



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

► To cite this version:

Eduardo M. Bruch, Pierre Vilela, Lu Yang, Alexandra Boyko, Norik Lexa-Sapart, et al.. Actinobacteria challenge the paradigm: A unique protein architecture for a well-known, central metabolic complex. *Proceedings of the National Academy of Sciences of the United States of America*, 2021, 118 (48), pp.e2112107118. 10.1073/pnas.2112107118 . pasteur-03458284

HAL Id: pasteur-03458284

<https://pasteur.hal.science/pasteur-03458284>

Submitted on 1 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

4 Eduardo M. Bruch^{1*}, Pierre Vilela¹, Lu Yang^{1,2}, Alexandra Boyko¹, Norik Lexa-Sapart¹,
5 Bertrand Raynal³, Pedro M. Alzari¹, Marco Bellinzoni^{1*}

15 *Correspondence: eduardobruch@gmail.com and marco.bellinzoni@pasteur.fr

20 §Present addresses: E.M.B: BioStructure and Biophysics, Integrated Drug Discovery, Sanofi
21 R&D, 13 Quai Jules Guesde, 94403 Vitry-sur-Seine Cedex, France; P.V.: IGBMC, 1 rue
22 Laurent Fries, 67404 Illkirch cedex, France; N.L.-S.: ALTEN, 221 bis Boulevard Jean Jaurès,
23 92514 Boulogne-Billancourt, France

27 **ABSTRACT**

28 α -ketoacid dehydrogenase complexes are large, tripartite enzymatic machineries carrying out
29 key reactions in central metabolism. Extremely conserved across the tree of life, they have so
30 far all considered to be structured around a high molecular weight hollow core, consisting of
31 up to 60 subunits of the acyltransferase component. We provide here evidence that
32 Actinobacteria break the rule by possessing an acetyltransferase component reduced to its
33 minimally active, trimeric unit, characterized by a unique C-terminal helix bearing an
34 actinobacterial-specific insertion that precludes larger protein oligomerization. This particular
35 feature, together with the presence of an *odhA* gene coding for both the decarboxylase and the
36 acyltransferase domains on the same polypeptide, is spread over Actinobacteria and reflects the
37 association of PDH and ODH into a single physical complex. Considering the central role of
38 the pyruvate and 2-oxoglutarate nodes in central metabolism, our findings pave the way to
39 both therapeutic and metabolic engineering applications.

40

41 **SIGNIFICANCE STATEMENT**

42 α -ketoacid dehydrogenase are large, evolutionary conserved multi-enzymatic complexes that
43 carry out key oxidative reactions in central metabolism. For decades, these complexes were
44 thought to share a similar organization in all aerobic organisms, structured around a hollow
45 core resulting from the high oligomeric arrangement of the acyltransferase component. Here
46 we show that Actinobacteria, one of the largest eubacterial phyla, break this ‘dogma’ owing
47 to a distinct trimeric acyltransferase core. We provide a structural explanation to this different
48 oligomeric architecture and show how this feature, related to a fusion of two genes coding for
49 2-oxoglutarate dehydrogenase, constitutes a trait of Actinobacteria and sheds new light on the
50 evolution of these metabolic complexes.

51

52 INTRODUCTION

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

79 Archaea and Eukarya. While the reasons for the presence (and evolutionary conservation) of
80 such huge macromolecular scaffolds remain unclear (9), active site coupling (transfer of acyl
81 groups between lipoyl domains within the core) has often been proposed as the major
82 advantage (1, 10). Also, despite the tripartite organization of these complexes as separate
83 E1/E2/E3 enzymes had always been considered as universal, *Corynebacterium glutamicum*
84 was shown to be deprived of an E2o component (specific to the ODH complex), and to rather
85 possess an E2o succinyl transferase domain fused to E1o, in a protein called OdhA (11, 12).
86 The same situation has then been confirmed for the model mycobacterium *M. smegmatis* (13).
87 As the lipoyl-binding domain of E2o is absent from the E2o-E1o fusion, the ODH activity
88 depends on functional lipoyl groups provided *in trans*, and proven to be supplied by E2p from
89 the PDH complex (14), therefore suggesting the presence of a mixed PDH/ODH
90 supercomplex. By using an integrative structural biology approach, we describe here how *C.*
91 *glutamicum* E2p, that was expected to serve as the core of the mixed complex, breaks the rule
92 about the oligomeric state of acyltransferase E2 enzymes reducing its size to the minimal,
93 catalytically active trimeric unit. We also provide evidence supporting these features as a
94 common trait of Actinobacteria.

95

96 **RESULTS**97 ***M. tuberculosis* and *C. glutamicum* E2p are both trimeric in solution**

98 To get insight into the oligomeric state of E2p in *Corynebacteriales*, full-length E2p from
99 both *M. tuberculosis* (*MtE2p_FL*) and *C. glutamicum* (*CgE2p_FL*) were successfully
100 expressed in *E. coli* and purified to homogeneity. Analytical ultracentrifugation (AUC) was
101 used to assess the oligomeric assembly of the proteins (Table 1 and Fig. S1). Strikingly, the
102 ratio between the estimated molecular weight in solution and the one predicted from the
103 sequence indicated a trimeric state in solution for both samples, with no experimental
104 evidence of higher oligomeric states. The same result was observed for the *CgE2p* catalytic
105 domain alone (*CgE2p_CD*, comprising residues Leu437-Leu675). No larger assembly,
106 suggestive of a 24-mer or 60-mer state, was observed for any of these proteins in the
107 conditions tested (Table 1). To check for proper functionality of the recombinant proteins, we
108 determined the specific transacylase activity of full-length proteins according to the
109 spectrophotometric assay by Hall and Weitzman (15) (Fig. S2). Both proteins were indeed
110 able to catalyze the transfer of acetyl units from acetyl-CoA to free dihydrolipoamide, in turn
111 generated by commercial dihydrolipoamide dehydrogenase with the concomitant oxidation of
112 NADH (Fig. S2), resulting in a specific activity of $17.3 \pm 0.9 \text{ nmol min}^{-1}\text{mg}^{-1}$ for *CgE2p_FL*
113 vs. $2.1 \pm 0.1 \text{ nmol min}^{-1}\text{mg}^{-1}$ for *MtE2p_FL*.

114

115 The absence of the expected higher oligomeric structure was further confirmed by size
116 exclusion chromatography coupled to SAXS analysis, where the molecular mass was
117 estimated using MW estimation tools available in the ATSAS suite (16, 17) (Table 1 and
118 Table S1). For both *CgE2p_FL* and *CgE2p_CD*, the trimeric form was predominant in
119 solution, whereas for *MtE2p_FL* the analysis showed a molecular weight estimation that
120 could correspond to a dimer of trimers (Table 1). However, no evidence of a cubic,
121 dodecameric or any other higher oligomeric state was observed for any of the three proteins.
122 For *CgE2p_CD*, constant radius of gyration (R_g) through the intensity peak indicated the

123 presence of a unique oligomeric species (Fig. S3); R_g and D_{max} experimental values were 27.7
124 Å and 84.7 Å, respectively (Table S1 and Fig. S3). The dimensionless Kratky plot showed a
125 gaussian bell shape that decreases to zero, suggestive of a folded, globular and compact
126 structure (Fig. S3), and consistently with the AUC results that indicated a trimeric *CgE2p_CD*
127 protein with low friction coefficient. For the two full-length proteins, *MtE2p_FL* and
128 *CgE2p_FL*, the dimensionless Kratky plot showed a non-gaussian bell shape followed by a
129 shoulder reaching a plateau above zero, consistent with a properly folded, but not globular,
130 multidomain protein with flexible linkers (Fig. S3). This result, together with the high friction
131 coefficient observed by AUC, was suggestive of well folded, highly elongated proteins in
132 both cases.

133

134 ***CgE2p_CD* possesses a unique C-terminal region characterized by an insertion that**
135 **blocks the protein in its trimeric form**

136 To understand the reasons for *CgE2p* behaving as a trimeric protein and not assuming a
137 higher oligomerization state, we characterized it structurally, starting by extensive
138 crystallization trials of its catalytic core (*CgE2p_CD*; Fig. 1a). Crystals were obtained from
139 the protein in the absence of ligands, in the presence of the acceptor substrate CoA and in a
140 ternary complex with CoA and dihydrolipoamide (Table S2). X-ray diffraction showed that
141 the samples belonged to three different crystallographic space groups, with resolutions
142 ranging from 2.1 to 1.35 Å. In all the crystal forms observed, *CgE2p_CD* showed the
143 expected, chloramphenicol acetyltransferase-like fold, with the same, obligate homotrimeric
144 assembly (Fig. 1b) first observed for the *A. vinelandii* orthologue (*AvE2p_CD*) (6). The active
145 site environment at the interface between subunits in the functional trimer was also conserved
146 (Fig. S4a), with His645 and Asp649 as the equivalent residues to His610 and Asn614 in
147 *AvE2p_CD*. However, the crystal structure provided no evidence of any higher-order
148 quaternary assembly. To investigate whether the trimeric assembly of *CgE2p_CD* was still
149 competent to substrate binding, as indicated by its detectable transacylase activity (3.9 ± 0.1
150 nmol min⁻¹mg⁻¹ with acetyl-CoA as the donor substrate), we proceeded to co-crystallization

151 trials with CoA, dihydrolipoamide (Lip(SH)₂), DTT (as an antioxidant) or their possible
152 combinations. Co-crystallization with CoA yielded high resolution diffracting crystals (Table
153 S2) which show a full-occupancy complex, with no significant conformational change with
154 respect to the *apo* structure. Overall, the structure was similar to the equivalent complex from
155 *A. vinelandii* (PDB 1EAD, (7)), with the adenosine 3'-phosphate moiety occupying basically
156 the same position in both enzymes, CoA assuming a 'IN' conformation with the pantetheine
157 chain entering the active site tunnel, and the guanidinium group of Arg480 (Arg450 in
158 AvE2p_CD) making a salt bridge with the CoA 3'-phosphate (Fig. S4b). Likewise, electron
159 density maps indicated oxidation of the CoA terminal thiol group, as in AvE2p (Fig. S4b).
160 In addition, a ternary complex was obtained by co-crystallization with CoA and free
161 dihydrolipoamide, similarly to AvE2p_CD (7), although it belonged to a different crystalline
162 form with respect to the CoA complex (Table S2). This structure showed the same, 'IN'
163 conformation for the CoA pantetheine chain and the dihydrolipoamide molecule in an
164 orientation close, albeit not identical to the one observed for the corresponding complex in
165 AvE2p_CD (PDB 1EAB; Fig. 1c, Fig. S5).

