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

23/12/2022

SET OF 180 ENZYMES FOR EXPERIMENTAL FOCUS

D2.4

VÍCTOR GUALLAR
BSC
Jordi Girona 31, BARCELONA 08034,Spain

Document information sheet

Work package:	WP2, Machine learning enzyme bioprospecting integrated into an industrial context
Authors:	BSC (Víctor Guallar), CSIC (Manuel Ferrer, Patricia Molina)
Document version:	2
Date:	23/12/2022
Starting date:	01/06/2021
Duration:	48 months
Lead beneficiary:	BSC
Participant(s):	BSC, CSIC, BANGOR, UHAM, UDUS, CNR, IST-ID, SCHOELLER, HENKEL, EVO
Dissemination Level:	Confidential, only for consortium's members (including the Commission Services)
Type	Other
Due date (months)	14
Contact details:	Víctor Guallar (victor.guallar@bsc.es), Manuel Ferrer (mferrer@icp.csic.es), Patricia Molina (patricia.molina@icp.csic.es)

Summary

SET OF 180 ENZYMES FOR EXPERIMENTAL FOCUS

1. Scope of Deliverable

This deliverable consists in a list of at least 180 full-length top priority candidate sequences encoding enzymes with high probability to fulfil manufacturer specifications (according to the initial proposal). These sequences have been selected in Task 2.4 by iterative and decision-making hierarchical procedure applied to the at least 1,000 full-length candidate sequences delivered in D2.3 (according to the initial proposal). The list of sequences pre-selected for experimental focus has been compiled in fasta or Excel tables deposited in the FuturEnzyme internal repository and also as a report that accompanies this deliverable (the present document) that also details the hierarchical procedure applied for the selection.

2. Reasons for the update

The first version of the Deliverable D2.4 was submitted in July 2022. This update is due to the fact that since the submission, the partners were able to pre-select a new set of sequences for experimental focus. This was a consequence of the increase number of sequences pre-selected in WP2-WP4 and that were detailed in Deliverables D2.2, D2.3, D3.3, D3.4, D4.2, D4.3, D4.4 and D4.6, the outcomes of which are summarized below. In November 2022, the Coordinator (Manuel Ferrer) contacted the Project Officer (Colombe Warin) to explain these circumstances and ask her to re-open the submission of this deliverable (amongst others), at which she agreed.

3. Origin of the deliverable

Along the already 18 months of project, several deliverables have been accomplished from which the present one nourishes or were key for their delivery. To be mentioned:

Deliverables in the frame of WP2:

- D2.2: Set of 250 000 sequences pre-selected (November 2021, updated December 2022)
  In this deliverable, information about the approximately 3.16 million sequences encoding target enzymes that were retrieved and pre-selected by a number of in silico methods, are detailed.
- D2.3: Set of 1000 enzymes selected using motif screens (May 2022; updated December 2022)
  In this deliverable, a number of bioinformatics and computational tools were applied that allow the pre-selection of approx. 1355 sequences encoding enzymes relevant to FuturEnzyme.
- D2.4: Set of 180 enzymes for experimental focus (July 2022; present update, December 2022)
  In this deliverable, at least 180 enzymes from the priority sequences retrieved in the frame of WP2 (deliverables D2.2, D2.3) and WP3 (deliverable D3.3), were preliminary selected to proceed with their cloning, synthesis, expression and characterization.

Deliverables in the frame of WP3:

- D3.3: Set of 100 clones, 10 isolates, 10 enzymes shortlisted for sequencing (March 2022; updated December 2022)
  In this deliverable, bio-resources available before the beginning of the project and newly generated during the project were screened by naïve/functional methods to identify those with interest for our project. Bio-resources include previous and new enzymes, environmental samples, isolates, enrichments, and clone libraries that were checked for the purpose of the present project, and the best selected ones sequenced and sequences with interest for our project were retrieved.
- D3.4: Sequence, activity, and stability datasets from best positive bio-resources (November 2022)
  This deliverable consists in the datasets informing about the sequences, performances and stabilities of best preselected bio-resources (isolates and clones).

Deliverables in the frame of WP4:

- D4.2: The FuturEnzyme portfolio of 1000 enzyme (recombinant/native/biomimetic) material, obtained (September 2022)
  In this deliverable, the expression, preparation and production of a set of protein samples of about 1000 enzymatic materials were undertaken by members of the consortium in a variety of hosts (heterologous or native) and vectors, cell-free systems, biomimetic metamorphosis systems, to name but a few, as well as genetically-engineered mutants and supramolecular-engineered (immobilized) enzymes generated in the frame of WP5.
- D4.3: Cell-free expression reported system developed (September 2022)
  In this deliverable, a cell-free expression system was developed that allow the production and detection of enzymatic activities in a high-throughput manner by skipping the step of recombinant expression.
- D4.4: Biomimetic protease production system, developed (September 2022, re-opened to be updated in month 30)
  In this deliverable a green chemical-system was designed that allows the production of enzymes with inherent problems of expression, particularly, biomimetic proteases.
- D4.6: The metadata on expression yield, activity and stability available (November 2022)
  This deliverable consists on the datasets informing about the expression yield, activity and stability of all enzymes generated in the project until month 18.

4. Methodology and Results

The starting point of this deliverable is multiple. From one side, the sequences obtained in the frame of WP2 (obtained after *in silico* screens of sequences, that included bioinformatics and computational methods) and from other side the sequences obtained in WP3 (obtained after functional screens of bio-resources), whose outcomes are listed in the deliverables mentioned in Section 3. A combination of both methods allowed the pre-selection of 678 enzymes (including microbes containing enzymes) for experimental focus (Annex **Table 1**). The ID name, the screening method by which each enzyme was retrieved, the enzyme class, the origin and the nature of the enzymatic material (native enzyme, mutant, immobilized preparation, biomimetic, protein extracts, etc.), the amino acid sequence, and the GPS coordinates (if available), have been extensively detailed in D4.6 “The metadata on expression yield, activity and stability available” (November 2022), and for this reason they are not repeated in this deliverable (see Deliverable D4.6 for details).

A hierarchical procedure was applied for the selection of the 678 enzymes (including microbes containing enzymes).

1. For those sequences retrieved from D2.3 by number of bioinformatics and computational tools (1355 candidates), only those with higher number of catalytic events, those computationally predicted to have thermal and stability features fitting manufacturer specifications (detailed in Deliverable D2.1 “Manufacturers' needs and specifications, protocol”), and those selected for validation of computational tools, were pre-selected for experimental focus.
2. For those sequences retrieved from D3.3 and 3.4 by a number of functional screens, only those having activity and stability phenotypes fitting manufacturer specifications (detailed in Deliverable D2.1 “Manufacturers' needs and specifications, protocol”) were pre-selected for experimental focus.

Annex 1

Annex Table 1_D2.4. Detailed information on expression, activity and stability of the available enzymes. Shown are, among other datasets: 1) Enzymatic activity; 2) Name of the candidate; 3) Screen method; 4) Expression host; 5) Expression level; 6) Amino acid sequence or genome sequencing status; 7) Origin; 8) Details of stability features including denaturing temperature (Td), detergent stability; 9) Details of activity features, including substrate profile, optimal temperature and pH, etc.; 10) Sequence homology. Document available under the designation *D4.6_Annex 1_Enzymes including stability data_FINAL* at the FuturEnzyme web intranet through the following QR code (password needed), in the section *Shared data, Datasets*: