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

30/11/2022

THE METADATA ON EXPRESSION YIELD, ACTIVITY AND STABILITY, AVAILABLE

D4.6

STEPHAN THIES
UDUS
Universitaetsstrasse 1, Düsseldorf 40225, Germany

Document information sheet

Work package:	WP4, Small-scale enzyme production and characterization.
Authors:	Stephan Thies (UDUS), Manuel Ferrer (CSIC)
Document version:	1
Date:	30/11/2022
WP starting date:	01/06/2021
WP duration:	40 months
WP lead:	UHAM
WP participant(s):	UHAM, CSIC, Bangor, UDUS, IST ID, CNR, FHNW, Bio_Ch, Eucodis
Deliv lead beneficiary:	UDUS
Dissemination Level:	Confidential, only for members of the consortium (including the Commission Services)
Type:	Data sets, microdata, etc.
Due date (months):	18
Contact details:	Stephan Thies (s.thies@fz-juelich.de), Manuel Ferrer (mferrer@icp.csic.es)

Summary

The metadata on expression yield, activity and stability available

1. Scope of Deliverable

This deliverable consists on the datasets informing about the expression yield, activity and stability of all enzymes generated in the project until month 18; the datasets will be continuously updated as new sets of enzymes become available. This information is available in the internal FuturEnzyme repository though a QR code.

2. Origin of the material

Along the already 18 months of the project, different deliverables have been accomplished from which the present one nourishes. To be mentioned:

Deliverables in the frame of WP2:

- D2.2_Set of 250 000 sequences pre-selected (November 2021; to be updated)
  In this deliverable, a set of about 3.2 sequences with interest for our project was retrieved by BLAST and HMM search, complemented with computational screens.
- D2.3_Set of 1000 enzymes selected using motif screens (May 2022; to be updated)
  In this deliverable, sequences retrieved in the frame of WP2 (deliverable D2.2) and WP3 (deliverable D3.3) were subjected to a filtering pipeline comprising the following criteria: confirmation of full-length sequence, presence and conservation of all proper domains and catalytic residues (along with MSA), the 3D structure modeling using AlphaFold 2.0, substrates (specified by the manufacturers) docking using Glide software (Schrödinger company) in the active site of the enzymes, and the substrate positioning around the active site with PELE (Protein Energy Landscape Exploration) software from BSC. The pipeline also includes the analysis of sequence coverage, the homology with reported similar sequences, the pair-wise similarity, and the network analysis using the MCL (Markov Cluster Algorithm) algorithm. The idea of this pipeline was to select the priority enzymes to work with in WP4.
- D2.4_Set of 180 enzymes for experimental focus (July 2022; to be updated)
  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; to be updated)
  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.

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.

3. Description of enzymatic materials from which datasets were generated

The **Table 1** (access provided in the Annex section), compiled all the relevant information regarding all the 678 enzymatic materials from which datasets were generated until month 18. This information includes: The ID name, the screening method by which each enzyme was retrieved (**Table 2**), the enzyme class (**Table 3**), the origin and the nature of the enzymatic material (native enzyme, mutant, immobilized preparation, biomimetic, protein extracts, etc.) (**Tables 4-5**), the expression host and the expression level (**Table 6**), the amino acid sequence, the GPS coordinates (if available), the physicochemical properties (denaturing temperature, optimal temperature and pH), and the activity towards a variety of standard commercial substrates and real substrates relevant for the project.

In summary, a total of 678 enzymatic preparations were generated and tested. The enzymes exhibited maximum amino acid sequence identities ranging from 100 to 36.9% to homologous proteins in public databases, with an average value (reported as %, with the interquartile range (IQR) in parentheses) of 88.6% (20.3%). The pairwise amino acid sequence identity for all enzymes ranged from 11.9 to 99.7%. Taken together, the primary sequence analysis suggests that the diversity of enzymes is not dominated by a particular type of protein or highly similar protein clusters, but consists of diverse non-redundant sequences assigned to multiple folds and subfamilies, which are distantly related to known homologues in many cases. The geographic distribution of the enzymes included in this deliverable is shown in **Figure 1**.