166

167 Overall, CgE2p_CD was very similar to the available structures of homologous
168 acyltransferase domains from other E2 enzymes, with RMSDs of 1.1-1.3 Å over 220-225
169 aligned residues, like E2p from *Bacillus stearothermophilus* (5) (PDB 1B5S, with which the
170 rms difference is 1.10 Å over 222 residues), SucB (E2o) from *E. coli* (18) (PDB 1SCZ, rmsd
171 1.29 Å over 211 residues) and E2b from *Bos taurus* (3)(PDB 2II4, average rmsd 1.3 Å over
172 221-223 aligned residues), despite the low sequence identity (lower than 25% in all cases;
173 Fig. S6). However, all previously known enzymes shared a higher order oligomeric assembly
174 displaying 432 or 532 symmetry, with the functional homotrimers acting as the 'building
175 blocks' that interact with each other through their C-terminal, 3₁₀ helices (1, 5). Interestingly,
176 this helix showed the most striking differences when CgE2p_CD was superimposed to
177 AvE2p_CD (Fig. 2a/2b). While in AvE2p_CD this trimer-trimer interacting helix (TTI helix)
178 was roughly perpendicular to the preceding helix and protruded outward positioning its

179 hydrophobic residues into a pocket on the facing trimer, in the case of the corynebacterial
180 enzyme the helix laid on the same monomer surface, with the C-terminal hydrophobic
181 residues occupying the same pocket in an intramolecular interaction (Fig. 2b). In other words,
182 the position of the C-terminal helix for *CgE2p_CD* coincided with the position of the same,
183 incoming helix from the facing trimer in the *A. vinelandii* cubic structure (Fig. 2b). As a
184 consequence, the ‘knobs and sockets’ interaction between trimers is replaced by an
185 intramolecular interaction involving the same features, hindering any trimer-trimer
186 intermolecular interaction and therefore forcing the protein in an homotrimeric state (Fig. 2c).

187

188 The above differences were also reflected by the distribution of the hydrophobic residues
189 (Fig. 3a), which in oligomeric *AvE2p* are clustered at the helix C-terminus (Fig. S6), whereas
190 in *CgE2p* are arranged at one side of the helix, thus preserving the mostly hydrophobic nature
191 of the helix-pocket interaction (Fig. 3b). This property is exemplified by the replacement of
192 Leu637 in *AvE2p* for a polar residue (Gln674) in *CgE2p*, which reflects the new amphipathic
193 nature of the helix. Most importantly, the different orientation of the terminal helix in
194 *CgE2p_CD* is due to the presence of a three-residue insertion (Phe669-Glu670-Gly671) in
195 *CgE2p* (Fig. 3a). This insertion allows the C-terminal helix to change its orientation (Fig. 3c)
196 and lay on the surface to reach the accommodating pocket *in cis* (Fig. 3b), thus replacing an
197 intermolecular interaction by an intramolecular one (Fig. 2c). To confirm that this structural
198 arrangement was not due to the truncated *CgE2p_CD* protein, a new *CgE2p* construct
199 including the peripheral subunit binding domain (PSBD), was also overexpressed in *E. coli*,
200 purified and crystallized, yielding a new, 2.5 Å resolution structure (Table S2) showing the
201 catalytic domain in addition to an N-terminal, 13-residue extension that represent part of the
202 flexible linker connecting to the PSBD (Fig. S7). Noteworthy, no difference in the oligomeric
203 state, nor in the position of the C-terminal, 3₁₀ helix was observed compared to the previous
204 structures (Fig. S7). In addition, the E2p ortholog from *M. tuberculosis*, *MtE2p*, also
205 possesses a three-residue insertion (Phe547-Glu548-Ala549) in an equivalent position just
206 preceding the predicted 3₁₀ terminal helix (Fig. 3a), suggesting that the Phenylalanine-

207 containing insertion (PCI) could be responsible of the lack of higher order oligomerization
208 states. To check this hypothesis, we generated a series of seven synthetic *CgE2p_CD*
209 constructs coding mutant versions of the catalytic domain in which the C-terminal end
210 bearing the PCI and following helix were replaced by the corresponding ones from *E. coli*
211 *SucB* (*EcE2o*), the closest E2 enzyme in terms of sequence homology having a known
212 structure (44% sequence identity for the catalytic domain; Fig. S6). The designed mutants
213 were distinguished by a few additional amino acid substitutions to adapt the binding pocket to
214 an intermolecular interaction, incrementing the number of substitutions (Fig. S8). The
215 construct expressing mutant M3 was selected for the best solubility profile upon
216 overexpression in *E. coli*; this mutein showed a polydisperse size-exclusion chromatography
217 profile that indicates the presence of a continuum of higher molecular weight species (Fig.
218 S9). However, its specific transacetylase activity ($3.2 \pm 0.6 \text{ nmol min}^{-1}\text{mg}^{-1}$) was not
219 significantly affected when compared to the wild-type construct expressing the catalytic core
220 ($3.9 \pm 0.1 \text{ nmol min}^{-1}\text{mg}^{-1}$).

221

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

223 To analyze the distribution of the PCI, we performed sequence analysis of the E2-coding
224 genes in representative genomes distributed within the phylum Actinobacteria (Fig. S10). Our
225 results show that the PCI was not only present in *Corynebacteriales*, including *M.*
226 *tuberculosis* and *C. glutamicum*, but more generally in sequences from two classes, namely
227 *Actinobacteria* (also called Actinomycetales that represent the biggest class inside the Phylum
228 Actinobacteria) and *Acidimicrobia* (Fig. S10). The high conservation of the Phe residue
229 present in the insertion emphasizes the importance of its interaction within the context of the
230 hydrophobic pocket. Other residues involved in H-bond interactions with the pocket (mainly
231 Glu670, Leu675 and Leu677) were also conserved (Fig. S10). Interestingly, although the PCI
232 is conserved, the TTI helix region has lost some of the C-terminal residues and even the full
233 C-terminal helix is absent in some of the organisms, in agreement with the absence of the
234 inter-trimer interactions in which this helix is involved.

235

236 **PCI and E2 oligomerization**

237 To confirm the relationship between the presence of the three-residue, actinobacterial-specific
238 C-terminal insertion (PCI) and the oligomerization order of E2 proteins from the
239 Actinobacteria class, we followed two parallel paths. First, we structurally analyzed the
240 orthologous protein from *Corynebacterium mustelae*, a recently described species from the
241 same genus (19), as an example of orthologue bearing the same insertion. The catalytic
242 domain of the predicted E2p enzyme from this species (*CmE2p_CD*, residues 428-666) was
243 crystallized in complex with CoA, and its structure solved by molecular replacement to 2.5 Å
244 resolution (Table S2). The crystal structure shows an homotrimeric arrangement of
245 *CmE2p_CD* (Fig. 3d) with no significant structural changes with respect to the *C. glutamicum*
246 counterpart (RMSD of 0.79 Å over the whole trimer), and without evidence of any relevant
247 higher order oligomeric state. In particular, the C-terminal 3₁₀ helix had the same relative
248 orientation as in *CgE2p_CD* (Fig. 3c) and showed the same intramolecular interaction over
249 the trimer surface, with the three-residue insertion (including the well-conserved Phe)
250 occupying an equivalent position in both structures (Fig. 3a and 3c). The active site
251 environment, including bound CoA, was also superimposable to *CgE2p_CD* (Fig. S11).

252

253 Likewise, we analyzed *MtE2b*, the predicted E2 component of the BCKDH (Branched Chain
254 KetoAcid Dehydrogenase) complex in *M. tuberculosis* (20). The rationale for the choice of
255 this protein came from two observations. First, the analysis of its primary sequence did not
256 show any insertion at the C-terminus, in contrast with *CgE2p*, *CmE2p* but also *MtE2p* from
257 the same species (Fig. 3a), suggesting that it may assume a higher oligomeric state. However,
258 an unpublished model of its catalytic domain issued from a structural genomics initiative (pdb
259 3L60) showed a trimeric arrangement, apparently challenging our hypothesis. We therefore
260 designed a similar construct to overexpress *MtE2b_CD* (residues 165-393) in *E. coli* with an
261 N-terminal, cleavable His₆ tag and purified the recombinant protein. Crystallization
262 experiments carried out after cleavage of the N-terminal His₆-tag yielded bipyramidal shaped

263 crystals belonging to the cubic space group F432 and diffracting X-rays up to 1.5 Å resolution
264 (Table S2). The crystal structure, solved by molecular replacement using the coordinates
265 available from the pdb, showed, unlike the 3L60 model, the same 24-mer cubic arrangement
266 first observed for AvE2p_CD or bovine E2b (3) (Fig. S12). The terminal 3₁₀ helix assumed
267 the same canonical conformation, being involved in the ‘knobs and sockets’ trimer-trimer
268 interaction. To verify the oligomerization state in solution of this construct as well as the
269 entire *MtE2b* protein, we carried out solution studies on both *MtE2b_FL* and *MtE2b_CD*.
270 Both AUC data and *ab-initio* models obtained from SAXS data were consistent with a large
271 assembly for both constructs (Table 1), indicative of the presence of the canonical 24-mer
272 cubic assembly (Fig. S1 and S13). We also found *MtE2b_FL* able to catalyze transacylase
273 activity using isobutyryl-CoA as the donor substrate, with specific activity values (2.5 ± 0.2
274 nmol min⁻¹mg⁻¹) comparable to those measured for *MtE2p_FL*, which, in contrast, is trimeric
275 (2.1 ± 0.1 nmol min⁻¹mg⁻¹ with acetyl-CoA as the donor substrate).

276

277 We then took advantage of the oligomerization state of *MtE2b* to analyze recombinant
278 proteins from both constructs by single-particle cryo-EM. Isolated particles indeed showed a
279 cubic shape structure for both *MtE2b_FL* and *MtE2b_CD* proteins, with a particle side of
280 approximately 13 nm (Fig. 4a and Fig. S14), consistently with the crystallographic data.
281 Several rounds of 2D-classification were performed to select a subset of good particles (Fig.
282 S15), where the class averages reflected the expected cubic shape and high-resolution features
283 such as defined α -helices (Fig. 4b), whereas particles showing apparent partial dissociation of
284 the cubic complex were discarded (Fig. S15). Several *ab-initio* models were calculated and,
285 after 3D classification, homogeneous refinement, local CTF refinement and density
286 modification, an EM map at 3.82 Å resolution was obtained, showing the expected cubic
287 assembly (Fig. 4c and Fig. S15). Real space refinement was then performed to fit the 24-mer,
288 cubic crystallographic assembly into the EM map, showing a very good fitting overall (Fig.
289 4d) including the TTI helix and its canonical orientation including the ‘knobs and sockets’

290 interactions (Fig. 4e). The ensemble of our structural and biophysical data therefore supports
291 a strong link between the three-residue C-terminal insertion (PCI), the orientation of the
292 terminal 3_{10} helix (TTI), and the E2 oligomeric state.

