The mobility of the cap domain is essential for the substrate promiscuity of a family IV esterase from sorghum rhizosphere microbiome

Marco Distaso[a,‡], Isabel Cea-Rama[b,‡], Cristina Coscolín[c], Tatyana N. Chernikova[a], Hai Tran[a], Manuel Ferrer[c,*], Julia Sanz-Aparicio[b,*], Peter N. Golyshin[a,*]

[a]*Centre for Environmental Biotechnology, Bangor University, LL57 2UW Bangor, UK*

[b]*IQFR, CSIC, 28006 Madrid, Spain*

[c]*ICP, CSIC, 28049 Madrid, Spain*

[‡]Equal contribution.

[*]Corresponding authors at: Institute of Physical Chemistry "Rocasolano", CSIC, Serrano 119, 28006 Madrid, Spain (J. Sanz-Aparicio). ICP, CSIC, Marie Curie 2, 28049 Madrid, Spain (M. Ferrer). Centre for Environmental Biotechnology, Bangor University, LL57 2UW Bangor, UK (P.N. Golyshin)

E-mail addresses: mferrer@icp.csic.es (M. Ferrer), xjulia@iqfr.csic.es (J. Sanz-Aparicio), p.golyshin@bangor.ac.uk (P.N. Golyshin)

Phone numbers: +34915854872 (M. Ferrer), +34915619400 (J. Sanz-Aparicio), +441248383629 (P.N. Golyshin)

Running title: The cap domain of an esterase from sorghum rhizosphere

ABSTRACT Metagenomics offers the possibility to screen for versatile biocatalysts. In this study, the microbial community of the *Sorghum bicolor* rhizosphere was spiked with technical cashew nut shell liquid, and after incubation, the eDNA was extracted and subsequently used to build a metagenomic library. We report the biochemical features and crystal structure of a novel esterase from the family IV, EH_0, retrieved from an uncultured sphingomonad after a functional screen in tributyrin agar plates. EH_0 (T_{opt}, 50 °C; T_d, 55.7 °C; pH_{opt}, 9.5) was stable in the presence of 10-20% v/v organic solvents and exhibited hydrolytic activity against *p*-nitrophenyl esters from acetate to palmitate, preferably butyrate (496.5 U mg^{-1}), and a large battery of 69 structurally different esters (up to 30.23 U mg^{-1}), including bis(2-hydroxyethyl)-terephthalate (0.16 ± 0.062 U mg^{-1}). This broad substrate ambiguity contrasts with the fact that EH_0 showed a long and narrow catalytic tunnel, whose access appears to be hindered by a cap domain. We demonstrated that this cap domain is a mobile element whose opening is controlled by two contiguous proline residues which, likely acting as possible hinges, allows for the entrance of the substrates. Together, this work provides a new role for the cap domain, which until now was thought to be immobile elements that contain hydrophobic patches involved in substrate pre-recognition and in turn substrate ambiguity within family IV esterases.

IMPORTANCE A better understanding of functional and structural enzymatic diversity allows addressing biotechnological challenges. Here, we studied the structural basis of the substrate promiscuity of EH_0, an ester hydrolase of family IV, isolated from a sample of the *Sorghum bicolor* rhizosphere microbiome exposed to technical cashew nut shell liquid. The analysis of EH_0 revealed the potential of the sorghum rhizosphere microbiome as a source of enzymes with interesting properties, such as pH and solvent tolerance and remarkably broad substrate

promiscuity. Its structure resembled those of homologous proteins from mesophilic *Parvibaculum* and *Erythrobacter* spp. and hyperthermophilic *Pyrobaculum* and *Sulfolobus* spp. and had a very narrow, single-entry access tunnel to the active site, access which is controlled by a capping domain whose mobility involves two contiguous prolines. These structural markers, distinct from those of other substrate promiscuous esterases, can help tuning substrate profiles beyond tunnel and active site engineering.

KEYWORDS Crystal structure, Esterase, α/β-fold hydrolase, metagenome, promiscuity, rhizosphere.

Currently, in the biotech industrial sector, there is a constant demand for novel biocatalysts with enhanced biochemical properties fulfilling industrial requirements to replace low-performance enzymes and dismiss chemical processes (1). From this perspective, the enzymatic diversity within the microbial communities represents a yet unexplored and exceptional resource for new biocatalysts (2). However, the impossibility of isolating and culturing 99% of the environmental microbes significantly restricts the exploration of this valuable taxonomic, genetic and enzymatic variability. The microbial diversity in soil is probably highest compared to other microbial communities, with an estimated amount of 10^3-10^5 microbial species in 1 g of soil (3), and thus, it can be a source of biocatalysts. Metagenomics has proven in the last two decades to be an effective method for exploring the microbial and enzyme diversity in any kind of environmental sample, including soil (4)(2)). For this purpose, bioinformatics (5) and so-called naïve screening methods (6) are available, with the latter providing a straightforward evaluation of the enzymatic activity without the need for previous knowledge about the coding gene.

71 Carboxylic ester hydrolases (EC 3.1.1.-) are a class of enzymes widely distributed in nature
72 that have attracted much attention because of their high sequence diversity, structural
73 versatility, catalytic versatility, robustness and lack of need for cofactors (7)(8)(9)(10).
74 Considering the importance of these classes of enzymes, it is not surprising that they are
75 priority targets to screen by metagenomics. For example, approximately 280,638 sequences
76 encoding carboxylic ester hydrolases have been cataloged (10), and more than 4,000 new
77 enzymes have been discovered through metagenomics (11); few of them show enhanced
78 biophysical properties, such as stability at extreme temperatures, salinity, pH and solvents
79 conditions, high catalytic efficiency, broad substrate specificity and regio- and
80 enantioselectivity (12)(13)(6)(14). Additionally, metagenomic studies on soil environments
81 significantly contributed to broadening the knowledge about this class of enzymes with the
82 reannotation of sequences known as hypothetical proteins and the definition of new esterase
83 families (15)(16).
84 Carboxylic ester hydrolases are also the most studied enzymes with respect to promiscuity
85 (17), including catalytic (18) and substrate (19) promiscuity. The question that arises is, would
86 it not be more efficient to have just a couple of promiscuous enzymes with the ability to
87 perform a wide range of reactions (19). Based on this principle, biotechnological industries
88 demand biocatalysts that are capable of transforming a wide range of substrates. Their
89 isolation has been solved through the application of metagenomic techniques that have
90 emerged as robust tools for the discovery of unknown biocatalysts, including ester hydrolases
91 with a broad substrate scope. This method, together with the understanding of the structural
92 mechanisms and engineering behind promiscuous enzymes, seems crucial for their industrial
93 implementation. In this context, our recent study of approximately 145 ester hydrolases has
94 revealed that members of family IV show higher levels of substrate promiscuity than those of

95 other families. This is the result of their capacity to adapt the topology of the large but
96 occluded active sites to a high variety of substrates but also to the presence of a cap domain
97 next to the entrance of the substrate tunnel (19). The cap domains were thought to be
98 immobile elements, unlike lid elements in lipases, that contain hydrophobic patches involved
99 in substrate pre-recognition which could contribute to the substrate promiscuity within
100 members of family IV (20).

101 The aim of this study was to mine carboxylic ester hydrolases of the soil rhizosphere of
102 sorghum plants after enrichment with cashew nut shell liquid (CNSL), an oily substance rich
103 in phenolics and fatty acids (21), with respect to novel biocatalysts. By applying an activity-
104 based strategy, four such enzymes were retrieved and characterized. We also reported the
105 crystal structure of one of these enzymes with an extraordinarily broad substrate specificity.
106 The structural information, complemented with the experimental analysis of the substrate
107 specificity of a series of mutants, provides additional mechanistic understanding of the
108 molecular basis for the role of the cap domain in the promiscuous character of members
109 comprising family IV esterases. Importantly, we have established the role of proline residues
110 at the end of the α2-helix as hinges of the movement of the cap domain allowing substrates
111 to enter the active center.

112

113 **Results**

114 **Identification of a carboxylic ester hydrolase.** The search for hydrolytic enzymes in the
115 metagenomics library SorRhizCNSL3 W (see details in Materials and Methods) was carried out
116 by applying a metagenomic functional screening approach. The strategy involved the use of
117 LB agar plates containing tributyrin (22). The pooling of 40 clones per well dispensed into 96-
118 well microtiter plates (approximately 3.8 k clones in one plate) facilitated the colony screening

119 at a high throughput. After screening more than 100,000 clones, 1 hit from the
120 SorRhizCNSL3 W library showed lipase/esterase activity. Fosmid insert was extracted using
121 the QUIAGEN Large-Construct Kit (QIAGEN, Hilden, Germany) according to the
122 manufacturer's protocol, digested with restriction enzymes for insert size estimation, and the
123 insert sequenced by Illumina technology. Upon the completion of sequencing, the reads were
124 quality-filtered and assembled to generate nonredundant meta-sequences, and genes were
125 predicted and annotated as described previously (12). One gene encoding a predicted
126 carboxylic ester hydrolase was identified. The gene was amplified with specific primers,
127 cloned into the p15TV-L vector and transformed into *E. coli* BL21(DE3) cells for expression of
128 the N-terminal (His)6-tagged proteins. The deduced amino acid sequence of the enzyme (324
129 amino acids long) was used for homology searches in the taxonomy and functional
130 assignment. A database search indicated that EH_0 showed 99% identity with the α/β
131 hydrolase enzyme from *Sphingomonas pruni* (protein ID WP_066587239), both of which are
132 classified as members of the α/β hydrolase-3 family (PF07859). The typical H-G-G-G motif and
133 the G-X-S-X-G catalytic motif are conserved, and EH_0 clustered together with family IV
134 esterase.

135 **Biochemical characterization.** The recombinant protein was successfully expressed in
136 soluble form and purified by nickel affinity chromatography. Purified protein was desalted by
137 ultrafiltration, and its enzymatic activity was assessed. Four model *p*-nitrophenyl (*p*-NP) ester
138 substrates with different chain lengths were used to determine the substrate specificity of
139 the enzymes and therefore determine whether the enzymes are in fact true lipases or
140 esterases. Lipases hydrolyze ester bonds of long-chain triglycerides more efficiently than
141 esterases, which instead exhibit the highest activity toward water soluble esters with short
142 fatty acid chains (23). The substrates used for the hydrolytic test were *p*-NP-acetate (C2), *p*-

143 NP-butyrate (C4), *p*-NP-dodecanoate (C12), and *p*-NP-palmitate (C16). The hydrolytic activity
144 was recorded under standard assay conditions (Fig. 1). EH_0 showed a specific activity of 496.5
145 U mg^{-1} for *p*-NP butyrate, which was the best substrate. Lower levels of activity were observed
146 with longer chain esters (C ≥ 12). The esterase followed the Michaelis–Menten kinetics, and
147 its kinetic parameters are reflected in Table 1. A comparison of the catalytic efficiency values
148 (k_{cat}/K_M) indicated a high reactivity toward *p*-NP-butyrate followed by *p*-NP-acetate.
149 Its voluminous (volume of the active site cavity: 5133 $Å^3$) but low exposed (solvent accessible
150 surface area (SASA): 5.07 over 100 dimensionless percentage) active site allows hydrolysis of
151 a broad range of 68 out of 96 structurally and chemically diverse esters (Table S1), as
152 determined by a pH indicator assay (pH 8.0, 30 °C). Phenyl acetate (30.23 U mg^{-1}) and glyceryl
153 tripropionate (29.43 U mg^{-1}), were the best substrates. We also found that EH_0 efficiently
154 hydrolyzed bis(2-hydroxyethyl)-terephthalate (BHET; 163.6 ± 6.2 U mg^{-1}), an intermediate in
155 the degradation of polyethylene terephthalate (PET) (24); HPLC analysis (Fig. S1), performed
156 as described (25), confirmed the hydrolysis of BHET to mono-(2-hydroxyethyl)-terephthalic
157 acid (MHET) but not to terephthalic acid (TA). However, using previously described conditions
158 (25), we found that the enzyme did not hydrolyze large plastic materials such as amorphous
159 and crystalline PET film and PET nanoparticles from Goodfellow. According to the number of
160 hydrolyzed esters, EH_0 can be thus considered as an esterase with a wide substrate ambiguity,
161 similar to other enzymes of the family IV (19)(20).
162 EH_0 showed maximal activity at 50 °C, retaining more than 80% of the maximum activity at
163 40-55 °C (Fig. 2A), suggesting that it is moderately thermostable. This was confirmed by
164 circular dichroism analysis, which revealed a denaturing temperature of 55.7 ± 0.2 °C (Fig. 2B).
165 Its optimal pH for activity was 9.5 (Fig. 2C). The effect on the enzymatic activity of organic
166 solvents at different concentrations was evaluated (Table S2). An activation effect was

167 observed for EH_0 when 10% methanol and 10-20% DMSO were added to the reaction mixture.
168 The presence of bivalent and trivalent cations did not have a remarkable positive effect on
169 the activity of the enzymes, which showed, in some cases, tolerance to high concentrations
170 of cations (Table S3). A prominent inhibiting effect was shown for all cations, except for
171 magnesium, which was well tolerated at 1-10 mM (<5% inhibition).

172 **EH_0 presents tight folding of its cap domain.** The crystal structure of wild type EH_0 was
173 obtained at 2.01 Å resolution, with the $P2_12_12_1$ space group and two crystallography-
174 independent molecules in the asymmetric unit. Molecular replacement was performed using
175 Est8 as a template (PDB code 4YPV) (26), and the final model was refined to a crystallographic
176 R-factor of 0.1717 and R-free of 0.2019 (Table S4). As with other reported family IV esterases,
177 EH_0 has an **α/β** hydrolase fold with two different domains, a cap domain (residues 1-43 and
178 208-229) and a catalytic domain (residues 44-207 and 230-324), constituted by a total of 9 **α**-
179 helices and 8 **β**-sheets (Fig. 3A). The catalytic domain was composed of a central **β**-sheet with
180 eight parallel **β**-strands (**β**1, **β**3, **β**4, **β**5, **β**6, **β**7, **β**8), except **β**2, which was antiparallel and
181 surrounded by five **α**-helices (**α**3, **α**4, **α**5, **α**8, **α**9). The cap domain involved four **α** helices
182 (**α**1, **α**2, **α**6, **α**7) (Fig. 3B). There were two *cis* peptides, Ala122-Pro123 and Trp127-Pro128,
183 located at the **β**4-**α**4 turn within the catalytic domain.

184 Analysis of EH_0 folding using the *DALI* server (27) was employed to search for homologous
185 proteins. The closest homologs are E53 isolated from *Erythrobacter longus*, with 46% identity
186 and RMSD of 2.3 Å on 296 Cα atoms (PDB code 7W8N) (28); Est8 isolated from *Parvibaculum*,
187 with 38% identity and RMSD of 2.5Å on 291 Cα atoms (PDB code 4YPV) (26); PestE isolated
188 from *Pyrobaculum calidifontis*, with 33% identity and RMSD of 2.6Å on 294 Cα atoms (PDB
189 code 3ZWQ) (29); and EstA isolated from *Sulfolobus islandicus REY15A,* with 31% identity and
190 RMSD of 2.6Å on 281 Cα atoms (PDB code 5LK6). Major differences are found in the cap

191 domain, with α2 and the long α7 being very close to the EH_0 active site and apparently
192 impeding the entrance of substrates (Fig. 3C). This was unexpected considering the broad
193 substrate ambiguity of esterase EH_0, approaching that of most promiscuous ones (19). Thus,
194 withdrawal of the cap domain seems a necessary requirement for allowing access of the bulky
195 substrates to the EH_0 catalytic site.

196 In agreement with this assumption, a substantial rearrangement of the cap domain was
197 previously described in the homolog esterase EST2 (having 31 % sequence identity), with its
198 M211S/R215L double mutant being trapped in the crystal in a conformation resembling the
199 open form of lipases (30). In this study, some flexibility at this region was assumed from the
200 two conformations adopted by loop α1-α2 observed in subunits A and B of EH_0 (Fig. 3D). In
201 addition, the inspection of residues within the cap domain showed a high number of proline
202 residues (Pro8, Pro21, Pro23, Pro31, Pro46 and Pro47) (Fig. 3B) that could confer flexibility to
203 the N-terminal cap domain allowing substrate entrance. This feature will be discussed below.

204 **EH_0 is a dimeric enzyme.** EH_0 is presented as a biological homodimer with approximate
205 dimensions of 6.3 x 4.9 x 2.7 nm, which is assembled in a twofold axis symmetry arrangement
206 (Fig. 4A) that buries 5.6% of its total surface area. Hydrogen bonds mainly involve β8 and α8,
207 while salt bridges involve motifs β8 and α9 (Fig. 4B and Table S5). Similar to its homolog PestE,
208 dimerization occurs through a tight interaction of β8 strands from both subunits. However,
209 while only the subsequent helices **α**12 were involved in the PestE interface (29) (Fig. 4C), both
210 the precedent **α**8 and the subsequent **α**9 helices make the EH_0 interface (Fig. 4D). These
211 different interactions observed in PestE and EH_0 were reflected in a different orientation
212 between the monomers that, nevertheless, present a similar distance between the catalytic
213 serines of 35-38 Å and an equivalent disposition of the tunnels giving access to the catalytic
214 site at two edges of the dimer (Fig. 4E, 4F).

215 **The peculiar EH_0 active site.** EH_0 has a long catalytic tunnel with a very narrow entrance
216 (approximately 16.8 Å depth) (Fig. 5A). The catalytic triad of EH_0 is formed by Ser161 (in the
217 conserved motif ^{159}G-D-S-A-G^{163} in the nucleophilic elbow), Asp253 (in the conserved motif
218 ^{253}D-P-I-R-D^{258}) and His283 (Fig. 5B). To analyze the active site, a series of soaking and
219 cocrystallization experiments were performed with different suicide inhibitors, all of which
220 were unsuccessful. A deep inspection of the active site showed that the nucleophilic Ser161
221 is hydrogen bonded to Glu226 from the close **α**7, and a movement of loop **β**3-**α**3, including
222 the oxoanion in the conserved motif ^{87}H-G-G-G^{90,} compared to EH_0 homologs (Fig. 5B).
223 Therefore, to remove this hydrogen link in an attempt to expand access to the active site
224 enabling ligand binding, the EH_{0E226A} mutant was designed by directed mutagenesis. However,
225 the crystals grown from this mutant failed to diffract, suggesting a high level of disorder,
226 putatively due to mobility of the cap. Interestingly, the activity profile of this mutant reveals
227 a faster hydrolysis of most of the esters (62 esters in total; from 70- to 1.2-fold; average: 6.1-
228 fold). Only 7 esters were hydrolyzed to a lower extent compared to the wild type (from 1.5-
229 to 17-fold; average: 4.6-fold). Interestingly, the EH_{0E226A} mutant was capable of accepting
230 large and voluminous esters such as dodecanoyl acetate, pentadecyl acetate, vinyl laurate,
231 methyl 2,5-dihydroxycinnamate, and ethyl 2-chlorobenzoate, which were not accepted by the
232 wild-type enzyme. This apparently supports that removal of the Ser161-Glu226 hydrogen
233 bond increases flexibility and enhances the enzymatic efficiency toward large esters either at
234 the acyl or alcohol sites.

235 Indeed, it appears clear that helix α7 must retract from the catalytic site to allow substrate
236 entrance, but this helix seems fixed by many atomic interactions. Close to the Ser161-Glu226
237 hydrogen bond, Tyr223 makes additional hydrogen bonds to Asp194 and Lys197 main chain,
238 from loop β6-α6, and with the side chain of Gln258 from α8 (Fig. 5C). A mutation EH_{0Y223A} was

239 designed by directed mutagenesis and found, unlike the EH_{0E226A} mutation, little effect on
240 conversion rates compared to the wild-type enzyme. However, this mutation allowed the
241 hydrolysis of large and voluminous esters, such as dodecanoyl acetate, pentadecyl acetate,
242 methyl 2,5-dihydroxycinnamate and ethyl 2-chlorobenzoate, which were also hydrolyzed by
243 the mutant EH_{0E226A}. Unlike EH_{0E226A}, the EH_{0Y223A} mutant was not unable to hydrolyze vinyl
244 laurate, but it was able to hydrolyze methyl 3-hydroxybenzoate, not accepted by the EH_{0E226A}
245 mutant. This suggested that Tyr223 may play a role in accepting large esters, particularly at
246 the acyl side, but may also play an additional role in substrate specificity different from that
247 of Glu226.

248 Additionally, at the beginning of α7, Trp218 is within a hydrophobic pocket surrounded by
249 residues Phe13, Ile17, Leu28 and Phe219 from the cap domain and Phe91 from the catalytic
250 domain, which is also anchored by interaction with loop α1-α2 and Pro23 (Fig. 5D). Thus, to
251 investigate the effect of the proposed cap motion in disrupting these tight molecular packing,
252 an ensemble refinement modeling was computed. The molecular dynamics simulation,
253 combined with the X-ray data, revealed that the region comprising α1 and α2 shows a wide
254 spectrum of possible pathways from more "open" to more "closed" conformations (Fig. 5E).
255 At one edge, Pro23 is in an extended α1-α2 loop far from Trp218, therefore releasing α7 that
256 consequently could retract from the catalytic pocket ("open conformation") (Fig. 5F). In fact,
257 the ensemble refinement models a very flexible conformation being even unstructured at
258 regions corresponding to α1 and α2. At the other edge, the second scenario is the entrapment
259 of Trp218 by Pro23 in loop α1-α2, hindering substrate entrance ("closed conformation") (Fig.
260 5G). This last scenario is equivalent to the 3D structure captured by crystallography (Fig. 5D).
261 Three prolines can be found within this α1-α2 loop, Pro21 (at the end of α1), Pro23 (at the
262 middle) and Pro31 (at the beginning of α2), which are probably behind the two different

263 conformations observed at this loop in both subunits within the asymmetric unit (Fig. 3B and
264 3D).

