

*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

### 26/05/2023

First version of the predictive online web completed

MS8

Victor 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) |
| **Document version:** | 1 |
| **Date:** | 31.05.2023 |
| **Starting date:** | 01/06/2021 |
| **Duration:** | 48 months |
| **Lead beneficiary:** | BSC |
| **Participant(s):** | BSC, CSIC, Bangor, UHAM, UDUS, Schoeller, Henkel, Evonik |
| **Dissemination Level:** | Confidential, only for consortium's members (including the Commission Services) |
| **Type** | Websites, patents filling, etc. |
| **Due date (months)** | 24 |
| **Contact details:** | Víctor Guallar (victor.guallar@bsc.es) |

Summary

[First version of the predictive online web completed 4](#_Toc135373828)

[1. Means of verification 4](#_Toc135373829)

[2. Online tool available 4](#_Toc135373830)

## First version of the predictive online web completed

## 1. Means of verification

Tool available to consortium - this milestone consists in a first version of the online web tool that through integrating motif descriptors and big datasets will allow screening enzymes of interest; this tool will be made avilable for the members of the consortium for trials.

## 2. Online tool available

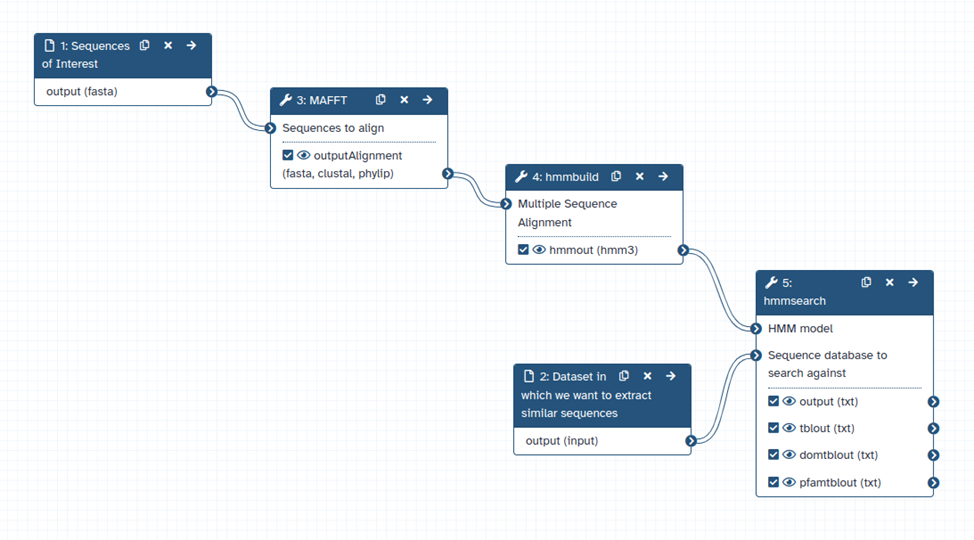
Milestone MS8 related to the development of an open-source, predictive software prototype to pre-select enzymes that meet key indicators requested by industrial sectors to be targeted in the Project, and possibly others. This software would revolutionise industry's ability to retrieve enzymes with the requested specifications from public or internal repositories. The strategy is correlating experimental metadata for a broad set of enzymes and theoretical descriptors/motifs obtained through bioinformatics and computational techniques. Based on that a predictive online platform will be developed.

The manufacturers' needs and especifications have been compiled (for details see Deliverable 2.1), and an extensive set of experimental meta-data have been produced (see Technical Report B, first reporting period) and new ones are being generated. Their integration has led to the discovery of a number of theoretical descriptors/motifs (see Technical Report of the first reporting period, and Deliverable D2.3), which were then used to built the first prototype of the predictive tool. As detailed in the Technical Report of the first reporting period, and thanks to the collaboration with our sister FNR project, OXIPRO, partner BSC decided to adopt the biocontainer BioExcel Building Blocks ([BioBBs](https://mmb.irbbarcelona.org/biobb/)) format for the individual modules, and the graphical implementation in Galaxy for the overall platform integration. BioBBs are a set of open-source, modular software tools for biomolecular simulations. The library offers a layer of interoperability between the wrapped tools, which makes them compatible and prepared to be directly interconnected to build complex biomolecular workflows. All building blocks (containers, o biocontainers) share a unique syntax, requiring input files, output files and input parameters (properties), irrespective of the program being wrapped. Containers will be distributed also via dockerhub and quay.io, to download and run the base container. A singularity distribution of some of them will be available for running in High-performance computing (HPC) environments and systems. After integrating the containers in the BioBBs library the user will be able to use them through the [Galaxy web page](https://usegalaxy.org/), with an easy and intuitive graphical interface. Galaxy is an open-source platform for data-intensive biomedical research, and it allows users to create and share workflows for data analysis and visualization. Workflows in Galaxy are created using the graphical user interface, and they can be shared with other users, enabling collaboration and reproducibility. Workflows can be used to automate complex analysis pipelines, reducing the need for manual intervention and enabling efficient, scalable data analysis. Implementing them is as easy as using a box representation for each tool and creating connections between them with arrows indicating the output of the first tool will be used as the input for the second tool (**Figure 1)**.

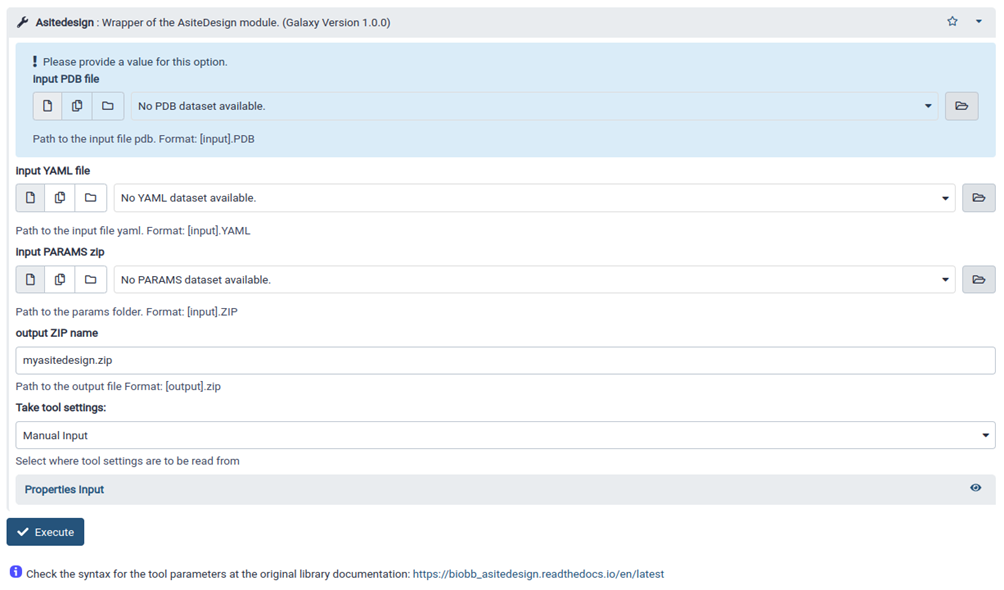
Based on the above, several biocontainers have already been developed, including the AHATool, EP-Pred and AsiteDesign ones, as detailed in the Technical Report of the first reporting period. For details see the recent reference reporting the machine learning [EP-Pred](https://www.mdpi.com/2218-273X/12/10/1529) ensemble classifier. The program can be downloaded in [GitHub etiur/EP-pred](https://github.com/etiur/EP-pred): A machine learning program to predict promiscuity of esterases. These software tools are also being integrated into online wrappers in Galaxy to facilitate their diffusion and usage among the consortium, in collaboration with OXIPRO (see AsiteDesign in **Figure 2**).

Based on the information provided in this document, we can consider Milestone MS8 be achieved. Future actions include: i) second round of enzyme prospecting (already ongoing); based on initial results, we are conducting further rounds of validation; ii) evaluate new experimental and computational correlations to guide a robust pre-selection based on these calculations; iii) further testing of concepts and applied applications of first draft of the consensus machine learning predictor (EP-Pred); iv) platform development. As indicated, the main development of the platform will involve the second part of the proposal in which we will integrate the different bio-containers being developed into a graphical web application (Galaxy).

Based on this information, we can consider Milestone MS8 achieved.



**Figure 1.** Example of a workflow being developed with Galaxy for bioprospecting.



**Figure 2.** Example of AsiteDesign wrapper in Galaxy.