293

294 **The presence of the PCI in actinobacterial E2p is related to the presence of OdhA**

295 Unlike other canonical E1 proteins, *C. glutamicum* OdhA has a unique domain architecture
296 characterized by the fusion of an E2 domain to E1o (12, 14), also found in mycobacterial
297 KGD (13) (Fig. 5a). Taking advantage of this unique E2-E1 fusion feature, we retrieved the
298 sequences of OdhA homologs from an extended list of available genomes (Table S3), seeking
299 for the same association of Pfam domains that identify the E2 and E1o domains, respectively
300 (Fig. 5a). This analysis revealed that, except for a few isolated cases, OdhA homologs are a
301 unique feature of Actinobacteria (Fig. 5b). Inside this Phylum, OdhA homologs are widely
302 present in the Actinomycetales class, except for members of the order Bifidobacteriales, and
303 the Acidimicrobiia class, but not in the classes Coriobacteria or Rubrobacteria (Fig. 5c). The
304 three known members from the class Thermoleophilia show a unique architecture, in which
305 the first domain is replaced by two lipoyl domains and a E3-binding domain (Fig. 5a), thus
306 bearing a full fusion of E2 and E1o enzymes. Unlike OdhA, such an enzyme, which is so far
307 uncharacterized, could perform both the oxidative decarboxylation and the transacylation
308 steps without the need for lipoyl domains provided *in trans*. Although speculative at this
309 stage, this hypothesis is coherent with the predicted absence, in this sequence, of the N-
310 terminal α -helical segment which is found in OdhA and in other E1o enzymes, and predicted
311 to be required for the interaction with E2 (Fig. 5a).

312

313 Noteworthy, the correlation between the PCI in E2 and the presence of an OdhA homolog
314 (devoid of a lipoyl-binding domain) could be observed for all the representative genomes
315 (Fig. 5c). Both features are present in classes Actinobacteria and Acidimicrobiia and absent in
316 the classes Thermoleophilia, Rubrobacteria or Coriobacteria. Since OdhA requires E2p to
317 perform its activity in the context of the PDH/ODH supercomplex, it is tempting to speculate

318 that loss of E2 higher oligomerization could be a feature evolved to optimize the assembly of
319 such a mixed complex, where an E2o-E1o fusion enzyme has to compete with E1p for the
320 same lipoyl groups provided by a shared E2p. Along the same line, this might explain why
321 this ‘minimal size’ E2p was only detected in the Actinobacteria phylum.

322

323 DISCUSSION

324 The ensemble of enzymatic reactions that make up central metabolism and concur to the
325 production of reducing equivalents in heterotrophic organisms is commonly considered to be
326 extremely conserved. The same has longtime been believed for the respective enzymes and
327 their assemblies, for which most of our current knowledge comes from studies made decades
328 ago on model organisms, which, for prokaryotes, were in most cases represented by *E. coli* or
329 *B. subtilis*. 2-oxoacid dehydrogenase complexes, which are key players in central
330 metabolism, were also object of intensive study during this ‘golden age’ of enzymology, with
331 most of our knowledge coming from seminal work carried out, over several decades, by the
332 groups led by renowned biochemists like Richard N. Perham (1), Lester Reed (21), and John
333 Guest (22). Their extensive biochemical work was then supported by structural evidence that
334 started being published during the ‘90s, initiated by the work performed in Wim Hol’s lab on
335 the diazotroph, Gram-negative model *A. vinelandii*, and the first determination of a crystal
336 structure of AvE2p_CD (6), which in turn confirmed observations dating back to the ‘60s on
337 *E. coli* E2p (23) and E2o (24). The following structural studies have mostly focused on the
338 PDH complexes, especially the ones from model species like *E. coli*, *B. stearothermophilus* or
339 their counterparts from human or yeast (1, 25, 26), while ODH and BCKDH complexes,
340 despite a number of biochemical studies (27), were largely assumed to share PDH-like
341 molecular structures. In all cases, 2-oxoacid dehydrogenase complexes have all been
342 considered to conform to the ‘rule’ of a hollow, large E2 core showing either cubic (432) or
343 icosahedral (532) symmetry, assembled according to an ensemble of equivalent, and quasi-
344 equivalent interactions involving the C-terminal end of each E2 monomer (5). The more
345 unusual 42-mer E2 assembly described in *T. acidophilum* (8), albeit deviating from the

346 standard symmetry, still possess a hollow core whose assembly is based on the same kind of
347 interactions between C-terminal helices, and thus does not represent a real exception to the
348 rule. Although active site coupling has generally been believed to be the main benefit arising
349 from the presence of large multimeric assemblies (1), questions were already raised in the
350 past, when the same team that described the *T. acidophilum* E2 structure showed that
351 precluding the formation of higher E2 oligomers did not lead to loss of catalytic activity (9),
352 in contrast with previous reports. Consistently with our results, recent work showed that PDH
353 and ODH activities in *C. glutamicum* are associated to molecular assemblies that are smaller
354 than in *E. coli*, proposing an hexameric E2p core for the PDH/ODH supercomplex from size
355 exclusion chromatography experiments (28). Our work here thus shows how
356 *Corynebacteriales*, and most *Actinomycetales*, break the ‘dogma’ regarding the three-
357 dimensional architecture of 2-oxoacid dehydrogenase complexes, providing a simple
358 structural explanation to the lack of larger E2 molecular assemblies, and at the same time
359 raises new questions about the origin and evolution of these enzymatic machineries.

360

361 Actinobacteria represent one of the largest and ubiquitous prokaryotic phyla, with more than
362 50 families. Its members include the *Streptomyces* genus, one of the largest bacterial genera
363 and a major source of antibiotics, as well as relevant human pathogens such as *M.*
364 *tuberculosis* and cell factories as *C. glutamicum*. Despite the wide applications of *C.*
365 *glutamicum* in the biotech industry, which range from the production of amino acids to that of
366 biofuels (29), relatively little is known about how *Corynebacteriales* regulate their central
367 metabolism. The evidence reported here supports a strong correlation between the presence of
368 E2o and E1o activities on a single OdhA-like polypeptide and the loss of highly oligomeric
369 structure of the PDH core, two properties that are most likely related to the dependence of
370 both the PDH and ODH activities on the lipoyl domains provided by a single E2p protein,
371 and, therefore, their coexistence in a mixed multienzyme machinery. The presence of such
372 kind of supercomplex, first hypothesized in 2006 for *C. glutamicum* (11), is in turn suggestive
373 of the presence of regulation mechanisms that might well be unique to Actinobacteria. In this

respect, the discovery that the FHA protein OdhI is a phospho-dependent regulator of OdhA in *C. glutamicum* (11, 30) paved the way to show that ODH activity in *Corynebacteriales* – including *M. tuberculosis* and *M. smegmatis* – is regulated by a signal transduction pathway triggered by the Ser/Thr kinase PknG in response to nutrient availability (31-34). So far, the presence of an FHA protein as a metabolic regulator, which in mycobacteria also acts on glutamate dehydrogenase and glutamate synthase (34, 35), has only been described in *Corynebacteriales* and might be related to the unusual architecture of the PDH/ODH supercomplex. An additional ODH allosteric regulator is acetyl-CoA, identified as an allosteric activator of *M. smegmatis* KGD (13) but seemingly common in *Corynebacteriales*, as it was earlier reported as an activator of corynebacterial ODH (36). As acetyl-CoA is the product of the PDH reaction, positive feedback mechanisms by product channeling from PDH to ODH in a common complex are possible. On the other hand, little is known about the regulation of the PDH reaction in Actinobacteria, apart from the known involvement of the RamB transcriptional regulator in the expression of the *aceE* (E1p) gene in *C. glutamicum* (37).

389

Recent developments in metabolomics have started to unravel the mechanisms of metabolic regulation in *M. tuberculosis* (38, 39), a particularly important facet as metabolic plasticity is a key capability of the pathogen to switch from active replication to dormant states. The strategy of targeting central metabolism for new drugs has been suggested by several groups, mainly due to the presence of unique *M. tuberculosis* pathways that are crucial for pathogen survival and latent state persistence within the host (38, 40, 41). The peculiar architecture of the PDH/ODH supercomplex might therefore be exploited for drug design purposes, especially considering that chemical inhibition of DlaT (*MtE2p*) has been shown to kill non-replicating mycobacteria (40). On the other hand, considering that both pyruvate and 2-oxoglutarate are essential metabolites and precursors of a number of compounds of biotechnological interest, many of which industrially produced via *C. glutamicum* (42), our findings also pave the way to metabolic engineering applications.

402

403 **MATERIALS AND METHODS**

404

405 ***Plasmid construction***

406 *C. glutamicum* open reading frames were amplified from genomic DNA by PCR and cloned
407 in pET-28a vector by restriction free cloning (43), while expression constructs for the other
408 proteins (*CmE2p*, *MtE2p*, *MtE2b*) were provided by Genscript (Leiden, the Netherlands). In
409 all cases, a sequence coding for a His₆-tag followed by the TEV protease cleavage site
410 (ENLYFQG) were fused to the N-terminus of the protein of interest. Constructs were made to
411 express: for *CgE2p* (Uniprot accession no. Q8NNJ2), the full-length protein (*CgE2p_FL*,
412 residues 1-675), the catalytic domain (*CgE2p_CD*, corresponding to residues 437-675), and
413 the catalytic core and the peripheral subunit binding domain (PSBD; *CgE2p-ΔLBDs*,
414 corresponding to residues 372 to 675). For *CmE2p* (Uniprot accession no. A0A0G3H170), the
415 catalytic core was expressed (*CmE2p_CD*, residues 428-666), while for *MtE2p* (Uniprot
416 accession no. P9WIS7), the full-length protein was expressed (*MtE2p_FL*, residues 1-553).
417 For *MtE2b* (Uniprot accession no. O06159), two constructs were built: the first one
418 corresponding to the full-length protein (*MtE2b_FL*, residues 1-393), and the second one
419 corresponding to the catalytic core region (*MtE2b_CD*, defined as residue 165-393). All
420 plasmids were verified by DNA sequencing.

421

422 ***Protein purification***

423 Each plasmid was transformed into *E. coli* BL21DE3 using standard protocols. In the case of
424 *CgE2p_FL* and *CgE2p_CD*, transformed cells were growth in LB medium at 37 °C until
425 optical density measure at 600 nm reach ~0.6. Afterwards, temperature was dropped to either
426 18 °C or 30 °C (Table S4) and incubated for 18 hours after adding 0.5 mM isopropyl-β-D-
427 thiogalactopyranoside (IPTG). For *MtE2p_FL*, *MtE2b_FL* or *MtE2b_CD*, or constructs
428 expressing mutants generated on *CgE2p_CD*, autoinduction medium adapted from Studier

429 (44) was used, with slight modifications. In this case, bacterial cells were grown at 37 °C for 4
430 hours, following by an overnight incubation at either to either 30 °C or 18°C (Table S4). Cells
431 were recovered by centrifugation and washed twice with the corresponding lysis buffer (Table
432 S4). Cell lysis was performed using a CF2 cell disruptor (Constant Systems Ltd., Daventry,
433 UK). The soluble fraction was separated from debris by centrifugation at 26,800 g for 1 hour.
434 The recombinant proteins were purified by Ni²⁺-IMAC on HisTrap columns (GE Healthcare).
435 Fractions, as confirmed by SDS-PAGE, containing the protein of interest were dialyzed
436 against GF buffer (see buffer table), supplemented with 1 mM dithiothreitol (DTT), for
437 12 hours at 18°C, after adding His-tagged TEV protease (45) at a 1:30 ratio (w/w) to remove
438 the N-terminal His₆ tag from the recombinant protein. The cleaved, recombinant proteins
439 were then separated from the His.tagged TEV protease (and the cut N-terminal portion) by
440 gravity-flow separation on Ni-NTA agarose (Qiagen). Finally, the recovered recombinant
441 proteins were further purified by size exclusion chromatography using a Sephadryl S-400
442 16/60 column (GE Healthcare). Fractions containing the protein were pulled, concentrated
443 and flash frozen in liquid nitrogen.