Table 2. Summary of methods by which each enzyme was retrieved.

Screen method	Nr
Homology screen	349
Functional screen	257
Homology and computational screen	47
Computational screen	18
Literature screen	4
Supramolecular screen	2
Biomimetic "metamorphosis"	1
TOTAL	678

Table 3. Summary of enzyme classes.

Enzyme class	Nr
Esterase/lipase	494
Glycosyl hydrolase	68
Hyaluronidase	41
Polyester hydrolase	29
Laccase/monooxygenase/oxidase	21
Esterase/lipase/polyester or plastic degrading hydrolase (i.e. PETase)	13
Amidase	6
PluriZymes	3
Protease/peptidase	3
TOTAL	678

Table 4. Summary of origin of each enzymatic material.

Origin	Nr
Metagenome (environmental sample)	271
Metagenome (public database)	184
Genome	155

Isolate	19
Isolate + Genome	19
Metagenome (enrichment)	16
Metagenome (anaerobic digestor)	11
Biomimetic	1
Immobilization	2
TOTAL	678

Table 5. Nature of the of the enzymatic material

Wild type or mutant	Nr
Wild type enzymatic material	662
Mutants	13
Biomimetic	1
Immobilized	2
TOTAL	678

Table 6. Summary of expression/production systems and expression/production level

Expression/production system	
Escherichia coli	623
Native hosts	52
Biomimetic "metamorphosis"	1
Immobilization	2
TOTAL	678
Expression behaviour	
Soluble (expression +)	570
Soluble (low expression)	60
Inclusion body	48
TOTAL	678
Percentage of soluble enzymes	93%

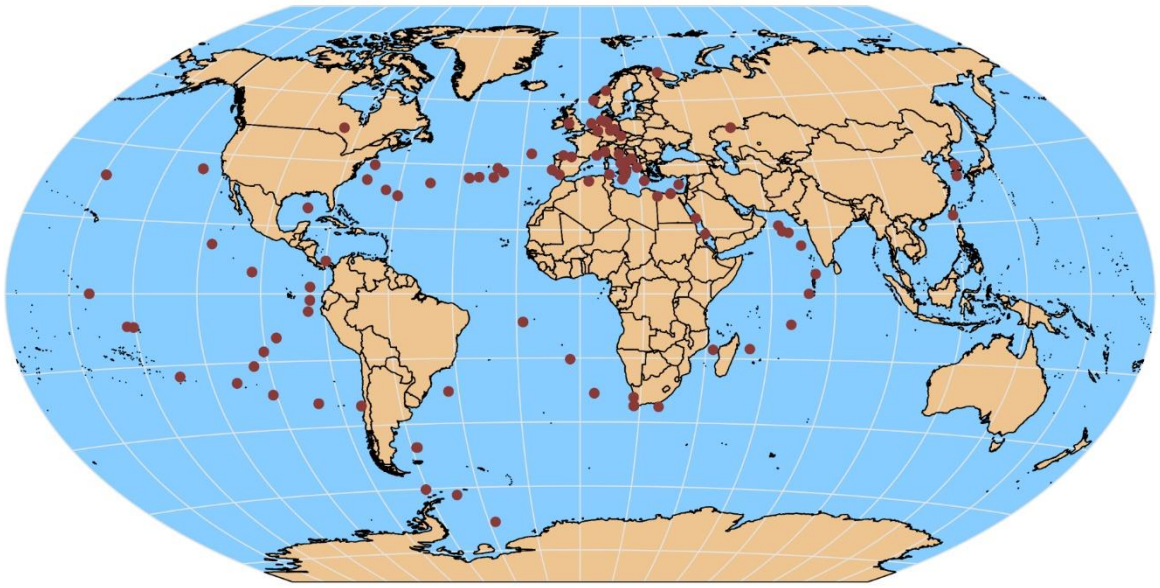

Figure 1. Geographic distribution of the enzymes included in this deliverable.

