



*Horizon 2020 Work programme*

Food Security, Sustainable Agriculture and Forestry, Marine, Maritime and Inland Water Research and the Bioeconomy

*Call*

H2020-FNR-2020: Food and Natural Resources

*Topic name*

FNR-16-2020: ENZYMES FOR MORE ENVIRONMENT-FRIENDLY CONSUMER PRODUCTS

*FuturEnzyme:*

Technologies of the Future for Low-Cost Enzymes for Environment-Friendly Products

Final ID: 101000327

21/11/2023



# DELIVERABLE D4.5: AT LEAST 9 ENZYME CRYSTAL STRUCTURES

D4.5

MANUEL FERRER  
CSIC  
Marie Curie 2, MADRID 28049, Spain

## Document information sheet

<b>Work package:</b>	WP4, Small-scale enzyme production and characterisation.
<b>Authors:</b>	Manuel Ferrer (CSIC)
<b>Document version:</b>	1
<b>Date:</b>	20/11/2023
<b>WP starting date:</b>	01/06/2021
<b>WP duration:</b>	40 months
<b>WP lead:</b>	CSIC
<b>WP participant(s):</b>	FHNW, CSIC, BSC, Bangor, UHAM, UDUS, Inofea AG, Biosynth
<b>Deliv lead beneficiary:</b>	CSIC
<b>Dissemination Level:</b>	Confidential, only for members of the consortium (including the Commission Services)
<b>Type:</b>	Other
<b>Due date (months):</b>	31
<b>Contact details:</b>	Manuel Ferrer (mferrer@icp.csic.es)

## Summary

At least 9 enzyme crystal structures.....	3
1. Scope of Deliverable.....	3
2. Introduction.....	3
3. Summary of Enzyme Structures and Models.....	3
4. Comprehensive Structural Analysis of Enzymes in FuturEnzyme: Experimental Data .....	4
4.1. PluriZymes .....	5
4.2. Biomimetic catalysts .....	10
4.3. Native Enzymes.....	11
5. Comprehensive Structural Analysis of Enzymes in FuturEnzyme: AlphaFold-Derived Data .....	20
6. Other Resolved Structures within the Framework of the Project .....	22
7. Conclusions.....	24
8. References .....	24
9. Annex.....	25

## At least 9 enzyme crystal structures

### 1. Scope of Deliverable

As stated in the original proposal, this deliverable consists in a set of crystal structures of best enzyme prototypes. This deliverable will be accompanied by a report detailing the structures and PDB codes and their features and characteristics.

### 2. Introduction

In the realm of biotechnological advancements, the elucidation of enzyme structures stands as a cornerstone, particularly in projects like FuturEnzyme, where the focus is on harnessing the potential of enzymes for innovative applications. The significance of obtaining detailed structural information of the enzymes under study cannot be overstated, not only for fundamental structure-function studies, within the WP4, but also as a foundation for future endeavors in genetic and supramolecular engineering (WP5). Indeed, engineering relies heavily on precise structural knowledge to design novel variants with superior catalytic properties.

Recognizing the paramount importance of these aspects, a dedicated effort has been made within the FuturEnzyme project to resolve the structures of enzymes that have emerged as promising candidates during the research phases. This deliverable, titled "At least 9 enzyme crystal structures," comprises a comprehensive set of crystal structures of the best enzyme prototypes identified in our research. Accompanying this deliverable is a detailed report that includes the Protein Data Bank (PDB) codes, the methodology used and brief outcomes.

In addition to our experimental endeavors to elucidate protein structures, we have also embraced the groundbreaking capabilities of AlphaFold in the FuturEnzyme project. This innovative computational tool has revolutionized the field of structural biology by predicting protein structures with remarkable accuracy. The integration of AlphaFold into our workflow is pivotal, considering the impracticality of experimentally resolving the structure of every enzyme of interest. The AlphaFold's predictive power not only complements our experimental efforts but also accelerates the pace at which we can gather crucial structural information. The ability to predict enzyme structures reliably opens up new avenues for exploring enzyme functionality, especially for those proteins where experimental structure determination remains elusive. By leveraging AlphaFold's advanced algorithms, we are able to expand our structural database significantly, enhancing our understanding of enzyme mechanisms and facilitating more informed strategies for enzyme modification. This integration of cutting-edge computational methods with traditional experimental techniques represents a holistic approach to enzyme research, ensuring that the FuturEnzyme project remains at the forefront of innovation in enzyme structural analysis.

Finally, in addition to generating our own structural data in this project through the structural resolution of proteins or AlphaFold models, we have also utilized resolved protein structures from existing literature. These structures have served as a foundation for the development and design of enhanced mutants, as well as for the application of the project's computational tools.

### 3. Summary of Enzyme Structures and Models

To date the following atomic coordinates, generated in the frame of the FuturEnzyme project, have been deposited in the Protein Data Bank:

1. 8BBI (<https://www.ncbi.nlm.nih.gov/Structure/pdb/8BBI>)  
PluriZyme Xyn11-mut1 from *Pseudothermotoga thermarum*
2. 7ZR3 (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7ZR3>)  
Ester-hydrolase EH<sub>0</sub> (Distaso et al. 2023)
3. 7PP3 (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7PP3>)  
Ester-hydrolase EH<sub>7</sub> (Cea-Rama et al. 2022)

4. 7PP8 (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7PP8>)  
Ester-hydrolase EH<sub>7</sub> with methyl 4-nitrophenyl hexylphosphonate (Cea-Rama et al. 2022)
5. 7PU6 (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7PU6>)  
Ester-hydrolase EH<sub>7</sub> complexed with 4-nitrophenyl hexylphosphonate (Cea-Rama et al. 2022)
6. 7QYG (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7QYG>)  
Transaminase TR<sub>2</sub> (Roda et al. 2022)
7. 7QYF (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7QYF>)  
PluriZyme variant TR<sub>2</sub>E<sub>2</sub> (Roda et al. 2022)
8. 7QX3 (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7QX3>)  
PluriZyme variant TR<sub>2</sub>E<sub>2</sub> complexed with EOS (Roda et al. 2022)
9. 7QX0 (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7QX0>)  
PluriZyme variant TR<sub>2</sub>E<sub>2</sub> complexed with PLP (Roda et al. 2022)
10. 8B4U (<https://www.ncbi.nlm.nih.gov/Structure/pdb/8B4U>)  
PET46, PETase enzyme from *Candidatus bathyarchaeota* (not published)
11. 8A2C (<https://www.ncbi.nlm.nih.gov/Structure/pdb/8A2C>)  
PET40 (S178A mutant), PETase enzyme from an unclassified *Amycolatopsis*
12. 7SPN (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7SPN>)  
Ester-hydrolase IS11, thermophilic (Distaso et al. 2023b)
13. 8PCT (<https://www.rcsb.org/structure/8PCT>)  
Biomimetic ester-hydrolase bEH<sub>3</sub> with 6-hexyl-[1,3,2]dioxaphosphepino[5,4-b:6,7-b']dipyridine 6-oxide (Fernandez-Lopez et al. 2023)
14. 8OTU (not yet released to PDB)  
PET44, PET-degrading enzyme from *Alkalilimnicola ehrlichii*.
15. 8U7F (not yet released to PDB)  
CIB12 GH1 from the extremely acidophilic archaeon *C. divulgatum*
16. 8U7G (not yet released to PDB)  
CIB13 GH1 from the extremely acidophilic archaeon *C. divulgatum*

In addition, in the course of the FuturEnzyme project, we have successfully generated 678 AlphaFold models, representing a significant advancement in our understanding of enzyme structures.

- Ester-hydrolases: 494
- Hyaluronidases and hyaluronate\_lyases: 102
- Peptidases: 44
- Amidases: 11
- Poly-ester hydrolases: 17

Finally, we have used at least 1 resolved protein structures from existing literature that served as a foundation for the development and design of enhanced mutants, as well as for the application of the project's computational tools.

- Non-catalytic pore-forming protein, Fragaceatoxin C: 4TSY and 3W9P.

#### 4. Comprehensive Structural Analysis of Enzymes in FuturEnzyme: Experimental Data

To date, 11 structures of 9 different enzymes have been generated. Two of these, EH<sub>0</sub> and EH<sub>7</sub>, correspond to: (i.) ester-hydrolases EH<sub>0</sub> and EH<sub>7</sub>, which have been studied as models to investigate the mechanisms of substrate specificity; (ii.) the transaminase TR<sub>2</sub> for the design of PluriZymes, enzymes with multiple active sites (one native and one artificial) using new computational tools; (iii.) poly-ester degrading enzymes IS11, PET46, PET44, which are relevant for the circularity of polyester-based textiles.

1. 7ZR3 (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7ZR3>)

- Ester-hydrolase EH<sub>0</sub> (Distaso et al. 2023)
2. 7PP3 (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7PP3>)  
Ester-hydrolase EH<sub>7</sub> (Cea-Rama et al. 2022)
  3. 7PP8 (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7PP8>)  
Ester-hydrolase EH<sub>7</sub> with methyl 4-nitrophenyl hexylphosphonate (Cea-Rama et al. 2022)
  4. 7PU6 (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7PU6>)  
Ester-hydrolase EH<sub>7</sub> complexed with 4-nitrophenyl hexylphosphonate (Cea-Rama et al. 2022)
  5. 7QYG (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7QYG>)  
Transaminase TR<sub>2</sub> (Roda et al. 2022)
  6. 8B4U (<https://www.ncbi.nlm.nih.gov/Structure/pdb/8B4U>)  
PET46, PETase enzyme from *Candidatus bathyarchaeota* (not published)
  7. 8A2C (<https://www.ncbi.nlm.nih.gov/Structure/pdb/8A2C>)  
PET40 (S178A mutant), PETase enzyme from an unclassified *Amycolatopsis*
  8. 7SPN (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7SPN>)  
Ester-hydrolase IS11, thermophilic poly-ester degrading hydrolase (Distaso et al. 2023b)
  9. 8OTU (not yet released to PDB)  
PET44, PET-degrading enzyme from *Alkalilimnicola ehrlichii*.
  10. 8U7F (not yet released to PDB)  
CIB12 GH1 from the extremely acidophilic archaeon *C. divulgatum*
  11. 8U7G (not yet released to PDB)  
CIB13 GH1 from the extremely acidophilic archaeon *C. divulgatum*

In addition, 4 structures of 2 different PluriZymes, enzymes with multiple mutations needed to introduced multiple active sites (one native and one artificial) using new computational tools, have been generated.

1. 8BBI (<https://www.ncbi.nlm.nih.gov/Structure/pdb/8BBI>)  
PluriZyme Xyn11-mut1 from *Pseudothermotoga thermarum*
2. 7QYF (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7QYF>)  
PluriZyme variant TR<sub>2</sub>E<sub>2</sub> (Roda et al. 2022)
3. 7QX3 (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7QX3>)  
PluriZyme variant TR<sub>2</sub>E<sub>2</sub> complexed with EOS (Roda et al. 2022)
4. 7QX0 (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7QX0>)  
PluriZyme variant TR<sub>2</sub>E<sub>2</sub> complexed with PLP (Roda et al. 2022)

Also, 1 structure of a biomimetic catalyst, that is, an ester hydrolase (EH<sub>3</sub>) to which a bipyridinyl derivative has been covalently bound to the active centre, have been generated.

1. 8PCT (<https://www.rcsb.org/structure/8PCT>)  
Biomimetic ester-hydrolase bEH<sub>3</sub> with 6-hexyl-[1,3,2]dioxaphosphepino[5,4-b:6,7-b']dipyridine 6-oxide (Fernandez-Lopez et al. 2023)

Below are the detailed methodologies and most relevant results pertaining to the structures outlined above.

#### 4.1. PluriZymes

The concept of PluriZyme involves the innovative design of new active sites within protein structures, including enzymes with one native active site. The process of adding artificial active sites is facilitated by the use of PELE, a software tool that enables the identification of suitable regions within a protein scaffold. These regions are selected based on their potential to accommodate a specific substrate that is targeted for transformation. Once a suitable site is identified, computational analysis is employed to mutate nearby amino acids. This mutation process is crucial as it allows for the incorporation of necessary amino acids that are essential for catalysis to occur. The approach is highly strategic, focusing on the precise modification of the protein's active site to achieve desired catalytic functions. The novelty of this approach compared to state

of art, can be summarized as follows: (i.) Targeted Design: Unlike traditional methods that often rely on random or semi-random mutagenesis, the PluriZyme approach uses targeted design. This specificity allows for more efficient and predictable modifications, leading to potentially higher success rates in enzyme design; (ii.) Computational Efficiency: The use of PELE software streamlines the process of identifying and modifying active sites. This computational approach can significantly reduce the time and resources needed compared to experimental trial-and-error methods; (iii.) Precision at the Molecular Level: The ability to mutate specific amino acids for accommodating substrates offers a level of precision that is not always achievable with other methods. This can lead to the creation of enzymes with highly specialized and novel catalytic abilities; (iv.) Potential for Complex Substrate Processing: Given its precision and specificity, the PluriZyme approach could be particularly effective in designing enzymes capable of processing complex or unusual substrates, which might be challenging with traditional enzyme engineering methods; and (v.) Integration with Existing Protein Scaffolds: This method allows for the modification of existing enzyme scaffolds, which can be advantageous over designing entirely new enzymes. It leverages the stability and well-characterized nature of existing proteins while imbuing them with new functionalities. In order to unambiguously demonstrate the design of PluriZymes, we have solved the structures of two PluriZymes, Xyn11-mut1 and TR<sub>2</sub>E<sub>2</sub>, that are detailed below.

### PluriZyme Xyn11-mut1

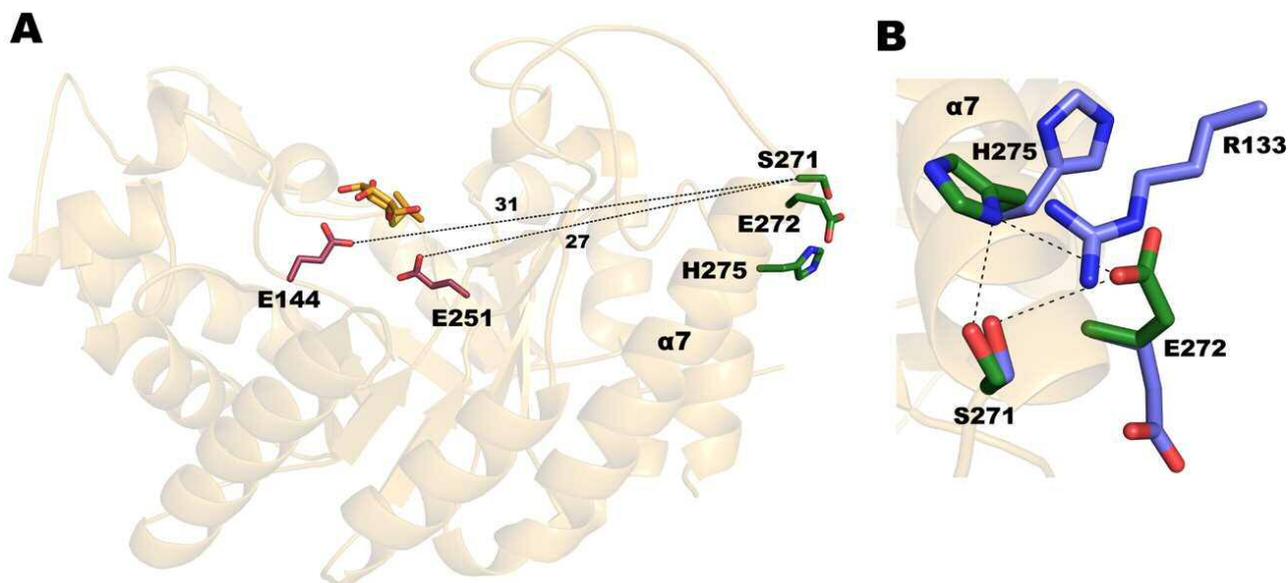
1. 8BBI (<https://www.ncbi.nlm.nih.gov/Structure/pdb/8BBI>)  
PluriZyme Xyn11-mut1 from *Pseudothermotoga thermarum*

### Results

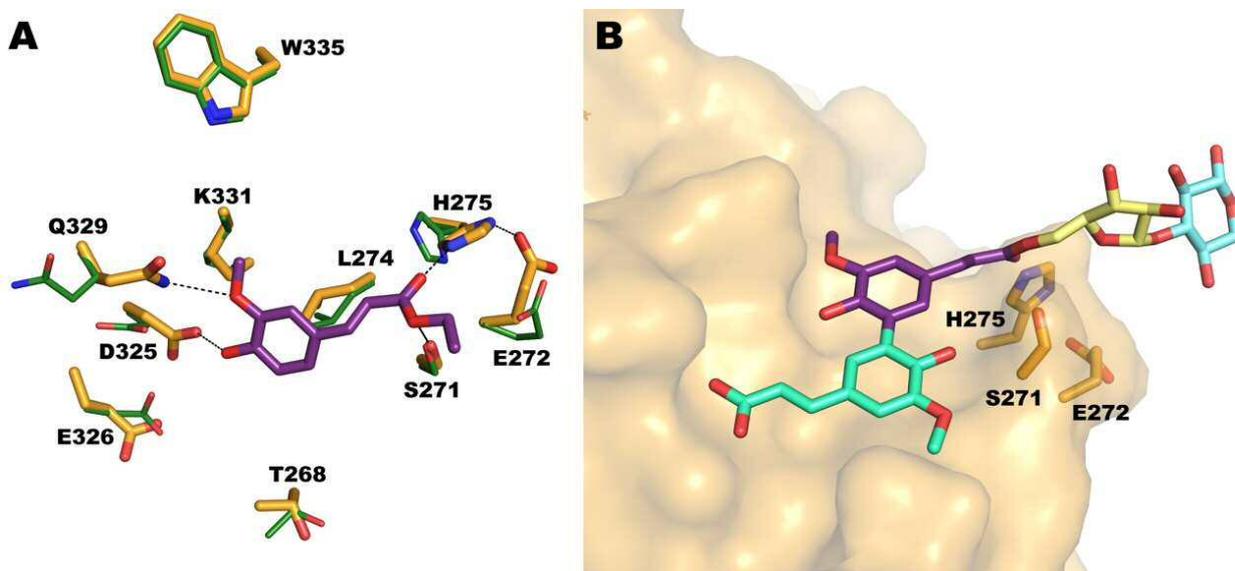
**Folding of Xyn11-Mut1.** The crystal structure of mutant 1 of Xyn11 was obtained by X-ray diffraction at 2.1 Å. It shows a TIM barrel architecture typical of GH10 xylanases. The intrinsic xylanase active site, located at the axis of the barrel, is formed by the pair Glu144 and Glu251 and includes a trapped IPTG molecule. Using the software PELE allowed to create a second artificial catalytic site located on the surface of the protein at 27 Å and 31 Å from Glu251 and Glu144, respectively (Figure 1A). Mutagenesis experiments were performed to create the double mutant (Xyn11-Mut1) L271S and K275H. The second artificial catalytic triad is formed by residues Ser271, Glu272 and His275, at the beginning of α7, and having esterase activity (Figure 1A). There are no significant structural differences upon superimposition of Xyn11 onto the Xyn11-Mut1 coordinates. However, as happened in the native crystal, Xyn11-Mut1 crystals contain two molecules in the asymmetric unit that present some conformational differences at the artificial esterase catalytic triad. Thus, only chain B presents the Ser271-Glu272-His275 triad in a proper conformation that displays the expected hydrogen bonding pattern conserved in reported esterases (Figure 1B). In chain A, Arg133, from a symmetry related molecule, is making a hydrogen bond to Ser271 producing a shift in Glu272 and His275 side-chains.

**Docking simulations.** Attempts to crystallize Xyn11-Mut1 with different irreversible inhibitors were performed, but the ligand was not captured in the crystal. Therefore, docking simulations were performed to depict the Xyn11-Mut1 artificial esterase site by using molecule B coordinates, which presents an active conformation of its catalytic triad. Ethyl 4'-hydroxy-3'-methoxycinnamate, an ethyl derivative of ferulic acid (PDB code 3PFB, in violet), was docked into molecule B using Autodock Vina, and residues underlining the catalytic site were considered flexible. Among all results obtained, the best solution shown in Figure 2A was selected on the basis of a productive interaction of the ligand with the catalytic triad. A series of minor conformational changes are observed in the flexible residues upon ligand binding to allocate the substrate. The ligand adopts the tetrahedral conformation typical of the intermediate of ester hydrolases where the nucleophilic serine is located under the carbonyl group to produce the attack, with the oxygen being close to two glycines that stabilize the formed oxoanion (Wang et al. 2018). However, in this case, it seems that the oxygen from the carbonyl group could be stabilized by His275 from the catalytic triad (Figure 2A) that, in turn, could be hydrogen linked to the catalytic Glu272. Additional hydrogen links could be stabilizing the ferulic moiety to Asp325 and Gln329 side-chains. Furthermore, to the best solution obtained by the docking simulations, arabinose, xylose and ferulic acid units can be manually attached to the ligand to illustrate the putative binding of xylan (Figure 2B). As shown in the figure, the polymeric xylose chain would extend far

from the protein surface therefore having no impediment to allocate its terminal ferulic moieties into the artificial esterase active site of Xyn11-Mut1.



**Figure 1.** A) Xyn11-Mut1 folding. The intrinsic xylanase catalytic pair is shown as raspberry sticks (Glu144 and Glu251) whereas the artificial secondary esterase triad is shown as green sticks (Ser271, Glu272 and His275). The trapped IPTG molecule is shown as orange sticks. B) Superimposition of chain A (blue) and chain B (green) at the artificial esterase secondary catalytic site. The catalytic triad of molecule B shows the proper hydrogen bond pattern conserved among esterases.



**Figure 2.** A) Best model obtained after automatic docking of ethyl 4'-hydroxy-3'-methoxycinnamate (in violet) within the artificial esterase active site of Xyn11-Mut1. The residues considered flexible are shown as sticks (in orange), highlighting the conformational changes from the free state (in forest). Hydrogen bonds are shown as dashes while covalent bond is shown as continuous line. B) Arabinose (in yellow), xylose (in cyan) and ferulic acid (in green) units have been manually built to the automatically docked substrate to illustrate the putative binding of a xylan template. The cavity of the artificial second active site is depicted by the surface of the protein (in orange). The catalytic triad is shown as orange sticks.

## Methodology

**Crystallization and X-ray structure determination of Xyn11-Mut1.** Initial crystallization conditions were explored by high-throughput techniques with a NanoDrop robot (Innovadyne Technologies) using 5 mg·mL<sup>-1</sup> protein concentrations in Tris (20 mM, pH 7) and NaCl (50 mM), protein reservoir ratios of 1:1 and commercial screens: Index (Hampton Research), JBScreen JCSG and JBScreen PACT (Jena Bioscience). After three days, bar shape crystals were grown in 20% PEG3350, 0.1M Bis-Tris-propano pH 8.5, 0.2M sodium formate and ratio 250nL of protein and 250nL of reservoir. For data collection, crystals were transferred to cryoprotectant solution consisting of mother liquor and glycerol (25% (v/v)) before being cooled in liquid nitrogen. Diffraction data were collected using synchrotron radiation on the XALOC beamline at ALBA (Cerdanyola del Vallés, Spain). Diffraction images were processed with XDS (Kabsch 2010) and merged using AIMLESS from the CCP4 package (Evans and Murshudov 2013). The crystal was indexed in the P<sub>2</sub><sub>1</sub>2<sub>1</sub>2<sub>1</sub> space group, with two molecules in the asymmetric unit and 51% solvent content within the unit cell. The data collection statistics are given in Table 1. The structure of Xyn11-Mut1 was solved by difference Fourier synthesis (Vagin and Teplyakov 1997) using the coordinates from Xyn11 as a template (PDB Code 7NL2). Crystallographic refinement was performed using the program REFMAC (Murshudov, Vagin, and Dodson 1997) within the CCP4 suite, with automatic local non-crystallographic symmetry (NCS). The free *R*-factor was calculated using a subset of 5% randomly selected structure-factor amplitudes that were excluded from automated refinement. Subsequently, heteroatoms were manually built into the electron density maps with Coot8 (Emsley et al. 2010), and water molecules were included in the model, which, combined with more rounds of restrained refinement, reached the *R* factors listed in Table 1. The figures were generated with PyMOL. The crystallographic statistics of Xyn11-Mut1 are listed in Table 1.

**Docking simulations.** The crystal structure of Xyn11-Mut1 was used as a template for the docking experiments. All resolved water molecules and other heteroatoms were removed from the structure prior to protein preparation and subsequent docking simulations. The coordinates of ethyl 4'-hydroxy-3'-methoxycinnamate, an ethyl derivative of ferulic acid, were obtained from the PDB deposition website with the PDB code 3PFB. Autodock Tools was used to create a .pdbqt file that is a modified PDB file containing the coordinates of the protein and substrate and additional information such as the atom type. Substrate docking was performed using AutoDock Vina (Trott and Olson 2009) with Ser271, Glu272, His275, Thr268, Leu274, Asp325, Glu326, Gln329, Lys331 and Trp335 defined as flexible side chains. The docking space was visually defined in Autodock Tools. The grid box around the catalytic triad with dimensions 26 Å X 26 Å x 28 Å was used to cover the entire substrate-binding site. Default parameters were defined during docking with exhaustiveness of 16. Calculations were performed using Lamarckian Genetic Algorithm (LGA) method. Among 20 calculated models, the best solution was chosen on the basis of the productive interaction of the ligand with the catalytic triad. Xylan template was manually built by attaching arabinose and xylose units to the best docked ethyl-ferulic acid using Coot8 (Emsley et al. 2010).

**Codes and accession numbers.** The sequence encoding Xyn11-Mut1 was deposited in UniProtKB with the accession number F7YXD6. The atomic coordinates and structure factors for the Xyn11-Mut1 structure have been deposited in the RCSB Protein Data Bank with accession codes 8BBI.

**Table 1. Crystallographic statistics of Xyn11-Mut1.**

Values in brackets are for the high-resolution shell	
Crystal data	Xyn11-Mut1
Space group	P <sub>2</sub> <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Unit cell parameters	
a (Å)	90.32
b (Å)	95.55
c (Å)	101.34
Data collection	
Beamline	XALOC(ALBA)
Temperature (K)	100

Wavelength (Å)	0.979185
Resolution (Å)	47.77-2.10 (2.06-2.01)
<b>Data processing</b>	
Total reflections	258521 (21236) (((17149))
Unique reflections	51806 (4197)
Multiplicity	5.0 (5.1)
Completeness (%)	99.9 (99.9)
Mean I/σ (I)	15.8 (2.6)
R <sub>merge</sub> (%)	6.4 (59.3)
R <sub>pim</sub> (%)	3.1 (29.0)
Molecules per ASU	2
<b>Refinement</b>	
R <sub>work</sub> /R <sub>free</sub> <sup>†††</sup> (%)	16.1/19.6
<b>N° of atoms/average B (Å<sup>2</sup>)</b>	5976/37.73
Macromolecule	5635/37.14
Ligands	66/74.39
Solvent	275/41.14
<b>Ramachandran plot (%)</b>	
Favored	97.9
Outliers	0.3
<b>RMS deviations</b>	
Bonds (Å)	0.009
Angles (°)	1.469
PDB accession code	8BBI

<sup>†</sup>R<sub>merge</sub> =  $\sum hkl \sum i | I_i(hkl) - [I(hkl)] | / \sum hkl \sum i I_i(hkl)$ , where  $I_i(hkl)$  is the  $i$ th measurement of reflection  $hkl$  and  $[I(hkl)]$  is the weighted mean of all measurements.

<sup>††</sup>R<sub>pim</sub> =  $\sum hkl [1/(N-1)] \sum i | I_i(hkl) - [I(hkl)] | / \sum hkl \sum i I_i(hkl)$ , where  $N$  is the redundancy for the  $hkl$  reflection.

<sup>†††</sup>R<sub>work</sub>/R<sub>free</sub> =  $\sum hkl | F_o - F_c | / \sum hkl | F_o |$ , where  $F_c$  is the calculated and  $F_o$  is the observed structure factor amplitude of reflection  $hkl$  for the working/free (5%) set, respectively.

## PluriZyme TR<sub>2</sub>E<sub>2</sub>

2. **7QYF** (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7QYF>)  
PluriZyme variant TR<sub>2</sub>E<sub>2</sub> (Roda et al. 2022)
3. **7QX3** (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7QX3>)  
PluriZyme variant TR<sub>2</sub>E<sub>2</sub> complexed with EOS (Roda et al. 2022)
4. **7QX0** (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7QX0>)  
PluriZyme variant TR<sub>2</sub>E<sub>2</sub> complexed with PLP (Roda et al. 2022)

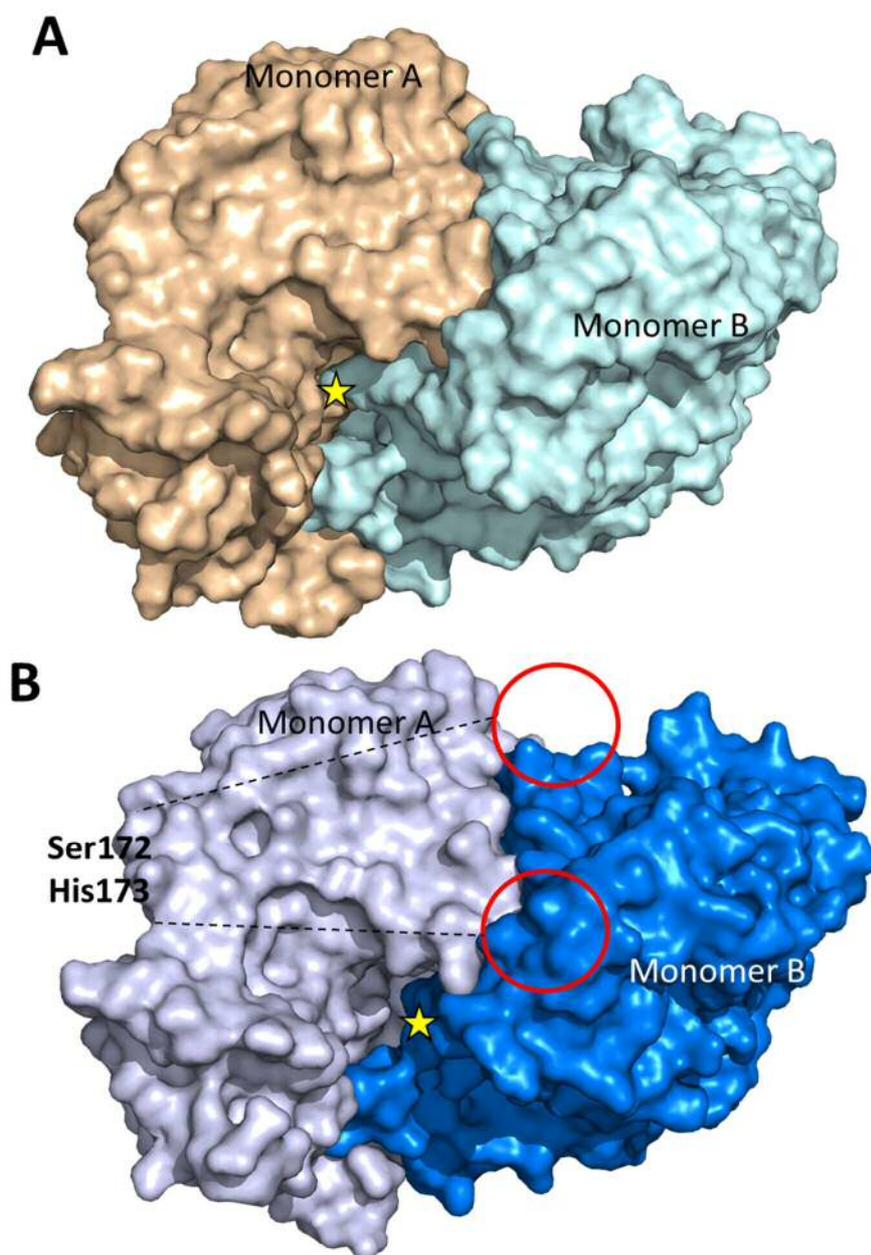
The 3 structures produced demonstrated that the design of *PluriZymes* with two biologically active sites supporting different chemistries, in particular esterase and transaminase activities, is feasible. For this purpose, we select as scaffold a class III ω-transaminase, TR<sub>2</sub>, isolated through metagenomic techniques, and added an artificial catalytic triad Ser-His-Asp supporting ester-hydrolysis. By using the computational methods implemented, a total of 20 mutants of the TR<sub>2</sub> enzyme were selected for investigations and synthesized; a priori, based on computational data, they may contain one active site supporting the original transaminase activity and a second active site with esterase activity. The mutants were produced and purified and their transaminase and esterase activity tested.

Only one mutant, TR<sub>2</sub>E<sub>2</sub> (with A172S and Q173H mutations) was confirmed as having both activities. The transaminase activity of TR<sub>2</sub>E<sub>2</sub> was unaffected by the mutations compared to that of the native TR<sub>2</sub>, while the introduced esterase active site astoundingly active with a wide range of chemically and structurally diverse esters. Functional and structural characterization of the native TR<sub>2</sub> (7QYF) and the PluriZyme TR<sub>2</sub>E<sub>2</sub>

(complexed with the inhibitor ethanolamine O-sulphate (EOS) and the cofactor pyridoxal-5'-phosphate (PLP); 7QX3 and 7QX0) confirmed the presence and functionality of both active sites. The structures of TR2 and TR<sub>2</sub>E<sub>2</sub> PluriZyme are shown in [Figure 3](#). For details methodological details and detailed results, see [Roda et al. 2022](#).

#### 4.2. Biomimetic catalysts

The structure of a biomimetic catalysts was solved to modify to what extent the properties of an enzyme can be tuned through incorporating an inhibitor that can bind metals to interchange the original activity for a new one. In this case, a chemical method called bioconjugation was used to modify the activity of an enzyme without manipulating the protein sequence by genetic engineering. In order to unambiguously demonstrate the design of such biomimetic catalyst, we have solved the structures of one such catalyst, bEH<sub>3</sub>, that is detailed below.



**Figure 3.** A) Surface representation of the TR<sub>2</sub> dimer with highlighted transaminase binding site (red circle). B) Surface representation of the TR<sub>2</sub>E<sub>2</sub> dimer with a highlighted transaminase binding site (red circle). The positions of the Ser172 and His173 mutations are colored magenta. The figure was created using PyMOL Version 2.3.2.

## Biomimetic catalyst bEH<sub>3</sub>

### 1. 8PCT (<https://www.rcsb.org/structure/8PCT>)

#### Biomimetic ester-hydrolase bEH<sub>3</sub> with 6-hexyl-[1,3,2]dioxaphosphepino[5,4-b:6,7-b']dipyridine 6-oxide (Fernandez-Lopez et al. 2023)

In brief, a metal-chelating irreversible inhibitor was designed and synthesized, formed by a bipyridine (byp) and a phosphonate group. Subsequently, the irreversible binding of the inhibitor to a serine-ester hydrolase was tested, with successful results for the ester-hydrolase EH<sub>3</sub>, an enzyme identified through a metagenomic search. The first step was demonstrating using mass spectrophotometry, kinetic, spectrophotometric, and finally using structural methods (Figure 4) that when the inhibitor and Cu<sup>2+</sup> salt (Cu(bpy) complex), were incubated together with EH<sub>3</sub>, the inhibitor was successfully bound to the enzyme, and the biomimetic bEH<sub>3</sub> was formed. Activity tests confirmed that through the use of the metal-chelating inhibitor, it was possible to achieve the transformation of the ester-hydrolase EH<sub>3</sub> into a biomimetic (bEH<sub>3</sub>) with oxidoreductase activity, in particular the decoloration of dyes relevant to textile sector. For details methodological details and detailed results, see Fernandez-Lopez et al. 2023.

## 4.3. Native Enzymes

The detailed structural data of the 7 enzymes for which structures were solved provided insights into their mechanism of action, substrate specificity, which are essential for tailoring enzymes to specific needs. Below, a short description is provided, with in deep results discussed in the corresponding publications.

### 12. 7ZR3 (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7ZR3>)

#### Ester-hydrolase EH<sub>0</sub> (Distaso et al. 2023a)

The structural bases of the activity and substrate promiscuity of EH<sub>0</sub> (GenBank ID MK791218), an ester-hydrolase from family IV, have been studied. EH<sub>0</sub> was isolated through metagenomics from a soil sample collected at the Henfaes Research Station, Abergwyngregyn, North Wales (53°14'N, 4°01'W), following its enrichment with technical cashew nut shell liquid (tCNSL). EH<sub>0</sub> exhibits broad substrate specificity, capable of hydrolyzing 68 esters from a collection of 96, but it is not stereoselective. The analysis of EH<sub>0</sub> has revealed its 3D structure (Figure 5), showing a very narrow access tunnel to the active site with a single entrance, unlike other ester-hydrolases described to date.

Additionally, one of the helices forming the cap domain ( $\alpha_7$ ) is located very close to the catalytic triad, in fact, forming a hydrogen bond with the catalytic serine (Glu226-Ser161). Therefore, the only possible way for substrates to enter the active center is through the opening of this cap domain. When comparing EH<sub>0</sub> with its homologues, it has been observed that this enzyme has two prolines adjacent to the end of the  $\alpha_2$  helix, suggesting their potential role as hinges for this movement that allows the entry of substrates into the active center. The flexibility of the cap domain has been modeled through molecular dynamics simulation by "ensemble refinement," showing the different conformations that the N-terminal part of this domain can adopt, ranging from a conformation similar to that captured by crystallography to very open conformations that allow access to bulky substrates. Key residues in the active site and the substrate access tunnel (positions 46, 223, and 226) have been identified, and a series of mutants with varying activity and specificity profiles have been designed and tested. The analysis of the activity profiles has shown that the proline at position 46 is one of the most critical residues, among those analyzed, for the entry and hydrolysis of bulky substrates, and its mutation to Ala allows for an expansion of substrate specificity. For details methodological details and detailed results, see Distaso et al. 2023a.

### 13. 7PP3 (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7PP3>)

#### Ester-hydrolase EH<sub>7</sub> (Cea-Rama et al. 2022)

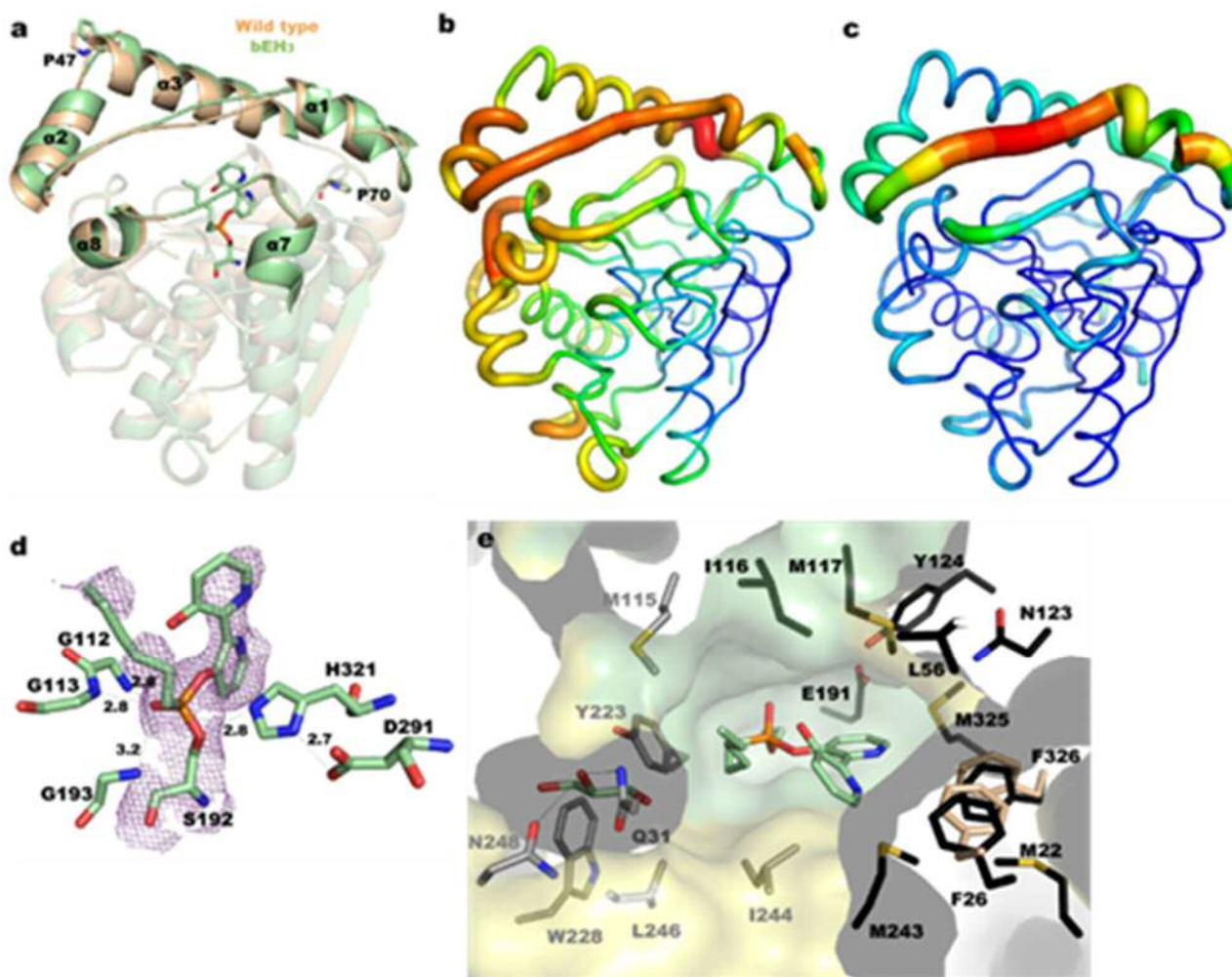
### 14. 7PP8 (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7PP8>)

#### Ester-hydrolase EH<sub>7</sub> with methyl 4-nitrophenyl hexylphosphonate (Cea-Rama et al. 2022)

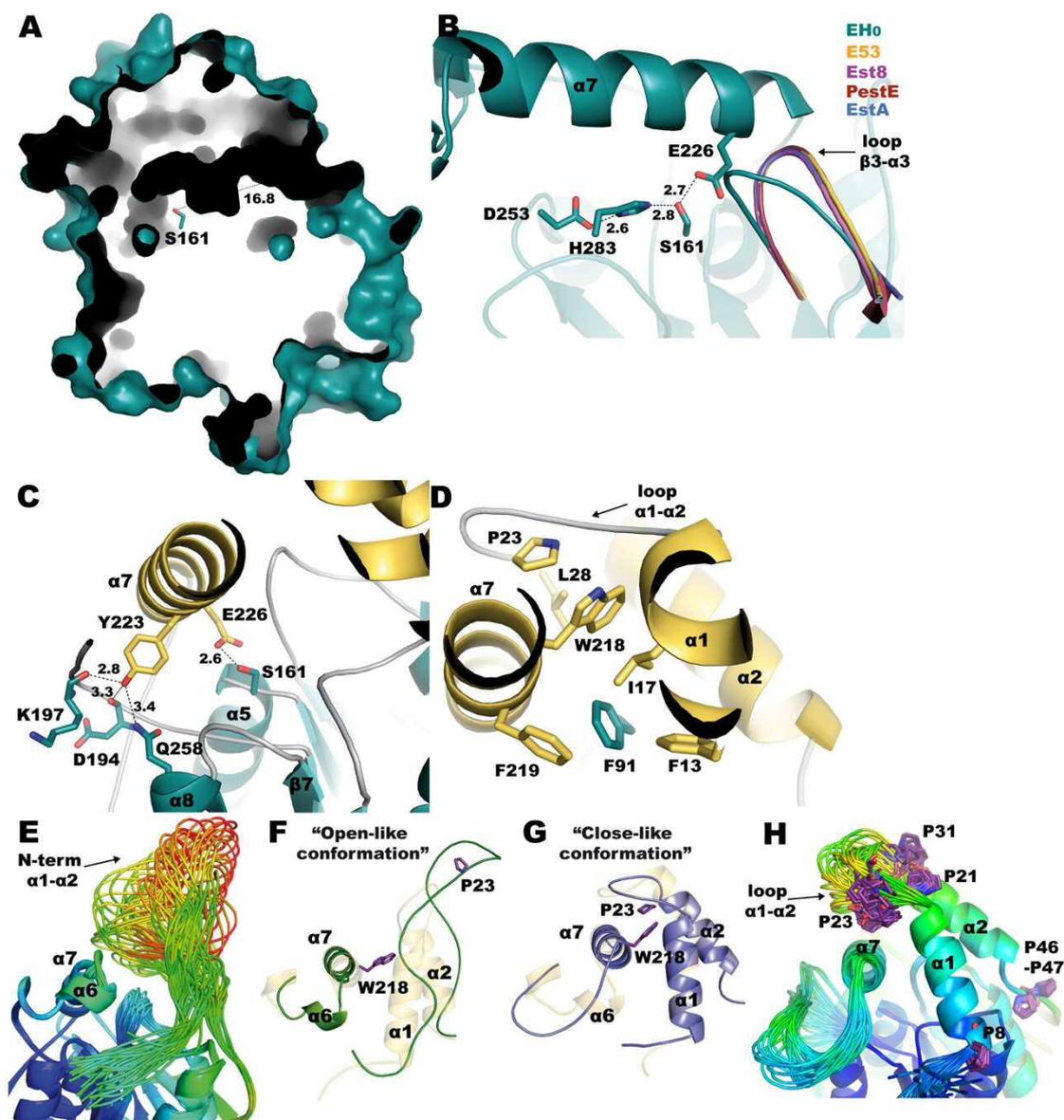
### 15. 7PU6 (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7PU6>)

#### Ester-hydrolase EH<sub>7</sub> complexed with 4-nitrophenyl hexylphosphonate (Cea-Rama et al. 2022)

We have also focused on the ester-hydrolase EH<sub>7</sub>, belonging to family VIII, which shows some similarities with certain class C  $\beta$ -lactamases. Specifically, this enzyme is included in the subfamily VIII.1, presenting a series of conserved motifs: S-X-X-K (where the catalytic Ser and Lys are located), Y-X-X (where the Tyr of the catalytic triad is found), and W-X-G. EH<sub>7</sub> is a promiscuous enzyme with broad substrate specificity, unusual not only among family VIII esterases but also among lipolytic enzymes: it is capable of hydrolyzing a wide array of 66 structurally different esters, including substrates that are intermediates in polyester degradation. The structural resolution of this enzyme in its native state (Figure 6) has allowed for the analysis of the molecular bases of its high promiscuity.



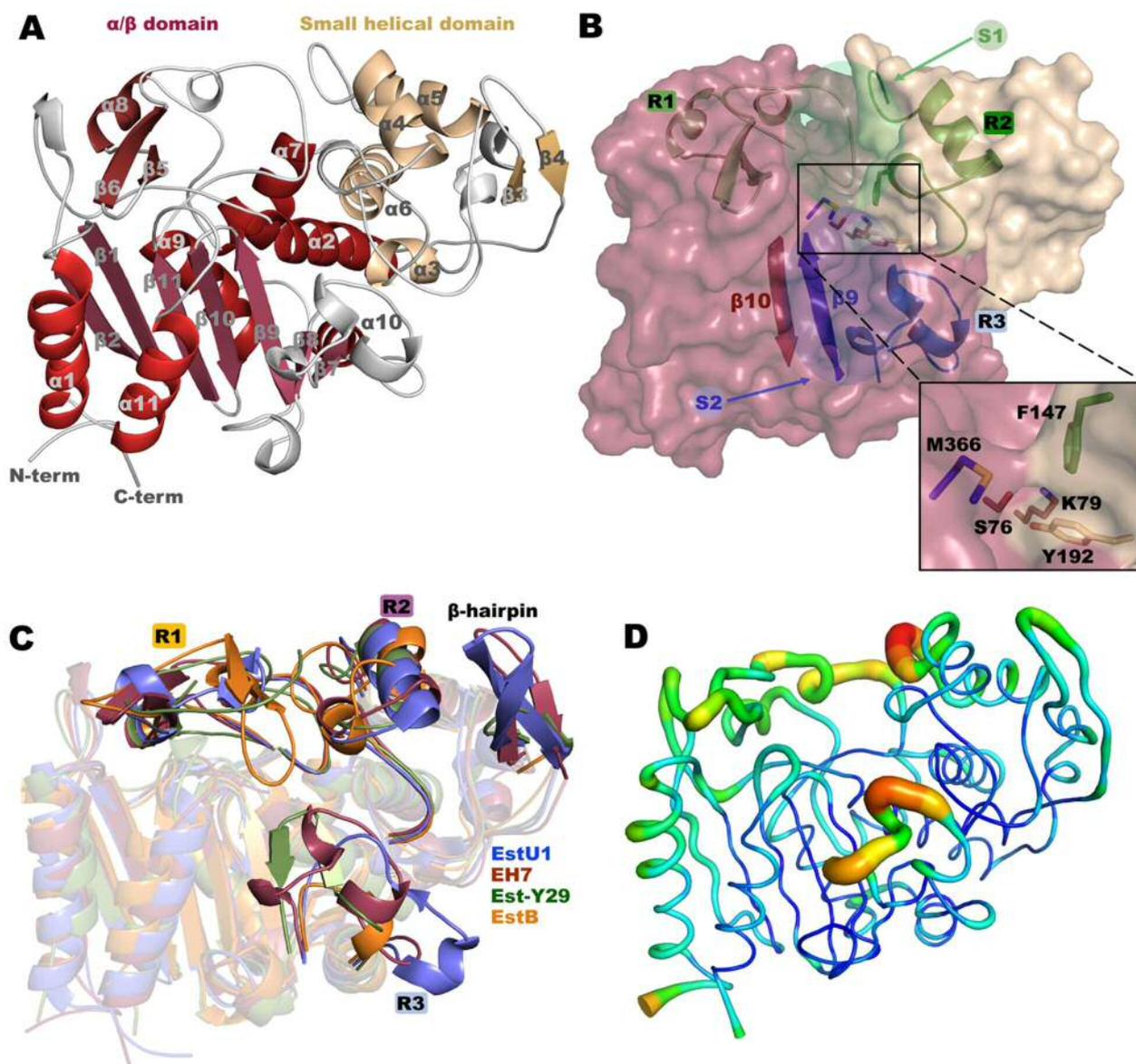
**Figure 4.** (a) Superimposition of unliganded EH<sub>3</sub> (wheat, PDB Code: 6SXP) onto bEH<sub>3</sub> (green, PDB Code: 7OWW), highlighting the structural changes observed for the CAP domain, where the catalytic domain is shown in transparency. Residues Pro47 and Pro70, acting as hinges, are shown as sticks. Representative B factors for the complex (b) and the unliganded enzyme (c), with values ranging from low (blue) to high (red). (d) Close-up view of the bEH<sub>3</sub> complex mimicking the tetrahedral intermediate of the reaction, and the Polder omit map calculated at the 2.5  $\sigma$  cut-off is shown in violet. The hydrogen bond network is shown as dashed lines, with atomic distances in Å. (e) Cross-section of the molecular surface of bEH<sub>3</sub>, with residues close to the alcohol moiety of the ligand coloured in black and residues near the acyl moiety coloured in grey. The residues composing the cap and catalytic domains are coloured yellow and green, respectively. A glycerol molecule captured at the acyl binding site is also shown. The positions of Phe26 and Phe326 in the free enzyme are shown by wheat sticks.



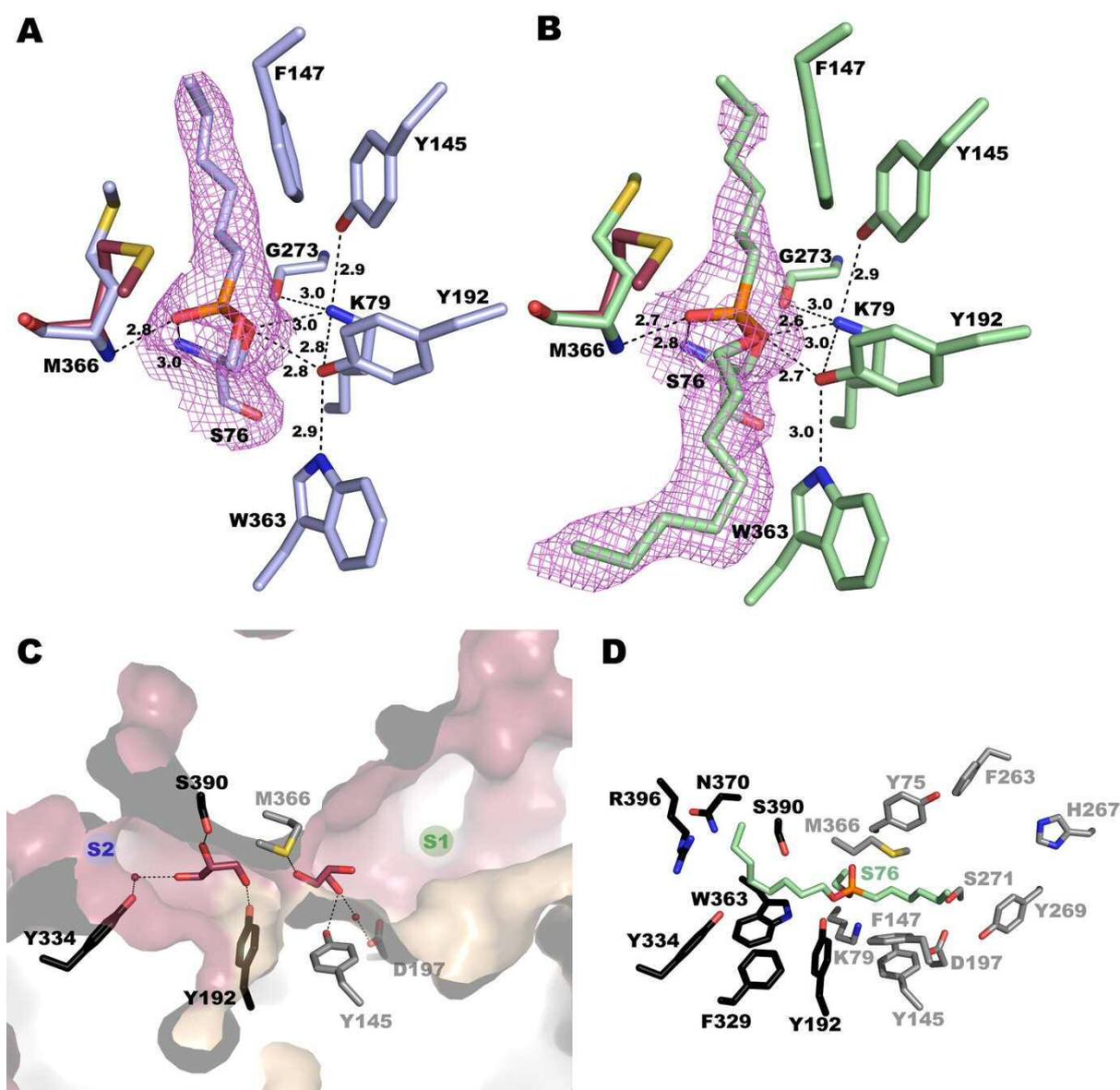
**Figure 5.** A) A tunnel 16.8 deep giving access to the catalytic triad. B) Hydrogen bond network among the residues making the catalytic triad of EH0 (teal). Movement of loop  $\beta 3-\alpha 3$  in EH0 (housing the oxoanion) with respect to its homologs E53 complexed with (4-nitrophenyl)hexanoate (yellow), Est8 (violet), PestE (raspberry), and EstA (slate). C) Hydrogen bond network of residues Glu226 and Tyr223 from  $\alpha 7$ . (D) Hydrophobic patch where Trp218 is located. The EH0 cap domain is shown in yellow, and the catalytic domain in teal (C and D). (E) All possible conformations of the region comprising  $\alpha 1-\alpha 2$  shown by the ensemble refinement of molecule A. The color code gives chain mobility, from low (blue) to high (red). The opposed scenarios of the region  $\alpha 1-\alpha 2$ , from an “open conformation” (F, green) to a “closed conformation” (G, slate). The cap domain seen in the crystal structure is shown in yellow. (H) All conformers shown by the ensemble refinement of molecule B.

The catalytic cavity is divided into two zones (S1 and S2). Comparing the secondary structure of EH7 with its homologues, a greater difference in the S1 zone has been observed, further divided into two regions R1 and R2 that flank both sides of the S1 catalytic tunnel. Both R1 and R2 are retracted from the active site, forming a wide cavity, which could justify EH7's high promiscuity. It has been observed that the position of the R1 region of the active site is associated with the presence or absence of  $\beta$ -lactamase activity due to the presence or absence of steric hindrance. In line with this hypothesis, EH7 exhibits this type of activity, using nitrocefin as a chromogenic model substrate. Likewise, the acyl transferase activity is associated with the W-X-G motif, which is present in EH7, a fact that was experimentally verified by following the synthesis of benzyl 3-oxobutanoate from benzyl alcohol and vinyl acetoacetate. Complexes of EH7 with derivatives of the suicide

inhibitors methyl (EH7-M-4NHP) and octyl 4-nitrophenylhexylphosphonate (EH7-O-4NHP) have also been obtained (Figure 7), allowing the identification of the acyl and alcohol zones of the catalytic cavities. Furthermore, based on structural data, different mutants of EH7 have been prepared through site-directed mutagenesis, finding that when the residue Tyr334, located at the end of one of the tunnels in the S2 region, is replaced by a less bulky residue, it is key for the hydrolysis of substrates with larger acyl/alcohol groups. For details methodological details and detailed results, see [Cea-Rama et al. 2022](#).



**Figure 6.** EH7 crystal structure. (A) General folding showing the  $\alpha/\beta$  domain (red) and the small helical domain (wheat). (B) Depiction of the EH7 surface showing the sites S1 (outlined by R1 and R2) and S2 (defined by  $\beta9$  and R3). Inset: zoom image of the catalytic pocket. Residues of the catalytic triad are: Ser76, Lys79 (located at the  $\alpha/\beta$  domain, raspberry) and Tyr192 (located at the small helical domain, wheat). Residues Met366 (located at the  $\beta9$ -  $\beta10$  loop, blue) and Phe147 (located in the R2 region, green) form the tunnel giving access to the catalytic pocket. (C) Superimposition of EH7 monomer (prune) and its homologues: EstU1 (slate, PDB code 4IVI), Est-Y29 (green, PDB code 4P6B) and EstB (orange, PDB code 1CI8). R1 and R2 regions present largest differences affecting the access to the S1 site. (D) EH7 B-factor values, from low (blue) to red (high). Most flexible regions are R2 and R3.



**Figure 7.** Active sites of EH7 complexed with M-4NHP (A, blue) and with O-4NHP (green, B). Hydrogen bonds networks is shown as dashed lines for M-4NHP and O-4NHP. Met366 for the wild type protein is shown as raspberry sticks. 2Fo-2Fc Electron maps contoured at 1.0  $\sigma$  for M-4NHP and O-4NHP (violet), respectively. C) A cross-section of the molecular surface of EH7 showing the catalytic pocket in the substrate-free wild type enzyme. The small helical domain in wheat whereas the  $\alpha/\beta$  domain in prune. D) A view of the phosphonate inhibitor complex, with residues bordering the alcohol (black) and acyl (grey) binding sites.

## 16. 7QYG (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7QYG>)

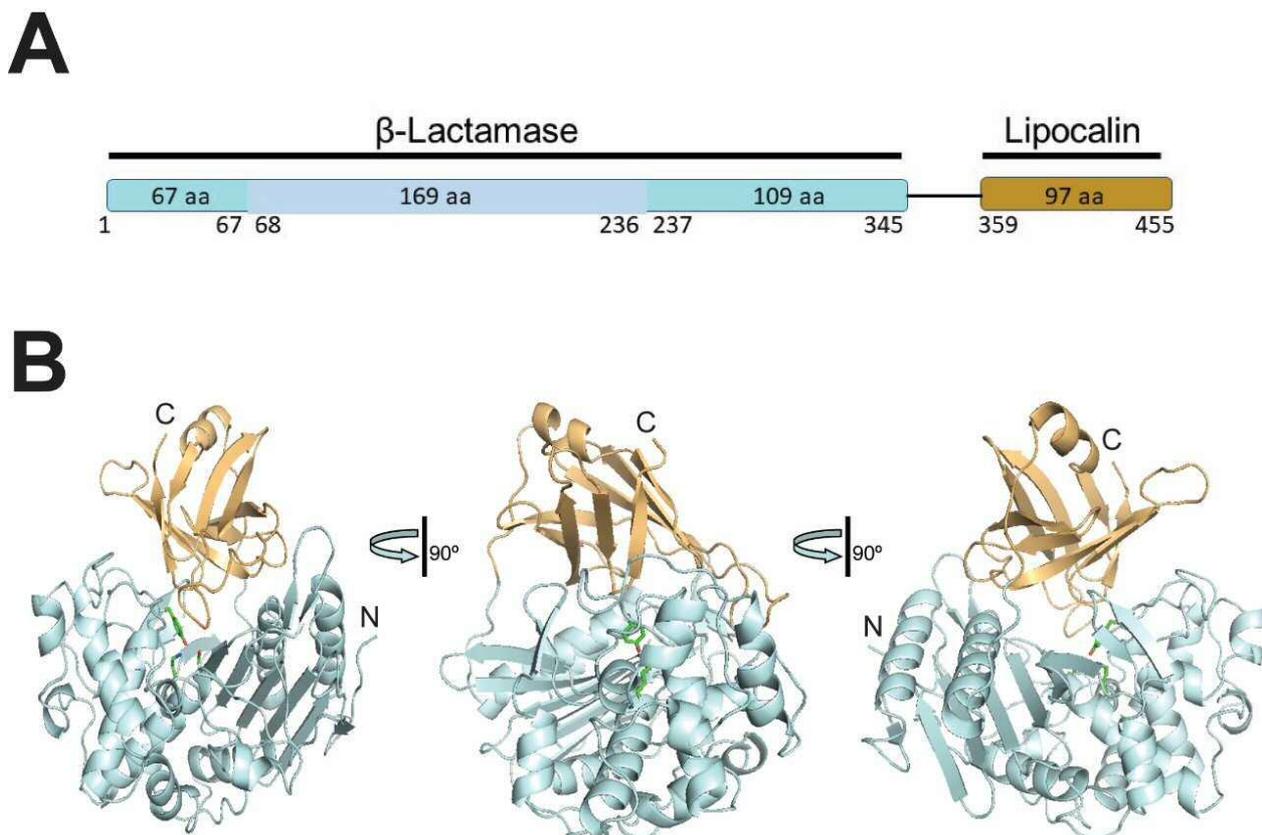
### Transaminase TR<sub>2</sub> (Roda et al. 2022)

We have also focused on the transaminase TR<sub>2</sub>, which was used as a target for the design of multi-purpose PluriZymes, as detailed earlier. TR<sub>2</sub> is a class III  $\omega$ -TA (GenBank acc. nr. MH588437) isolated from the beach acidic pool on Vulcano Island and most likely derived from a bacterium of the Acidihalobacter genus. We aimed to crystallize and solve the structure of TR<sub>2</sub> (see Figure 3A) to help the design of such PluriZymes. For details methodological details and detailed results, see Roda et al. 2022.

## 17. 7SPN (<https://www.ncbi.nlm.nih.gov/Structure/pdb/7SPN>)

### Ester-hydrolase IS11, thermophilic poly-ester degrading hydrolase (Distaso et al. 2023b)

We have also focused on IS11, a poly-ester hydrolase from microbial communities of terrestrial hydrothermal vents on the volcanic island of Ischia, Italy. The catalytic cleft of IS11 included catalytic Ser68, Lys71, Tyr160, and Asn162, while the lipocalin domain enclosed the catalytic cleft like a lid and contributed to substrate binding (Figure 8). Our study identified novel thermotolerant carboxylesterases with a broad substrate range, including polyesters and mycotoxins, for potential applications. The analysis of the crystal structure allowed us to reveal the key residues and domains for hydrolyzing poly-esters. In particular, the catalytic cleft of IS11 contained several hydrophobic and polar residues potentially involved in substrate binding (Asp126, Phe128, Trp158, Asn304, Ile307, and Leu309) and a lipocalin domain that participates in substrate binding rather than in oligomerization, as in other reported enzymes containing such a domain. For details methodological details and detailed results, see [Distaso et al. 2023b](#).



**Figure 8.** Crystal structure of IS11. A) Schematic representation of the IS11 domains: the N-terminal  $\beta$ -lactamase related Ser hydrolase domain is colored cyan with all-helical sub-domain shown in light blue, whereas the C-terminal lipocalin domain in orange. B) overall fold of the IS11 protomer shown in three views related by  $90^\circ$  rotations. The protein domains are shown as ribbon diagrams with the core domain ( $\beta$ -lactamase) colored pale cyan, whereas the C-terminal lipocalin domain is colored light orange. The position of the active site is indicated by the side chains of catalytic Ser68, Lys71, and Tyr160, whereas the protein N- and C-terminal ends are labeled (N and C).

#### 18. 8B4U (<https://www.ncbi.nlm.nih.gov/Structure/pdb/8B4U>)

##### PET46, PETase enzyme from *Candidatus bathyarchaeota* (not published)

PET46 from *Candidatus Bathyarchaeota* was nominated for end-of-life fabric recycling because availability of experimental structural data and activity on PET polyester fabric. We designed and characterized PET46 (NCBI accession RLI42440.1), the first enzyme from archaeal origin reported to hydrolyze PET polymer. The enzyme is encoded in the metagenome-assembled genome (MAG) of the *Candidatus Bathyarchaeota* archaeon B1\_G2, a member of the TACK group that was found at the Guaymas Basin. The experimentally established crystal structure of the protein was found to be similar to bacterial PET-degrading enzymes, but reveals several unique features, in agreement with the capacity of PET46 to hydrolyse not only PET polymers but also PET-trimer (3PET), BHET and MHET. Based on the structural analysis and docking analyses (Figure 9),

positions A46, A140 and A147 and a lid domain were found relevant for poly-ester degradation and thermo-stability.

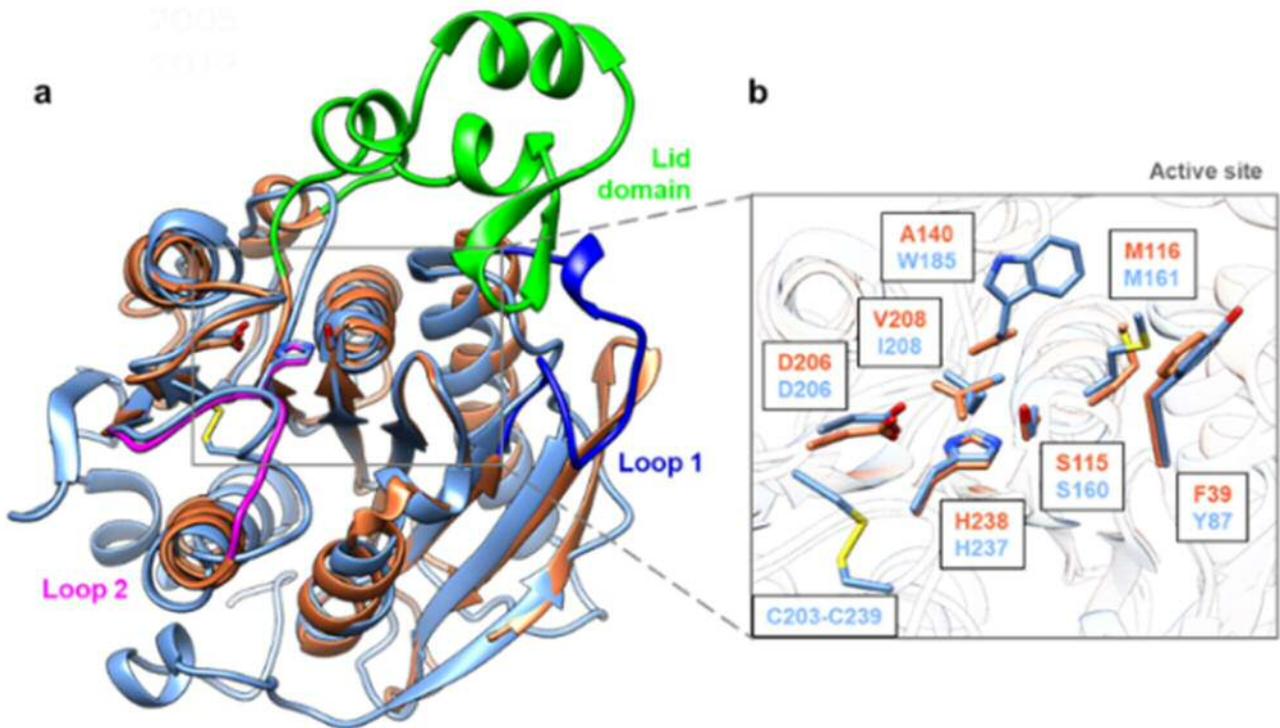


Figure 9. PET46 uses the lid domain to effectively degrade MHET, BHET and 3PET. A) General structure of PETase enzyme from *Candidatus bathyarchaeota*. B) Details of aminoacids found to be relevant.

### 19. 8OTU (not yet released to PDB)

#### PET44, PET-degrading enzyme from *Alkalilimnicola ehrlichii*

PET44 from *Alkalilimnicola ehrlichii* was also nominated as a poly-ester hydrolase that deserve structural analysis. The structural details are shown in Figure 10.

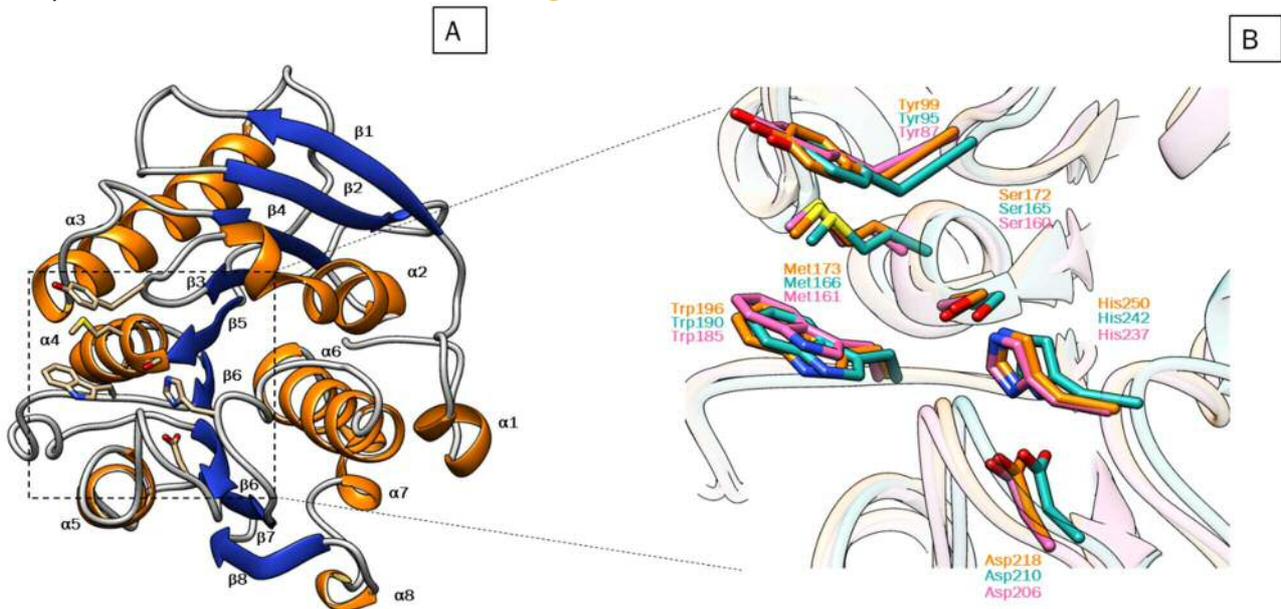


Figure 10. Crystal structure of PET44 (PDB code not yet released) (A) with active site alignment of LCC (PDB code 4EB0) and IsPETase (PDB code 6EQE) (B). Displayed is the overall structure of PET44 with its characteristic features.

The protein consists of 8  $\alpha$ -helices (orange) and 9  $\beta$ -sheets (blue). In figure B, the catalytic triad (Ser 172, Asp218, His250) and the binding triad (Tyr99, Met173, Trp196) are shown in orange. The catalytic triad (Ser165, Asp210, His242) and the binding triad (Tyr95, Met166, Trp190) for the LCC are displayed in green. The catalytic triad (Ser160, Asp206, His237) and binding triad (Tyr87, Met161, Trp185) of IsPETase are displayed in pink.

**20. 8U7F (not yet released to PDB)**

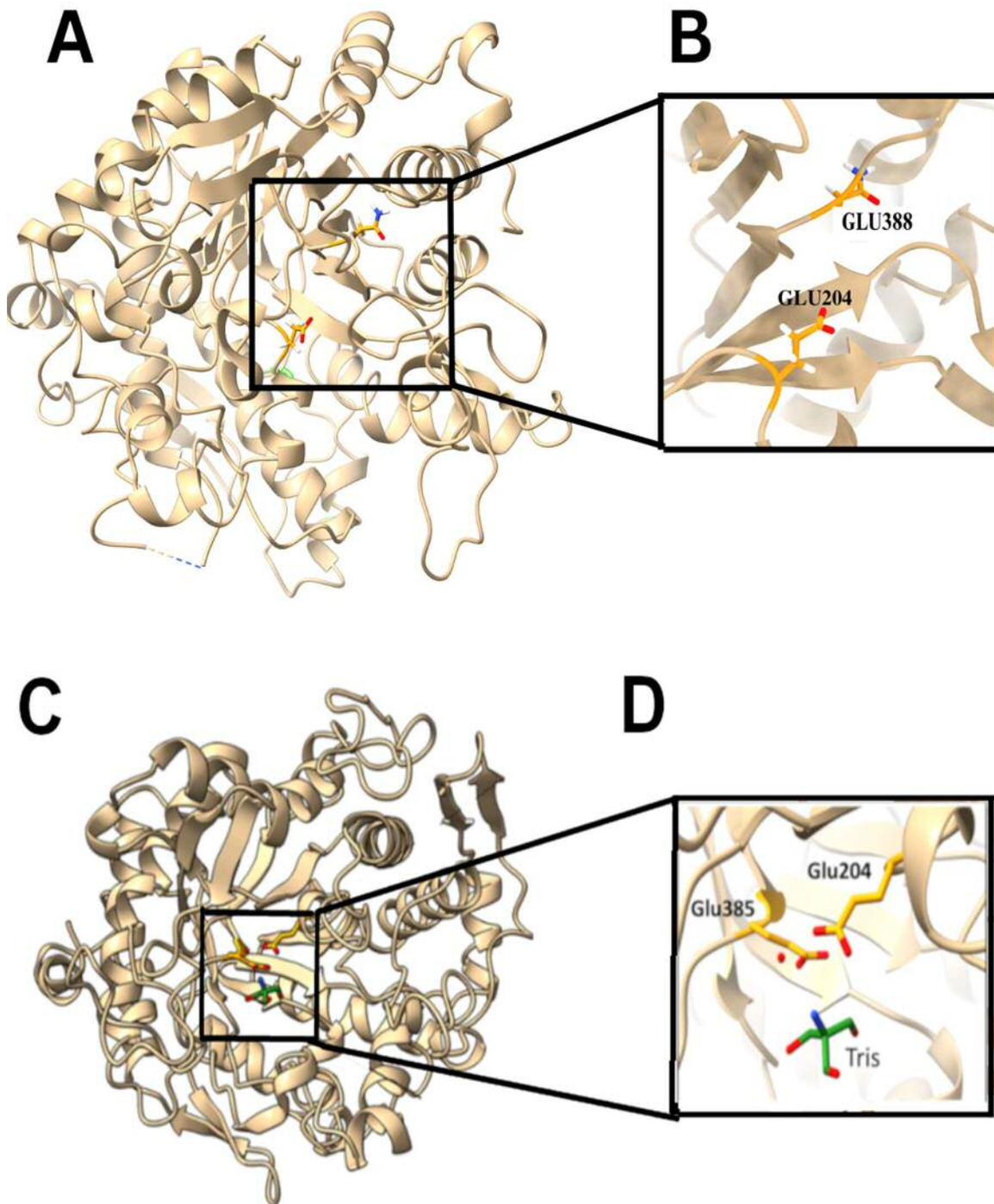
**CIB12 GH1 from the extremely acidophilic archaeon *C. divulgatum***

**21. 8U7G (not yet released to PDB)**

**CIB13 GH1 from the extremely acidophilic archaeon *C. divulgatum***

Two Glycosyl hydrolases (GH) from family 1 (GH1) from the extremely acidophilic archaeon *C. divulgatum*. *Cuniculiplasma divulgatum* S5 was isolated from the Cantareras AMD site in Spain. They were selected during the bioprospecting in WP3 because GH were considered also as secondary priority target for detergent and textile applications. Crystal structures of both CIB12 and CIB13 have been also determined and deposited to the PDB database with accession codes 8U7F and 8U7G, respectively. Structurally characterised enzymes from the GH1 family usually have a classical ( $\alpha/\beta$ )<sub>8</sub>-TIM barrel fold with two catalytic glutamate residues spatially conserved in the active site, with one of the glutamates acting as a nucleophile. Accordingly, the crystal structures of CIB12 and CIB13 revealed a classical ( $\alpha/\beta$ )<sub>8</sub> TIM barrel fold where 8  $\alpha$ -helices cover the internal 8  $\beta$ -strands forming the “barrel shape” (Figure 11A and C). Both were crystallised as homodimers which typically form a stable complex composed of two protein subunits. These subunits are monomers (protomers) that interact with each other via hydrogen bonds and other interactions which makes them more stable. Dimers can be stabilised by non-covalent interactions such as H-bonding, disulfide bridges, and electrostatic interactions. Dimer association constants which are also known as dimerization can vary significantly across wide value range from nanomolar to picomolar range. Solvent exposed  $\beta$ -strands can also play an important role in protein dimerization.

In CIB13 and CIB12, two short side  $\beta$ -strands from each monomer are providing 4 hydrogen bonds between two subunits including Leu456-Trp478 and Lys453-Glu474 for CIB12 and Lys450-Glu471 and Trp475-Ile453 for CIB13. In addition, the residues located on alpha helices add two more hydrogen bonds between the protomers: Tyr69-Glu465 for CIB12 and Arg461-Glu462 for CIB13, whereas no cysteine bridges were identified in the contact area of both homodimers. The crystal structures of both dimers also revealed the presence of additional electron density in each monomer located near the predicted catalytic glutamates, which were interpreted as glycerol (two molecules in the CIB12 dimer and one molecule in CIB13) and tris(hydroxymethyl)aminomethane (Tris, one molecule in the CIB13 dimer). Therefore, the crystal structures of both CIB12 and CIB13 confirmed the position of their catalytic glutamates (Glu204 and Glu388 for CIB12 and Glu204 and Glu385 in CIB13) predicted based on protein sequence analysis. In CIB12, the bound glycerol molecule is located close to His148 (3.1-3.1 Å), Trp149 (4.4-4.5 Å), Glu204 (3.3-3.9 Å), Tyr321 (3.9-4.3 Å), Glu388 (2.7-2.8 Å), Glu433 (2.9-3.1 Å), Trp426 (3.6-4.1 Å), and Trp434 (3.4-3.6 Å), whereas in CIB13 the bound ligands (glycerol and Tris) are positioned near His148 (3.2-5.6 Å), Trp149 (3.7-4.2 Å), Glu204 (3.2-3.5 Å), Tyr320 (2.6-3.3 Å), Glu385 (2.8-2.9 Å), Trp423 (3.2-3.7 Å), Glu430 (2.5-3.4 Å), and Trp431 (2.9-3.2 Å). In both enzymes, the substrate binding pockets are wide open and mostly hydrophilic (Figure 11A) with a few hydrophobic patches formed by Trp149, Tyr321, and Tyr341 in CIB12. In CIB13, there are two hydrophobic patches formed by Tyr320, Phe340, Leu341, Trp359, Trp432, as well as by Trp30, Trp33, Trp149, Val207, and Trp431 (Figure 11B).

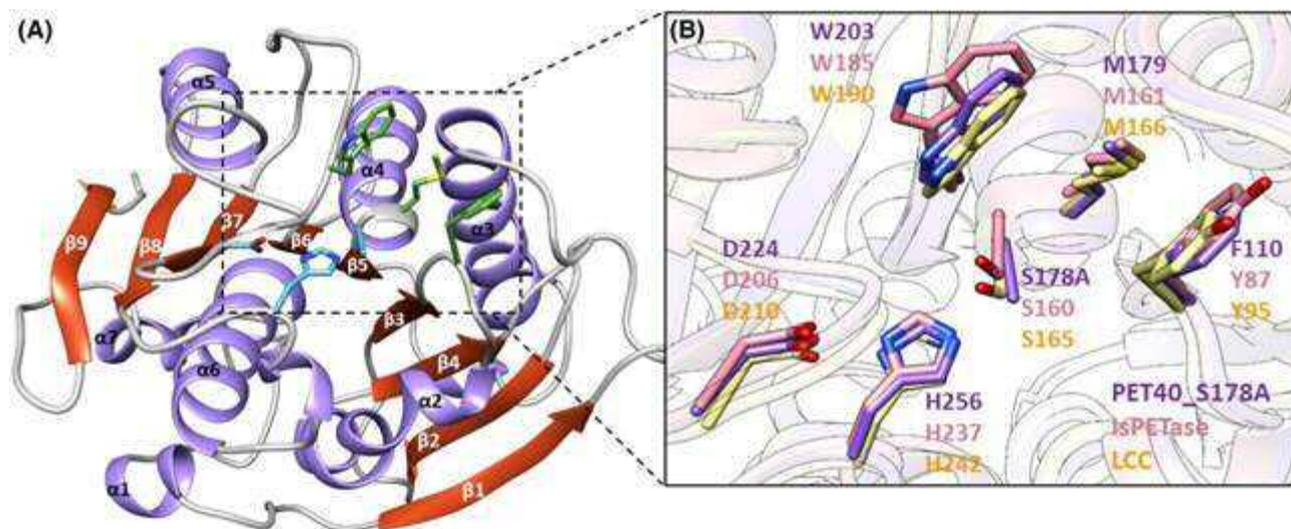


**Figure 11.** Crystal structures of CIB12 and CIB13: overall fold of the CIB12 (A) and CIB13 (C) monomers showing a TIM barrel fold. (B and D), Close-up view of the active site pockets with catalytic glutamates (B, CIB12; D, CIB13).

**22. 8A2C (<https://www.ncbi.nlm.nih.gov/Structure/pdb/8A2C>)  
 PET40 (S178A mutant), PETase enzyme from an unclassified *Amycolatopsis***

PET40 is a cutinase-like enzyme derived from a GC-rich Gram-positive *Amycolatopsis* species. The enzyme's crystal structure was determined, revealing stability across diverse conditions. PET40 exhibited high substrate promiscuity, showcasing hydrolytic activity on PET as well as various other substrates, including  $\delta$ -lactones,  $\beta$ -lactam antibiotics, polyester-polyurethane Impranil® DLN, and para-nitrophenyl esters. Molecular docking studies provided insights into PET40's versatile binding modes. The enzyme's ability to

maintain activity in the presence of sodium dodecyl sulphate (SDS) further highlighted its potential for diverse applications, including detergent and textile applications. The structure of PET40 is shown in Figure 12.



**Figure 12.** Crystal structure of PET40\_S178A and alignment with known IsPETase and LCC. Cartoon representation showing the secondary structure elements and the active site (box; A). The catalytically active amino acids are colored in blue, and the substrate binding site is colored green. In the solved crystal structure the catalytic serine was mutated to alanine. (B) The residues of the active site of PET40\_S178A (purple), IsPETase (pink; PDB code 6EQE), and LCC (yellow; PDB code 4EB0) are overlaid. The amino acids related to the catalytic activity are labeled. The catalytic triad is conserved within all three structures (Ser; Asp; His), but PET40 has Phe instead of Tyr at the first position of the substrate binding site.

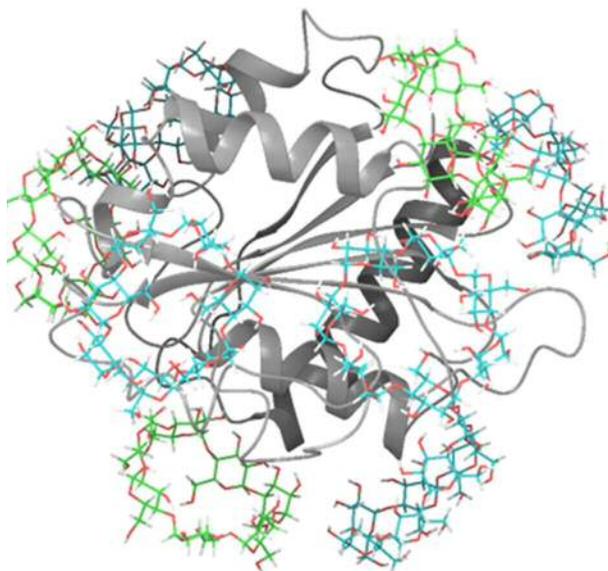
## 5. Comprehensive Structural Analysis of Enzymes in FuturEnzyme: AlphaFold-Derived Data

For structures lacking an experimentally determined solution, we utilized the AlphaFold 2.0 deep learning algorithm to generate a model. We employed five pre-defined pTM models from AlphaFold, each fine-tuned to provide not only structure predictions but also predicted TM-score (pTM) and predicted aligned error (PAE) values. The maximum template search date was 2021-05-14 with full databases preset. We then selected the best model based on the pTM score. The chosen model, ranked first, was used for subsequent analyses. The Predicted Template Modeling Score (pTM) is a metric to evaluate the structural congruency between two folded protein structures, as detailed in Jumper et al 2021 and Varadi et al 2022. AlphaFold assigns a confidence estimate, denoted as pLDDT, to each residue, ranging from 0 to 100, reflecting the model's predicted score on the IDDT-C $\alpha$  metric. This confidence measure is stored in the B-factor fields of the mmCIF and PDB files available for download, with higher pLDDT values indicating increased confidence (in contrast to B-factors, where higher values suggest lower confidence). Regions with pLDDT < 70 are classified as low confidence, necessitating careful interpretation. To enhance simulation quality, regions with extreme values indicative of lower quality are excluded. In the bioprospecting context, our methodology encompassed a sequence-based similarity search within the AlphaFold Database using the Protein Basic Local Alignment Search Tool (BLASTP). This tool compared the provided protein sequence query with sequences in the AlphaFold Database, resulting in a list of AlphaFold predictions exhibiting similarity to the user-provided sequence. Subsequently, proteins were retrieved from the database, each possessing precomputed AlphaFold models.

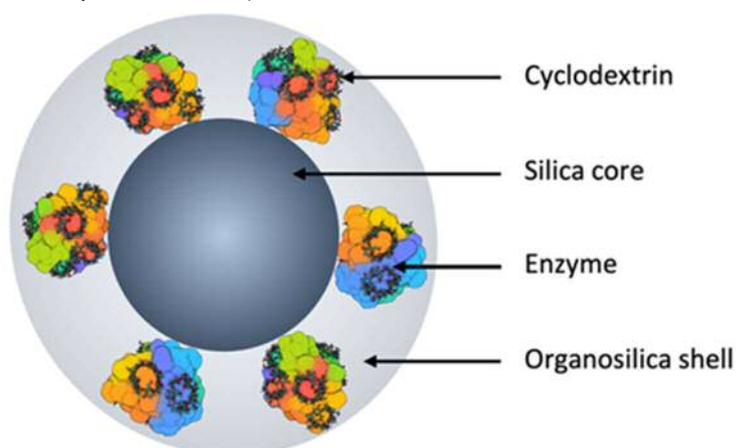
In brief, we constructed 678 AlphaFold2 models and checked the structural quality of the model with ProSA-web that compares the structure models with experimentally determined proteins from Protein Data Bank and estimates a Z-score for each model, the lower the better. The following 16 AlphaFold2 models have been created by AlphaFoldDB by UDUS partner: D3XB96, Q0VT77, Q0VLP6, A0A218VKW9, A0A1H6BYA1, A0A1H5S3T6, G1ARD1, A0A0K0LBH0 and A0A0K0LCF8 for ester-hydrolases, and A0A1H1UUH5,

AOA031MKR8, AOA1B0Z6Y3, AOA2P4EZRO, AOA1H1PFY8, AOA1H1RJ19 and AOA1S8DKK6 for poly-ester hydrolases. The represent models for the following enzymes Dim\_#008, Abo\_Est3, Abo\_LipD, Paes\_TB035, Paes\_TB045, Paes\_Est3, ED30, TBec304, TBec310, Psab\_PE-H, Pbau\_PE-H, Ppel\_PE-H (PpelaLip), Poce\_PE-H, Pxin\_PE-H, Plit\_PE-H and Ppac\_PE-H, some of which were selected as priority targets for pre-industrial validations, after the model and experimental data (Dim\_#008, Paes\_TB035, TBec304, Pbau\_PE-H, Poce\_PE-H). The other 662 AlphaFold2 models have been created by partners BSC and CISC and are available at [Annex, Figure 1](#), and have been used for several objectives: (i.) to determine in WP2 through molecular modeling and calculations of catalytic events the enzymes to pre-select for undergoing synthesis and characterization; (ii.) to stablish mutations through rational design or advance computational tools or define the basis for supramolecular engineering in WP5. Relevant examples include the lipases Lip9, Lip5 and Polur1, and the hyaluronidases VD\_PL9, VD\_PL22, VA\_PL9 and Hyal\_HRDSV\_2334 selected as priority targets for pre-industrial validations, to cite some relevant examples. For details see [Annex, Table 1](#).

Below, an example of the lipase Lip9 selected as priority target for textile and detergent applications is shown (see [Figure 13](#)). The [Figure 13](#) shows how through the analysis of the AlphaFold structure of lipase Lip9 and PELE simulations the distinct binding sites for  $\beta$ -cyclodextrin ( $\beta$ -CD) are identified. This allows the stability of the lipase Lip9 through exploiting the artificial protein chaperone properties of  $\beta$ -CD, when the Lip9-  $\beta$ -CD are embedded in an ultrathin organosilica layer ([Figure 14](#)). This supramolecular engineering approach belong to the activities performed in WP5.

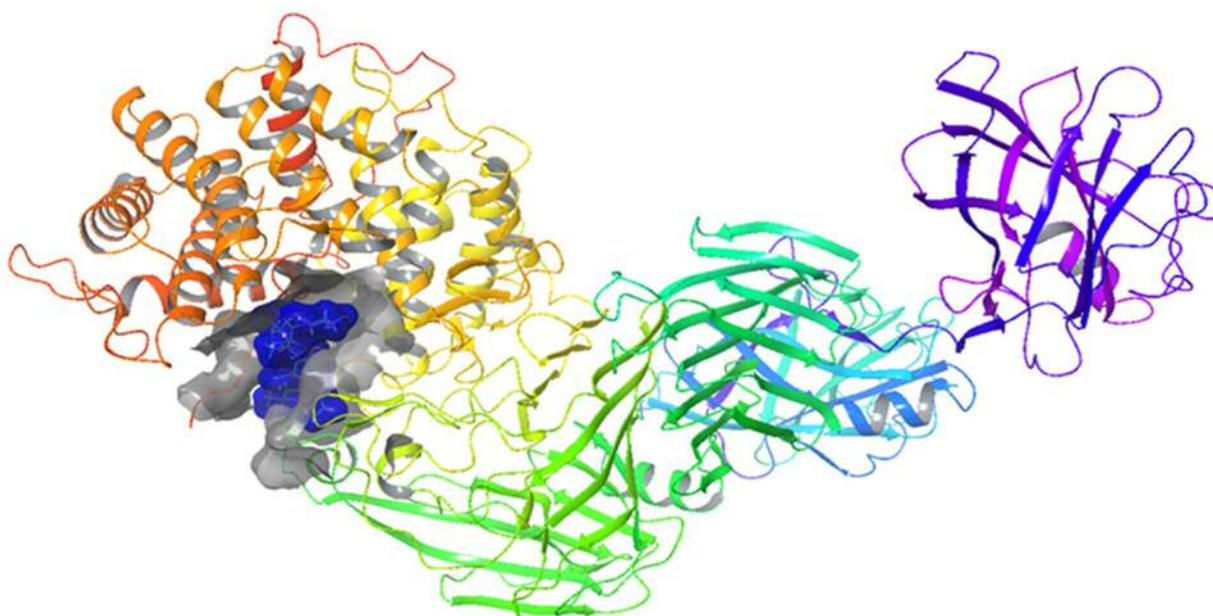
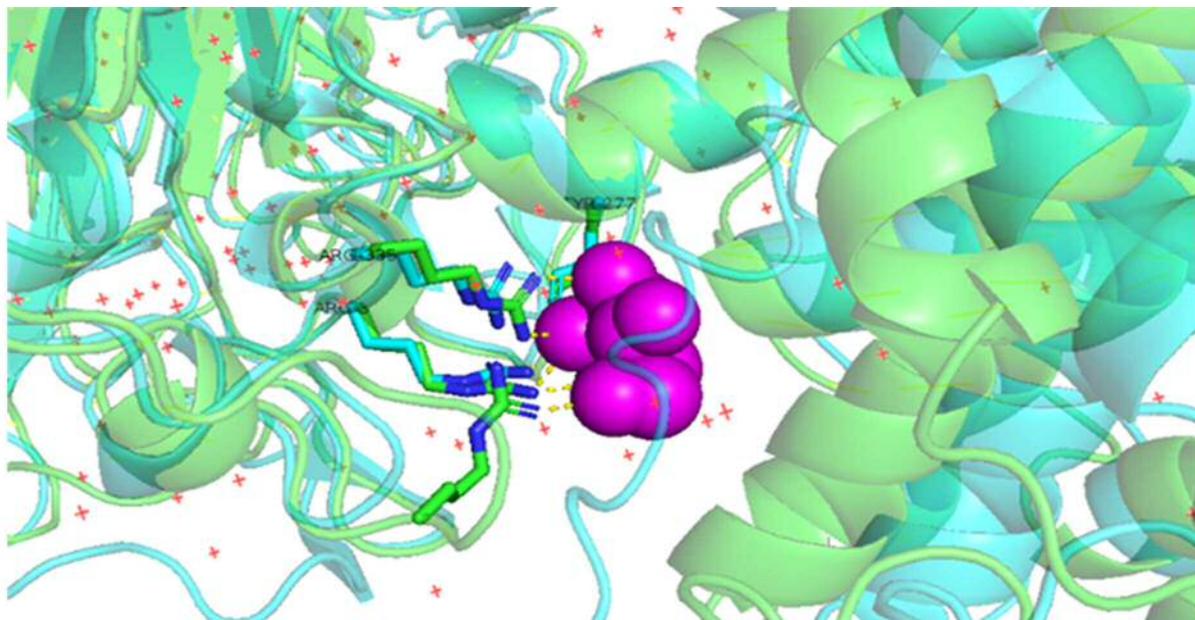


[Figure 13](#). Visualisation of the Lip9 bound to 8  $\beta$ -CD, showcasing the binding sites identified in this subsequent simulation round in cyan, and the previously selected in green. The AlphaFold model is shown (<https://alphafold.ebi.ac.uk/entry/A0A4Y9JWV9>).



[Figure 14](#). Schematic representation of chaperone-stabilised Lip9 enzyme embedded in an organosilica layer.

In [Figure 15](#), the example of the AlphaFold model for hyaluronidase LC1Hm\_4133, selected as priority target for cosmetic applications, is also shown. The model was used to see which amino acids are in contact with the substrate, and confirm the presumptive capacity of this enzyme to degrade hyaluronic acid. This supramolecular engineering approach belongs to the activities performed in WP2.



[Figure 15](#). AlphaFold model of hyaluronidase LC1Hm\_4133 to see which amino acids are in contact with the substrate. We performed a docking (swissdock) with hyaluronic acid obtained from chemspider directed against residue Arg331 atom Ccz with a 10 angstroms window.

## 6. Other Resolved Structures within the Framework of the Project

In addition to generating our own structural data in this project through the structural resolution of proteins or AlphaFold models, we have also utilized resolved protein structures from existing literature. These structures have served as a foundation for the development and design of enhanced mutants, as well as for the application of the project's computational tools. Below, a relevant example is highlighted.

## 1. 4TSY and 3W9P

Homo-octamer biological assembly crystal structure of FraC (4TSY;

<https://www.ncbi.nlm.nih.gov/Structure/pdb/4TSY>) and its monomer soluble structure (3W9P;

<https://www.ncbi.nlm.nih.gov/Structure/pdb/3W9P>)

Finally, 1 structure of a resolved structure of a non-catalytic pore-forming protein, namely Fragaceatoxin C (FraC) from *Actinia fragacea*, have been used as starting point for implementing novel enzymedesigns. More in details, during the Project we used computational tools to introduce, following the approach for the PluriZyme design, into a non-catalytic pore-forming protein, namely Fragaceatoxin C (FraC) from *Actinia fragacea*, a hydrolysing active centers consisting on an artificial catalytic triad (Ser, Asp, His) and an oxyanion hole to stabilize the charge that appears during hydrolysis; we though introducing such elements with a geometry similar to that of a benchmark PETase, thinking that by having a similar arrangement it would be possible to transform a non-catalytic nanopore into a polyethylene terephthalate (PET)-capturing and degrading one. Following this approach, with computational methods, we found in the three-dimensional structure described in the literature of FraC (4TSY; <https://www.ncbi.nlm.nih.gov/Structure/pdb/4TSY>), two potential binding sites for PET and strategically position new amino acids for effective hydrolysis (Figure 16).

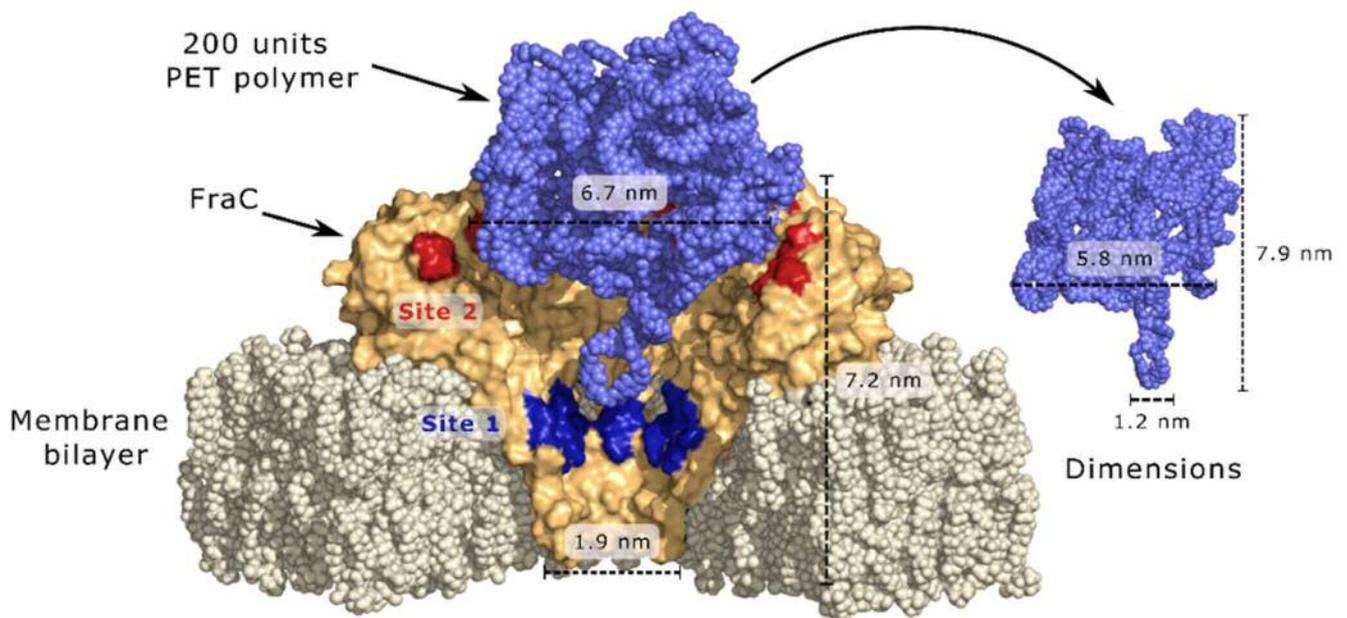


Figure 16. Graphical representation of a schematic coupling between the pore and the PET particle polymer. The artificial sites 1 (FraC<sub>m1</sub>) and 2 (FraC<sub>m2</sub>) are color coded.

Through these structural and computational analysis we have designed two engineered FraC proteins, FraC<sub>m1</sub> (Ser21, His20, Asp17) and FraC<sub>m2</sub> (Ser38, His175, Glu40) that when assembled into nanopores (npFraC<sub>m1</sub> and npFraC<sub>m2</sub>) efficiently capture and degrade PET microplastics at room temperature. The structural analysis further helped deeping into the degradation mechanism, and found that PET particles exhibit dynamic shapes with multiple protrusions, and that it's not the particles themselves that enter the nanopores, but rather these protrusions. As the protrusions break down, the particle size decreases, reducing their ability to enter the nanopores and undergo degradation. Depending on the placement of this artificial active center and the accessibility of the protrusions, we achieved varied degradation pathways for the nanoparticles. Based on these results, a PluriZyme with two PET-degrading active sites, FraC<sub>m3</sub>, have been generated and the experimental characterization is in progress. For methodological details and detailed results, see Robles-Martín et al 2023.

## 7. Conclusions

To date, we have successfully generated: (i.) 11 structures for 9 enzymes; (ii.) 4 structures of 2 different PluriZymes; (iii.) 1 structure of a biomimetic catalyst: an ester hydrolase (EH3) that has been covalently modified with a bipyridinyl derivative at its active center; (iv.) 678 models using AlphaFold. Moreover, the resolved structure of a non-catalytic pore-forming protein, Fragaceatoxin C (FraC) from *Actinia fragacea*, has been utilized as a foundation for pioneering new enzyme designs. Based on these accomplishments, we consider Deliverable D4.5 to have been successfully achieved.

## 8. References

- Cea-Rama I, Coscolín C, Gonzalez-Alfonso JL, Raj J, Vasiljević M, Plou FJ, Ferrer M, Sanz-Aparicio J. Crystal structure of a family VIII  $\beta$ -lactamase fold hydrolase reveals the molecular mechanism for its broad substrate scope. *FEBS J.* 2022; 289(21):6714-6730.
- Distaso M, Cea-Rama I, Coscolín C, Chernikova TN, Tran H, Ferrer M, Sanz-Aparicio J, Golyshin PN. The Mobility of the cap domain is essential for the substrate promiscuity of a family IV esterase from *Sorghum* rhizosphere microbiome. *Appl Environ Microbiol.* 2023a; 89(1):e0180722.
- Distaso MA, Chernikova TN, Bargiela R, Coscolín C, Stogios P, Gonzalez-Alfonso JL, Lemak S, Khusnutdinova AN, Plou FJ, Evdokimova E, Savchenko A, Lunev EA, Yakimov MM, Golyshina OV, Ferrer M, Yakunin AF, Golyshin PN. Thermophilic Carboxylesterases from hydrothermal vents of the volcanic island of ischia active on synthetic and biobased polymers and mycotoxins. *Appl Environ Microbiol.* 2023b; 89(2):e0170422.
- Fernandez-Lopez L, Cea-Rama I, Alvarez-Malmagro J, Ressmann AK, Gonzalez-Alfonso JL, Coscolín C, Shahgaldian P, Plou FJ, Modregger J, Pita M, Sanz-Aparicio J, Ferrer M. Transforming an esterase into an enantioselective catecholase through bioconjugation of a versatile metal-chelating inhibitor. *Chem Commun (Camb).* 2023; 59(62):9469-9472.
- Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr.* 2010; 66(Pt 4):486-501.
- Evans PR, Murshudov GN. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr.* 2013; 69(Pt 7):1204-14.
- Jumper, J et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021; 596(7873):583-589.
- Kabsch W. XDS. *Acta Crystallogr D Biol Crystallogr.* 2010; 66(Pt 2):125-32.
- Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr.* 1997; 53(Pt 3):240-55.
- Robles-Martín, A., Amigot-Sánchez, R., Fernandez-Lopez, L. et al. Sub-micro- and nano-sized polyethylene terephthalate deconstruction with engineered protein nanopores. *Nat Catal* (2023). <https://doi.org/10.1038/s41929-023-01048-6>.
- Roda S, Fernandez-Lopez L, Benedens M, Bollinger A, Thies S, Schumacher J, Coscolín C, Kazemi M, Santiago G, Gertzen CGW, Gonzalez-Alfonso JL, Plou FJ, Jaeger KE, Smits SHJ, Ferrer M, Guallar V. A Plurizyme with transaminase and hydrolase activity catalyzes cascade reactions. *Angew Chem Int Ed Engl.* 2022; 61(37):e202207344.
- Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* 2010; 31(2):455-61.
- Vagin A, Teplyakov A. MOLREP: An Automated Program for Molecular Replacement. *J Appl Cryst.* 1997; 30, 1022-1025.
- Varadi, M et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* 2022; 50(D1):D439-D444.
- Wang D, Zou L, Jin Q, Hou J, Ge G, Yang L. Human carboxylesterases: a comprehensive review. *Acta Pharm Sin B.* 2018; 8(5):699-712.

## 9. Annex

**Table 1.** List of the 678 proteins for which AlphaFold models were constructed.

See intranet's project website in [www.futureenzyme.eu](http://www.futureenzyme.eu) -> login -> private-area -> shared-data -> Annex -> Annex\_Table 1 for D4.5\_At least 9 enzyme crystal structures

**Figure 1.** Alphafold2 models created by AlphafoldDB. To visualize a PDB (Protein Data Bank) structure, one can use various molecular visualization software tools such as PyMOL, Chimera, or Maestro. The process typically involves importing the PDB file into the chosen software. In most visualization software, users can open the application and use the 'File' menu to load the PDB file. Once loaded, users can manipulate the molecular structure with commands for zooming, rotating, and translating to gain different perspectives. Additionally, when visualizing an AlphaFold-generated PDB structure, the pLDDT (predicted local distance difference test) values can be effectively displayed by color-coding the structure based on its B-factor values. This is achievable within visualization tools, as the B-factor field in PDB files corresponds to the pLDDT confidence measure. By employing a color gradient, users can discern regions with varying levels of confidence, which allows for the identification of loops or extremities at the N- or C-terminals, where lower confidence is typically observed. This visualization method offers valuable insights into the reliability of AlphaFold predictions, enabling a more comprehensive analysis of the modeled protein structure.

See intranet's project website in [www.futureenzyme.eu](http://www.futureenzyme.eu) -> login -> private-area -> shared-data -> Annex -> Annex\_Figure 1 for D4.5\_At least 9 enzyme crystal structures