444

445 *Transacylase assay*

446 Transacylase activity of *CgE2p*, *MtE2p* and *MtE2b* was determined by the spectrophotometric
447 method first described by Hall and Weitzman (15), with slight modifications, following
448 transacylation in the reverse direction from an acyl-CoA donor to free dihydrolipoamide,
449 generated by the dihydrolipoamide dehydrogenase reaction (Fig. S2). The assay mixture
450 contained 0.1 M Tris-HCl pH 7.6, 10 mM MgCl₂, 1 mM EDTA, 0.4 mM NADH, 0.2 mM
451 lipoamide, 0.25 U of lipoamide dehydrogenase from bovine intestinal mucosa (EC 1.8.1.4,
452 Sigma-Aldrich L-6777) and a E2 protein: 22.5 µg of *CgE2p_FL* (1.6 µM), 300-600 µg of
453 *MtE2p_FL* (26.2 – 52.4 µM) or 120-240 µg of *MtE2b_FL* (14.6 – 29.2 µM). For catalytic
454 domain constructs, quantities added were 110 µg for *CgE2p_CD* (21.1 µM) or 69 µg for the
455 M3 mutant (12.1 µM). The approach to equilibrium in lipoamide dehydrogenase reaction and

456 the background rate of NADH oxidation was followed for 30 minutes, then transacylation
457 was started by adding 0.15 mM acetyl-CoA (Fig. S2) or 1 mM isobutyryl-CoA to the assay
458 mixture for both E2p proteins or *MtE2b_FL*, respectively, to a final reaction volume of 200
459 µl. Subtraction of endogenous rate of NADH oxidation from the final rate after acyl-CoA
460 addition gave a measure for the dihydrolipoamide transacylase activity. The linear region of
461 the progress curves, generally limited to 1 min (Fig. S2) was used to calculate the steady state
462 velocities. Extinction coefficient of NADH used for calculations is 6.22 mM⁻¹cm⁻¹.
463 Lipoamide dehydrogenase, as well as acyl-CoA substrates were not limiting components, as
464 increasing of their concentrations did not affect the final transacylase activity. The
465 measurements were performed at least as triplicates, and resulting activities presented in nmol
466 min⁻¹mg⁻¹ of protein.

467

468 ***SEC-SAXS and ab initio modelling***

469 SAXS experiments were performed either at the SOLEIL synchrotron (Saint-Aubin, France)
470 on the SWING beamline, or at the ESRF (Grenoble, France) on BM29 (*CgE2p_CD* sample).
471 Experimental settings are provided in Table S1. The protein samples were injected, through
472 an Agilent HPLC system, into a size exclusion chromatography column (SuperoseTM 6
473 Increase 5/150 GL, GE Healthcare Bio-Sciences) previously equilibrated with the
474 corresponding buffer (Table S4) and eluted directly into the SAXS flow-through capillary cell
475 at a flow rate of 0.2 ml/min. Frames were collected continuously during the whole protein
476 elution time.

477

478 Frames were analyzed using either the FOXTROT software developed at Synchrotron Soleil
479 (for data collected on Swing), or the RAW software (46). Frames corresponding to the buffer
480 were averaged and subtracted from sample frames. The forward scattering I(0) and R_g for
481 each sample subtracted frame were derived from Guinier approximation; the I(0) vs frame
482 curve was used to select the region corresponding to the protein's elution peak (Fig. S3).
483 Frames corresponding to the elution peak and with constant R_g were averaged. All subsequent

484 data processing was performed using software part of the ATSAS suite (16), starting from
485 PRIMUS which was used to recalculate I(0) and R_g, while GNOM was used to calculate the
486 pair-distance distribution function P(r) and Dmax (17). The POROD volume and other tools
487 available in ATSAS were used to estimate the molecular weights of the proteins (47). The
488 dimensionless Kratky plot was calculated according to previous work (48, 49), using the
489 value determined by Guinier analysis.

490 *Ab initio* modelling was also performed using software from the ATSAS suite. For
491 *MtE2b_CD*, 10 independent models were generated using DAMMIF, with no imposed
492 symmetry. In both cases best models were obtained assuming oblate anisometry. The models
493 were then compared and align with DAMSEL and DAMSUP and output files were generated
494 using DAMAVER, DAMFILT and DAMSTART, all within ATSAS. Finally, the
495 DAMSTART output was provided to DAMMIN to obtain a dummy atom model fitting the
496 data.

497

498 ***Analytical Ultracentrifugation (AUC)***

499 Sedimentation velocity (SV) analytical ultracentrifugation assays were performed using a
500 Beckman Coulter ProteomeLab XL-I analytical ultracentrifuge equipped with UV-Vis
501 absorbance and Raleigh interference detection systems, using the 8-hole Beckman An-50Ti
502 rotor at 20°C. Experiments were performed at 10,000 rpm for *MtE2b_CD* and *MtE2b_FL* and
503 36,000 rpm for *MtE2p_FL*, *CgE2p_FL* and *CgE2p_CD*. Three concentrations of each protein
504 were prepared for this experiment in their corresponding buffer and loaded into analytical
505 ultracentrifugation cells. During the run SV was followed using by measuring absorbance at
506 290 nm. SEDFIT 15.01 (available at <http://analyticalultracentrifugation.com>; (50)) was used
507 to calculate the sedimentation coefficient distribution C(s), then corrected to standard
508 conditions to get the final standard values. These coefficients were plotted as a function of the
509 different concentrations and an extrapolation to zero concentration was made to obtain the
510 standard value for the main oligomer. From these values, molecular mass and friction ratio
511 were obtained.

512

513 ***Crystallization***

514 The crystallization experiments were performed at either 18°C or 4 °C by the sitting drop
515 vapor diffusion technique in 96-well plates, according to established protocols at the
516 Crystallography Core Facility of the Institut Pasteur (51). Crystallization conditions were: for
517 *CgE2p_CD* in apo form, 0.1 M Hepes-Na pH 7.5, 5% PEG4000, 30% 2-methyl-2,4-pentane
518 diol (MPD), at 4°C; for *CgE2p_CD* in complex with CoA, 0.1 M Hepes-Na pH 7.5, 1.56 M
519 tri-sodium citrate, at 4°C; for *CgE2p_CD* in ternary complex with CoA and
520 dihydrolipoamide, 0.1 M Tris-HCl pH 7.3, 1.62 M tri-sodium citrate, at 18°C; for
521 *CgE2p_ALBDs*, 30% PEG1500 at 18°C; for *CmE2p_CD* in complex with CoA, 0.1 M Hepes-
522 Na pH 7.5, 150 mM NaCl, 30% PEG4000, at 18°C; for *MtE2b_CD*, 0.1M Imidazole pH 6.5,
523 4 M NH₄⁺ acetate at 18°C.

524

525 ***Data collection and structure solution***

526 Diffraction data were acquired from crystals maintained at 100 K, either at the SOLEIL
527 Synchrotron on the beamline Proxima-1 or Proxima-2A, or on the automated beamline
528 MASSIF-1 at the ESRF (52). The data were processed with XDS (53), run through either
529 XDSME (available at <https://github.com/legrandp/xdsme>) or autoPROC (54), and scaled with
530 Aimless, available from the CCP4 suite (55), or STARANISO as provided within autoPROC.
531 The structures were solved by molecular replacement through the program PHASER (56).
532 Coordinates of the *E. coli* E2p catalytic domain (PDB entry 4N72; (57)) were used as the
533 search model to first solve the structure of *CgE2p_CD* in apo form, which, in turn, served as
534 molecular replacement search model for the following datasets, including *CmE2p_CD*, but
535 not *MtE2b_CD* whose structure was solved by using the previously released coordinates of
536 the same protein (pdb entry 3L60). All rebuilding and adjustments of the models were
537 performed with COOT (58). The refinement was carried out with BUSTER, applying local
538 structure similarity restraints for non-crystallography symmetry (NCS) (59) and a
539 Translation-Libration-Screw (TLS) model. ‘Polder’ maps for the identification of ligand

540 density were calculated with the specific tool within the PHENIX suite (60), while chemical
541 dictionaries for ligands were generated with the Grade server (<http://grade.globalphasing.org>).
542 Validation of models was performed with MOLPROBITY (61) and the validation tools in
543 PHENIX (62). Data collection, refinement, model statistics and pdb accession codes for
544 coordinates and structure factors are indicated in Table S2. Graphical representations were
545 rendered with Pymol (63).

546

547 ***CryoEM***

548

549 For grid preparation 3.5 µl of 0,2 mg/ml *MtE2b_CD* were applied into a Lacey Carbon Film
550 on 200 Mesh Copper Grids. Grid vitrification was performed in a Vitrobot apparatus (Thermo
551 Fisher Scientific) with 95% humidity, ashless filter paper (Standard Vitrobot Filter Paper,
552 Ø55/20mm, Grade 595; Electron Microscopy Sciences) and using a blot time of 3.5 s, while
553 blotting force was set to zero. Grids were stored in liquid nitrogen and collected in a 200 keV
554 Thermo Scientific Glacios® Cryo-Transmission Electron Microscope. Images were recorded
555 using a Falcon III direct electron detector camera in linear mode. The images were collected
556 using EPU software 2.2.0.65REL (Thermo Fisher Scientific), with a defocus range from -0.8
557 µm to -2.6 µm. The pixel size was set to 0.96 Å. 37 frames per movie were collected, with an
558 overall dose of 42.5 e-/Å².