4. Methodology

Materials. Enzymatic materials were produced and expression yields were determined using the host and strategies described in the deliverable D4.2 "The FuturEnzyme portfolio of 1000 enzyme (recombinant, native, biomimetic) material, obtained".

Biochemical assays. Hydrolytic activity was determined, in multi-titer plates using appropriated spectrophotometers and other analytical techniques, by using multiple methods and substrates detailed in deliverable D3.2 "Standard assays, analytics and calculations for monitoring enzymatic performance". In brief, these methods include:

1) Measuring the amount of *p*-nitrophenol released by catalytic hydrolysis of *p*-nitrophenyl (*p*-NP) esters and glycosides through modified methods reported by partners (see publications at the Zenodo Community). *p*-NP substrates include:

   i) *p*-NP esters such as *p*-NP acetate, *p*-NP propionate, *p*-NP butyrate, *p*-NP hexanoate, *p*-NP octanoate, *p*-NP decanoate, *p*-NP dodecanoate, *p*-NP myristate and *p*-NP palmitate

   ii) *p*-NP sugars as *p*-NP L/D-galactopyranoside, *p*-NP beta-D-galactopyranoside, *p*-NP L/D-glucopyranoside, *p*-NP beta-D-glucopyranoside, *p*-NP L/D-mannopyranoside, *p*-NP L/D-maltopyranoside, *p*-NP beta-D-maltopyranoside, *p*-NP L/D-xylopyranoside, *p*-NP beta-D-xylopyranoside, *p*-NP L/D-glucuronide, *p*-NP beta-D-galactopyranoside tetra-acetate, *p*-NP L/D-ribofuranoside, *p*-NP N-acetyl-beta-D-glucosaminide, *p*-NP L-arabinopyranoside, *p*-NP beta-D-cellobiopyranoside, *p*-NP L-fucopyranoside, *p*-NP L-rhamnofuranoside, *p*-NP beta-D-arabinofuranoside, *p*-NP L/D-maltohexanoside, and *p*-NP N-acetyl-beta-D-galactosaminide.
2) Measuring the amount of acid released after the hydrolysis of short, medium and long-chain triglycerides, coconout oil, palm oil, and olive oil emulsified with gum arabic.
3) Measuring hydrolytic activity towards oil and fat stains from real-life stained swatches (**Figure 2**) outlined in D2.1 "Manufacturers' needs and specifications", using a high-throughput free fatty acid concentration measurements NEFA kit. These substrates, relevant for detergent applications, included:
   - C-S-61, Beef fat, coloured with Sudan red on Cotton - 45 cm width
   - PC-09, Pigment with oil (below 60°C) on Polyester/Cotton - 90 cm width
   - C-S-05S, Mayonnaise with carbon black on Cotton - 90 cm width
   - C-S-10, Butterfat with colourant on Cotton - 45 cm width
   - P-S-16, Lipstick, pink on Polyester/Cotton - 45 cm width
   - C-S-17, Fluid make-up on Cotton - 90 cm width
   - PC-S-132, High discriminative sebum BEY with pigment on Polyester/Cotton - 45 cm width.

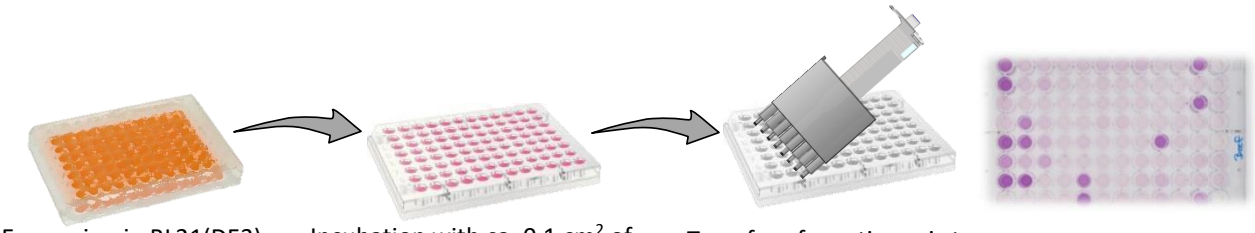

Figure 2. Workflow of assaying activity towards standard stained fabrics and assay plate showing active enzymes.

