *Correlation between odhA like genes and the presence of the PCI in E2p***
544 A representative pool of genomes from the larger orders inside actinobacteria class (namely
545 Corynebacteriales, Streptomycetales, Propionibacteriales, Micromonosporales, Microccocales
546 and Bifidobacteriales) and representatives from the other classes inside the Phylum
547 actinobacteria were selected. For each genome, the corresponding E2 sequences were identified
548 using its corresponding Pfam architecture, namely Pfam00364 (Lypoil binding domain),
549 Pfam02817 (E3 binding domain), Pfam00198 (2-oxoacid dH). For each genome sequence
550 alignments were performed as previously described to assess the existence of the PCI (as
551 described previously).
552
553 **ACKNOWLEDGMENTS**
554 This work was funded by the ANR projects SUPERCPLX (ANR-13-JSV8-0003) and
555 METACTINO (ANR-18-CE92-0003), both granted to M.B., and by institutional grants from
556 the Institut Pasteur and the CNRS. We are grateful to Ahmed Haouz, Patrick Weber and Cédric
557 Pissis (Crystallography Core Facility, Institut Pasteur) for carrying out robot-driven
558 crystallization screenings. We also acknowledge Sébastien Brûlé (Molecular Biophysics Core
559 Facility) for his assistance in AUC experiments, and Francesca Gubellini (Structural

Microbiology) for her help in the sample preparation for negative staining EM grids. The NanoImaging Core at Institut Pasteur is acknowledged for support with sample preparation, image acquisition and analysis, and its whole staff (Jean-Marie Winter, Stéphane Tachon and Matthijn Vos) is gratefully acknowledged for the supportive help. The NanoImaging Core was created with the help of a grant from the French Government's 'Investissements d'Avenir' program (EQUIPEX CACSICE – "Centre d'analyse de systèmes complexes dans les environnements complexes", ANR-11-EQPX-0008). L.Y. was funded by CSC (China Scholarship Council) and is part of the Pasteur - Paris University (PPU) International PhD program. The PPU project has received funding from the CNBG (China National Biotec Group Company Limited). We also acknowledge the synchrotron sources Soleil (Saint-Aubin, France), and ESRF (Grenoble, France) for granting access to their facilities, and their staff for helpful assistance on the respective beamlines. We would also like to thank Juan José Pierella Karlusich (ENS, Paris) and Aurélien Thureau (Synchrotron Soleil) for many helpful discussions.

REFERENCES

1. Perham, R. N. Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions. *Annu. Rev. Biochem.* **69,** 961–1004 (2000).

2. Milne, J. L. S. *et al.* Molecular structure of a 9-MDa icosahedral pyruvate dehydrogenase subcomplex containing the E2 and E3 enzymes using cryoelectron microscopy. *J. Biol. Chem.* **281,** 4364–4370 (2006).

3. Kato, M. *et al.* A synchronized substrate-gating mechanism revealed by cubic-core structure of the bovine branched-chain alpha-ketoacid dehydrogenase complex. *EMBO J.* **25,** 5983–5994 (2006).

4. Milne, J. L. S. *et al.* Molecular architecture and mechanism of an icosahedral pyruvate dehydrogenase complex: a multifunctional catalytic machine. *EMBO J.* **21,** 5587–5598 (2002).

587 5. Izard, T. *et al.* Principles of quasi-equivalence and Euclidean geometry govern the
588 assembly of cubic and dodecahedral cores of pyruvate dehydrogenase complexes. *Proc. Natl.*
589 *Acad. Sci. U S A* **96,** 1240–1245 (1999).
590 6. Mattevi, A. *et al.* Atomic structure of the cubic core of the pyruvate dehydrogenase
591 multienzyme complex. *Science* **255,** 1544–1550 (1992).
592 7. Mattevi, A., Obmolova, G., Kalk, K. H., Teplyakov, A. & Hol, W. G. J.
593 Crystallographic Analysis of Substrate Binding and Catalysis in Dihydrolipoyl Transacetylase
594 (E2p). *Biochemistry* **32,** 3887–3901 (1993).
595 8. Marrott, N. L. *et al.* The catalytic core of an archaeal 2-oxoacid dehydrogenase
596 multienzyme complex is a 42-mer protein assembly. *FEBS J.* **279,** 713–723 (2012).
597 9. Marrott, N. L. *et al.* Why are the 2-oxoacid dehydrogenase complexes so large?
598 Generation of an active trimeric complex. *Biochem J* **463,** 405–412 (2014).
599 10. Byron, O. & Lindsay, J. G. The pyruvate dehydrogenase complex and related
600 assemblies in health and disease. in *Macromolecular Protein Complexes* **83,** 523–550 (Springer
601 International Publishing, 2017).
602 11. Niebisch, A., Kabus, A., Schultz, C., Weil, B. & Bott, M. Corynebacterial protein
603 kinase G controls 2-oxoglutarate dehydrogenase activity via the phosphorylation status of the
604 OdhI protein. *J. Biol. Chem.* **281,** 12300–12307 (2006).
605 12. Wagner, T., Bellinzoni, M., Wehenkel, A., O'Hare, H. M. & Alzari, P. M. Functional
606 plasticity and allosteric regulation of α-ketoglutarate decarboxylase in central mycobacterial
607 metabolism. *Chem. Biol.* **18,** 1011–1020 (2011).
608 13. Hoffelder, M., Raasch, K., van Ooyen, J. & Eggeling, L. The E2 domain of OdhA of
609 *Corynebacterium glutamicum* has succinyltransferase activity dependent on lipoyl residues of
610 the acetyltransferase AceF. *J. Bacteriol.* **192,** 5203–5211 (2010).
611 14. Franke, D. *et al.* ATSAS 2.8: a comprehensive data analysis suite for small-angle
612 scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50,** 1212–1225 (2017).

613 15. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I.
614 PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl.*
615 *Crystallogr.* **36,** 1277–1282 (2003).
616 16. Knapp, J. E. *et al.* Crystal structure of the truncated cubic core component of the
617 *Escherichia coli* 2-oxoglutarate dehydrogenase multienzyme complex. *J. Mol. Biol.* **280,** 668
618 (1998).
619 17. Rückert, C., Eimer, J., Winkler, A. & Tauch, A. Complete Genome Sequence of the
620 Type Strain *Corynebacterium mustelae* DSM 45274, Isolated from Various Tissues of a Male
621 Ferret with Lethal Sepsis. *Genome Announc.* **3,** 871 (2015).
622 18. Venugopal, A. *et al.* Virulence of *Mycobacterium tuberculosis* depends on lipoamide
623 dehydrogenase, a member of three multienzyme complexes. *Cell Host Microbe* **9,** 21–31
624 (2011).
625 19. Usuda, Y. *et al.* Molecular cloning of the *Corynebacterium glutamicum*
626 ('*Brevibacterium lactofermentum*' AJ12036) odhA gene encoding a novel type of 2-
627 oxoglutarate dehydrogenase. *Microbiology* **142,** 3347–3354 (1996).
628 20. Reed, L. J. From lipoic acid to multi-enzyme complexes. *Protein Sci.* **7,** 220–224
629 (1998).
630 21. Guest, J. R., Angier, S. J. & Russell, G. C. Structure, expression, and protein
631 engineering of the pyruvate dehydrogenase complex of *Escherichia coli*. *Ann Ny Acad Sci* **573,**
632 76–99 (1989).
633 22. Reed, L. J., Fernandez-Moran, H., Koike, M. & Willms, C. R. Electron microscopic
634 and biochemical studies of pyruvate dehydrogenase complex of *Escherichia coli*. *Science* **145,**
635 930–932 (1964).
636 23. DeRosier, D. J., Oliver, R. M. & Reed, L. J. Crystallization and preliminary structural
637 analysis of dihydrolipoyl transsuccinylase, the core of the 2-oxoglutarate dehydrogenase
638 complex. *Proc. Natl. Acad. Sci. U S A* **68,** 1135–1137 (1971).
639 24. Patel, M. S., Nemeria, N. S., Furey, W. & Jordan, F. The pyruvate dehydrogenase
640 complexes: structure-based function and regulation. *J. Biol. Chem.* **289,** 16615–16623 (2014).

641 25. Zhou, Z. H., McCarthy, D. B., O'Connor, C. M., Reed, L. J. & Stoops, J. K. The
642 remarkable structural and functional organization of the eukaryotic pyruvate dehydrogenase
643 complexes. *Proc. Natl. Acad. Sci. U S A* **98,** 14802–14807 (2001).
644 26. Bunik, V. & Artiukhov, A. *Vitamin-Dependent Multienzyme Complexes of 2-Oxo Acid*
645 *Dehydrogenases: Structure, Function, Regulation and Medical Implications*. (Nova Science
646 Publishers, 2017).
647 27. Lewin, G. R. *et al.* Evolution and Ecology of Actinobacteria and Their Bioenergy
648 Applications. *Annu. Rev. Microbiol.* **70,** 235–254 (2016).
649 28. Rieck, B. *et al.* PknG senses amino acid availability to control metabolism and
650 virulence of *Mycobacterium tuberculosis*. *PLoS Pathog* **13,** e1006399 (2017).
651 29. Ventura, M. *et al.* GarA is an essential regulator of metabolism in *Mycobacterium*
652 *tuberculosis*. *Mol. Microbiol.* **90,** 356–366 (2013).
653 30. O'Hare, H. M. *et al.* Regulation of glutamate metabolism by protein kinases in
654 mycobacteria. *Mol. Microbiol.* **70,** 1408–1423 (2008).
655 31. Wagner, T. *et al.* Structural insights into the functional versatility of an FHA domain
656 protein in mycobacterial signaling. *Science Signaling* **12,** eaav9504 (2019).
657 32. Shiio, I. & Ujigawatakeda, K. Presence and Regulation of Alpha-Ketoglutarate
658 Dehydrogenase Complex in a Glutamate-Producing Bacterium, *Brevibacterium flavum*. *Agr.*
659 *Biol. Chem. Tokyo* **44,** 1897–1904 (1980).
660 33. Blombach, B., Cramer, A., Eikmanns, B. J. & Schreiner, M. RamB Is an Activator of
661 the Pyruvate Dehydrogenase Complex Subunit E1p Gene in *Corynebacterium glutamicum*. *J.*
662 *Mol. Microbiol. Biotechnol.* **16,** 236–239 (2009).
663 34. Rhee, K. Y. *et al.* Central carbon metabolism in *Mycobacterium tuberculosis*: an
664 unexpected frontier. *Trends Microbiol.* **19,** 307–314 (2011).
665 35. de Carvalho, L. P. S. *et al.* Metabolomics of *Mycobacterium tuberculosis* reveals
666 compartmentalized co-catabolism of carbon substrates. *Chem. Biol.* **17,** 1122–1131 (2010).
667 36. Bryk, R. *et al.* Selective killing of nonreplicating mycobacteria. *Cell Host Microbe* **3,**
668 137–145 (2008).