559

560 Movies frames were aligned using Motiocor2 (64, 65), and CTF estimation was performed
561 using Gctf (66). Movies were imported and analyzed in Cryosparc v2.15 (67). Using the
562 curate exposure feature, 1235 out of 1476 movies, followed by 2483 out of 3533 movies from
563 a second data collection were selected for further analysis. A first round of manual picking
564 was performed and after two rounds of 2D classification, the three most populated classes
565 were selected for template-based particle picking against a dataset containing the selected
566 micrographs. The selected particles were cleaned using the ‘inspect pick’ function of
567 Cryosparc and three rounds of 2D classification and a final dataset of 40805 and 168705 good

568 particles, corresponding to collection session 1 and 2, respectively, was selected. A round of
569 *ab initio* and 3D classification into 4 classes was then performed to select a subset of 111891
570 particles that could be refined to 5.8 Å. Following local CTF refinement and imposing
571 octahedral O3 symmetry, the dataset could be refined to 3.92 Å. Finally, density modification
572 was applied using the tool resolve_cryo_em from the PHENIX suite (68), reaching a final
573 map at 3.82 Å resolution. Model fitting into cryo-EM maps was performed using the
574 programs UCSF Chimera (69) and phenix.real_space_refine (70). The final, sharpened map
575 from Cryosparc as well as the density modified map from resolve_cryo_em were deposited to
576 the EMDB under the accession code 11600. Figures were generated and rendered with UCSF
577 Chimera.

578

579 **Bioinformatic analysis**

580 Multiple sequence alignments were performed using MUSCLE
581 (<https://www.ebi.ac.uk/Tools/msa/muscle/>) (71). Further alignment details were carried out in
582 AliView version 1.26 (72). Sequences from *MtE2b*, *MtE2p*, *CgE2p*, *CmE2p* and E2 proteins
583 with known structure were compared to the representative sequences to account for the PCI
584 presence; the software ESPript 3.0 (<http://escript.ibcp.fr>) (73) was then used to generate
585 alignment pictures providing secondary structure elements from a reference structure. The
586 presence of the PCI was analyzed in the region corresponding to the C-terminal α-helix
587 (residues 630 to 637 in AvE2p).

588 To identify *odhA*-like genes, the gene search tool from IMG/M was used (Integrated
589 Microbial Genomes & Microbiomes system v.5.0) (74); to simplify the analysis, only
590 genomes annotated as finished were selected. For each genome, *odhA*-like genes were
591 identified by their conserved Pfam domain associations, including: PF16078 ('2-
592 oxogl_dehyd_N'), PF00198 ('2-oxoacid_dH'), PF00676 ('E1-dh'), PF02279 ('Transket_pyr')
593 and PF16870 ('OxoGdeHyase_C'). All the finished genomes corresponding to the phylum
594 Actinobacteria and two non-finished genomes corresponding to the class Thermoleophilia
595 were selected from IMG/M available genomes (only one finished genome was available for

596 Thermoleophilia). After elimination of duplicated genomes, *odhA*-like genes and E2-E1 like
597 genes were identified using their corresponding Pfam architecture. For the selected genomes,
598 a tree was created using the JGI distance tree tool, using the alignment of 16S genes based on
599 the SILVA database and the ‘dnadist’ and ‘neighbor’ tools from the PHYLIP package
600 (available at <https://evolution.genetics.washington.edu/phylip/>). Figures were built using the
601 Interactive Tree Of Life (iTOL) v4 online tool (75).

602 To analyze whether any correlation could be detected between *odhA*-like genes and the
603 presence of the PCI in E2p genes, a representative pool of genomes from the larger orders
604 inside Actinobacteria class (namely Corynebacteriales, Streptomycetales, Propionibacteriales,
605 Micromonosporales, Micrococcales and Bifidobacteriales), as well as representatives from the
606 other classes inside the Phylum actinobacteria were selected. For each genome, predicted E2
607 proteins were identified using the signature pfam domains (see above), and aligned.

608

609 ACKNOWLEDGMENTS

610 This work was funded by the ANR projects SUPERCPLX (ANR-13-JSV8-0003) and
611 METACTINO (ANR-18-CE92-0003), both granted to M.B., and by institutional grants from
612 the Institut Pasteur and the CNRS. We are grateful to the core facilities at Institut Pasteur
613 C2RT (Centre for Technological Resources and Research), in particular to A. Haouz, P.
614 Weber and C. Pissis (Crystallography), S. Brûlé and B. Baron (Molecular Biophysics), M.
615 Matondo, T. Chaze and T. Douché (Proteomics), J.-M. Winter, S. Tachon and M. Vos
616 (NanoImaging). We also gratefully acknowledge F. Gubellini (Structural Microbiology Unit)
617 for her help in EM sample preparation, J.J. Pierella Karlusich (ENS, Paris), A. Thureau
618 (Synchrotron Soleil) and V. Bunik (Lomonosov University, Moscow) for many helpful
619 discussions. The NanoImaging Core at Institut Pasteur was created with the help of a grant
620 from the French Government’s ‘Investissements d’Avenir’ program (EQUIPEX CACSICE –
621 “Centre d’analyse de systèmes complexes dans les environnements complexes”, ANR-11-
622 EQPX-0008), and is acknowledged for support with cryo-EM sample preparation, image
623 acquisition and analysis. We also acknowledge the synchrotron sources Soleil (Saint-Aubin,

624 France), and ESRF (Grenoble, France) for granting access to their facilities, and their staff for
625 helpful assistance on the respective beamlines. L.Y. and A.B. are both part of the Pasteur –
626 Paris University (PPU) International PhD program; L.Y. was funded by the Wuhan Institute
627 of Biological Products Co. Ltd. (Wuhan, People's Republic of China), subsidiary company of
628 China National Biotec Group Company Limited, and by a doctoral fellowship from the China
629 Scholarship Council (CSC).

630

631 REFERENCES

- 632 1. R. N. Perham, Swinging arms and swinging domains in multifunctional enzymes:
633 catalytic machines for multistep reactions. *Annu. Rev. Biochem.* **69**, 961–1004 (2000).
- 634 2. J. L. S. Milne, *et al.*, Molecular structure of a 9-MDa icosahedral pyruvate
635 dehydrogenase subcomplex containing the E2 and E3 enzymes using cryoelectron
636 microscopy. *J. Biol. Chem.* **281**, 4364–4370 (2006).
- 637 3. M. Kato, *et al.*, A synchronized substrate-gating mechanism revealed by cubic-core
638 structure of the bovine branched-chain alpha-ketoacid dehydrogenase complex. *EMBO J.* **25**,
639 5983–5994 (2006).
- 640 4. J. L. S. Milne, *et al.*, Molecular architecture and mechanism of an icosahedral
641 pyruvate dehydrogenase complex: a multifunctional catalytic machine. *EMBO J.* **21**,
642 5598 (2002).
- 643 5. T. Izard, *et al.*, Principles of quasi-equivalence and Euclidean geometry govern the
644 assembly of cubic and dodecahedral cores of pyruvate dehydrogenase complexes.
645 *Proc. Natl. Acad. Sci. U. S. A.* **96**, 1240–1245 (1999).
- 646 6. A. Mattevi, *et al.*, Atomic structure of the cubic core of the pyruvate dehydrogenase
647 multienzyme complex. *Science* **255**, 1544–1550 (1992).
- 648 7. A. Mattevi, G. Obmolova, K. H. Kalk, A. Teplyakov, W. G. J. Hol, Crystallographic
649 Analysis of Substrate Binding and Catalysis in Dihydrolipoyl Transacetylase (E2p).
650 *Biochemistry* **32**, 3887–3901 (1993).
- 651 8. N. L. Marrott, *et al.*, The catalytic core of an archaeal 2-oxoacid dehydrogenase
652 multienzyme complex is a 42-mer protein assembly. *FEBS J.* **279**, 713–723 (2012).
- 653 9. N. L. Marrott, *et al.*, Why are the 2-oxoacid dehydrogenase complexes so large?
654 Generation of an active trimeric complex. *Biochem. J.* **463**, 405–412 (2014).
- 655 10. O. Byron, J. G. Lindsay, “The Pyruvate Dehydrogenase Complex and Related
656 Assemblies in Health and Disease” in *Macromolecular Protein Complexes*,
657 Subcellular Biochemistry. (Springer International Publishing, 2017), pp. 523–550.
- 658 11. A. Niebisch, A. Kabus, C. Schultz, B. Weil, M. Bott, Corynebacterial protein kinase G
659 controls 2-oxoglutarate dehydrogenase activity via the phosphorylation status of the
660 OdhI protein. *J. Biol. Chem.* **281**, 12300–12307 (2006).

- 661 12. Y. Usuda, *et al.*, Molecular cloning of the *Corynebacterium glutamicum*
662 (“*Brevibacterium lactofermentum*” AJ12036) *odhA* gene encoding a novel type of 2-
663 oxoglutarate dehydrogenase. *Microbiology* **142**, 3347–3354 (1996).
- 664 13. T. Wagner, M. Bellinzoni, A. Wehenkel, H. M. O'Hare, P. M. Alzari, Functional
665 plasticity and allosteric regulation of α -ketoglutarate decarboxylase in central
666 mycobacterial metabolism. *Chem. Biol.* **18**, 1011–1020 (2011).
- 667 14. M. Hoffelder, K. Raasch, J. van Ooyen, L. Eggeling, The E2 domain of OdhA of
668 *Corynebacterium glutamicum* has succinyltransferase activity dependent on lipoyl
669 residues of the acetyltransferase AceF. *J. Bacteriol.* **192**, 5203–5211 (2010).
- 670 15. E. R. Hall, P. D. Weitzman, A continuous spectrophotometric assay for the
671 transacylase (E2) component of pyruvate and alpha-oxoglutarate dehydrogenase
672 enzyme complexes. *Anal. Biochem.* **62**, 286–290 (1974).
- 673 16. D. Franke, *et al.*, ATSAS 2.8: a comprehensive data analysis suite for small-angle
674 scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).
- 675 17. P. V. Konarev, V. V. Volkov, A. V. Sokolova, M. H. J. Koch, D. I. Svergun,
676 PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl*
677 *Crystallogr* **36**, 1277–1282 (2003).
- 678 18. J. E. Knapp, *et al.*, Crystal structure of the truncated cubic core component of the
679 Escherichia coli 2-oxoglutarate dehydrogenase multienzyme complex. *J. Mol. Biol.*
680 **280**, 668 (1998).
- 681 19. C. Rückert, J. Eimer, A. Winkler, A. Tauch, Complete Genome Sequence of the Type
682 Strain *Corynebacterium mustelae* DSM 45274, Isolated from Various Tissues of a
683 Male Ferret with Lethal Sepsis. *Genome Announc.* **3**, 871 (2015).
- 684 20. A. Venugopal, *et al.*, Virulence of *Mycobacterium tuberculosis* depends on lipoamide
685 dehydrogenase, a member of three multienzyme complexes. *Cell Host Microbe* **9**, 21–
686 31 (2011).
- 687 21. L. J. Reed, From lipoic acid to multi-enzyme complexes. *Protein Sci.* **7**, 220–224
688 (1998).
- 689 22. J. R. Guest, S. J. Angier, G. C. Russell, Structure, expression, and protein engineering
690 of the pyruvate dehydrogenase complex of *Escherichia coli*. *Ann Ny Acad Sci* **573**,
691 76–99 (1989).
- 692 23. L. J. Reed, H. Fernandez-Moran, M. Koike, C. R. Willms, Electron microscopic and
693 biochemical studies of pyruvate dehydrogenase complex of *Escherichia coli*. *Science*
694 **145**, 930–932 (1964).
- 695 24. D. J. DeRosier, R. M. Oliver, L. J. Reed, Crystallization and preliminary structural
696 analysis of dihydrolipoyl transsuccinylase, the core of the 2-oxoglutarate
697 dehydrogenase complex. *Proc. Natl. Acad. Sci. U. S. A.* **68**, 1135–1137 (1971).
- 698 25. M. S. Patel, N. S. Nemeria, W. Furey, F. Jordan, The pyruvate dehydrogenase
699 complexes: structure-based function and regulation. *J. Biol. Chem.* **289**, 16615–16623
700 (2014).