4) Measuring the cleaning efficiency, as for real-life stained swatches (**Figure 2**), of residual spinning oils of synthetic materials (PES, PA and EL) / outlined in D2.1 "Manufacturers' needs and specifications", using a high-throughput free fatty acid concentration measurements NEFA kit. These substrates, relevant for textile applications, included:
   - 61488, 61488Z ROH, 61488Z VORB, 92% PA, 8% EL 180g/m^2
   - 61988, 61988F1 ROH, 61988F1 VORB, 92% PA, 8% EL 280g/m^2 – fatty acid esters, fatty acid amides, little silicone

- 67007, 67007 ROH, 67007 VORB, 88% PA,12% EL 135g/m^2 – mineral oil, paraffin, fatty acid ethoxylates, fatty acid amides, silicone
- 3X58, 2X34G ROH, 3X58 VORB, 100% PES 100g/m^2 – fatty acid esters, mineral oil, paraffin
- 66299, 5237/00 ROH, 92% CO, 8% EL 240g/m^2
- E03130, E03130 ROH, E03130 VORB, 80%PA6, 20%EL – fatty acid ethoxylates, fatty acid amides, silicone.

5) Measuring the degradation of polymeric sugars and hyaluronic-based substrates outlined in D2.1 "Manufacturers' needs and specifications", using colorimetric and HPLC methods. These substrates, relevant for cosmetic applications, included:
   - α-cyclodextrin, β-cyclodextrin, ϒ-cyclodextrin, acarbose, pullulan, maltotriose, L-lactose, D-xylan, gum arabic, maltose, pectin, sucrose, trehalose, gellan gum, agarose, chitin, dextrin, dextran, amylose, starch, xanthan gum, hyaluronate, carboxymethyl cellulose, hyaluronic acid in MDa (Sigma-Aldrich)
   - Hyacare, High molecular weight (HMW, 700 kDa) hyaluronic acid produced after fermentation with *B. subtilis*
   - Hyacare50, Low molecular weight (LMW, 50 kDa) hyaluronic acid HyaCare® 50.
6) Measuring the degradation of dyes and related substrates outlined in D2.1 "Manufacturers' needs and specifications", using colorimetric methods. These substrates, relevant for textile applications, included:
   - ABTS, nitrophenyldiamine and sinapic acid
   - Real-life dye BEMAPLEX Black D-HF
7) Measuring the degradation of polyesters and plastics outlined in D2.1 "Manufacturers' needs and specifications", using colorimetric (i.e., turbidity decrease) and HPLC-UV/Vis analysis methods. These substrates, relevant for textile applications, included:
   - MHET (mono(2-hydroxyethyl) terephthalic acid), BHET (Bis(2-Hydroxyethyl) terephthalic acid), PET (polyethylene terephthalate), pre-treated amorphous PET foil, milled amorphous PET (UDUS) and pre-treated PET fabric, PET nanoparticles, Impranil DLN, polymeric polyesteramide and polylactic acid
   - Real-life pre-treated PET fabrics (Schoeller).

