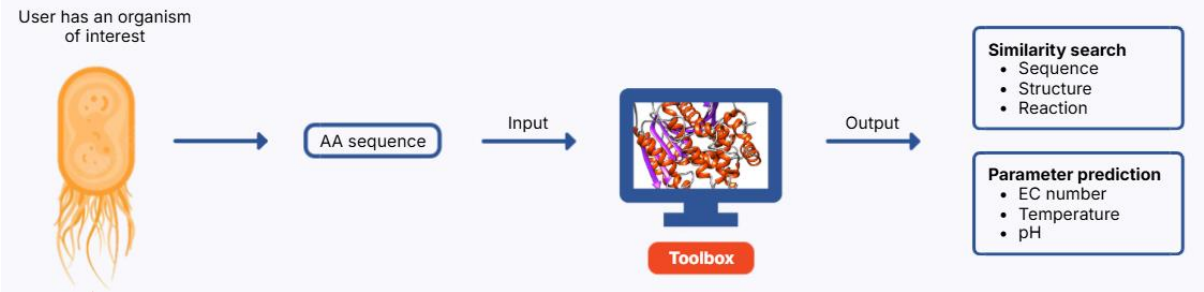
Use case: novel xylanases

To showcase the potential of the Enzymares toolbox, we present two examples that combine its output with data from experimental research. These examples, drawn from case studies within the Enzymares project, highlight how the Enzymares toolbox can support the discovery and characterization of novel xylanases for specialized applications. This case study comprises the potential of the toolbox to aid in predicting the characteristics of novel xylanases, on the one hand and to search for novel xylanases for a specific industrial application, on the other hand.

Predicting the characteristics of a novel xylanase

The Enzymares toolbox can contribute to predicting the characteristics of a novel xylanase without the need for time-consuming laborious experiments. For the illustration, we use the bacteria *Cellulophaga baltica,* which is available in the biobanks curated by the UGent-LM and included in the Enzymares project. The bacteria was cultivated and the resulting culture was screened for xylanase activity. This resulted in a positive outcome. Next, whole genome annotation of the bacteria was performed, which provided the needed information as input for the Enzymares toolbox. With this information, the toolbox can now be used to identify and predict the characteristics of the xylanase of interest.



The Enzymares toolbox identified different potential xylanases in *Cellulophaga baltica* based on the sequence similarity with the whole genome sequence and the xylanase sequences from two annotation tools, as illustrated in Table 1. Note that sequence similarity with the xylanases from the annotation tools not by default means all identified enzymes are xylanases.

genome_id	protein_id	sequence	length
R30316	76594.15.peg.748	MTKIVIGCFLLFLEVGLLTAQDT…	307
R30316	76594.15.peg.846	MKSITPYLLMAFMFLSLNLKAQT…	989
R30316	76594.15.peg.1608	MKKSYVCCLLFFCLYALNGQHIT…	402
R30316	76594.15.peg.2398	MLRIAPLFLMLLTFFSCKTEKKEV…	379
R30316	76594.15.peg.2402	MIKISKFLIATFSLLVAVSCSNNDPGEDR…	453
R30316	76594.15.peg.3221	MFTFLLGGMLLVNCSTEDADTSTNL…	254
R30316	76594.15.peg.459	MIHLTTTLQRITFFLGLIFVTHFSFSA…	311
R30316	76594.15.peg.1483	MKKSNFKQRVNACLFFTSLLCSSVVL…	588
R30316	76594.15.peg.2398	MLRIAPLFLMLLTFFSCKTEKKEVENE…	379
R30316	76594.15.peg.2402	MIKISKFLIATFSLLVAVSCSNNDPGEDR…	453
R30316	76594.15.peg.2877	MRIPLFKTVFFLSLFLLISCSSDAEPEI…	365

Table 1: The output from the Enzymares toolbox after whole genome annotation and a sequence similarity search with defined xylanases. The annotation tools used were mainly SwissProt and PATRICK. The first column shows the genome identifier from the biobank of UGent-LM followed by the second column which shows the open reading frame of the potential xylanases. The third column provides the sequence of the open reading frame which was shortened for illustrative purposes. The last column provides the length of the amino acid sequence corresponding to the potential xylanases.

After identifying enzymes with high sequence similarity with xylanases, the Enzymares toolbox can predict the candidates' characteristics, such as EC number, temperature optima, temperature stability and pH optima. Table 2 shows the output for the different predicted parameters of the potential xylanases identified above. For this specific example, two xylanases could be identified, as indicated by their predicted EC number 3.2.1.8 as shown in the third column of Table 2. They had both a predicted optimal temperature of approximately 40°C. Their predicted optimal pH was around neutral pH.

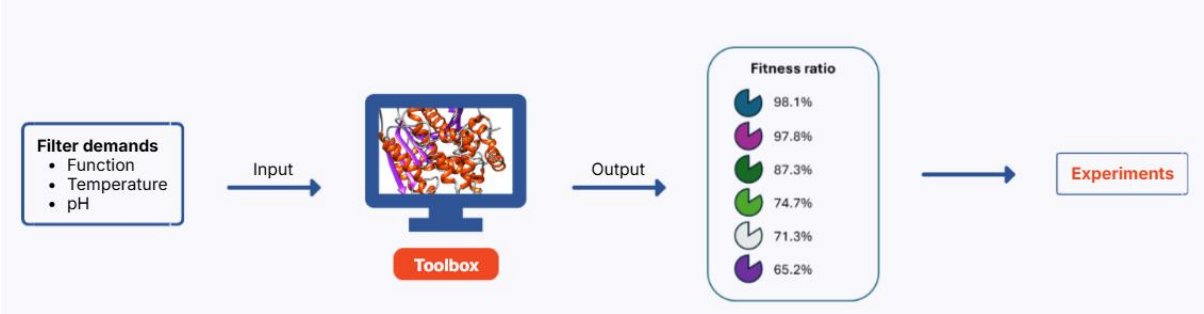
protein_id	length	predicted_EC	probability_EC	predicted_T_{opt}	predicted pH
76594.15.peg.748	307	3.1.1.3	0.55	31.81	8.31
76594.15.peg.846	989	3.2.1.4	0.22	32.89	6.09
76594.15.peg.1608	402	3.1.1.108	0.18	31.12	7.28
76594.15.peg.2398	379	3.2.1.8	1.00	39.85	7.29
76594.15.peg.2402	453	3.2.1.8	0.99	39.37	7.58
76594.15.peg.3221	254	3.1.22.1	0.45	50.05	6.86
76594.15.peg.459	311	3.2.1.39	0.05	24.02	8.09
76594.15.peg.1483	588	3.2.1.4	0.98	41.38	6.42
76594.15.peg.2877	365	3.2.1.4	0.98	36.85	6.41

Table 2: The output from the Enzymares toolbox after the prediction of the EC number, temperature optima and pH optima. The position of the open reading frame in the genome for the potential xylanases and the sequence length could be found in the first and second column. The third column showed the predicted EC number and the probability of this prediction which was shown in column four. The last two column comprise the predicted optimal temperature and pH. The two identified xylanases are highlighted in green.

Within the Enzymares project, experimental verification of the output of the Enzymares toolbox was performed. The xylanase activity was verified by activity-based protein profiling. This methodology utilizes fluorophore-conjugated activity-based probes to label active xylanases. Moreover, the temperature optima and pH optima can also be verified using these probes. For more information on this methodology, you can reach out to the Laboratory of Microbiology (UGent) for their expertise

Searching for a suitable xylanase candidate for your application

A second approach with which the Enzymares toolbox can aid, is in the search for novel enzymes to answer specific needs for your application. Here, we present a second example, focusing on xylanases, for how the Enzymares toolbox can extensively limit the amount of experimental work needed to find an enzyme for your application. With this approach, a suitable candidate can be found by defining the preferred characteristics (for example, EC number, temperature optima, pH optima) as the input for the Enzymares toolbox.



In this example, we want to search for a novel xylanase which shows potential to degrade highly substituted wheat arabinoxylan as this could be of relevance to produce prebiotic arabinoxylan oligosaccharides. This can be done by looking for a xylanase which shows high active site sequence similarity with the xylanase (glycoside hydrolase family 5) from *Clostridium thermocellum* as this xylanase has shown potential to degrade highly substituted wheat arabinoxylan. This search provided 17 potential candidates, as shown in Table 3.

UniProt accession	length	similarity	xylanase_tagger	probability xylanase_tagger	predicted_T_{opt}	predicted_pH_{opt}
A3DHG6	948	100	3.2.1.8	0.51	57	6.0
W4VBS9	623	95	3.2.1.8	0.98	57	6.1
A0AAD5SZ58	302	53.2	0	0.00	47	4.6
A0AAD5T7R4	1013	53.9	0	0.00	37	6.2
A0AAD5T063	1031	53.2	0	0.00	33	6.3
A0A139AI41	396	56.3	0	0.00	45	6.2
A0A139APP3	443	56.3	0	0.00	45	5.5
A0AAD5SYI4	397	52.8	0	0.00	53	4.0
A0A139A5T2	331	52.1	0	0.00	45	4.9
A0A927F7G4	1506	49.6	0	0.00	40	6.8

B5JIL5	1507	46.8	0	0.00	40	6.7
A0A5C6AKU1	2898	52.5	0	0.00	43	6.3
A0A556QSA6	780	45.7	0	0.00	37	6.3
A0A139ACR3	236	52.6	0	0.00	44	4.8
A0AAD5SR64	242	43.2	0	0.00	45	5.0
W0IYK7	1084	42.4	0	0.00	41	6.4
A0A556QSZ8	631	40.4	0	0.00	39	5.9

Table 3: Output of the Enzymares toolbox after sequence similarity search and the prediction of xylanase activity, optimal temperature and optimal pH of the resulting candidates. The UniProt accession number of the potential candidates is indicated in column one, their amino acid sequence length in column two and their similarity to the template sequence in column three. The output of the xylanase tagger and its probability is shown in respective column four and five. The last two column showed the predicted optimal temperature and pH. The promising xylanase candidates are highlighted in green.

Next, the search request can be further refined to meet our needs for a specific application. Here we choose to look for a xylanase which would be active during the fermentation phase of the bread-making process. For this goal, a xylanase tagger was used which was developed by Computational System Biology (KU Leuven) to specifically identify xylanases amongst the candidates. Secondly, the preferred optimum temperature and pH were defined which were respectively 30-40°C and a pH of 5.0. By refining these filters, suitable candidates can be obtained. In total, the Enzymares toolbox offered two candidates which could be of interest for this specific application as they have predicted xylanase activity and temperature and pH optima close to our preferred optimal conditions as shown in Table 3. With this simple search using the Enzymares toolbox, we were able to limit the amount of literature review, time and lab supplies required to typically select appropriate candidates for experimental screening.

The verification of this application of the Enzymares toolbox was also included in the Enzymares project. A screening methodology was designed to screen multiple candidates simultaneously in a simple and fast manner by measuring their catalytic activity using spectrophotometry. The screening methodology can be altered to the temperature, pH or substrate of interest. For more information on this methodology, you can reach out to the Laboratory of Food Chemistry and Biochemistry (KU Leuven).

To conclude, the Enzymares toolbox offers a large collection of different datasets and predictive tools. In this example, it was illustrated how the toolbox can aid in both the identification and characterization of novel xylanase as well as help in finding suitable candidates for an application of interest. As a result, the Enzymares toolbox successfully limits the amount of time spent on literature review and experimental work offering you faster the information you need to move towards your goal.