669 37. Gouzy, A. *et al. Mycobacterium tuberculosis* exploits asparagine to assimilate nitrogen
670 and resist acid stress during infection. *PLoS Pathog.* **10,** e1003928 (2014).
671 38. Maksymiuk, C., Balakrishnan, A., Bryk, R., Rhee, K. Y. & Nathan, C. F. E1 of α-
672 ketoglutarate dehydrogenase defends *Mycobacterium tuberculosis* against glutamate
673 anaplerosis and nitroxidative stress. *Proc. Natl. Acad. Sci. U S A* **112,** E5834–43 (2015).
674 39. Unger, T., Jacobovitch, Y., Dantes, A., Bernheim, R. & Peleg, Y. Applications of the
675 Restriction Free (RF) cloning procedure for molecular manipulations and protein expression.
676 *J. Struct. Biol.* **172,** 34–44 (2010).
677 40. van den Berg, S., Lofdahl, P. A., Hard, T. & Berglund, H. Improved solubility of TEV
678 protease by directed evolution. *J. Biotechnol.* **121,** 291–298 (2006).
679 41. Hajizadeh, N. R., Franke, D., Jeffries, C. M. & Svergun, D. I. Consensus Bayesian
680 assessment of protein molecular mass from solution X-ray scattering data. *Sci. Rep.* **8,** 7204–
681 13 (2018).
682 42. Pérez, J., Vachette, P., Russo, D., Desmadril, M. & Durand, D. Heat-induced unfolding
683 of neocarzinostatin, a small all-beta protein investigated by small-angle X-ray scattering. *J.*
684 *Mol. Biol.* **308,** 721–743 (2001).
685 43. Durand, D. *et al.* NADPH oxidase activator p67(phox) behaves in solution as a
686 multidomain protein with semi-flexible linkers. *J. Struct. Biol.* **169,** 45–53 (2010).
687 44. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity
688 ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78,** 1606–1619 (2000).
689 45. Weber, P. *et al.* High-Throughput Crystallization Pipeline at the Crystallography Core
690 Facility of the Institut Pasteur. *Molecules* **24,** (2019).
691 46. Bowler, M. W. *et al.* MASSIF-1: a beamline dedicated to the fully automatic
692 characterization and data collection from crystals of biological macromolecules. *J. Synchrotron*
693 *Radiat.* **22,** 1540–1547 (2015).
694 47. Kabsch, W. XDS. *Acta Crystallog.r D Biol. Crystallog.r* **66,** 125–132 (2010).
695 48. Vonrhein, C. *et al.* Data processing and analysis with the autoPROC toolbox. *Acta*
696 *Crystallogr D Biol. Crystallogr.* **67,** 293–302 (2011).

697 49. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta*
698 *Crystallog.r D Biol. Crystallogr.* **67,** 235–242 (2011).
699 50. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674
700 (2007).
701 51. Wang, J. *et al.* Structure and function of the catalytic domain of the dihydrolipoyl
702 acetyltransferase component in *Escherichia coli* pyruvate dehydrogenase complex. *J. Biol.*
703 *Chem.* **289,** 15215–15230 (2014).
704 52. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of
705 Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501 (2010).
706 53. Smart, O. S. *et al.* Exploiting structure similarity in refinement: automated NCS and
707 target-structure restraints in BUSTER. *Acta Crystallogr. D Biol. Crystallogr.* **68,** 368–380
708 (2012).
709 54. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom
710 structure validation. *Protein Science* **27,** 293–315 (2018).
711 55. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons
712 and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75,** 861–877
713 (2019).
714 56. Schrödinger, LLC. The PyMOL Molecular Graphics System. (2012).
715 57. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for
716 improved cryo-electron microscopy. *Nat. Meth.* **14,** 331–332 (2017).
717 58. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-
718 atomic-resolution single-particle cryo-EM. *Nat. Meth.* **10,** 584–590 (2013).
719 59. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193,** 1–
720 12 (2016).
721 60. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms
722 for rapid unsupervised cryo-EM structure determination. *Nat. Meth.* **14,** 290–296 (2017).
723 61. Ramírez-Aportela, E. *et al.* Automatic local resolution-based sharpening of cryo-EM
724 maps. *Bioinformatics* **36,** 765–772 (2020).