265 Furthermore, as seen in Fig. 3B, Pro46 and Pro47 are potential hinges that would involve
266 flexibility of a larger region of the EH_0 cap domain, including the whole N-terminal peptide
267 chain up to the end of α2. This is consistent with the more "open" conformations resulting
268 from the ensemble refinement shown in Fig. 5F. In fact, the sequence comparison of EH_0 to
269 its closest homologs reveals that only EH_0 has two sequential prolines at this region and that
270 Pro46 is unique to EH_0 (Fig. 6), which could be the reason behind the high EH_0 promiscuity.
271 Interestingly, EST2 also presents the two contiguous Pro residues (Pro38, Pro39), which can
272 explain the "open" conformation captured in the crystal mentioned above (30). The mutation
273 EH_{0P46A} was designed by directed mutagenesis. It is interesting to highlight that crystals grown
274 from this mutant EH_{0P46A} failed to diffract, which may suggest that the flexibility associated
275 with Pro46 is essential for structural stability. Analysis of the activity profile showed that
276 Pro46 is a critical residue for the entry and hydrolysis of bulky substrates, as its mutation by
277 Ala extends the substrate specificity from 68 to 84 esters. Additionally, the hydrolytic rate
278 increased from 1.2- to 18000-fold (average: 335-fold) for most esters. This mutant was also
279 able to hydrolyze large glyceryl trioctanoate and 2,4-dichlorophenyl 2,4-dichlorobenzoate,
280 which were not hydrolyzed by the wild-type enzyme or the EH_{0E226A} and H_{0Y223A} mutants.

281 Structural details of the EH_0 active site and assignment of the acyl and alcohol moieties
282 were explored by comparison with its homologs E53 complexed with 4-nitrophenyl
283 hexanoate (PDB code 6KEU) (28) and EH_{1AB1} in complex with a derivative of methyl 4-
284 nitrophenylhexylphosphonate (PDB code 6RB0) (31). However, the acyl and alcohol moieties
285 of these complexes are located in opposite sites (Fig. 5H, 5I). Therefore, as we could not
286 obtain complexes from EH_0, the activity experiments were crucial to correctly assign

287 acyl/alcohol moieties. As mentioned above, experimental evidence demonstrated that
288 Tyr223 produces a steric hindrance at the acyl moiety, and consequently, the acyl and alcohol
289 sites correspond to those observed in EH_{1AB1} (Fig. 5I). On the basis of this assumption, the acyl
290 binding site seems to be a small cavity bordered by the Tyr199 and Ile255 side chains, which
291 produce steric hindrance for substrates with large acyl moieties. The long and narrow alcohol
292 binding site is surrounded by both hydrophobic (Phe13, Phe91, Val92, Leu288) and
293 hydrophilic residues (Asp45, His99, Asp160, Tyr190, Thr287), with Pro46 and Pro47 being at
294 the entrance of the tunnel (Fig. 5J). Most residues at the acyl and alcohol moieties are
295 conserved among EH_0 homologs, with the exception of Met40, Pro46, Tyr199, Glu226 and
296 Thr287. Remarkably, the bulky Met40 and Tyr199 residues are substituted by smaller residues
297 in the EH_0 homologs. As previously mentioned, Pro46 is unique in EH_0, while Glu226 is
298 substituted by a conserved Asp, and finally, most homologs show an Asn residue instead of
299 Thr287 (Fig. 6). Therefore, as the retraction of the cap domain must be performed to allow
300 substrate entrance, residues Phe91, Tyr190 and Tyr199 from the catalytic domain, which are
301 located close to the catalytic triad, seem essential for substrate specificity (Fig. 5J).
302
303 **DISCUSSION**
304 In this study, the microbial community of the *Sorghum bicolor* rhizosphere was exposed to
305 a chemical treatment prior to eDNA extraction to construct a metagenomic library. Plant roots
306 can secrete exudates composed of a large variety of compounds into the soil, some of which
307 may play important roles in the rhizosphere (32)(33), and with effects that involve multiple
308 targets, including soil microorganisms. This is why an amendment of the soil with technical
309 cashew nut shell liquid (tCNSL), containing a mixture of phenolic compounds with long
310 aliphatic side chains (up to C22:0), was carried out directly in the rhizosphere and, later, for

311 three weeks under control laboratory conditions, as we were interested in screening lipases.
312 The corresponding metagenomic clone library (SorRhizCNSL3W) was constructed, which
313 based on their size guarantees a fair representation of the preenriched microbial assemblage.
314 Functional screening of the fosmid library was carried out by the incorporation of specific
315 substrates in the growth medium, which allows the identification of the corresponding
316 enzymatic activity encoded by a metagenomic clone. One positive hit for lipolytic activities
317 was identified and affiliated with a bacterium of the order *Sphingomonadales*. The deduced
318 amino acid sequence showed the α/β hydrolase fold, which is common to several hydrolytic
319 enzymes of widely different phylogenetic origins and catalytic functions. Hydrolytic assays
320 conducted on esters with short and long acyl chains ultimately defined the metagenomic
321 enzyme as an esterase. Phylogenetic and structural analysis allowed us to classify the esterase
322 according to the previously described hormone-sensitive lipase (HDL) type IV within at least
323 35 families and 11 true lipase subfamilies known to date (10)(23)(34). This family is reported
324 to contain ester hydrolases with high levels of substrate ambiguity (19), a characteristic also
325 observed for EH_0, which demonstrated an outstanding response when its hydrolytic activity
326 was tested toward structurally different ester substrates, being among the top promiscuous
327 ester-hydrolase enzyme characterized (19). This enzyme houses a very long and narrow
328 catalytic pocket, where helix α7 is very close to the catalytic triad with residue Glu226, making
329 a direct hydrogen bond to the catalytic nucleophile Ser161. Therefore, it appeared clear that
330 the cap domain must retract to allow substrate entrance to the active site. The movement of
331 this domain was modeled by molecular dynamics simulation through the ensemble
332 refinement procedure. This strategy showed a broad range of putative conformations at the
333 cap domain, with Pro46 and Pro47 likely acting as hinges conferring a high plasticity to the N-
334 terminal region of the cap. Remarkably, the presence of these two sequential prolines is a

335 unique feature of EH_0 compared to its homologs, and other substrate-promiscuous members
336 of the family IV (19). Mutational analysis confirmed the role of one of these prolines in the
337 access and hydrolysis of large and voluminous substrates and thus in the increase in the
338 substrate promiscuity level.

339 Biochemical characterization of the novel esterase revealed interesting properties that
340 make the enzyme a good candidate for industrial applications. The activity of the enzyme was
341 in most cases stimulated in the presence of 10% organic solvents, particularly in 10%
342 methanol and DMSO. The use of organic solvents offers more advantages over canonical
343 aqueous biocatalysis for various reasons: higher solubility of hydrophobic substrates, minor
344 risk of contamination, and higher thermal stability (35)(36)(37). EH_0 has a potential advantage
345 in applications that require alkaline conditions due to its ability to act at the optimal pH of
346 9.0. Temperature-controlled tests indicated a mesophilic/slightly thermophilic profile of the
347 esterase as expected from the original habitat at moderate temperatures. In addition, the
348 structure of EH_0 is similar (31-33% identity) to those of extremophiles, namely, *Pyrobaculum*
349 (3ZWQ) and *Sulfolobus* (5LK6) species.

350 In summary, the present study evaluated the conformational plasticity of the cap domain
351 in members of Family IV, and the role of two sequential prolines as putative structural factor
352 regulating their broader substrate specificity compared to other members of the 35 families
353 and 11 true lipase subfamilies reported (10). This high molecular flexibility is markedly
354 different to that found in other family IV esterases and a family VIII β-lactamase fold hydrolase
355 (EH_7) which has been recently shown to be highly substrate-promiscuous. In this case, the
356 broad substrate ambiguity is given by the presence of a more opened and exposed S1 site
357 having no steric hindrance for the entrance of substrates to the active site, and more flexible

358 R1, R2 and R3 regions allowing the binding of a wide spectrum of substrates into the active
359 site.
360
361 **CONCLUSIONS**
362 Activity-based metagenomics approach was used to study the microbial enzyme diversity
363 in rhizosphere soil of *Sorghum* plants amended with CNSL soil. A novel esterase was found,
364 which possessed a broadest substrate promiscuity in combination with a significant pH and
365 solvent tolerance. This work is crucial for deciphering structural markers responsible for the
366 outstanding broad substrate specificity of EH_0, which makes this enzyme an interesting
367 candidate for biotechnological applications. This work further provides important insights
368 into the role of cap domains and its contribution to the diverse selectivity profiles and thus
369 versatility of family IV esterases/lipases toward diverse applications.
370
371 **MATERIALS AND METHODS**
372 **Plant material and outdoor seed germination.** Seeds of *Sorghum bicolor* genotype BTx623
373 were obtained from the Agricultural Research Service of United States Department of
374 Agriculture (USDA) (Georgia, US) as a gift. Field soil was sampled from the Henfaes Research
375 Centre (53°14'21.0"N, 4°01'06.5"W, Gwynedd, Wales) in September 2014. The soil sample
376 was composed of a mixture of five topsoil samples collected from randomly selected positions
377 in the field. The soil was air-dried, mixed thoroughly, and stored at room temperature for use
378 in subsequent experiments. Two L pots were filled with soil, and two seeds of *S. bicolor*
379 BTx623 were planted per pot. Plants were cultivated in a greenhouse at 20 °C, and the soil
380 moisture content was maintained with tap water.