Determination of enzyme stabilities. To investigate stability and activities, the protocols detailed in deliverable D3.2 "Standard assays, analytics and calculations for monitoring enzymatic performance" were applied. The denaturation temperatures were determined by circular dichroism (CD) or nano-differential scanning fluorimetry (nanoDSF). CD spectra were acquired between 190 and 270 nm with a Jasco J-720 spectropolarimeter equipped with a Peltier temperature controller in a 0.1-mm cell at 25°C. The spectra were analyzed, and melting temperature (Tm) values were determined at 220 nm between 10 and 85°C at a rate of 30°C per hour at appropriate concentrations and buffers that ensure protein stability. Tm, and the standard deviation of the linear fit, was calculated by fitting the ellipticity (mdeg) at 220 nm at each of the different temperatures. Protein melting curves measured by nanoDSF were determined using a Prometheus device (NanoTemper Technologies, Inc.), according to the manufacturer's recommendation, with the purified enzymes at protein concentrations of 4-8 mg/mL. Detergent stability was assessed by incubating and assaying enzymatic materials in washing liquor prepared according to D2.1 "Manufacturers' needs and specifications" using model esterase substrates. Alternatively, enzymes were incubated with laboratory-standard detergents like SDS. Stability towards water-miscible organic solvents was determined as described in Bollinger *et al*. (Appl Environ Microbiol. 2020;86(9):e00106-20).

5. Enzymes assessed for expression conditions

As summarized in **Table 6**, a total of 678 enzymatic materials were generated and tested. They include: i) 623 enzymes which were assessed (after cloning or gene synthesis) for heterologous expression; ii) 52 enzymatic extracts from native hosts, the presence of enzymatic activities was confirmed by activity check; iii) one biomimetic with protease activity; and iv) two immobilized preparations. Strategies were established that allowed for soluble protein expression. From small- to medium-scale expression and purification culture volumes of 1-5000 mL, and in two cases 10 L, yields of 0.1 – 100 mg/L culture were obtained (See D4.2 "The FuturEnzyme portfolio of 1000 enzymes, recombinant, native, biomimetic, material, obtained"). As hosts, *E.*

coli (vectors pBXCH, pBXNH3, pEBP, pET21a, pET22b, pET-45b(+), pET-46 Ek/LIC, pQE306, p15TVL, pMCSG53, and pMGS68, to cite the most significant ones) and *Bacillus subtilis* (vector, pBSMul) were used. Protein materials from a number of native hosts were produced, including *Bacillus licheniformis*, *Bacillus subtilis subsp. Spizizenii*, *Bacillus vallismortis*, *Burkholderia multivorans*, *Citrobacter freundii*, *Haloarcula hispánica, Haloarcula salina HA08-SCL, Halobacterium salinarum, Haloferax mediterranei HA22-RCL, Halolactibacillus miurensis F4-4, Halorhabdus utahensis BNXHr, Halorhabdus utahensis SVXHr, Marinobacter sp., Martelella sp., Microbacterium oxydans, Nanohalococcus occultus SVXNc, Paenibacillus peoriae, Proteus mirabilis, Pseudomonas guineae, Pseudomonas protegens, Pseudomonas stutzeri, Psychrobacter celer, Serratia quinivorans, Stutzerimonas stutzeri,* and *Vibrio diabolicus*, to cite some. Overall, the percentage of soluble proteins was ca. 93% of the total. Immobilized preparations (particularly those of Lip9 and Polur1 lipases), were prepared at ca. 50-60 mg scale. This shows that the different protocols applied are effective for small-scale production and testing.

6. Enzymes with activity towards application-relevant substrates

For activity assessment, the partners focused on substrates of relevance for the envisaged application, which are summarized in Section 4. Many enzyme candidates were identified in the first place by functional screens on relevant substrates (naphthyl esters, triglycerides, HMW hyaluronic acid, Impranil-DLN polyester, etc.) using long-established screening strategies applied in the frame of WP3 (for details see deliverable D3.2 “Standard assays, analytics and calculations for monitoring enzymatic performance”). However, given the progress in sequence-analysis tools during the last decade and the developments within FuturEnzyme (e.g. AlphaFold, Hidden Markov-Model searches, prediction of the probability of catalytic events, etc.), about 60% of the FuturEnzyme collection of enzymes was identified by such sequence-based prediction tools (see **Table 1**).

As detailed in **Table 1** (access provided in the Annex section), 44 of the enzymatic preparations were relevant for cosmetics applications (i.e., degradation of sugars, polymeric sugars and hyaluronic acid), 591 for both detergent and textile applications (i.e., hydrolysis and degradation of a variety of esters and oils in emulsions, and cleaning of stained swatches and raw fabrics), and 43 for textiles applications (i.e., degradation of dyes and related compounds, and PET-like fabrics).

- For detergents application, hydrolysis of fatty stains has been identified as a key enzyme activity for safe detergents (oils are likewise a component to be removed during textile finishing, see below). A set of 27 enzymatic preparations have been identified with activity towards at least one standardized stained swatch, and 19 to hydrolyze pure oils in emulsions.
- Breaking down HMW hyaluronic acid to the desired size is the addressed application regarding the cosmetic application relevant to the project. A total of 39 enzymatic preparations showed, in initial assays, activity towards hyaluronic acid.
- For textile applications, one main and two secondary assays are targeted. The main one consists of the removal of residual spinning oils of synthetic materials. The second one consists in end-of-life degradation of synthetic fabrics; for example, polyester (mainly PET) is an important raw material for fabric production, and PET-degrading enzymes may help to tailor the properties of fabrics or to an option to recycle production residues or with end-of-life textiles. The third one consists in the removal of dyes and related chemicals. A total of 51 enzymes turned positives for the first two assays (44 of which exhibited activity on PET), and 20 for the third assay.

7. Enzymes able to withstand challenging conditions during application

Many processes in the industry run at elevated temperatures, e.g. to avoid costly cooling steps. Therefore, enzymes to be applied in textiles need to withstand comparably high temperatures. About 282 ester-hydrolases in the FuturEnzyme collection were assessed regarding their melting point or denaturing temperature. A total of 12 showed a melting point/denaturing temperature higher than 70°C, 26 higher than 60°C, and 63 higher than 50°C (**Table 7**), posing a starting point for protein engineering. The maximal melting point was 85°C, and the minimal about 18°C. Opposite, in many cases, enzymes working at low temperatures (i.e. 20-40°C in applications relevant to the project) are needed; within the enzyme collection, at least 218

enzymes were found to have a melting point/denaturing temperature lower than 50°C (**Table 7**). Determination of optimal temperatures and pH revealed, among the enzymes for which the thermal and pH profiles were obtained, a high number of enzymes being most active at temperatures below 50°C (**Table 8**) and neutral to slightly alkaline pH (**Table 9**). Note that the enzymes for textile applications relevant to the project should perform at an industrial scale in water, and therefore the pH profile of tested enzymes may fit to this condition.

Table 7. Thermal stability of the hydrolases in the FuturEnzyme collection.

Thermostability	Nr
Enzymes with Tm <50°C	219
Enzymes with Tm >50°C	63
Enzymes with Tm >60°C	26
Enzymes with Tm >70°C	12
Maximal melting point	80°C
No data	393

Table 8. Temperature at which enzymatic preparations showed maximal activity.

Optimal temperature	Nr
Enzymes with Topt <50°C	260
Enzymes with Topt >50°C	48
Enzymes with Topt >60°C	21
Enzymes with Topt >70°C	7
Maximal Topt	70°C
No data	369

Table 9. pH at which enzymatic preparations showed maximal activity.

Optimal pH	Nr
Enzymes with pHopt <7	4
Enzymes with pHopt >7	64
Enzymes with pHopt >8	56
Enzymes with pHopt >8	28
Maximal pHopt	10
No data	611

Enzymes for detergent have to work in the presence of surfactants, preferably in the washing liquor, provided by partner Henkel (a solution surfactant in real-life working concentration). A total of 397 enzymes relevant for detergent applications have been assessed so far for activity in the so-called washing liquor. A set of 53 (ca. 13%) tested enzymatic preparations showed considerable residual activity in this medium.

8. Annex

Table 1. 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*: