



## D 6.21 (Activity 6.12, CNR-IBBA)

**Harmonized, distributed IBISBA-IT platforms for metagenomic analysis, biomolecule production and structural/functional characterization in support of the accelerated generation of customized bioprocesses.**

**30/06/2025**



Finanziato  
dall'Unione europea  
NextGenerationEU



Ministero  
dell'Università  
e della Ricerca



**Italiadomani**  
PIANO NAZIONALE  
DI RIPRESA E RESILIENZA



<b>Deliverable number:</b>	D6.21
<b>Work package:</b>	WP6 – Terrestrial Biosphere
<b>Intermediate Objective:</b>	Third draft
<b>Deliverable type:</b>	X Document, report
	<input type="checkbox"/> Websites, patent filings, videos, etc.
	<input type="checkbox"/> Other: please specify .....
<b>Dissemination level:</b>	X Public
	<input type="checkbox"/> Restricted
<b>Estimated delivery (bimester):</b>	B14
<b>Actual delivery date:</b>	30/06/2025
<b>Author(s) (Partner-OU):</b>	AUTHORS: Caterina Catalano, Anna Paola Casazza, Aldo Ceriotti, Paola Cremonesi, Barbara Menin, Emanuela Pedrazzini, Pietro Roversi, Stefano Santabarbara, Bianca Castiglioni (CNR-IBBA)
<b>Reviewed by:</b>	<b>ITINERIS Executive Board</b>
<b>Note:</b>	

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System - CUP B53C22002150006 (D.D. n. 130/2022)  
 Funded by EU - Next Generation EU  
 Mission 4 “Education and Research” - Component 2: “From research to business” -  
 Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”



## Table of contents

<i>INTRODUCTION</i> .....	4
<i>1. METAGENOMICS PLATFORM</i> .....	4
<i>1.1 GRIDION SEQUENCING PLATFORM (OXFORD NANOPORE TECHNOLOGIES)</i> .....	5
<i>1.2 BACTERIAL CELL SORTER FOR SINGLE CELL SEQUENCING – B. SIGHT (CYTENA)</i> .....	6
<i>1.3 LIQUID DISPENSER FOR NANOVOLUMES – I.DOT (DISPENDIX)</i> .....	6
<i>1.4 SYSTEM FOR NUCLEIC ACID PURIFICATION - G.PURE (DISPENDIX)</i> .....	7
<i>1.5 AGILENT 2100 BIOANALYZER SYSTEM</i> .....	8
<i>2. BIOMOLECULE PRODUCTION PLATFORM</i> .....	9
<i>2.1 BIOLECTOR XT MICROBIOREACTOR (BECKMAN COULTER)</i> .....	10
<i>2.2 MULTI-CULTIVATOR MC 1000-OD REACTOR SYSTEM (PSI)</i> .....	11
<i>2.3 LGEM LAB-25 TUBULAR GLASS PHOTOBIOREACTOR</i> .....	11
<i>2.4 CCD DETECTOR FOR SPECTROFLUORIMETRY</i> .....	12
<i>2.5 PHOTON COUNTING SPECTROFLUORIMETER</i> .....	12
<i>2.6 PREPARATIVE HPLC SYSTEMS - 1260 INFINITY II PREPARATIVE LC SYSTEM</i> .....	13
<i>2.7 HIGH PRESSURE CELL HOMOGENISER</i> .....	13
<i>3. STRUCTURAL/FUNCTIONAL CHARACTERIZATION PLATFORM</i> .....	14
<i>3.1 STACKABLE REFRIGERATED INCUBATORS- INNOVA S44I (EPPENDORF)</i> .....	17
<i>3.2 INVERTED MICROSCOPE – AXIO OBSERVER 7M (ZEISS)</i> .....	17
<i>3.3 PLANT GROWTH CHAMBER – FITOCLIMA 1200 (ARALAB)</i> .....	17
<i>3.4 ULTRACENTRIFUGE - OPTIMA MAX- XP (BECKMAN COULTER)</i> .....	18
<i>3.5 MASS PHOTOMETER TWOMP REFEYN</i> .....	18
<i>3.6 FAST PROTEIN LIQUID CHROMATOGRAPHY (FPLC) SYSTEM - ÄKTA PURE MICRO..</i>	19
<i>4. DATA STORAGE AND HARMONISATION OF THE PLATFORM</i> .....	20
<i>4.1 SERVER</i>	21

## INTRODUCTION

A metagenomic platform is being established through the acquisition of a coordinated system for bacterial single cell isolation and DNA library preparation, and a high throughput sequencing system. For biomolecule production, a high throughput microbioreactor and a set of bacterial and algal cultivation systems and monitoring equipment were acquired. Structural/functional characterization is supported through the acquisition of an FPLC for macromolecular purification and a mass photometer for *in vitro* biophysical measurements of the distribution of masses in solution. Standardization of the data produced by these platforms will be functional to the integration and harmonization of the platforms within the IBISBA research infrastructure.

### 1. METAGENOMICS PLATFORM

The metagenomics platform provides scientific and technical support for the metagenomic and/or metabarcoding analyses of different environmental samples (soil, water, gut, food, etc).

The metagenomic platform comprises a coordinated system for isolating single bacterial cells (B.Sight, Cytena) and preparing DNA libraries. Additionally, it incorporates two distinct high-throughput sequencing systems (MiSeq, Illumina, and Oxford Nanopore GridION). The resulting datasets offer a comprehensive perspective on the genetic and functional landscape of the bacterial community, facilitating detailed analysis of microbial diversity, gene content, and potential functional capabilities.

Metagenomics analyses include:

- 1) DNA extraction and quality control: DNA is extracted directly from samples using mostly commercial kits with magnetic beads to specifically capture DNA while excluding organic inhibitors.
- 2) Library preparation: For shotgun metagenomics, whole DNA is prepared for sequencing using the Illumina Nextera XT DNA library prep kit. Samples are barcoded and mixed together for sequencing. For metabarcoding libraries are prepared following the 16S Metagenomic Sequencing Library Preparation Protocol (Illumina, San Diego, CA, USA).
- 3) Shotgun metagenomics: The libraries are sequenced using Illumina technologies for paired end reads. The number of samples multiplexed depends on the sample type.
- 4) 16S rRNA and other target genes sequencing by Next generation sequencing using Illumina technology (Metabarcoding): The libraries obtained are quantified by Real-Time PCR with KAPA Library Quantification Kits (Kapa Biosystems, Inc., MA, USA), pooled in equimolar proportion and sequenced by MiSeq (Illumina) instrument with 2×250-base paired-end runs. Bioinformatics analyses allow microorganisms (bacteria, yeast, fungi) identification.
- 5) Long read metagenomics sequencing in real time by Oxford Nanopore GridION (Metagenomics). Bioinformatic analysis enables taxonomic identification, gene discovery and annotation, and metabolic reconstruction.

Metabarcoding analyses consist of advanced microbial community analysis through next generation sequencing/massively parallel approaches directly from environmental samples. Metabarcoding using targeted amplicons such as 16S rRNA genes is the most popular method for



characterizing microbial communities of interest. Other amplicon targets can be used for other groups of microbes, such as ITS for fungi, 18S rRNA for eukaryotes, or RdRP for RNA viruses. Shotgun metagenomics is the approach used to analyze all the genes present in a microbial community, providing not just phylogenetic analysis but also insight into the functional capabilities of each species within the community.

Below is a description of the metagenomics platform instruments, and a list of the main data formats generated by the single equipment.