381 **Enrichment with CNSL.** Three grams of technical cashew nut shell liquid (tCNSL) dissolved
382 in 70% ethanol was added to a pot of 20-day-old plants and thoroughly mixed with the soil.
383 After 60 days, plants were pulled out of the pot, and the soil was shaken off; samples of
384 rhizosphere soil attached to the plant roots were then brushed off and collected. tCNSL was
385 provided by the BioComposites Centre at Bangor University (Wales, UK). Three biological
386 replicates of laboratory microcosm enrichment were set up in conical 1 L Erlenmeyer flasks
387 by mixing 10 g of the collected rhizosphere soil with 300 mL of sterile Murashige Skoog Basal
388 medium (Sigma) and 10 mg/L cycloheximide. tCNSL was dissolved in 70% ethanol and added
389 to the medium to a final concentration of 0.1 g/L; flasks were incubated at 20 °C in an orbital
390 shaker. 50 g of oil slurry was sampled every seven days, and fresh tCNSL-containing (0.1 g/kg)
391 medium was added to replace the volume of the medium.
392 **Extraction of DNA and generation of metagenomic library.** Samples collected after 3
393 weeks of flask microcosm enrichment were used for the construction of fosmid metagenomic
394 libraries. Environmental DNA was extracted using the Meta-G-Nome DNA Isolation Kit
395 (Epicenter Biotechnologies; WI, USA) according to the manufacturer's instructions. Briefly, 50
396 ml of the soil suspension from the flask enrichment was centrifuged at 400 x g for 5 min. The
397 supernatant was filtered through 0.45 µm and 0.22 µm membrane filters. This procedure was
398 repeated with the initial soil sample four times: the remaining soil was resuspended in PBS
399 and centrifuged, and the supernatant was filtered as before. Filters were combined, and the
400 sediment on the filter was resuspended in extraction buffer and collected. DNA extraction
401 was carried out according to the protocol described by the manufacturer. The quality of the
402 extracted DNA was evaluated on agarose gel and quantified with the Quant-iT dsDNA Assay
403 Kit (Invitrogen) on a Cary Eclipse fluorimeter (Varian/Agilent) according to the
404 manufacturer's instructions. The extracted metagenomic DNA was used to prepare two

405 different metagenomic fosmid libraries using the CopyControl™ Fosmid Library Production Kit
406 (Epicenter). DNA was end-repaired to generate blunt-ended, 5'-phosphorylated double-
407 stranded DNA using reagents included in the kit according to the manufacturer's instructions.
408 Subsequently, fragments of 30-40 kb were selected by electrophoresis and recovered from a
409 low melting point agarose gel using GELase 50X buffer, and GELase enzyme preparation was
410 also included in the kit. Nucleic acid fragments were then ligated to the linearized CopyControl
411 pCC2FOS vector in a ligation reaction performed at room temperature for 4 hours, according
412 to the manufacturer's instructions. After in vitro packaging into phage lambda (MaxPlax™
413 Lambda Packaging Extract, Epicenter), the transfected phage T1-resistant EPI300™-T1R *E. coli*
414 cells were spread on Luria-Bertani (LB) agar medium (hereinafter, unless mentioned
415 otherwise, the agar content was 1.5% w/vol) containing 12.5 µg/ml chloramphenicol and
416 incubated at 37 °C overnight to determine the titer of the phage particles. The resulting
417 library, SorRhizCNSL3 W, has an estimated titer of $1.5x10^6$ clones. For long-term storage, the
418 library was plated onto solid LB medium with 12.5 µg/ml chloramphenicol, and after
419 overnight growth, colonies were washed off from the agar surface using LB broth with 20%
420 (v/v) sterile glycerol, and aliquots were stored at −80 °C.

421 **Screening metagenomic libraries: agar-based methods.** Fosmid clones obtained by plating
422 the constructed libraries on LB agar plates were arrayed in 384-microtiter plates (1
423 clone/well) or alternatively in 96-microtiter plates (pools of approximately 40 clones/well)
424 containing LB medium and chloramphenicol (12.5 µg/ml). The plates were incubated at 37 °C
425 overnight, and the day after replication, the plates were produced and used in the screening
426 assay. Glycerol (20% v/v, final concentration) was added to the original plates, which were
427 stored at -80 °C. Gel diffusion and colorimetric assays were adopted for the screening of the
428 desired activities. The detection of lipase/esterase activity was carried out on LB agar

429 supplemented with chloramphenicol (12.5 µg/ml), fosmid autoinduction solution (2 ml/L)
430 (Epicenter), and 0.3% (v/v) tributyrin emulsified with gum arabic (2:1, v/v) by sonication. The
431 previously prepared microtiter plates were printed on the surface of large (22.5 cm x 22.5 cm)
432 LB agar plates using 384-pin polypropylene replicators and incubated for 18-48 hours at 37
433 °C. Lipolytic activity was identified as a clear zone around the colonies where tributyrin was
434 hydrolyzed (12).

435 **Extraction of fosmids, DNA sequencing and annotation.** The fosmid DNA of the positive
436 clone was extracted using the QIAGEN Plasmid purification kit (QIAGEN). To reduce the host
437 chromosomal *E. coli* DNA contamination, the sample was treated with ATP-dependent
438 exonuclease (Epicenter). The purity and approximate size of the cloned fragment were
439 assessed by agarose gel electrophoresis after endonuclease digestion simultaneously with
440 *BamHI* and *XbaI* (New England Biolabs, in NEBuffer 3.1 at 37° for 1 hour using 1 U of enzyme
441 per 1 µg DNA). DNA concentration was quantified using the Quant-iT dsDNA Assay Kit
442 (Invitrogen), and DNA sequencing was then outsourced to Fidelity Systems (NJ, USA) for
443 shotgun sequencing using the Illumina MiSeq platform. GeneMark software (38) was
444 employed to predict protein coding regions from the sequences of each assembled contig,
445 and deduced amino acid sequences were annotated via BLASTP and the PSI-BLAST tool (39).

446 **Cloning, expression and purification of proteins.** The selected nucleotide sequence was
447 amplified by PCR using Herculase II Fusion Enzyme (Agilent, USA) with specific oligonucleotide
448 primer pairs incorporating p15TV-L adapters. The corresponding fosmid was used as a
449 template to amplify the target genes. The primers used to amplify the esterase gene
450 characterized in this study were as follows: EH_{0F},
451 TTGTATTTCCAGGGCATGACCGAGCTCTTCGTCCGC; EH_{0R},
452 CAAGCTTCGTCATCATGCCGCCGCCTGTGCCATC. PCR products were visualized on a 1% TAE

453 agarose gel and purified using the Nucleospin PCR Clean-up Kit (Macherey-Nagel) following
454 the manufacturer's instructions. Purified PCR products were cloned into the p15TV-L vector,
455 transformed into *E. coli* NovaBlue GigaSingles™ Competent Cells (Novagen, Germany) and
456 plated on LB agar with 100 µg/ml ampicillin. The correctness of the DNA sequence was then
457 verified by Sanger sequencing at Macrogen Ltd. (Amsterdam, The Netherlands). 3-D models
458 of the proteins were generated by Phyre2. The intensive mode attempts to create a complete
459 full-length model of a sequence through a combination of multiple template modeling and
460 simplified ab initio folding simulation (40). The nucleotide and amino acid sequences of the
461 selected nucleotide sequences are available GenBank under ref. MK791218. For recombinant
462 protein expression, the plasmids were transformed into *E. coli* BL21(DE3) cells and
463 subsequently plated on LB agar with 100 µg/ml ampicillin. To confirm the esterase activity of
464 recombinant proteins, *E. coli* clones harboring the recombinant plasmid were streaked onto
465 LB agar plates containing 0.5% tributyrin and 0.5 mM IPTG, or purified enzymes were spotted
466 directly on the agar. The plates were then incubated at 37 °C overnight and visually inspected
467 for the presence of signs of substrate degradation. *E. coli* clones were grown at 37 °C to an
468 absorbance of 0.8 at 600 nm, induced with 0.5 mM isopropyl-β-D-galactopyranoside (IPTG),
469 and allowed to grow overnight at 20 °C with shaking. Cells were harvested by centrifugation
470 at 5000 x g for 30 min at 4 °C. For purification of recombinant protein, the following protocol
471 was applied. Cell pellets were resuspended in cold binding buffer (50 mM HEPES pH 7.5, 400
472 mM NaCl, 5% glycerol, 0.5% Triton-X, 6 mM imidazole pH 7.5, 1 mM β-mercaptoethanol, 0.5
473 mM PMSF) and extracted by sonication. The lysates were then centrifuged at 22,000 x g for
474 30 min at 4 °C, and the supernatant was purified by affinity chromatography using Ni-NTA His-
475 bind resin (Novagen). The column packed with the resin was equilibrated with binding buffer,
476 and after the addition of supernatant, it was washed with 6 volumes of wash buffer (50 mM

477 HEPES pH 7.5, 400 mM NaCl, 5% glycerol, 0.5% Triton-X, 26 mM imidazole pH 7.5, 1 mM β-
478 mercaptoethanol, 0.5 mM PMSF) to remove nonspecifically bound proteins. His-tagged
479 proteins were then eluted with elution buffer (50 mM HEPES pH 7.5, 400 mM NaCl, 5%
480 glycerol, 0.5% Triton-X, 266 mM imidazole pH 7.5, 1 mM β-mercaptoethanol, 0.5 mM PMSF).
481 The size and purity of the proteins were estimated by SDS–PAGE. Protein solutions were
482 desalted through the Amicon Ultra15 10K centrifugal filter device. Protein concentrations
483 were determined using the Bradford Reagent (Sigma) and the BioMate™ 3S
484 spectrophotometer (Thermo Scientific, USA).

485 **Biochemical assays.** Hydrolytic activity was determined by measuring the amount of *p*-
486 nitrophenol released by catalytic hydrolysis of *p*-nitrophenyl (*p*-NP) esters through a modified
487 method of Gupta et al. (41). Stock solutions of *p*-NP esters (100 mM *p*-NP acetate, 100 mM *p*-
488 NP butyrate, 20 mM *p*-NP dodecanoate, and 20 mM *p*-NP palmitate) were prepared in
489 DMSO/acetonitrile (1:1 v/v). Unless stated otherwise, the enzymatic assay was performed
490 under standard conditions in a 1 ml reaction (50 mM potassium phosphate buffer pH 7.0,
491 0.3% (v/v) Triton X-100, 1 mM substrate) under agitation in a water bath at 50 °C until
492 complete substrate solubilization, and an appropriate volume of purified enzyme (0.25 µg)
493 was then added to start the reaction. The reaction mixture was incubated at 30 °C for 5 min
494 and then measured at 410 nm in a BioMate™ 3S spectrophotometer (Thermo Scientific, USA).
495 The incubation time for *p*-NP dodecanoate and *p*-NP palmitate was extended to 15 minutes.
496 All experiments were performed in triplicate, and a blank with denatured enzyme was
497 included. The concentration of product was calculated by linear regression equation given on
498 the standard curve performed by the reference Compound *p*-nithophenol (Sigma). One unit
499 of enzyme activity was defined as 1 µmol of p-nithrophenol produced per minute under the
500 assay conditions.

501 The optimal pH for enzyme activity was evaluated with *p*-NP butyrate by performing the
502 assay in different buffers, specifically 20 mM sodium acetate buffer (pH 4.0), sodium citrate
503 buffer (pH 5.5), potassium phosphate buffer (pH 7.0), and Tris-HCl (pH 8.0-9.0). The enzyme
504 reactions were stopped by adding 1 ml of cold stop solution (100 mM K-phosphate buffer pH
505 7.0, 10 mM EDTA) to neutralize the pH and avoid changes in the equilibrium between *p*-
506 nitrophenol and the deprotonated form *p*-nitrophenoxide, which would result in a decrease
507 in absorption at the applied wavelength of 410 nm (42). The optimal enzymatic temperature
508 was investigated with *p*-NP butyrate by performing the hydrolytic assay at different
509 temperatures under standard conditions (see above). To determine the denaturation
510 temperature, circular dichroism (CD) spectra were acquired between 190 and 270 nm with a
511 Jasco J-720 spectropolarimeter equipped with a Peltier temperature controller in a 0.1-mm
512 cell at 25 °C. The spectra were analyzed, and denaturation temperature (T_d) values were
513 determined at 220 nm between 10 and 85 °C at a rate of 30 °C per hour in 40 mM HEPES
514 buffer at pH 7.0. CD measurements were performed at pH 7.0 and not at the optimal pH (8.5-
515 9.0) to ensure protein stability. A protein concentration of 0.5 $mg \cdot mL^{-1}$ was used. T_d (and the
516 standard deviation of the linear fit) was calculated by fitting the ellipticity (mdeg) at 220 nm
517 at each of the different temperatures using a 5-parameter sigmoid fit with SigmaPlot 13.0.
518 Stability in organic solvents was assayed with *p*-NP butyrate under standard conditions in the
519 presence of 10-20-40-60% (v/v) of the water-miscible organic solvents ethanol, methanol,
520 isopropanol, acetonitrile, and DMSO. The effect of cations was investigated with *p*-NP
521 butyrate under standard conditions by the addition of $MgCl_2$, $CuCl_2$, $FeCl_3$, $CoCl_2$, $CaCl_2$, $MnCl_2$,
522 and $ZnSO_4$ at concentrations in the range of 1 to 10 mM. In all cases, the measured values
523 were then expressed as the relative activity in comparison to the control reaction performed
524 under standard conditions. Kinetic parameters were determined under standard conditions

525 over a range of substrate concentrations (0.05-10 mM) and calculated by nonlinear regression
526 analysis of raw data fit to the Michaelis–Menten function using GraphPad Prism software
527 (version 6.0).