- 701 26. Z. H. Zhou, D. B. McCarthy, C. M. O'Connor, L. J. Reed, J. K. Stoops, The
702 remarkable structural and functional organization of the eukaryotic pyruvate
703 dehydrogenase complexes. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 14802–14807 (2001).
- 704 27. V. Bunik, A. Artiukhov, *Vitamin-Dependent Multienzyme Complexes of 2-Oxo Acid*
705 *Dehydrogenases: Structure, Function, Regulation and Medical Implications* (Nova
706 Science Publishers, 2017).
- 707 28. H. Kinugawa, *et al.*, In vitroreconstitution and characterization of pyruvate
708 dehydrogenase and 2-oxoglutarate dehydrogenase hybrid complex from
709 *Corynebacterium glutamicum*. *Microbiologyopen* **5**, 593–14 (2020).
- 710 29. G. R. Lewin, *et al.*, Evolution and Ecology of Actinobacteria and Their Bioenergy
711 Applications. *Annu. Rev. Microbiol.* **70**, 235–254 (2016).
- 712 30. C. Schultz, A. Niebisch, L. Gebel, M. Bott, Glutamate production by *Corynebacterium*
713 *glutamicum*: dependence on the oxoglutarate dehydrogenase inhibitor protein OdhI
714 and protein kinase PknG. *Appl. Microbiol. Biotechnol.* **76**, 691–700 (2007).
- 715 31. N. Bhattacharyya, *et al.*, An Aspartate-Specific Solute-Binding Protein Regulates
716 Protein Kinase G Activity To Control Glutamate Metabolism in Mycobacteria. *mBio*
717 **9**, 349 (2018).
- 718 32. B. Rieck, *et al.*, PknG senses amino acid availability to control metabolism and
719 virulence of *Mycobacterium tuberculosis*. *PLoS Pathog* **13**, e1006399 (2017).
- 720 33. M. Ventura, *et al.*, GarA is an essential regulator of metabolism in *Mycobacterium*
721 *tuberculosis*. *Mol. Microbiol.* **90**, 356–366 (2013).
- 722 34. H. M. O'Hare, *et al.*, Regulation of glutamate metabolism by protein kinases in
723 mycobacteria. *Mol. Microbiol.* **70**, 1408–1423 (2008).
- 724 35. T. J. Nott, *et al.*, An intramolecular switch regulates phosphoindependent FHA domain
725 interactions in *Mycobacterium tuberculosis*. *Science Signaling* **2**, ra12–ra12 (2009).
- 726 36. I. Shiio, K. Ujigawatakeda, Presence and Regulation of Alpha-Ketoglutarate
727 Dehydrogenase Complex in a Glutamate-Producing Bacterium, *Brevibacterium*
728 *flavum*. *Agr Biol Chem Tokyo* **44**, 1897–1904 (1980).
- 729 37. B. Blombach, A. Cramer, B. J. Eikmanns, M. Schreiner, RamB Is an Activator of the
730 Pyruvate Dehydrogenase Complex Subunit E1p Gene in *Corynebacterium*
731 *glutamicum*. *J. Mol. Microbiol. Biotechnol.* **16**, 236–239 (2009).
- 732 38. K. Y. Rhee, *et al.*, Central carbon metabolism in *Mycobacterium tuberculosis*: an
733 unexpected frontier. *Trends Microbiol.* **19**, 307–314 (2011).
- 734 39. L. P. S. de Carvalho, *et al.*, Metabolomics of *Mycobacterium tuberculosis* reveals
735 compartmentalized co-catabolism of carbon substrates. *Chem. Biol.* **17**, 1122–1131
736 (2010).
- 737 40. R. Bryk, *et al.*, Selective killing of nonreplicating mycobacteria. *Cell Host Microbe* **3**,
738 137–145 (2008).
- 739 41. A. Gouzy, *et al.*, *Mycobacterium tuberculosis* exploits asparagine to assimilate
740 nitrogen and resist acid stress during infection. *PLoS Pathog.* **10**, e1003928 (2014).

- 741 42. S. Wieschalka, B. Blombach, M. Bott, B. J. Eikmanns, Bio-based production of
742 organic acids with *Corynebacterium glutamicum*. *Microb. Biotechnol.* **6**, 87–102
743 (2013).
- 744 43. T. Unger, Y. Jacobovitch, A. Dantes, R. Bernheim, Y. Peleg, Applications of the
745 Restriction Free (RF) cloning procedure for molecular manipulations and protein
746 expression. *J. Struct. Biol.* **172**, 34–44 (2010).
- 747 44. F. W. Studier, Protein production by auto-induction in high density shaking cultures.
748 *Protein Expr. Purif.* **41**, 207–234 (2005).
- 749 45. S. van den Berg, P. A. Lofdahl, T. Hard, H. Berglund, Improved solubility of TEV
750 protease by directed evolution. *J. Biotechnol.* **121**, 291–298 (2006).
- 751 46. J. B. Hopkins, R. E. Gilliland, S. Skou, BioXTAS RAW: improvements to a free open-
752 source program for small-angle X-ray scattering data reduction and analysis. *J. Appl.*
753 *Cryst* **50**, 1–9 (2017).
- 754 47. N. R. Hajizadeh, D. Franke, C. M. Jeffries, D. I. Svergun, Consensus Bayesian
755 assessment of protein molecular mass from solution X-ray scattering data. *Scientific*
756 *Reports* **8**, 7204–13 (2018).
- 757 48. J. Pérez, P. Vachette, D. Russo, M. Desmadril, D. Durand, Heat-induced unfolding of
758 neocarzinostatin, a small all-beta protein investigated by small-angle X-ray scattering.
759 *J. Mol. Biol.* **308**, 721–743 (2001).
- 760 49. D. Durand, *et al.*, NADPH oxidase activator p67(phox) behaves in solution as a
761 multidomain protein with semi-flexible linkers. *J. Struct. Biol.* **169**, 45–53 (2010).
- 762 50. P. Schuck, Size-distribution analysis of macromolecules by sedimentation velocity
763 ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).
- 764 51. P. Weber, *et al.*, High-Throughput Crystallization Pipeline at the Crystallography Core
765 Facility of the Institut Pasteur. *Molecules* **24** (2019).
- 766 52. M. W. Bowler, *et al.*, MASSIF-1: a beamline dedicated to the fully automatic
767 characterization and data collection from crystals of biological macromolecules. *J.*
768 *Synchrotron Radiat.* **22**, 1540–1547 (2015).
- 769 53. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
- 770 54. C. Vonrhein, *et al.*, Data processing and analysis with the autoPROC toolbox. *Acta*
771 *Crystallogr. D Biol. Crystallogr.* **67**, 293–302 (2011).
- 772 55. M. D. Winn, *et al.*, Overview of the CCP4 suite and current developments. *Acta*
773 *Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
- 774 56. A. J. McCoy, *et al.*, Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–
775 674 (2007).
- 776 57. J. Wang, *et al.*, Structure and function of the catalytic domain of the dihydrolipoyl
777 acetyltransferase component in *Escherichia coli* pyruvate dehydrogenase complex. *J.*
778 *Biol. Chem.* **289**, 15215–15230 (2014).
- 779 58. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot.
780 *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

- 781 59. O. S. Smart, *et al.*, Exploiting structure similarity in refinement: automated NCS and
782 target-structure restraints in BUSTER. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 368–
783 380 (2012).
- 784 60. D. Liebschner, *et al.*, Polder maps: improving OMIT maps by excluding bulk solvent.
785 *Acta Crystallogr. D Struct. Biol.* **73**, 148–157 (2017).
- 786 61. C. J. Williams, *et al.*, MolProbity: More and better reference data for improved all-
787 atom structure validation. *Protein Science* **27**, 293–315 (2018).
- 788 62. D. Liebschner, *et al.*, Macromolecular structure determination using X-rays, neutrons
789 and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**,
790 861–877 (2019).
- 791 63. Schrödinger, LLC., The PyMOL Molecular Graphics System (2012).
- 792 64. S. Q. Zheng, *et al.*, MotionCor2: anisotropic correction of beam-induced motion for
793 improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
- 794 65. X. Li, *et al.*, Electron counting and beam-induced motion correction enable near-
795 atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013).
- 796 66. K. Zhang, Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–
797 12 (2016).
- 798 67. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: algorithms for
799 rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296
800 (2017).
- 801 68. E. Ramírez-Aportela, *et al.*, Automatic local resolution-based sharpening of cryo-EM
802 maps. *Bioinformatics* **36**, 765–772 (2020).
- 803 69. E. F. Pettersen, *et al.*, UCSF Chimera--a visualization system for exploratory research
804 and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- 805 70. P. V. Afonine, *et al.*, Real-space refinement in PHENIX for cryo-EM and
806 crystallography. *Acta Crystallogr. D Struct. Biol.* **74**, 531–544 (2018).
- 807 71. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high
808 throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
- 809 72. A. Larsson, AliView: a fast and lightweight alignment viewer and editor for large
810 datasets. *Bioinformatics* **30**, 3276–3278 (2014).
- 811 73. X. Robert, P. Gouet, Deciphering key features in protein structures with the new
812 ENDscript server. *Nucleic Acids Res.* **42**, W320–4 (2014).
- 813 74. I.-M. A. Chen, *et al.*, IMG/M v.5.0: an integrated data management and comparative
814 analysis system for microbial genomes and microbiomes. *Nucleic Acids Res.* **47**,
815 D666–D677 (2019).
- 816 75. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v4: recent updates and new
817 developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).

818

819 **AUTHOR CONTRIBUTIONS**

820 E.M.B., P.V., L.Y. and N.L.-S. generated expression plasmids and produced recombinant
821 proteins; E.M.B., P.V., L.Y. and M.B. collected X-ray diffraction data; M.B. solved and
822 refined crystallographic structures; E.M.B., P.V., N.L-S. and B.R. collected and analyzed
823 AUC data; A.B. purified proteins and performed kinetic assays; E.M.B. and B.R. collected
824 and analyzed SAXS data; E.M.B. collected and analyzed cryo-EM data, and performed
825 phylogeny analysis; P.M.A. provided advice on data interpretation and contributed to the
826 manuscript; M.B. designed and supervised research; E.M.B. and M.B. wrote the paper. All
827 authors reviewed the manuscript and agreed on its content.

828 **FIGURE LEGENDS**

829 **Figure 1.** Crystal structure of *CgE2p_CD* showing a trimeric protein with no trimer-trimer
830 interactions. **a)** Schematic representation of the domain architecture of *CgE2p*. **b)** Ribbon
831 representation of the apo form of *CgE2p_CD* showing its homotrimeric structure, with
832 protomers highlighted in orange, green and yellow, respectively. **c)** Zoomed view of the
833 active site region of ternary complex formed by *CgE2p_CD* with its substrates CoA-SH and
834 dihydrolipoamide (Lip(SH)₂). The ‘polder’ omit electron density map for both ligands,
835 contoured at the 4 σ level, is shown as a green mesh.

836

837 **Figure 2.** A unique, C-terminal trimer-trimer interaction (TTI) helix orientation is responsible
838 for the lack of higher order oligomerization in *CgE2p*. **a)** Ribbon representation of the
839 trimeric *CgE2p_CD* (blue) superimposed to *AvE2p_CD* (brown), which forms a 24-mer cubic
840 assembly (PDB 1EAB). **b)** Zoomed view on the superimposed trimer-trimer interaction
841 region, showing how the TTI helix in *CgE2p* occupies the equivalent position of the incoming
842 TTI helix in *AvE2p*. **c)** Schematic representation of the trimer-trimer interactions in *AvE2p*
843 exemplifying canonical E2 enzymes (left), with different colors representing protomers and
844 the protruding extension representing the C-terminal TTI helix; on the right, the situation
845 observed in *CgE2p*, where the TTI helix is involved in intramolecular interactions that hinder
846 trimer-trimer associations.

847

848 **Figure 3.** A three residues insertion (PCI) is responsible for the unique TTI helix orientation
849 that hinders inter-trimer interactions. **a)** Sequence alignment, limited to the C-terminal region,
850 between several E2 enzymes of known structure and the actinobacterial proteins studied in
851 this work, *i.e.* *MtE2b*, *MtE2p*, *CmE2p* and *CgE2p*. **b)** Cartoon representation of the *CgE2p*
852 TTI helix (yellow) over its binding pocket, colored according to its surface electrostatic
853 charges. **c)** Structural superimposition of the same C-terminal region shown in (a), showing
854 the distinct orientation of the TTI helix in *CgE2p_CD* (yellow) and *CmE2p_CD* (cyan),
855 described in this work. Please note that the same helix assumes the canonical orientation in

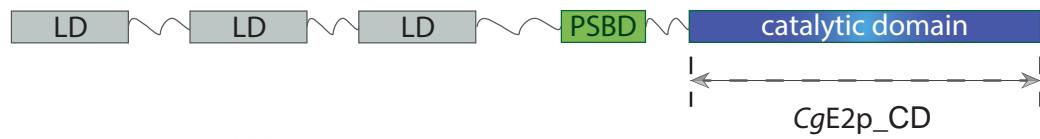
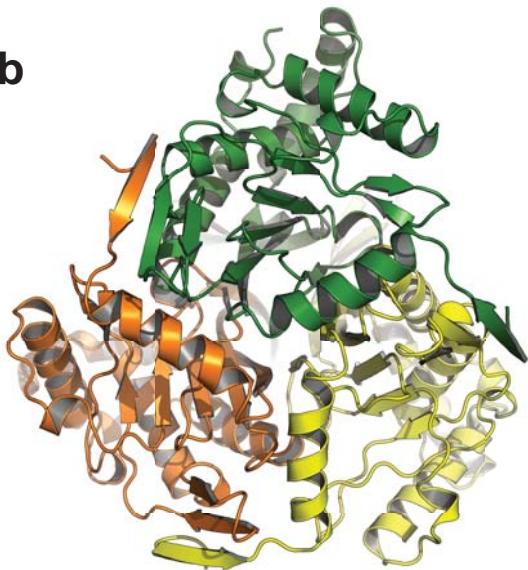
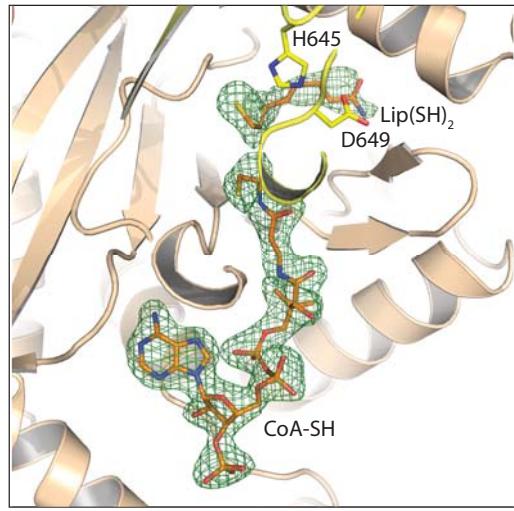
856 *MtE2b* (grey). **d**) Ribbon representation of the homotrimeric structure of CmE2p_CD, with
857 protomers indicated by distinct colors.

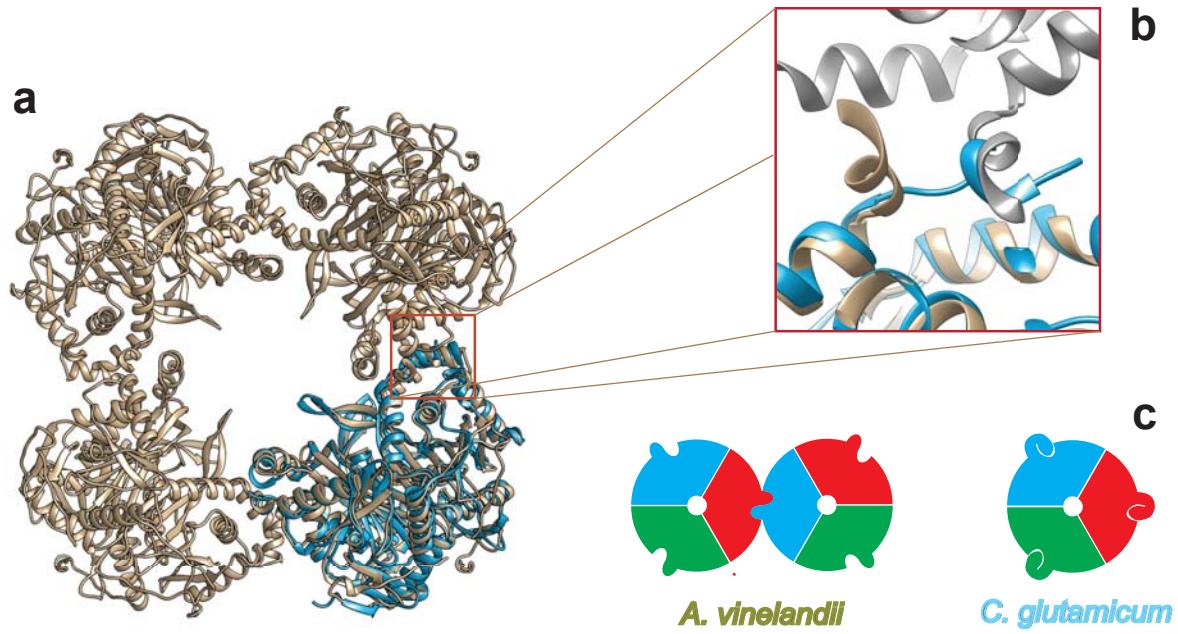
858

859 **Figure 4.** Cryo-EM structure of *MtE2b_CD*. **a)** Micrograph showing individual cubic
860 particles. **b)** Representative 2D classes showing different particle orientations. **c)** Electron
861 microscopy map of *MtE2b_CD* showing a cubic, 24-mer assembly with octahedral symmetry.
862 **d)** Cubic assembly of the same *MtE2b_CD* derived from its 1.5 Å resolution crystal structure
863 (PDB 6ZZN, this work), reconstituted from crystallographic symmetry and fitted into the
864 electron microscopy map. **e)** Zoomed view on the trimer-trimer interaction region, showing
865 the TTI helix adopting the canonical orientation observed in other E2 enzymes.

866

867 **Figure 5.** Distribution of OdhA orthologues and PCI-bearing E2 enzymes in the
868 Actinobacteria phylum. **a)** Schematic representation of the different domain associations.
869 Green blocks indicate Pfam domains associated to homodimeric E1o enzymes, while blue
870 blocks refer to domains composing E2 enzymes. The respective Pfam accession numbers are
871 indicated in the Material and Methods section. LBD: lipoyl-binding domain; PS: peripheral
872 subunit binding domain. **b)** Distribution of OdhA orthologues (red dots) or full E2o-E1o
873 fusion enzymes (LBD-PS-E2-E1; blue dots) in the phylum Actinobacteria. Branches
874 highlighted in black correspond to genomes selected to be analyzed for the PCI presence in
875 E2. **c)** Representative sequences from the actinobacterial tree, showing the correlation
876 between the presence of OdhA-like enzymes or full E2o-E1o fusion enzymes, and PCI-
877 bearing vs. canonical E2 enzymes.

a**b****c**



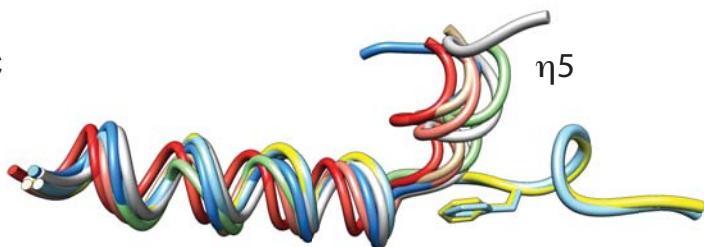
a

A.vinelandi_E2p (pdb: 1EAB)

A.vinelandii_E2p(pdb:1EAB)
H.sapiens_E2p(pdb:3B8K)
E.coli_E2o(pdb:1SCZ)
B.stearothermophilus_E2p(pdb:1B5S)
T.acidophilum_E2b(pdb:3RQC)
M.tuberculosis_E2b
M.tuberculosis_E2p
C.mustelae_E2p
C.glutamicum_E2p

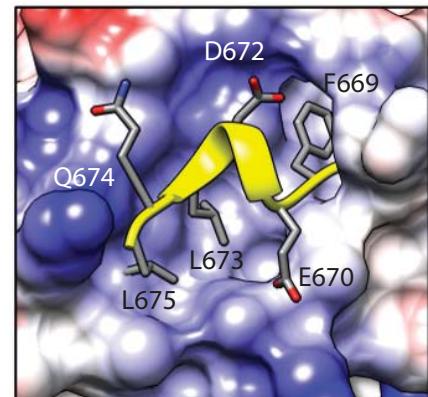
115
 llllllll ll...lll
 630
 KRLGDL LADIR . . AILL . .
 AEFRKY LEKPI . . TMLL . .
 VTIKEELLE DPT . . RLLMDV . .
 NHIKRLLS DPE . . LLMM . .
 VDLKKVIE DPN . . AI IYEI . .
 CELRDLIE SPE . . TALLD . .
 TTICKHRL EGAFAEDLGL . .
 TTICKDRLE TANFADLAL . .
 TTICKDRLE TANFEGDLQL . .

C

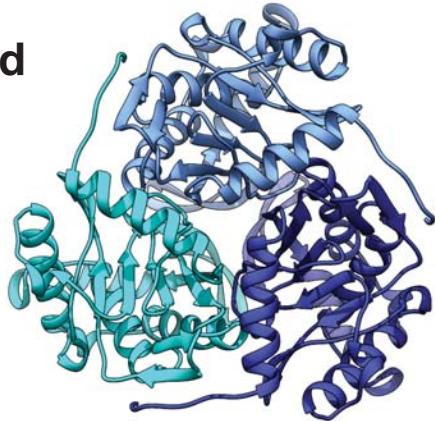


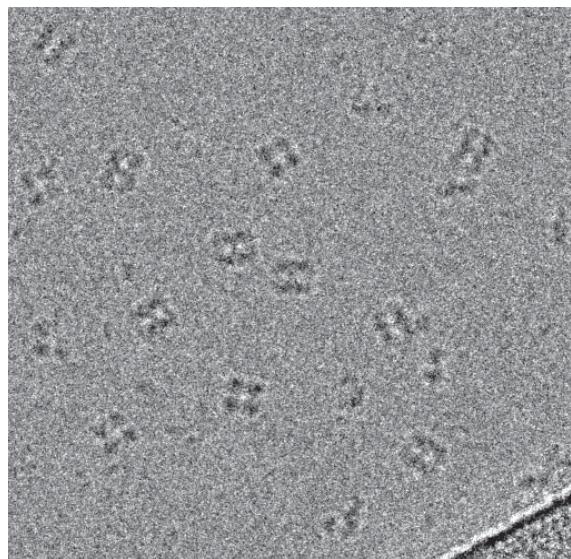
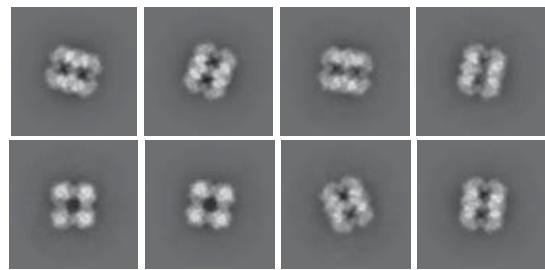
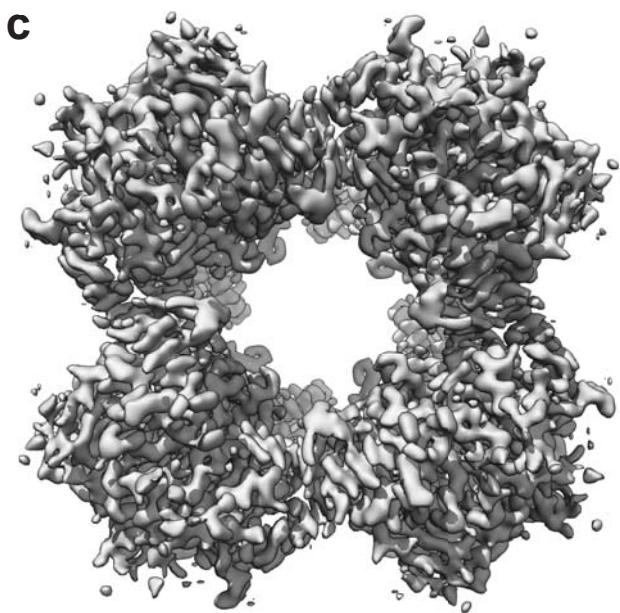
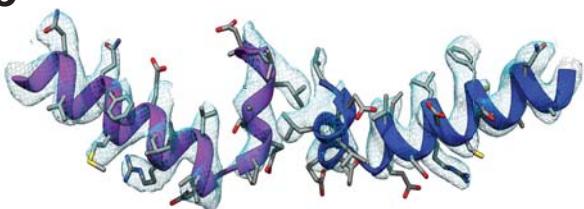
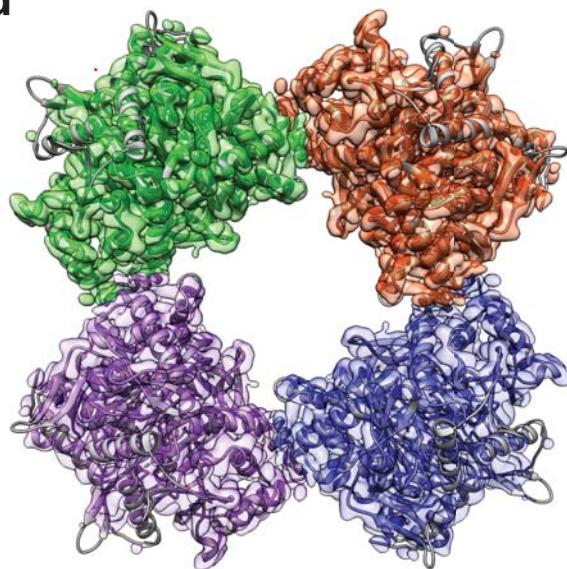
- *A. vinelandii* E2p (PDB 1EAB)
 - *H. sapiens* E2p (PDB 3B8K)
 - *E. coli* E2o (PDB 1SCZ)
 - *B. stearothermophilus* E2p (PDB 1B5S)
 - *T. acidophilum* E2b (PDB 3RQC)
 - *C. glutamicum* E2p (this work)
 - *C. mustelae* E2p (this work)
 - *M. tuberculosis* E2b (this work)

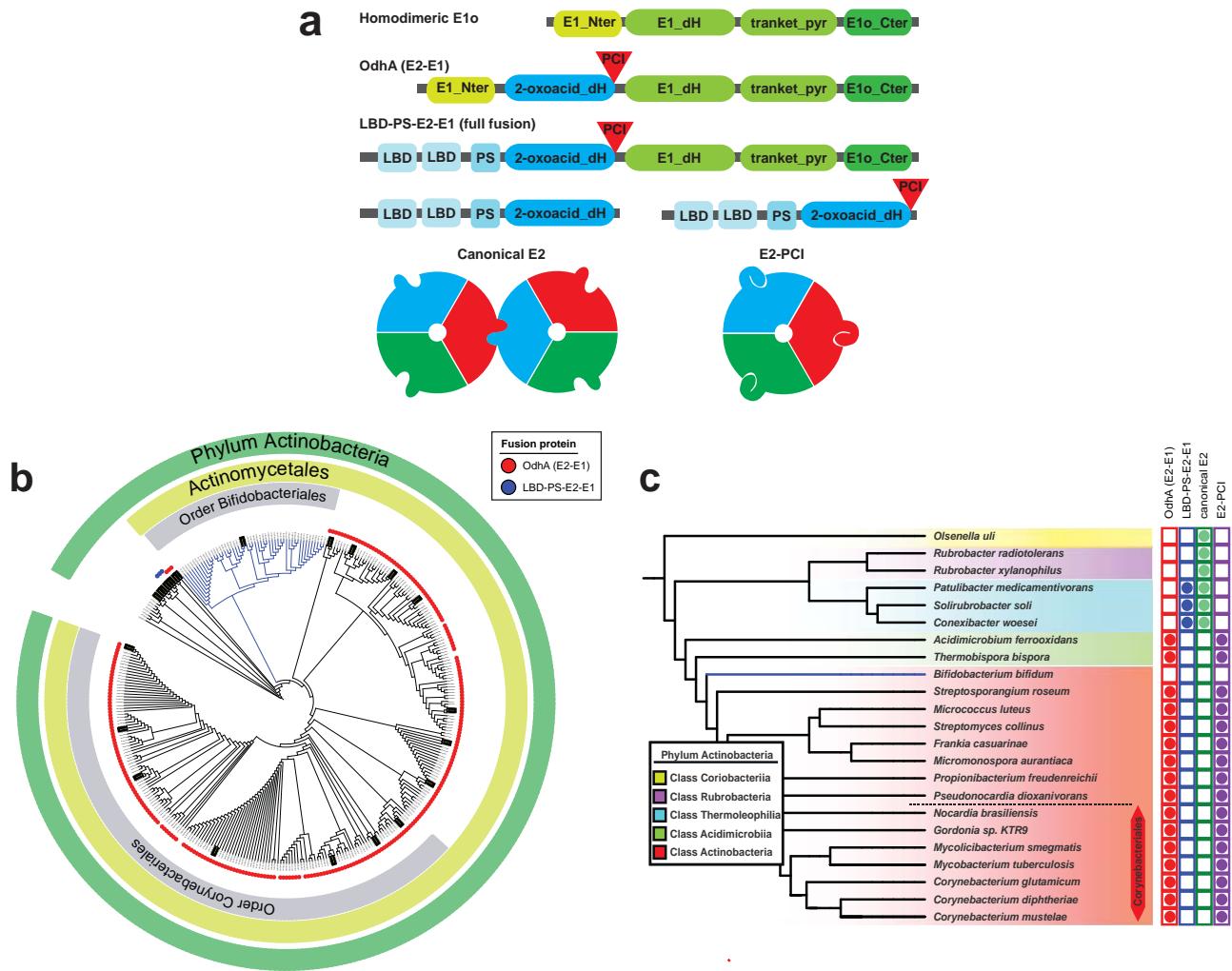
b



d



a**b****c****e****d**



| Sample | UNIPROT accession no. | Construct boundaries (residue number) | Predicted molecular mass (kDa) | AUC molecular mass estimation (kDa) | Extrapolated frictional ratio | Estimated number of protomers (AUC) | SAXS molecular mass estimation (Bayesian) (kDa) | Estimated number of protomers (SAXS) |
|-----------------|-----------------------|---------------------------------------|--------------------------------|-------------------------------------|-------------------------------|-------------------------------------|---|--------------------------------------|
| <i>CgE2p_FL</i> | Q8NNJ2 | 1-635 | 70.9 | 221 | 2.85 | 3.1 | 243 | 3.4 |
| <i>CgE2p_CD</i> | Q8NNJ2 | 437-635 | 25.4 | 78 | 1.40 | 3.1 | 77 | 3.0 |
| <i>MtE2p_FL</i> | P9WIS7 | 1-553 | 57.1 | 141 | 1.71 | 2.5 | 318 | 5.6 |
| <i>MtE2b_FL</i> | O06159 | 1-393 | 41.2 | 731 | 1.63 | 17.7 | 1179 | 28.6 |
| <i>MtE2b_CD</i> | O06159 | 165-393 | 24.6 | 600 | 1.48 | 24.3 | 479 | 19.4 |

Table 1. Constructs, predicted molecular masses and estimated properties from AUC and SAXS.