725 62. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research
726 and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
727 63. Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and
728 crystallography. *Acta Crystallogr. D Struct. Biol.* **74,** 531–544 (2018).
729 64. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high
730 throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).
731 65. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large
732 datasets. *Bioinformatics* **30,** 3276–3278 (2014).
733 66. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new
734 ENDscript server. *Nucleic Acids Res.* **42,** W320–4 (2014).
735 67. Chen, I.-M. A. *et al.* IMG/M v.5.0: an integrated data management and comparative
736 analysis system for microbial genomes and microbiomes. *Nucleic Acids Re.s* **47,** D666–D677
737 (2019).
738 68. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v4: recent updates and new
739 developments. *Nucleic Acids Res.* **47,** W256–W259 (2019).
740
741
742 **AUTHOR CONTRIBUTIONS**
743 E.M.B., P.V., N.L.-S. and L.Y. generated expression plasmids and produced recombinant
744 proteins; E.M.B., P.V., L.Y. and M.B. collected X-ray diffraction data; M.B. solved and refined
745 crystallographic structures; E.M.B., P.V., N.L-S. and B.R. collected and analyzed AUC data;
746 E.M.B. and B.R. collected and analyzed SAXS data; E.M.B. collected and analyzed cryo-EM
747 data, and performed phylogeny analysis; P.M.A. provided advice on data interpretation and
748 contributed to the manuscript; M.B. supervised the work; E.M.B. and M.B. wrote the paper.
749 All authors reviewed the manuscript and agreed on its content.
750
751

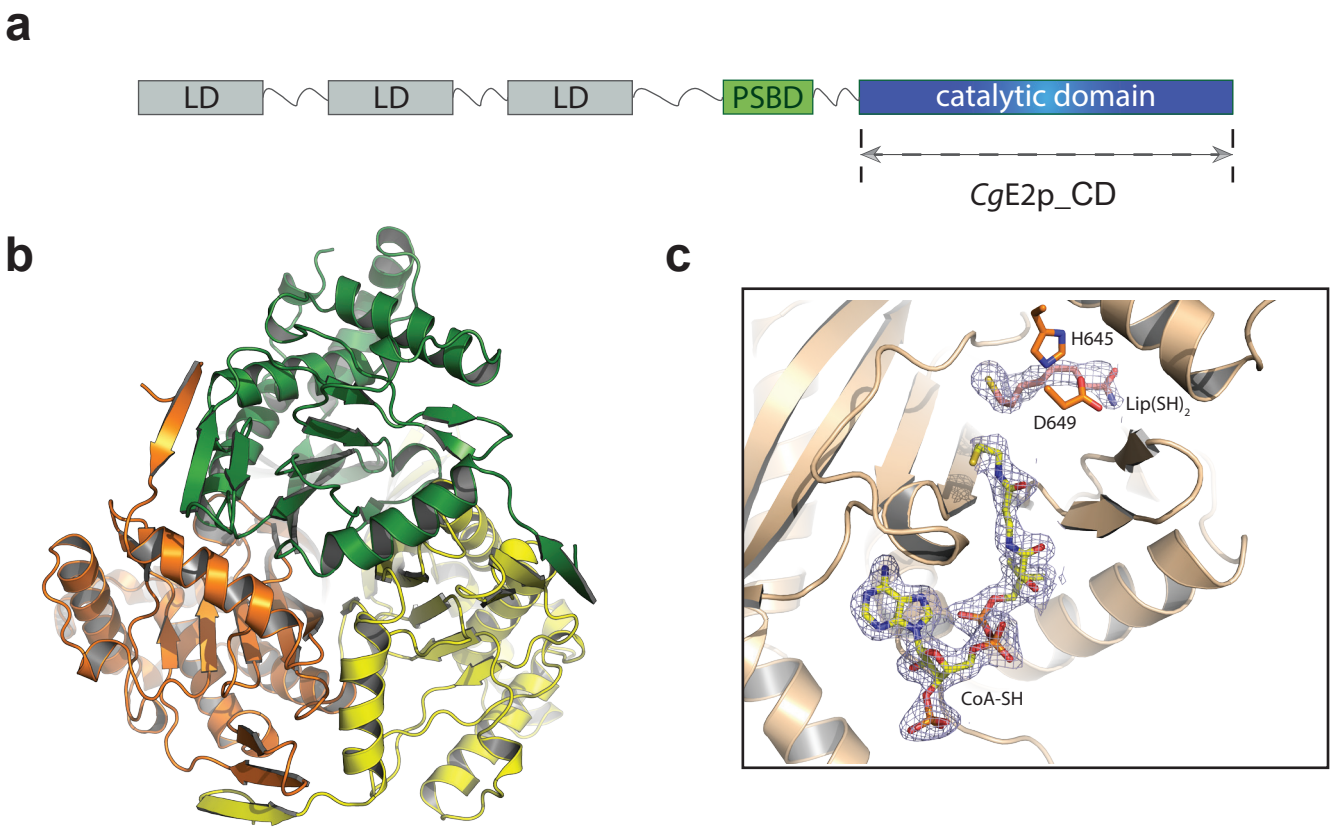


Figure 1: Crystal structure of *Cg*E2p_CD shows a trimeric protein without inter-trimer interactions.

a- Schematic representation of the domain architechture of *Cg*E2p_FL. The arrow indicates the region selected for the *Cg*E2p_CD construct. **b**- Ribbon representation of the apo form of *Cg*E2p_CD showing its homotrimeric structure, with monomers highlighted in orange, green and yellow, respectively. **c**- Zoomed view of the active site region of ternary complex formed by *Cg*E2p_CD with its sustrates CoA and dihydrolipoamide ($Lip(SH)_2$). The 2mFo-DFc Fourier electron density for both ligands, contoured at the 1σ level, is shown as a blue mesh.

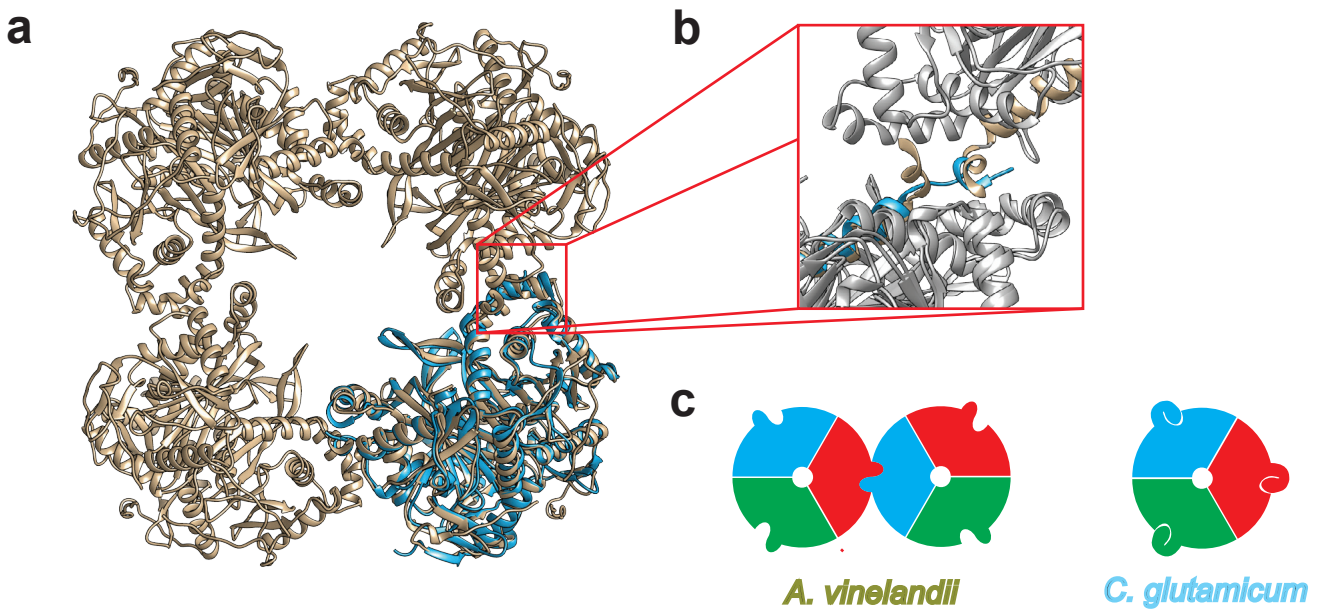


Figure 2: A unique TTI helix orientation is responsable for the absence of inter-trimer interaction.

a- Ribbon representation of the trimeric *Cg*E2p_CD (cyan) superimposed to *Av*E2p_CD, which forms a 24-mer cubic crystal structure (PDB entry 1EAB). **b**- Zoomed view of the trimer-trimer interaction region showing how the distinct conformation of the TTI helix in *Cg*E2p locates the helix in the position of the incoming TTI helix of *Av*E2p. **c**- On the left, schematic representation of the inter-trimer interaction in *Av*E2p, with different colours for each monomer and the protruding extension representing the C-terminal TTI, allowing the assembly of the 24-mer cube. On the right, the intramolecular interaction of the same TTI helix in *Cg*E2p, hindering trimer-trimer interactions and therefore blocking the protein in a trimeric state.