528 The hydrolysis of esters other than *p*NP esters, including bis(2-hydroxyethyl)-terephthalate
529 (BHET), was assayed using a pH indicator assay in 384-well plates at 30 °C and pH 8.0 in a
530 Synergy HT Multi-Mode Microplate Reader in continuous mode at 550 nm over 24 h
531 (extinction coefficient of phenol red, 8450 $M^{-1}cm^{-1}$). The acid produced after ester bond
532 cleavage by the hydrolytic enzyme induced a color change in the pH indicator that was
533 measured at 550 nm. The experimental conditions were as detailed previously (43), with the
534 absence of activity defined as at least a twofold background signal. Briefly, the reaction
535 conditions for 384-well plates (ref. 781162, Greiner Bio-One GmbH, Kremsmünster, Austria)
536 were as follows: protein, 0.5 µg per well; ester, 20 mM; T: 30 °C; pH, 8.0 (5 mM 4-(2-
537 hydroxyethyl)-1-piperazinepropanesulfonic acid (EPPS) buffer, plus 0.45 Phenol Red®);
538 reaction volume, 40 µl. The reactions were performed in triplicate, and datasets were
539 collected from a Synergy HT Multi-Mode Microplate Reader with Gen5 2.00 software (Biotek).
540 One unit (U) of enzyme activity was defined as the amount of enzyme required to transform
541 1 µmol of substrate in 1 min under the assay conditions.

542 **Crystallization and X-ray structure determination of EH_0**. Initial crystallization conditions
543 were explored by high-throughput techniques with a NanoDrop robot (Innovadyne
544 Technologies) using 24 $mg \cdot mL^{-1}$ protein concentrations in HEPES (40 mM, pH 7, NaCl 50 mM),
545 protein reservoir ratios of 1:1, 1.5:1 and 2:1, and commercial screens Crystal Screen I and II,
546 SaltRx, Index (Hampton Research), JBScreen Classic, JBScreen JCSG, JBScreen PACT (Jena
547 Bioscience). Further optimizations were carried out, and bar-shaped crystals of EH_0 were
548 grown after one day of mixing 1.2 µL of a mixture of protein (1 µL, 24 $mg \cdot mL^{-1}$) and seeds (0.2

549 µL, 1:100) with guanidine hydrochloride (0.2 µL, 0.1 M) and reservoir (0.5 µL, 11%
550 polyethylene glycol 8000, 100 mM Bis-Tris pH 5.5, 100 mM ammonium acetate). For data
551 collection, crystals were transferred to cryoprotectant solution consisting of mother liquour
552 and glycerol (20% (v/v)) before being cooled in liquid nitrogen. Diffraction data were collected
553 using synchrotron radiation on the XALOC beamline at ALBA (Cerdanyola del Vallés, Spain).
554 Diffraction images were processed with XDS (44) and merged using AIMLESS from the CCP4
555 package (45). The crystal was indexed in the $P2_12_12_1$ space group, with two molecules in the
556 asymmetric unit and 62% solvent content within the unit cell. The data collection statistics
557 are given in Table S4. The structure of EH_0 was solved by molecular replacement with MOLREP
558 (46) using the coordinates from Est8 as a template (PDB Code 4YPV). Crystallographic
559 refinement was performed using the program REFMAC (47) within the CCP4 suite, with NCS
560 (noncrystallography symmetry) medium restraints and excluding residues 17-36. The free *R*-
561 factor was calculated using a subset of 5% randomly selected structure-factor amplitudes that
562 were excluded from automated refinement. Subsequently, heteroatoms were manually built
563 into the electron density maps with Coot8 (48), and water molecules were included in the
564 model, which, combined with more rounds of restrained refinement, reached the *R* factors
565 listed in Table S4. The figures were generated with PyMOL. The crystallographic statistics of
566 EH_0 are listed in Table S4. To extract dynamical details from the X-ray data, the coordinates
567 of EH_0 were first refined using PHENIX (49) and then were used as input models for a time-
568 averaged molecular dynamics refinement as implemented in the Phenix.ensemble-
569 refinement routine, which was performed as described previously (49).

570 **Codes and accession numbers.** The sequence encoding EH_0 was deposited in GenBank
571 with the accession number MK791218. The atomic coordinates and structure factors for the
572 EH_0 structure have been deposited in the RCSB Protein Data Bank with accession codes 7ZR3.

SUPPLEMENTARY MATERIAL

Supplemental material is available online only.

Table S1

Table S2

Table S3

Table S4

Table S5

Figure S1

ACKNOWLEDGMENTS

This study was conducted under the auspices of the FuturEnzyme Project funded by the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 101000327. We also acknowledge financial support under Grants PID2020-112758RB-I00 (M.F.), PDC2021-121534-I00 (M.F.), and PID2019-105838RB-C33 (J.S-A.) from the Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación (AEI) (Digital Object Identifier 10.13039/501100011033), Fondo Europeo de Desarrollo Regional (FEDER) and the European Union ("NextGenerationEU/PRTR"), and Grant 2020AEP061 (M.F.) from the Agencia Estatal CSIC. C. Coscolín thanks the Ministerio de Economía y Competitividad and FEDER for a PhD fellowship (Grant BES-2015-073829). P.N.G. acknowledges the Sêr Cymru programme partly funded by ERDF through the Welsh Government for the support of the project BioPOL4Life, the project 'Plastic Vectors' funded by the Natural Environment Research Council UK (NERC), Grant No. NE/S004548/N and the Centre for Environmental Biotechnology Project co-funded by the European Regional Development Fund (ERDF) through the Welsh Government. The authors acknowledge David Almendral and Ruth Matesanz for supporting the circular dichroism analysis and Jose L. Gonzalez-Alfonso and Francisco J. Plou for supporting the HPLC analysis. We thank the staff of the Synchrotron Radiation Source at Alba (Barcelona, Spain) for assistance at the BL13-XALOC beamline.

We declare no conflict of interest.

600

601 P.N.G., M.F. and J.S-A. developed the conceptual framework, M.D., C.C., T.N.C. and H.T. developed the
602 experimental design, I.C-R. and J.S-A. performed the crystallization and X-ray structure determinations, analysis
603 and interpretation, M.D., M.F. and P.N.G. interpreted the experimental data, P.N.G., M.F. and J.S-A. wrote the
604 initial manuscript. All authors edited the manuscript. All authors read and approved the final manuscript.

605

606 **References**

607

608 1. Intasian P, Prakinee K, Phintha A, Trisrivirat D, Weeranoppanant N, Wongnate T, Chaiyen P. 2021.
609 Enzymes, in vivo biocatalysis, and metabolic engineering for enabling a circular economy and
610 sustainability. Chem Rev 121:10367–10451.

611 2. Ferrer M, Méndez-García C, Bargiela R, Chow J, Alonso S, García-Moyano A, Bjerga GEK, Steen IH,
612 Schwabe T, Blom C, Vester J, Weckbecker A, Shahgaldian P, De Carvalho CCCR, Meskys R, Zanaroli G,
613 Glöckner FO, Fernández-Guerra A, Thambisetty S, De La Calle F, Golyshina O V., Yakimov MM, Jaeger KE,
614 Yakunin AF, Streit WR, McMeel O, Calewaert JB, Tonné N, Golyshin PN. 2019. Decoding the ocean's
615 microbiological secrets for marine enzyme biodiscovery. FEMS Microbiol Lett 366:1–7.

616 3. Schloss PD, Handelsman J. 2006. Toward a census of bacteria in soil. PLoS Comput Biol 2:0786–0793.

617 4. Handelsman J. 2005. Metagenomics: Application of genomics to uncultured microorganisms. Microbiol
618 Mol Biol Rev 69:195–195.

619 5. Barzkar N, Sohail M, Tamadoni Jahromi S, Gozari M, Poormozaffar S, Nahavandi R, Hafezieh M. 2021.
620 Marine Bacterial Esterases: Emerging biocatalysts for industrial applications. Appl Biochem Biotechnol
621 193:1187–1214.

622 6. Martínez-Martínez M, Alcaide M, Tchigvintsev A, Reva O, Polaina J, Bargiela R, Guazzaroni ME, Chicote
623 Á, Canet A, Valero F, Eguizabal ER, Guerrero M del C, Yakunin AF, Ferrer M. 2013. Biochemical diversity
624 of carboxyl esterases and lipases from lake arreo (Spain): A metagenomic approach. Appl Environ
625 Microbiol 79:3553–3562.

626 7. Jaeger KE, Eggert T. 2002. Lipases for biotechnology. Curr Opin Biotechnol 13:390–397.

627 8. Jochens H, Hesseler M, Stiba K, Padhi SK, Kazlauskas RJ, Bornscheuer UT. 2011. Protein Engineering of

628 α/β-Hydrolase Fold Enzymes. ChemBioChem 12:1508–1517.
629 9. De Godoy Daiha K, Angeli R, De Oliveira SD, Almeida RV. 2015. Are lipases still important biocatalysts? A
630 study of scientific publications and patents for technological forecasting. PLoS One 10:1–20.
631 10. Bauer TL, Buchholz PCF, Pleiss J. 2020. The modular structure of α/β-hydrolases. FEBS J 287:1035–1053.
632 11. Ferrer M, Martínez-Martínez M, Bargiela R, Streit WR, Golyshina O V., Golyshin PN. 2016. Estimating the
633 success of enzyme bioprospecting through metagenomics: Current status and future trends. Microb
634 Biotechnol 9:22–34.
635 12. Placido A, Hai T, Ferrer M, Chernikova TN, Distaso M, Armstrong D, Yakunin AF, Toshchakov S V., Yakimov
636 MM, Kublanov I V., Golyshina O V., Pesole G, Ceci LR, Golyshin PN. 2015. Diversity of hydrolases from
637 hydrothermal vent sediments of the Levante Bay, Vulcano Island (Aeolian archipelago) identified by
638 activity-based metagenomics and biochemical characterization of new esterases and an
639 arabinopyranosidase. Appl Microbiol Biotechnol 99:10031–10046.
640 13. Tchigvintsev A, Tran H, Popovic A, Kovacic F, Brown G, Flick R, Hajighasemi M, Egorova O, Somody JC,
641 Tchigvintsev D, Khusnutdinova A, Chernikova TN, Golyshina O V., Yakimov MM, Savchenko A, Golyshin
642 PN, Jaeger KE, Yakunin AF. 2015. The environment shapes microbial enzymes: five cold-active and salt-
643 resistant carboxylesterases from marine metagenomes. Appl Microbiol Biotechnol 99:2165–2178.
644 14. Lopez-Lopez O, Cerdan ME, Gonzalez Siso MI. 2014. New Extremophilic Lipases and Esterases from
645 Metagenomics. Curr Protein Pept Sci 15:445–455.
646 15. Lee MH, Hong KS, Malhotra S, Park JH, Hwang EC, Choi HK, Kim YS, Tao W, Lee SW. 2010. A new esterase
647 EstD2 isolated from plant rhizosphere soil metagenome. Appl Microbiol Biotechnol 88:1125–1134.
648 16. Faoro H, Glogauer A, Souza EM, Rigo LU, Cruz LM, Monteiro RA, Pedrosa FO. 2011. Identification of a
649 new lipase family in the Brazilian Atlantic Forest soil metagenome. Environ Microbiol Rep 3:750–755.
650 17. Hult K, Berglund P. 2007. Enzyme promiscuity: mechanism and applications. Trends Biotechnol 25:231–
651 238.
652 18. Alcaide M, Tornés J, Stogios PJ, Xu X, Gertler C, Leo RDI, Bargiela R, Lafraya Á, Guazzaroni ME, López-
653 Cortés N, Chernikova TN, Golyshina O V., Nechitaylo TY, Plumeier I, Pieper DH, Yakimov MM, Savchenko
654 A, Golyshin PN, Ferrer M. 2013. Single residues dictate the co-evolution of dual esterases: MCP
655 hydrolases from the α/β hydrolase family. Biochem J 454:157–166.
656 19. Martínez-Martínez M, Coscolín C, Santiago G, Chow J, Stogios PJ, Bargiela R, Gertler C, Navarro-

657 Fernández J, Bollinger A, Thies S, Méndez-García C, Popovic A, Brown G, Chernikova TN, García-Moyano
658 A, Bjerga GEK, Pérez-García P, Hai T, Del Pozo M V., Stokke R, Steen IH, Cui H, Xu X, Nocek BP, Alcaide M,
659 Distaso M, Mesa V, Peláez AI, Sánchez J, Buchholz PCF, Pleiss J, Fernández-Guerra A, Glöckner FO,
660 Golyshina O V., Yakimov MM, Savchenko A, Jaeger KE, Yakunin AF, Streit WR, Golyshin PN, Guallar V,
661 Ferrer M. 2018. Determinants and prediction of esterase substrate promiscuity patterns. ACS Chem Biol
662 13:225–234.
663 20. Höppner A, Bollinger A, Kobus S, Thies S, Coscolín C, Ferrer M, Jaeger KE, Smits SHJ. 2021. Crystal
664 structures of a novel family IV esterase in free and substrate-bound form. FEBS J 288:3570–3584.
665 21. Kumar PP, Paramashivappa R, Vithayathil PJ, Rao PVS, Rao AS. 2002. Process for isolation of cardanol
666 from technical cashew (Anacardium occidentale l.) Nut shell liquid. J Agric Food Chem 50:4705–4708.
667 22. Reyes-Duarte D, Ferrer M, García-Arellano H. 2012. Functional-based screening methods for lipases,
668 esterases, and phospholipases in metagenomic libraries. Lipases and phospholipases 861:101–113.
669 23. Arpigny JL, Jaeger KE. 1999. Bacterial lipolytic enzymes: Classification and properties. Biochem J
670 343:177–183.
671 24. Yoshida S, Hiraga K, Takehana T, Taniguchi I, Yamaji H, Maeda Y, Toyohara K, Miyamoto K, Kimura Y, Oda
672 K. 2016. A bacterium that degrades and assimilates poly(ethylene terephthalate). Sci (American Assoc
673 Adv Sci 351:1196–1199.
674 25. Bollinger A, Thies S, Knieps-Grünhagen E, Gertzen C, Kobus S, Höppner A, Ferrer M, Gohlke H, Smits SHJ,
675 Jaeger KE. 2020. A novel polyester hydrolase from the marine bacterium *Pseudomonas aestusnigri* –
676 Structural and functional insights. Front Microbiol 11:1–16.
677 26. Pereira MR, Maester TC, Mercaldi GF, de Macedo Lemos EG, Hyvönen M, Balan A. 2017. From a
678 metagenomic source to a high-resolution structure of a novel alkaline esterase. Appl Microbiol
679 Biotechnol 101:4935–4949.
680 27. Holm L. 2020. Using Dali for protein structure comparison. Struct Bioinforma 2112:29–42.
681 28. Ding Y, Nie L, Yang XC, Li Y, Huo YY, Li Z, Gao Y, Cui HL, Li J, Xu XW. 2022. Mechanism and structural
682 insights into a novel esterase, E53, isolated from *Erythrobacter longus*. Front Microbiol 12:1–12.
683 29. Palm GJ, Fernández-Álvaro E, Bogdanović X, Bartsch S, Sczodrok J, Singh RK, Böttcher D, Atomi H,
684 Bornscheuer UT, Hinrichs W. 2011. The crystal structure of an esterase from the hyperthermophilic
685 microorganism *Pyrobaculum calidifontis* VA1 explains its enantioselectivity. Appl Microbiol Biotechnol

686 91:1061–1072.
687 30. De Simone G, Menchise V, Alterio V, Mandrich L, Rossi M, Manco G, Pedone C. 2004. The crystal structure
688 of an EST2 mutant unveils structural insights on the H group of the carboxylesterase/lipase family. J Mol
689 Biol 343:137–146.
690 31. Alonso S, Santiago G, Cea-Rama I, Fernandez-Lopez L, Coscolín C, Modregger J, Ressmann AK, Martínez-
691 Martínez M, Marrero H, Bargiela R, Pita M, Gonzalez-Alfonso JL, Briand ML, Rojo D, Barbas C, Plou FJ,
692 Golyshin PN, Shahgaldian P, Sanz-Aparicio J, Guallar V, Ferrer M. 2020. Genetically engineered proteins
693 with two active sites for enhanced biocatalysis and synergistic chemo- and biocatalysis. Nat Catal 3:319–
694 328.
695 32. Bais HP, Weir TL, Perry LG, Gilroy S, Vivanco JM. 2006. The role of root exudates in rhizosphere
696 interactions with plants and other organisms. Annu Rev Plant Biol 57:233–266.
697 33. Bertin C, Yang X, Weston LA. 2003. The role of root exudates and allelochemicals in the rhizosphere.
698 Plant Soil 256:67–83.
699 34. Hitch TCA, Clavel T. 2019. A proposed update for the classification and description of bacterial lipolytic
700 enzymes. PeerJ 2019.
701 35. Sellek GA, Chaudhuri JB. 1999. Biocatalysis in organic media using enzymes from extremophiles. Enzyme
702 Microb Technol 25:471–482.
703 36. Zaks A, Klibanov AM. 1988. Enzymatic catalysis in nonaqueous solvents. J Biol Chem 263:3194–3201.
704 37. Sharma S, Kanwar SS. 2014. Organic solvent tolerant lipases and applications. Sci World J 2014.
705 38. Lukashin A V., Borodovsky M. 1998. GeneMark.hmm: New solutions for gene finding. Nucleic Acids Res
706 26:1107–1115.
707 39. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and
708 PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389–3402.
709 40. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein
710 modeling, prediction and analysis. Nat Protoc 10:845–858.
711 41. Gupta N, Rathi P, Gupta R. 2002. Simplified para-nitrophenyl palmitate assay for lipases and esterases.
712 Anal Biochem 311:98–99.
713 42. Bowers GN, McComb RB, Christensen RG, Schaffer R. 1980. High-purity 4-nitrophenol: purification,
714 characterization, and specifications for use as a spectrophotometric reference material. Clin Chem

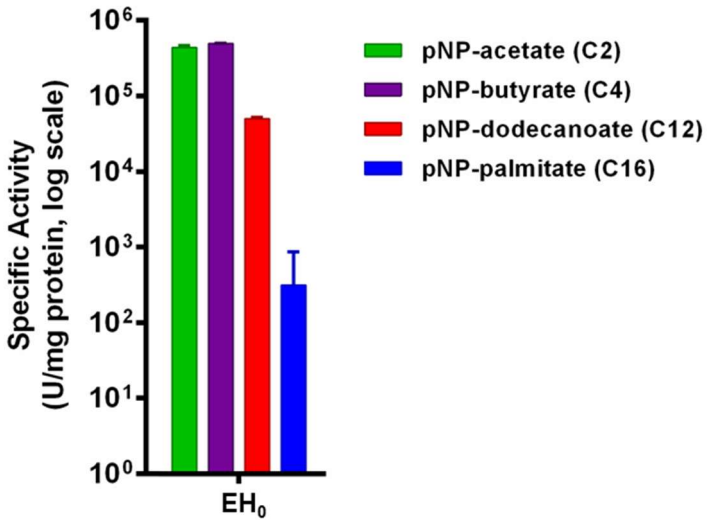
715 26:724–729.
716 43. Giunta CI, Cea-Rama I, Alonso S, Briand ML, Bargiela R, Coscolín C, Corvini PFX, Ferrer M, Sanz-Aparicio
717 J, Shahgaldian P. 2020. Tuning the properties of natural promiscuous enzymes by engineering their nano-
718 environment. ACS Nano 14:17652–17664.
719 44. Kabsch W. 2010. research papers XDS research papers. Acta Crystallogr Sect D Biol Crystallogr 66:125–
720 132.
721 45. Evans PR, Murshudov GN. 2013. How good are my data and what is the resolution? Acta Crystallogr Sect
722 D Biol Crystallogr 69:1204–1214.
723 46. Vagin A, Teplyakov A. 1997. MOLREP: an automated program for molecular replacement. J Appl
724 Crystallogr 30:1022–1025.
725 47. Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-
726 likelihood method. Acta Crystallogr Sect D, Biol Crystallogr 53:240–255.
727 48. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of coot. Acta Crystallogr
728 Sect D Biol Crystallogr 66:486–501.
729 49. Tom Burnley B, Afonine P V., Adams PD, Gros P. 2012. Modelling dynamics in protein crystal structures
730 by ensemble refinement. Elife 2012:1–29.
731

732 TABLE 1. Kinetic parameters of EH_0 against *p*NP-esters. The results are the mean ± SD of triplicates.

Substrate	K_M (mM)	V_{max} (μM min^{-1})	k_{cat} (min^{-1})	k_{cat}/K_M (s^{-1} mM^{-1})
*p*N-acetate	1.61 ± 0.12	291.50 ± 7.62	41.64 ± 1.09	0.430
*p*NP-butyrate	1.08 ± 0.18	252.60 ± 13.84	36.09 ± 1.98	0.558
*p*NP-dodecanoate	1.00 ± 0.28	25.56 ± 1.80	3.65 ± 0.26	0.061