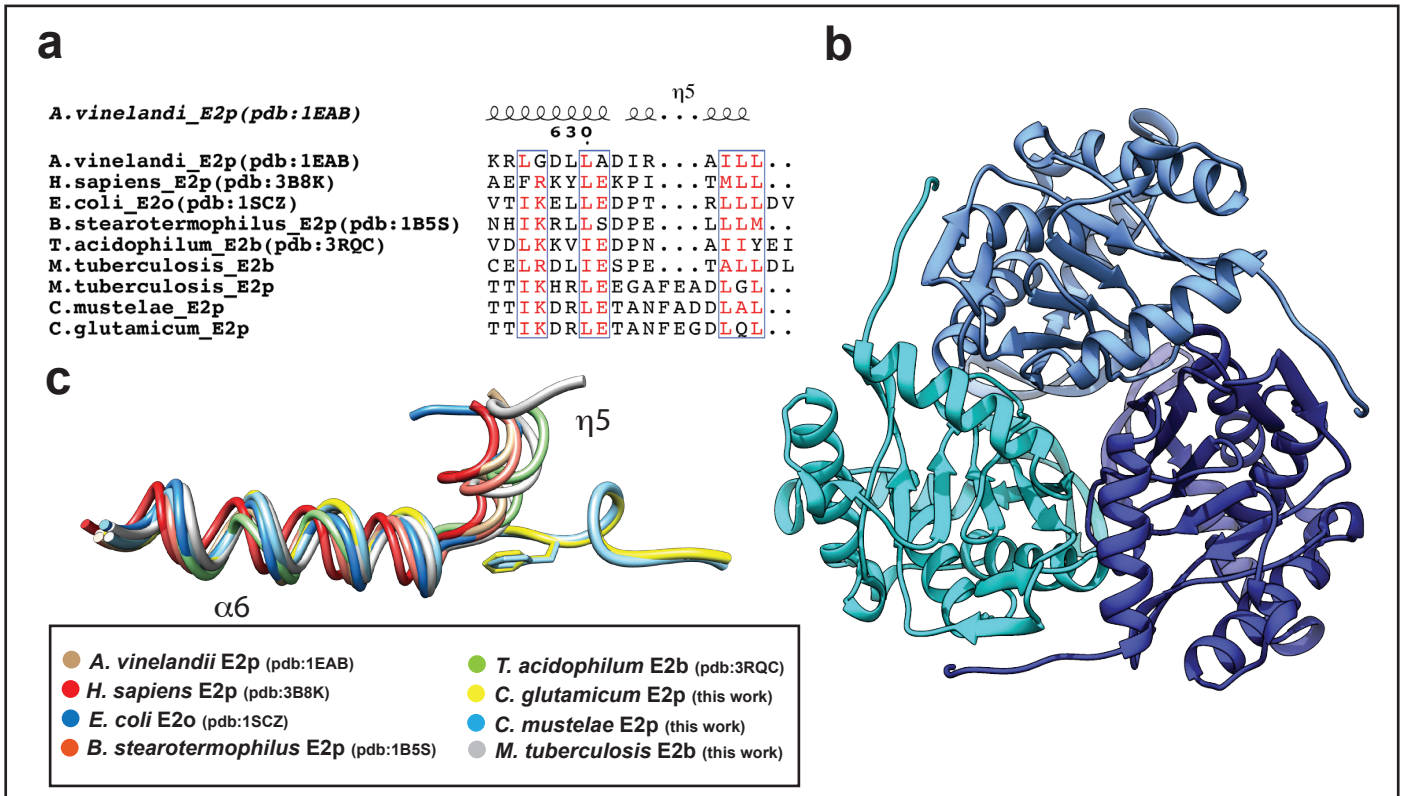


Figure 3: A three residues insertion (PCI) is responsable for the unique TTI helix orientation that blocks inter-trimer interactions.

a- Sequence alignment, limited to the C-terminal region, between several E2 enzymes of known structure with: *Mt*E2b, *Mt*E2p, *Cg*E2p and *Cm*E2p (this work). **b**- Trimeric structure of *Cm*E2p_CD shown in ribbon representation with different tones of blue per monomer. **c**- Licorized representation corresponding to the C-terminal region of the structural alignment of *Cg*E2p_CD (cyan) and *Cm*E2p_CD (yellow) with: E2p from *A.vinelandii*, *H. sapiens*, *E. coli, B. stearothemophilus* and *M. smegmatis*, E2o from *E. coli* and E2b from *T. acidophilum.*

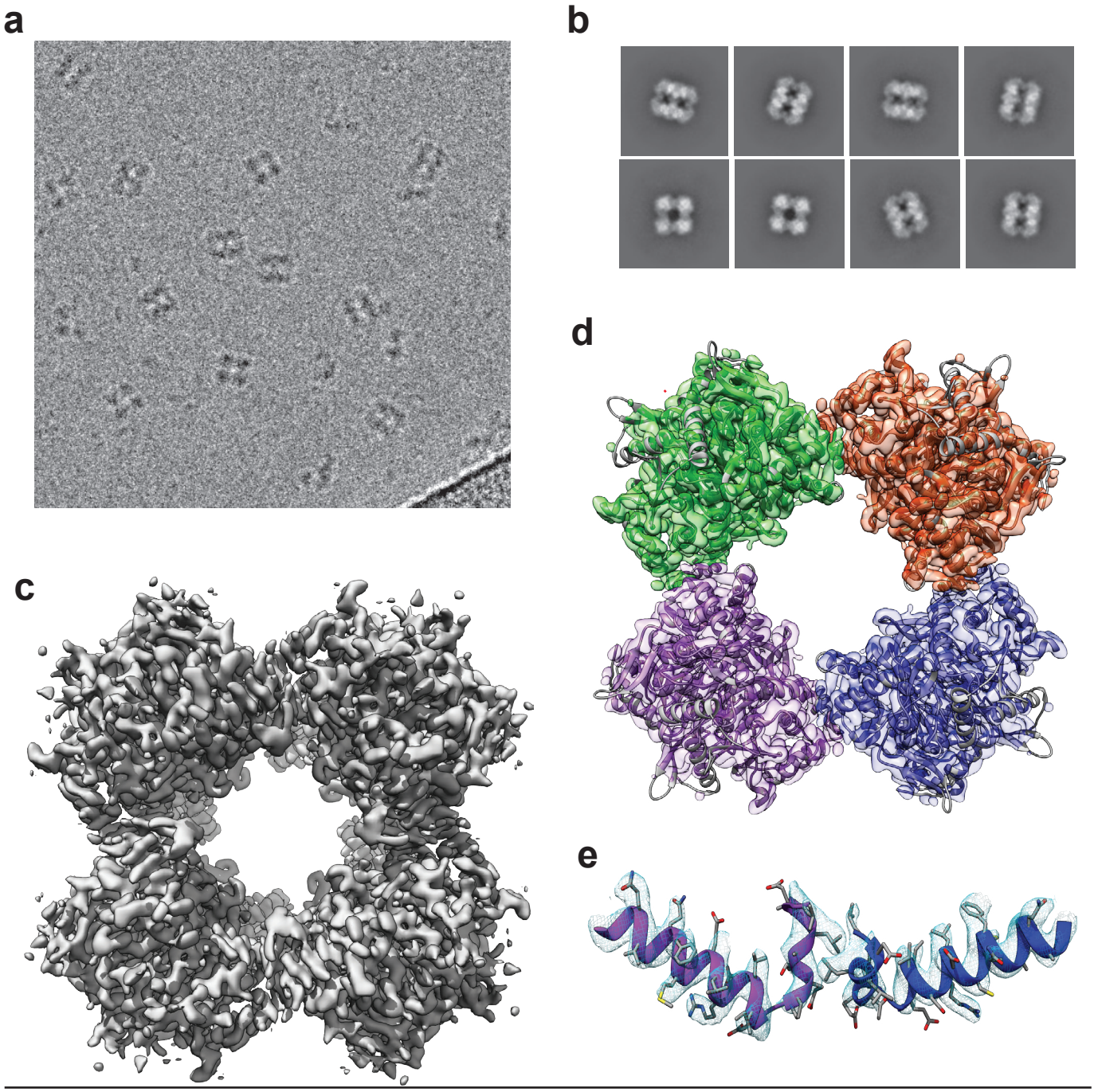
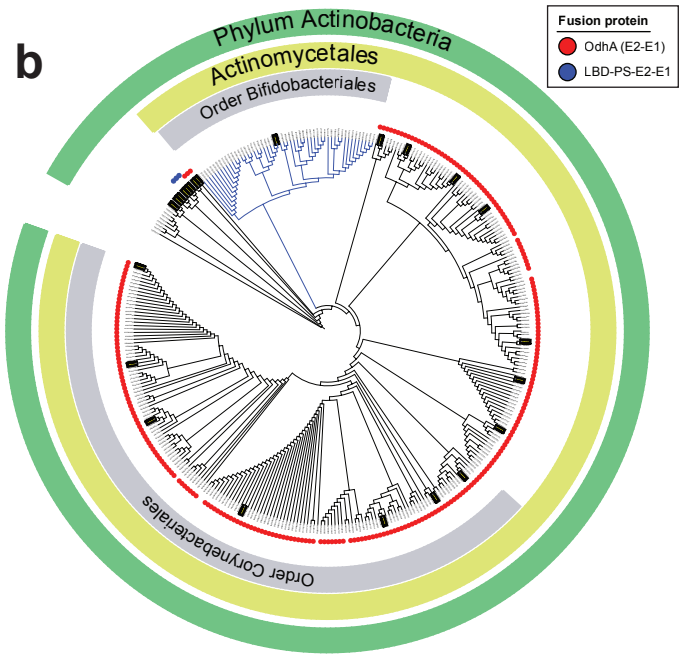
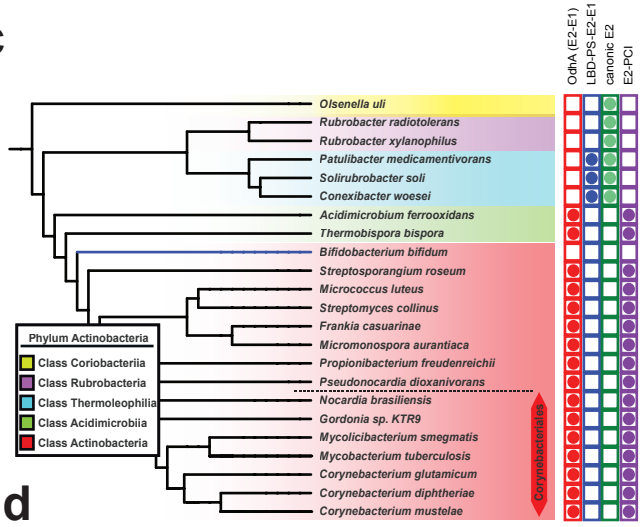


Figure 4: CryoEM structure of *Mt*E2b_CD, an actinobacterial E2b enzyme lacking the PCI, shows canonical TTI helix and quaternary structure assembly. **a**- CryoEM micrograph showing individual particles of *Mt*E2b_CD. **b**- Representative 2D classes showing different particle orientations. **c**- CryoEM map of *Mt*E2b_CD showing its cubic 24-mer with octahedral symmetry. **d**- Crystal structure of the same E2b_CD, extended to the full 24-mer cube (this work), fitted into the EM electron density. **e**- Zoomed region corresponding to the EM electron density of the C-terminal section of the protein, showing the canonical TTI helix orientation observed in other E2 enzymes.

a

Domain/Phylum	Number of finished genomes in JGI	Number of genomes with odhA-like gen (pfam: 00198, 00676, 02779, 16870)	Percentage of genomes with odhA-like gen
Archae	173	0	0
Eukaryota	37	0	0
Bacteria(non Actino)	2957	9	0,3
Actinobacteria	356	280	78,6





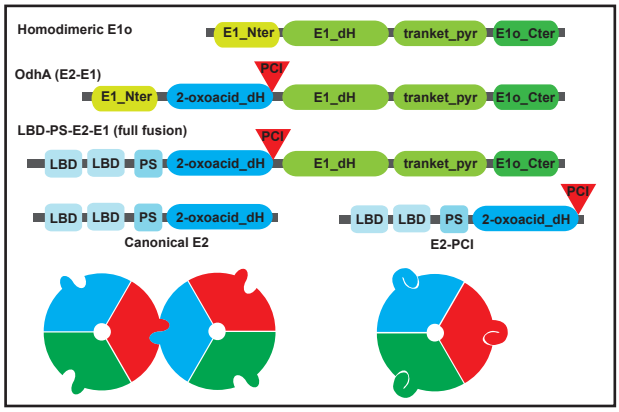


Figure 5: Distribution of OdhA homologs and PCI inside the phylum Actinobacteria.

a- Table showing the distribution of OdhA homologous found using the conservation of domain architecture predicted from the associated pfams. **b**- Distribution of OdhA homologous (red dots) or the full fusion LBD-PS-E2-E1 (blue dots) inside the phylum Actinobacteria. Branches highlighted in black correspond to genomes selected to be and analyzed for the PCI presence in E2. **c**- Sequences selected from the actinobacterial tree to verify the presence of the PCI.The table shows the correlation between the presence of OdhA and the PCI in E2p. **d**- Domain architecture of E1o, OdhA, the full fusion (LBD-PS-E2-E1) and E2p represented by the pfams used to identify the proteins. Green pfams are associated to homodimeric E1o (Pfam16078:2-oxogl_dehyd_N, Pfam00676:E1-dh, Pfam02279:Transket_pyr and Pfam16870:OxoGdeHyase_C) while blue pfams are associated to canonical E2 (Pfam00364:Lypoil binding domain, Pfam02817:E3 binding domain and Pfam00198:2-oxoacid dH).

Protein	locus name	UNIPROT code	Residues	MW monomer from sequence (kDa)	MW estimaton AUC (kDa)	Frictional ratio extrapolated	Number of oligomers (AUC)	MW estimaton SAXS Bayessian (kDa)	Number of oligomers (SAXS)
Cg E2p_FL	cg2421	Q8NNJ2	FL	70,91	221	2,85	3,1	243	3,4
Cg E2p_CD	cg2421	Q8NNJ2	437-675	25,40	78	1,40	3,1	77	3,0
Mt E2p_FL	Rv2215	P9WIS7	FL	57,09	141	1,71	2,5	318	5,6
Mt E2b_FL	Rv2495c	O06159	FL	41,21	731	1,63	17,7	1179	28,6
Mt E2b_CD	Rv2495c	O06159	165-393	24,67	600	1,48	24,3	479	19,4

Table 1: Oligomerization state of proteins in solution determined by AUC and SAXS

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	$P6_322$	$P2_13$	$P6_4$	$P6_322$	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 / σ(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 (°)	0.91	1.04	0.91	0.96	0.91	0.93
PDB entry code	6ZZI	6ZZJ	6ZZK	6ZZL	6ZZM	6ZZN

Table 2. Crystallographic data collection and refinement statistics (highest resolution shell in parenthesis).