733

734



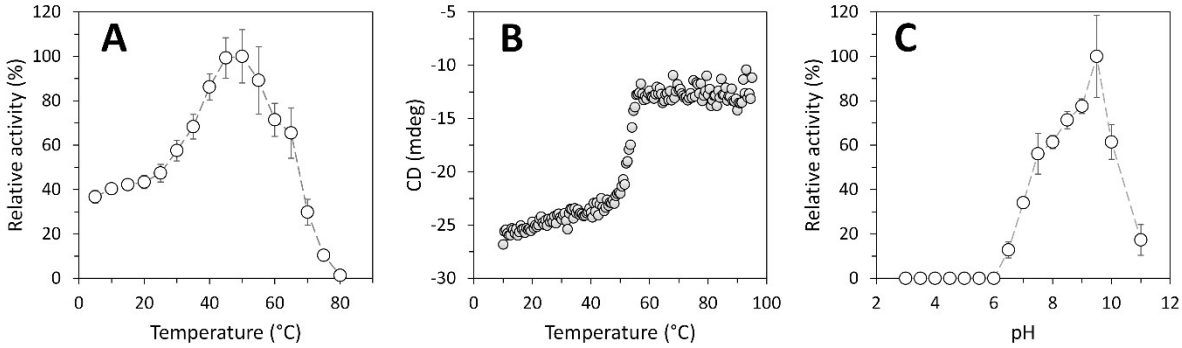
735

736 FIG 1. Substrate selectivity of EH_0 against a set of Model *p*-nitrophenyl esters of different acyl chain lengths.

737 Specific activities (mean ± SD of triplicates) are shown on a log_{10} scale. Reactions contained 1 mM of the

738 corresponding *p*NP-ester and were conducted under standard conditions described in Materials and Methods.

739



740

741 FIG 2. Optimal parameters for the activity and stability of purified EH_0. (A) Temperature profile. (B) The thermal
742 denaturation curve of EH_0 at pH 7.0 was measured by ellipticity changes at 220 nm and obtained at different
743 temperatures. (C) The pH profile. The maximal activity was defined as 100%, and the relative activity is shown as
744 the percentage of maximal activity (mean ± SD of triplicates) determined under standard reaction conditions
745 with *p*NP butyrate as the substrate. Graphics were created with SigmaPlot version 14.0 (the data were not fitted
746 to any model).

747

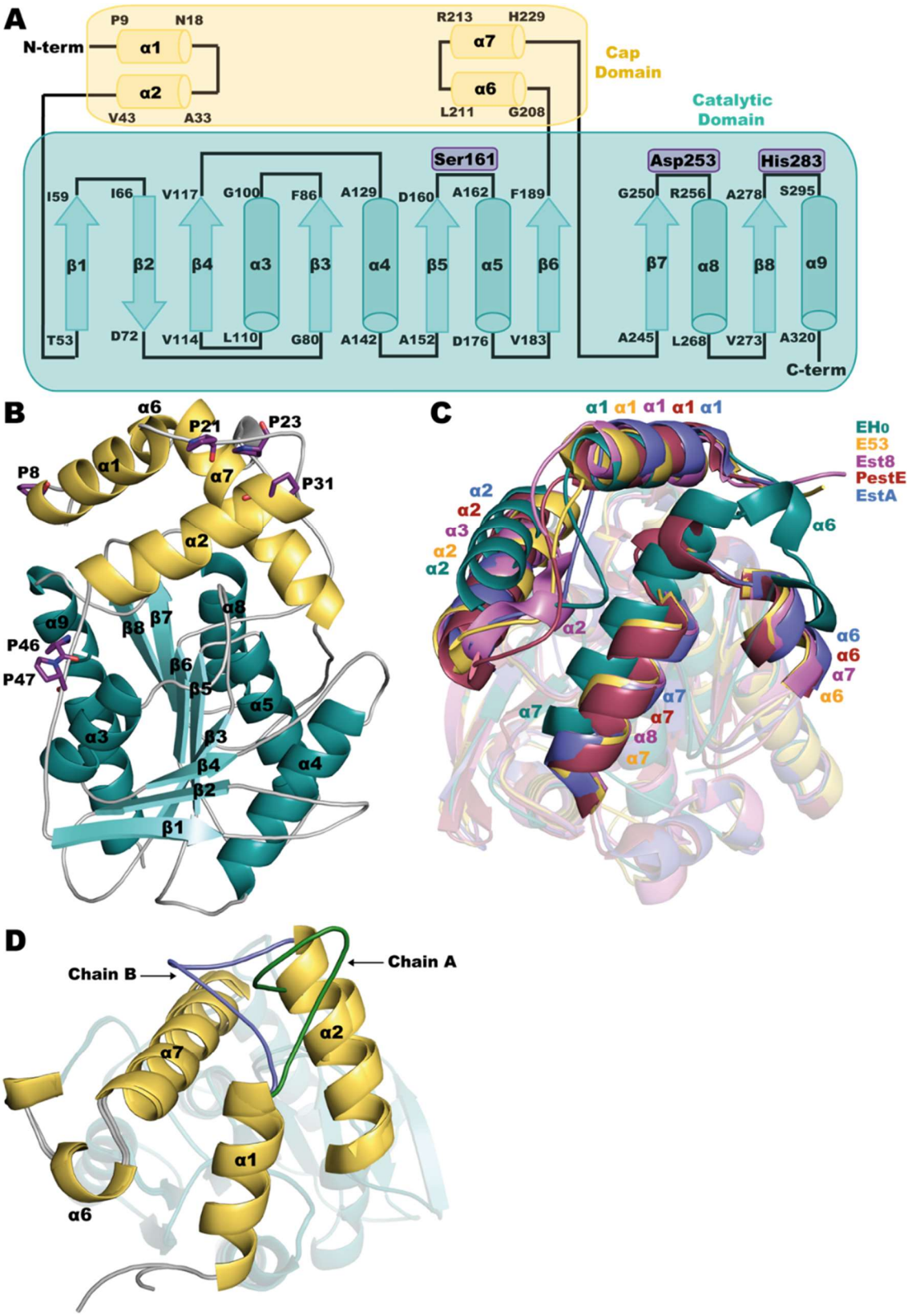
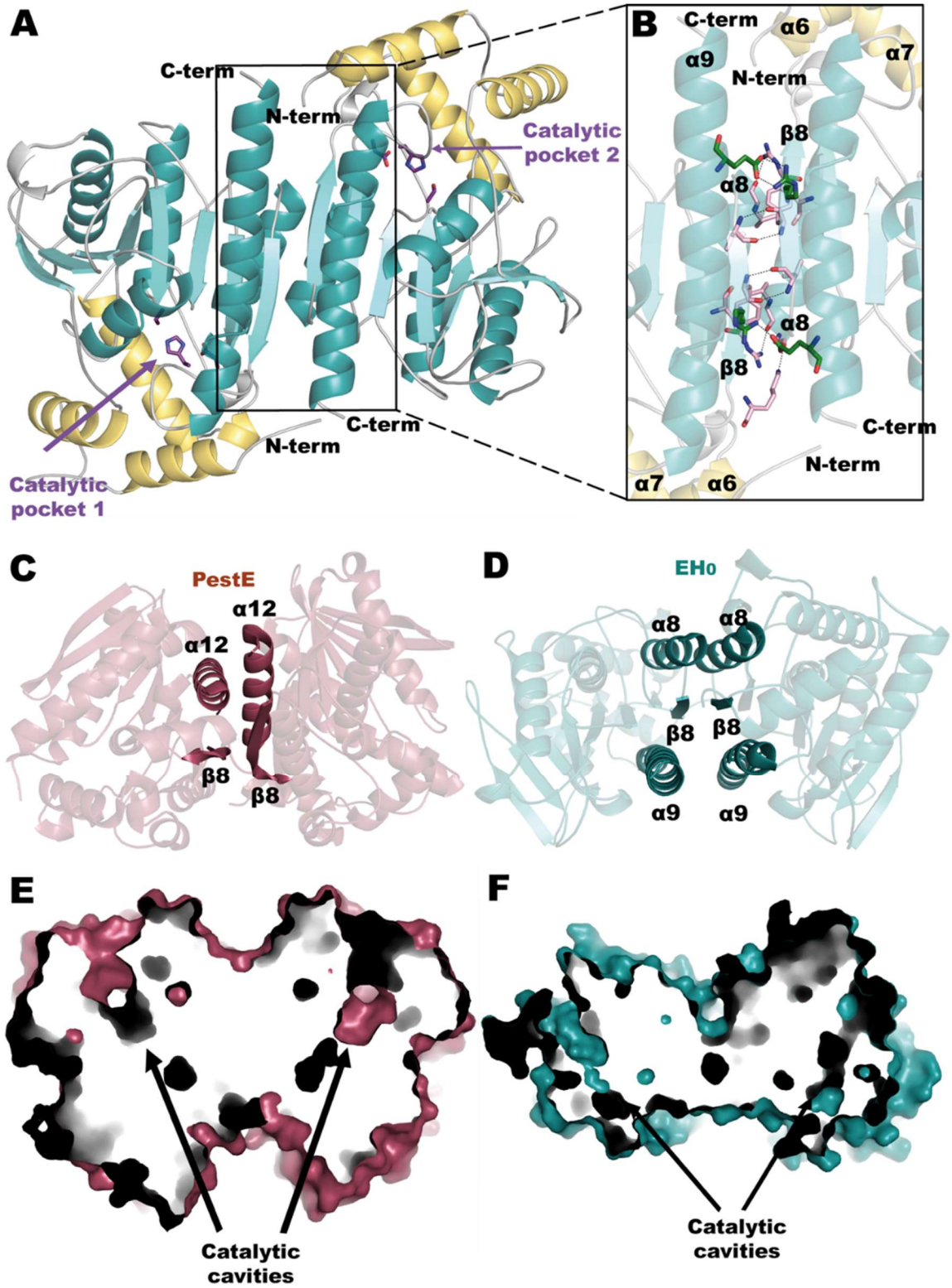


FIG 3. A) Topology diagram of EH_0. The catalytic triad is highlighted in violet. B) EH_0 folding, with the cap domain in yellow and the catalytic domain in teal. The prolines within the N-terminal cap domain are depicted as violet sticks. C) Superposition of EH_0 and its homologs: E53 (yellow, PDB code 7W8N), Est8 (violet, PDB code 4YPV),

752 PestE (raspberry, PDB code 3ZWQ) and EstA (slate, PDB code 5LK6). D) Superposition of subunits A and B showing
753 different conformations of loop α1-α2, with chain A in green and chain B in blue. Cap/catalytic domains follow
754 the color code described in B.

755



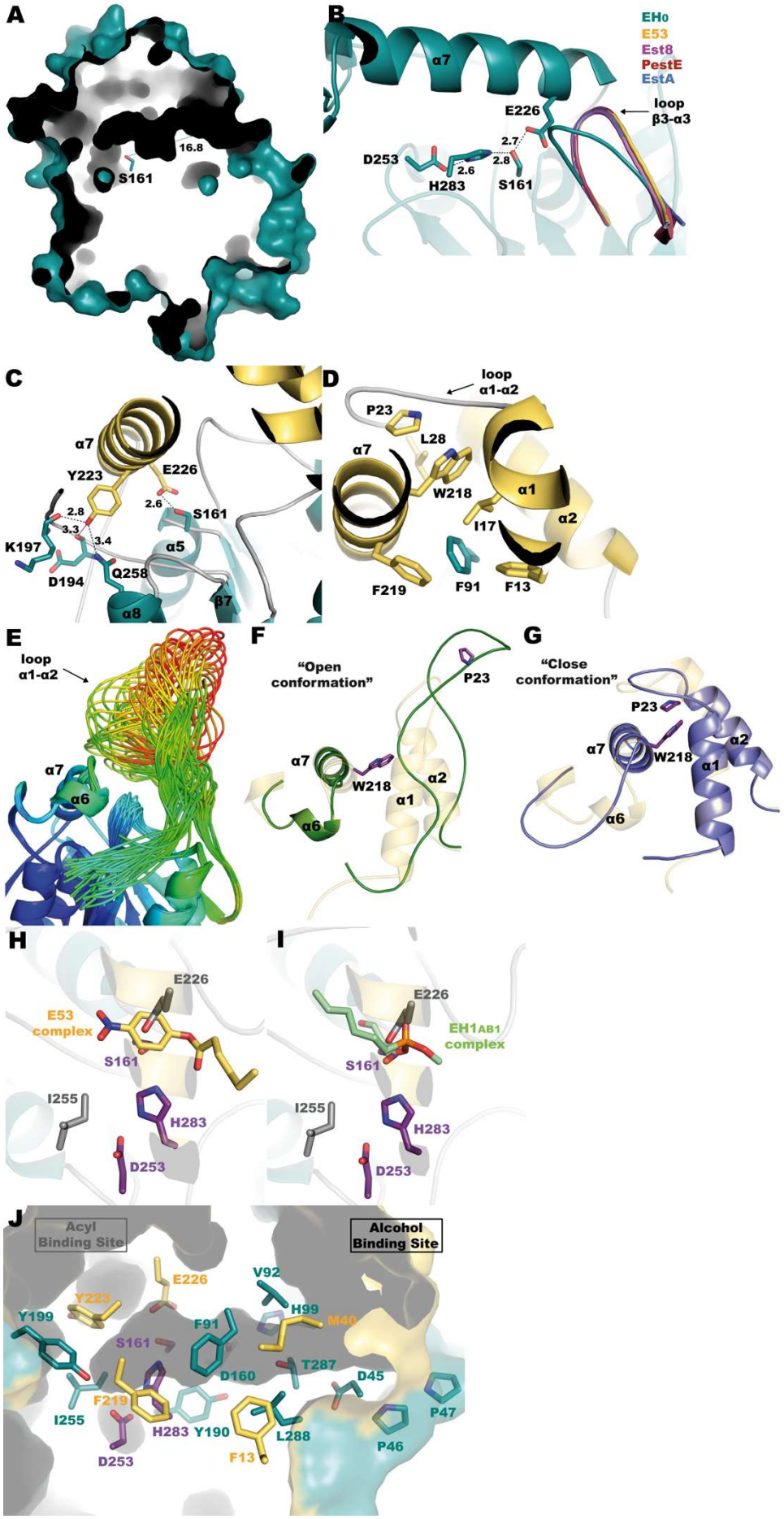
756

757

758 FIG 4. A) EH_0 homodimer showing the catalytic triad as sticks (violet). Both subunits are related by a twofold axis

759 that is perpendicular to the view shown and situates the two active sites on opposite faces of the dimer. B) A

760 magnified image of the dimeric interface that covers 722.9$Å^2$ (approximately 5.6% of the total surface), showing
761 hydrogen bonds (pink) and salt bridges (green). The Cap domain is colored yellow, whereas the catalytic domain
762 is shown in teal. C) Dimeric interface of the PestE homolog where the interactions are made by β8 and α12. D)
763 Dimeric interface of EH_0, where the interactions are made by α8, β8 and α9. Disposition of catalytic tunnels from
764 both subunits in PestE (E) and EH_0 (F).
765


A
16.8
S161
B
EH0
E53
Est8
PestE
EstA
α7
E226
loop
β3-α3
2.7
2.8
2.6
D253
H283
S161
C
α7
Y223
E226
2.6
S161
2.8
3.3
3.4
K197
α5
D194
Q258
α8
β7
D
loop
α1-α2
P23
L28
α7
W218
α1
α2
I17
F219
F91
F13
E
loop
α1-α2
α7
α6
F
"Open
conformation"
P23
α7
W218
α2
α1
α6
G
"Close
conformation"
P23
α7
W218
α2
α1
α6
H
E226
E53
complex
S161
I255
H283
D253
I
E226
EH1AB1
complex
S161
I255
H283
D253
J
Acyl
Binding Site
Alcohol
Binding Site
E226
V92
Y223
Y199
H99
M40
F91
S161
T287
D160
D45
F219
I255
H283
Y190
L288
P47
P46
D253
F13

766

767 FIG 5. A) A tunnel 16.8 deep giving access to the catalytic triad. B) Hydrogen bond network among the residues
768 making the catalytic triad of EH_0 (teal). Movement of loop β3-α3 in EH_0 (housing the oxoanion) with respect to
769 its homologs E53 complexed with (4-nitrophenyl)hexanoate (yellow), Est8 (violet), PestE (raspberry) and EstA
770 (slate). C) Hydrogen bond network of residues Glu226 and Tyr223 from α7. D) Hydrophobic patch where Trp218
771 is located. The EH_0 cap domain is shown in yellow and the catalytic domain in teal (C and D.)E) All possible
772 conformations of the region comprising α1-α2 shown by the ensemble refinement. Color code gives chain
773 mobility, from low (blue) to high (red). The opposed scenarios of the region α1-α2, from an "open conformation"
774 (F, green) to a "closed conformation" (G, slate). The cap domain seen in the crystal structure is shown in yellow.
775 H and I) Superimposition of EH_0 onto E53 complexed with 4-nitrophenyl)hexanoate (yellow) and EH1AB1 in
776 complex with a derivative of methyl 4-nitrophenylhexylphosphonate (green), showing only the ligands. The
777 catalytic triad of EH_0 is shown as violet sticks, and residues Glu226 and Ile255 are shown as gray sticks. The EH_0
778 cap domain is shown in yellow and the catalytic domain in teal. J) Active site cavity of EH_0 with acyl and alcohol
779 moieties. Residues from the cap domain are colored yellow, and residues from the catalytic domain are colored
780 teal. The catalytic triad is shown in violet.
781

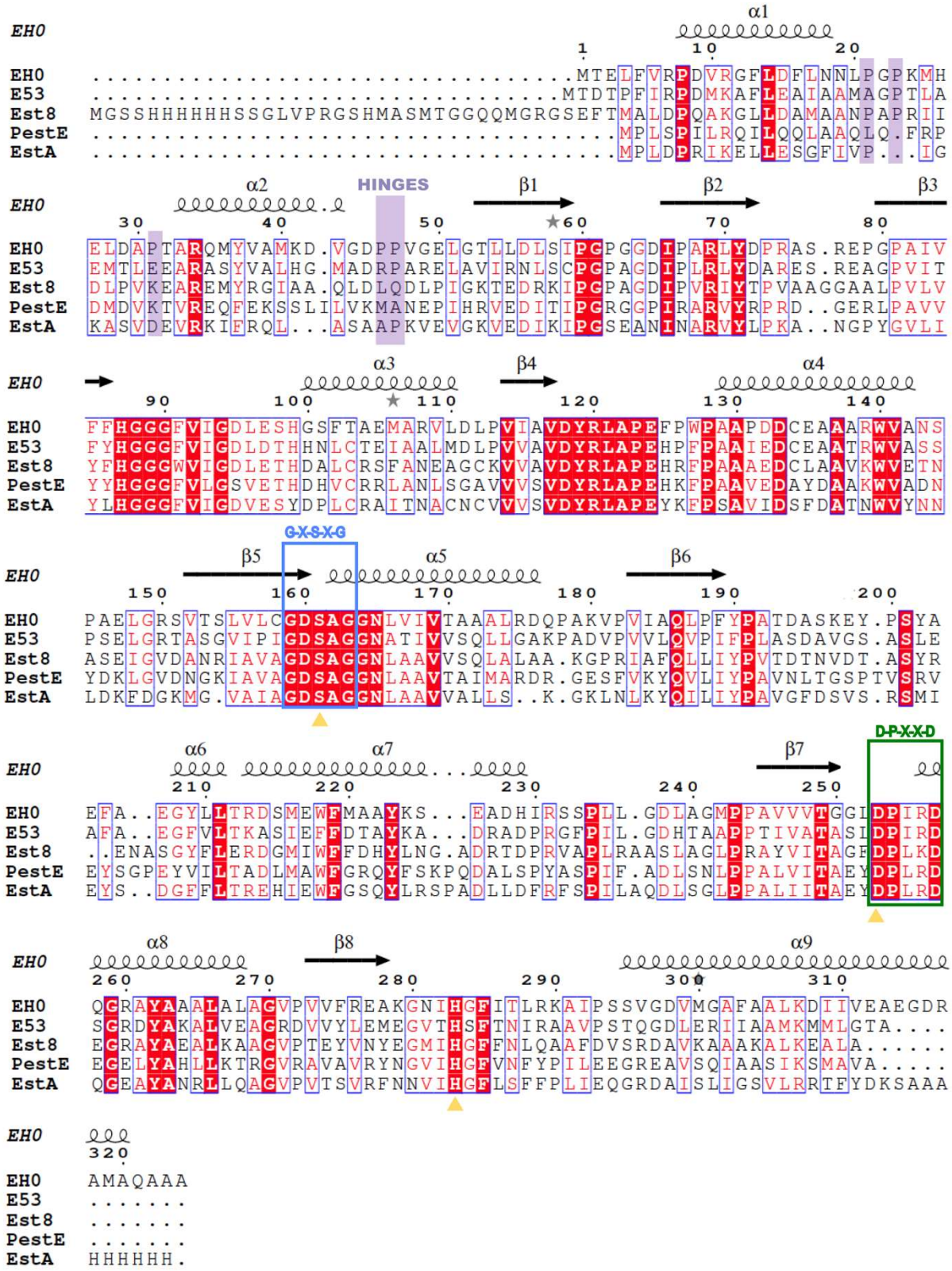


FIG 6. A) Multiple sequence alignment of EH_0 and its homologs E53, Est8, PestE and EstA. Highlighted in boxes are the conserved motif G-X-S-X-G (in blue), including the catalytic serine, and the D-P-X-X-D (in green), including

785 the catalytic aspartic acid. The prolines from loop α1-α2 (Pro21, Pro23 and Pro31) and the potential hinges (Pro46

786 and Pro47) are underlined in purple. The catalytic triad residues are shown with yellow triangles.