## 1.1 GRIDION SEQUENCING PLATFORM (OXFORD NANOPORE TECHNOLOGIES)

GridION is a DNA sequencing platform developed by Oxford Nanopore Technologies, designed to provide rapid and scalable genomic analysis. GridION uses a technique known as “Nanopore sequencing”, where individual DNA or RNA molecules are threaded through a nanopore (a very small protein or synthetic pore embedded in a membrane) and an electrical current is applied as they pass. Changes in the current are detected and used to identify the sequence of nucleotides (A, T, C, G) that make up the DNA or RNA strand. Unlike traditional sequencing methods that read short fragments of DNA (often requiring complex data assembly afterwards), nanopore technology enables “long-read sequencing”, which can read entire strands of DNA in their natural, unbroken form. This makes it particularly useful for sequencing complex genomes and for detecting structural variations, large insertions, deletions, and complex genomic rearrangements. The GridION system can be configured to run up to 5 flow cells simultaneously, each capable of sequencing large amounts of data in real time. Oxford Nanopore provides the “MinKNOW” software platform to manage the sequencing run, monitor the health of the flow cells, and collect data.

### *Summary of common data formats output by this instrument*

- **.FAST5:** Raw Signal Data, Contains raw electrical signals from nanopores.
- **.FASTQ:** Base-called Data, Sequence data with quality scores.
- **.FASTA:** Base-called Data, Sequence data (no quality scores).
- **.SAM:** Alignment Data, Aligned sequence data in text format.
- **.BAM:** Alignment Data (Binary), Binary format for aligned sequence data.
- **.CRAM:** Compressed Alignment Data, Compressed version of .bam.
- **.JSON:** Metadata, Run details and sequencing statistics.
- **.CSV:** Metadata, Run details in CSV format.
- **.XML:** Metadata, Run details in XML format.
- **.events:** Event Data, Raw electrical event data used for advanced analysis.
- **.HDF5:** Event Data, Hierarchical format for storing event data.
- **.FCS:** Fluorescence Data, Optional fluorescence data used in some experiments.

## 1.2 BACTERIAL CELL SORTER FOR SINGLE CELL SEQUENCING – B. SIGHT (CYTENA)

CYTENA's B. SIGHT is a platform for isolating and sequencing single bacterial cells from complex samples.

CYTENA's patented single-cell dispensing technology uses disposable cartridges capable of holding 5-80  $\mu\text{L}$  of cell suspension.

The dispensing technology is combined with high-resolution optical microscopy and real-time image analysis to ensure the dispensing of individual bacteria. B. SIGHT is also equipped with a high-precision dispensing system, Automated Offset Correction (AOC), whereby individual bacteria are precisely isolated in the center of PCR wells in 50 pL droplets, enabling downstream miniaturization for the preparation of Next Generation Sequencing (NGS) libraries.

In addition, B. SIGHT is equipped with a dual-camera system for simultaneous detection of high-resolution bright-field and fluorescence images. These images can be used for fluorescent sorting: fluorescently labelled cells can facilitate the identification and isolation of cells of interest from samples for subsequent downstream analysis.

The B. SIGHT platform generates structured datasets primarily to support the downstream management, analysis and processing of single cell data, such as single cell RNA sequencing (scRNA-seq), single cell DNA sequencing or other genomic techniques. Formats vary depending on the specific application and the nature of the experiment.

Datasets typically include detailed information on cell isolation, sequencing results, and associated metadata.

### *Summary of common data formats output by this instrument*

- **CSV/TSV:** for simple and tabular datasets (e.g. cell isolation, metadata, gene expression matrices).
- **JPEG image**

These formats offer the flexibility to store and process both the **raw data** generated during the cell dispensing and sequencing process, as well as the **metadata** and quality control information essential for accurate analysis and interpretation of single-cell experiments.

## LIQUID DISPENSER FOR NANOVOLUMES – I.DOT (DISPENDIX)

The I.DOT Dispendix is a precision liquid handling device designed to automate the dispensing of liquids in small volumes for laboratory experiments. The I.DOT uses a patented non-contact pressure-based technology, minimizing consumables usage and reducing plastic waste. It can dispense volumes as low as 8 nL up to 500  $\mu\text{L}$ . It handles various liquids—water-based, organic, proteins, nucleic acids, and cells—for applications like NGS, qPCR, and single-cell library preparation. It is often used in conjunction with systems like the BioLector XT to streamline workflows in biological and biochemical research.

The **I.DOT (Dispendix)** liquid dispenser for nanovolumes generates several key data formats to track and manage dispensing actions with precision.

### *Summary of common data formats output from by this instrument*

- **CSV:** A simple, tabular format capturing details like time, volume dispensed, location, reagent used, and operator. It is commonly used for logging dispensing events in a straightforward, human-readable format.
- **JSON:** A structured format for storing dispensing actions, typically used for integrating data with other systems or applications. It captures timestamped dispensing events, volumes, and locations in a flexible and easy-to-process format.
- **XML:** A hierarchical format often used for complex data exchange, storing dispensing actions with detailed metadata such as volume, reagent, and operator in a structured, machine-readable format.
- **Plain Text Log Files:** A simple log of real-time operations, recording events such as dispensing actions, errors, or system status in a text-based format for audit trails or troubleshooting.
- **HDF5:** A high-performance format used for storing large datasets, typically in high-throughput environments, where complex, multi-dimensional data is tracked over time.
- **Labware/Plate Maps:** Data related to the layout of dispensing locations (e.g., microplates), often in CSV, JSON, or XML format, specifying the wells, reagents, and volumes used.

These formats facilitate data management, traceability, and integration with other lab systems, enhancing the overall efficiency and reliability of experiments that rely on precise liquid handling at nanoliter scales.

## 1.4 SYSTEM FOR NUCLEIC ACID PURIFICATION - G.PURE (DISPENDIX)

In our metagenomics platform, this instrument plays a crucial role in isolating single microbial cells from complex environmental samples. The system's advanced microfluidic technology isolates single cells with high precision, ensuring that each droplet or microwell contains only one viable cell from the diverse microbial population. Single cell isolation is essential for studying microbial communities at a deeper level, as it allows for analysis of the genome or transcriptome of individual species within the microbiome. Once isolated, these single cells can be processed for downstream applications such as single-cell genome sequencing (SCG) or transcriptomics, providing a clearer picture of each microorganism's genetic makeup and role in the ecosystem. This approach significantly improves metagenomic analysis by combining the depth of single-cell sequencing with the broad coverage of metagenomic studies, helping to uncover the functional roles of specific microbes within a community.

### *Summary of Common Data Formats output by this instrument*

- **CSV:** These are often used for storing raw data from the purification process, such as sample IDs, well numbers, purification conditions, and final nucleic acid concentrations. These files are easy to import into data analysis software or laboratory information management systems (LIMS).
- **Log Files:** These contain detailed records of the system's operation, including error messages, operational events, and process start/stop times. Log files are typically generated for system diagnostics and troubleshooting.
- **Protocol Files:** These files contain information about the specific purification protocol used, such as reagent volumes, incubation times, and centrifugation speeds. This data helps ensure reproducibility and traceability in the process.
- **XML Files:** The G.PURE system may also generate XML files that contain structured data about the purification steps, sample metadata, and quality control results. XML is often used for integration with other laboratory systems.
- **Report Files:** These are often generated in PDF or HTML format and summarize the results of the purification run, including pass/fail information, yield, and any other performance indicators relevant to the experiment.

These data formats ensure that G.PURE can support the entire process of single-cell isolation, from initial tracking and quality control to final genomic or transcriptomic analysis. The ability to export and integrate this data in standard formats makes it compatible with a variety of downstream analysis tools and platforms.

## 1.5 AGILENT 2100 BIOANALYZER SYSTEM

The Agilent 2100 Bioanalyzer system is used for quality control and analysis of nucleic acids, proteins, and cells. It is used in molecular biology, genomics, and proteomics for applications such as RNA and DNA analysis, protein characterization, and sample integrity and purity assessment. The system performs electrophoresis on microfluidic chips, which allow for high sensitivity and minimal sample volumes (typically 1-5  $\mu\text{L}$  per sample), separating nucleic acids (DNA/RNA), proteins, or cells based on size, charge, and other molecular characteristics.

The Bioanalyzer uses laser-induced fluorescence (LIF) detection for real-time analysis during electrophoresis. Fluorescent dyes are used to stain nucleic acids or proteins, and the system measures the emitted fluorescence to determine analyte size, concentration, and quality.

The instrument is paired with Agilent's Bioanalyzer software, which provides data analysis tools to generate gel-like images and electropherograms, perform quantification, and assess sample quality.

The Agilent 2100 Bioanalyzer System generates a variety of data formats to meet different needs in sample analysis.

### *Summary of common data formats output by this instrument*

- **BAS** for raw data
- **XML** for structured, machine-readable data
- **CSV** for tabular, numerical data
- **PDF** for formal reports
- **TXT** for basic text results
- **PPT** for presentation files
- **FCS** for flow cytometry data

These formats allow for flexible data handling, reporting, integration with other systems, and downstream analysis in a range of biological and chemical research applications.

## 2. BIOMOLECULE PRODUCTION PLATFORM

The biomolecule production platform is designed with a view to allow the implementation of sustainable and efficient biotechnological processes for the synthesis of high-value biomolecules in wild-type, metabolically engineered microbial strains and plant cell lines.

The biomolecules platform offers the following expertise:

- 1) Design, generation, validation and characterization of metabolically engineered microbial strains to produce high-value metabolites.
- 2) Optimization of key parameters of microbial and cell lines growth processes (e.g. pH, T, dissolved nutrients, light if photoautotrophic.) in batch assays and bioreactors.
- 3) Extraction, chemical analysis and quantification of metabolites produced.
- 4) Process scale-up at lab-scale in photobioreactor system (available only for biotechnological processes based on photosynthetic microorganisms).

The biomolecule production platform includes a high throughput microbioreactor (BioLector XT Microbioreactor) equipped with different modules for working with light, anaerobiosis or modulated O<sub>2</sub> conditions, for the rapid optimization of photosynthetic and fermentation processes. In addition, for lab-scale screening experiments of phototrophic organisms, such as algae, cyanobacteria and plant cells, the platform is equipped with a Multi-Cultivator MC 1000-OD reactor system (PSI), equipped with a gas mixer and a cooling unit. For scaling up of algal biotechnological processes, the platform is equipped with a 25L horizontal tubular photobioreactor, namely the Lgem Lab-25 tubular glass PBR.

Regarding the analysis of the metabolites produced, the platform is equipped with the necessary instrumentation for fine characterization based on chromatographic analysis and optical spectroscopy. Preparative and analytical HPLC instrumentation is available for purification, analysis and quantification of target compounds, while with regard to the optical characterization, the platform includes a spectrofluorometer (UV-Vis) and a customized fluorimeter equipped with a CDD camera for higher resolution ( $\lambda > 600$  nm) for the study of molecules which emission extends up to the near-infrared.

Below is a description of the platform's instruments, including a homogenizer for cell disruption, and a list of the main data formats generated by the single piece of equipment.

## 2.1 BIOLECTOR XT MICROBIOREACTOR (BECKMAN COULTER)

The BioLector XT instrument is a bench-top microbioreactor platform performing high-throughput cultivations in combination with online-monitoring of the most common cultivation parameters such as biomass, pH value, oxygen saturation of the liquid phase (DO), and fluorescence intensity of various fluorescing molecules or proteins. The BioLector XT instrument uses 48 well microtiter plates (MTPs) in Society for Biomolecular Screening (SBS)/Society for Laboratory Automation and Screening (SLAS) format, which offers the possibility of automation. Environmental parameters such as temperature and shaking speed are controlled.

The system has been upgraded with additional modules:

- The microfluidics (MF) module allows the possibility of individual pH control and fed-batch operations for each single cultivation well.
- The O<sub>2</sub> down-regulation module, for fermentation under microaerophilic conditions. It consists of an additional gas connection for nitrogen on the backside of the BioLector and an oxygen sensor inside the cultivation chamber. In this way, an oxygen level between 2% and 21% can be generated in the chamber.
- The O<sub>2</sub> up-regulation module, for fermentation with O<sub>2</sub>-enriched air. It consists of an additional gas connection for oxygen on the backside of the BioLector, as well as an oxygen sensor inside the cultivation chamber. In this way, an oxygen level between 21% and 35% can be generated in the chamber.
- The CO<sub>2</sub> up-regulation module, for cultivation under CO<sub>2</sub>-controlled conditions. This module consists of an additional gas connection for CO<sub>2</sub> on the backside of the BioLector and a CO<sub>2</sub> sensor inside the cultivation chamber. In this way, a CO<sub>2</sub> level between 0% and 10% can be generated in the chamber.
- The anaerobic module enables strict anaerobic fermentation processes combined with a controlled, low nitrogen gas flow rate.

- The Light Array Module provides customizable light settings of 400-700 nm (nanometer) within the photosynthetic spectrum, including day and nighttime lighting.

These instruments may only be used with microorganisms of safety levels 1 and 2 (Risk Group 1 and Risk Group 2) as defined by the World Health Organization (WHO).

### *Summary of Common Data Formats output by this instrument*

- **CSV:** Raw time-series data with columns for time, OD, pH, DO, and other measured parameters. This format is commonly used for analysis in software like Excel, R, or Python.
- **Excel Files:** For more complex datasets, Excel files with multiple sheets might be used, each corresponding to different parameters or conditions.

## 2.2 MULTI-CULTIVATOR MC 1000-OD REACTOR SYSTEM (PSI)

The Multi-Cultivator MC 1000-OD, developed by PSI (Process Systems International), is intended for small-scale screening experiments of phototrophic organisms such as algae, cyanobacteria, and plant cells. Also, other microorganisms, for example heterotrophic bacteria and yeast grown in suspension, may enjoy this type of cultivator. The special design of the instrument allows for fast comparative studies of various organisms, mutants, or cultivation conditions. This instrument allows combining up to 8 different LED colors (from UV-violet 405 nm to far red 730 nm) within each cultivation slot. Each color in each slot can be individually controlled, and thus customized spectra may be defined by the user.

### *Summary of Common Data Formats output by this instrument*

- **CSV, Excel (XLSX):** For tabular data (time-series, sensor measurements).
- **JSON, XML:** For structured data exchange or API integration.
- **HDF5, NetCDF:** For large, high-dimensional time-series data.
- **PNG, JPEG, PDF:** For visual reports and graphs.
- **SQL-based formats:** For structured storage in databases.

## 2.3 LGEM LAB-25 TUBULAR GLASS PHOTOBIOREACTOR

The Lgem Lab-25 tubular glass photobioreactor (PBR) is designed as a reliable research tool for producing ultra-clean inoculum material. The unit comprises one vertical tubular helix made from transparent glass pipes, a glass circulation vessel, instrumentation, and LED illumination. All are mounted on a moveable stainless steel (316) frame. The reactor helix and circulation vessel can be heat treated with 100°C/212°F hot water or steam. Sterilization is a must-have for starter material production.

### *Summary of Common Data Formats output by this instrument*

- **CSV, Excel (XLSX):** For tabular sensor data and experimental logs.
- **JSON, XML:** For structured data exchange and settings/configuration data.
- **HDF5, NetCDF:** For large, multidimensional time-series datasets.
- **PNG, JPEG:** For graphical reports and visualizations.
- **SQL:** For relational storage of data in a database.

## 2.4 CCD DETECTOR FOR SPECTROFLUORIMETRY

CCD cameras are highly sensitive detectors employed for the study of very weakly emitting fluorophores. The Princeton Pylon-100BX liquid nitrogen cooled CCD camera with the coupled spectrographs Princeton HRS-300S-NI allow recording with very high sensitivity and signal-to-noise ratios the emission spectra of many biological fluorescent molecules including intact unicellular photosynthetic organisms. The CCD detector and spectrograph are part of a laboratory-assembled fluorimeter, where a 400 W Xenon lamp filtered through a monochromator acts as the excitation source. The covered emission spectral range extends from ~600 nm to ~900 nm, whereas the excitation is tuneable from 300 nm to 900 nm. The instrument is controlled by the LightField software, also provided by Princeton.

### *Summary of Common Data Formats output by this instrument*

- **.spc:** Princeton proprietary format for direct data acquisition
- **.txt:** ASCII compatible data export format

## 2.5 PHOTON COUNTING SPECTROFLUORIMETER

Fluorimetry is a very diffused technique for the analysis of a wide range of molecules/samples and could even be applied in vivo. The PC1 Photon Counting Spectrofluorometer from ISS Inc. operates in a single photon counting mode, therefore allowing very high sensitivity and linearity of the recorded signals. The instrument covers both the UV and visible range of the spectrum, both in the excitation and emission detection modes. It is moreover equipped with polarizers so that anisotropy measurements can be also performed. These are applicable to many biological issues ranging from the study of isolated metabolites, drugs, chromophores, proteins and their respective interactions including folding. Further, these could be explored by varying the temperature of the sample. A feature allowed and controlled by the instrument itself. The instrument is controlled by a dedicated ISS Inc. piece of software that also allows the analysis of the data.

### *Summary of Common Data Formats*

- **.txt:** ASCII compatible data export format

## 2.6 PREPARATIVE HPLC SYSTEMS - 1260 INFINITY II PREPARATIVE LC SYSTEM

The 1260 Infinity II Preparative Liquid Chromatography (LC) System from Agilent technologies is a high-performance system designed for large-scale purification, analysis and quantification of target compounds (proteins, peptides, pharmaceuticals, natural products, and other biomolecules). This preparative version is capable of handling large sample volumes and high flow rates (up to 200 mL/min or more), making it suitable for purifying milligram to gram quantities of compounds. The system includes high-pressure, solvent delivery pumps that provide precise control of overflow rates, ensuring consistent elution profiles for large-scale separations. The 1260 Infinity II Preparative LC offers a range of detectors, such as UV-Vis detectors, fluorescence detectors, and refractive index detectors, enabling precise monitoring of the separation process in real-time. The system is equipped with a fraction collector that automatically collects eluted fractions at specific time intervals or based on detected peaks. This allows for the separation of target compounds into individual fractions for further analysis or use. It is controlled using Agilent OpenLAB CDS (Chromatography Data System) software, which provides an intuitive interface for setting up methods, monitoring real-time data, and controlling hardware components. The software can also be used to analyze chromatograms and optimize separation conditions.

### *Summary of Key Data Formats output by this instrument:*

- **CSV:** Simple, exportable data format for raw chromatogram data.
- **.D:** Agilent proprietary format containing detailed chromatographic data and analysis.
- **.M:** Method file for storing and recalling experimental conditions.
- **XML:** Open format for exchanging chromatographic and method data.
- **PDF:** Final, formatted reports for presentation and documentation.
- **.RSD:** Raw spectral data, typically from detectors like UV-Vis.
- **.DAT:** Detailed data format for raw and processed chromatographic data within OpenLAB CDS.
- **.TLC:** Fraction collection data for preparative LC applications.

## 2.7 HIGH PRESSURE CELL HOMOGENISER

The SPCH-EP-10-60 (Stansted Homogenising Systems Ltd, UK) cell disrupter/homogeniser allows to adjust process pressure from 10 MPa (1,500 psi) to 400 MPa (58,000 psi), offering the widest scope of applications. The instrument's fully adjustable pressure range and its high-pressure capability give excellent performance for many different cell types. Installed options comprise an after cooler/heat exchanger - liquid cooled (LC-HX) and a cell body heat exchanger (CB-HX) to treat the product post homogenization and for heating/cooling the cell through a DC-2030 chiller, respectively. The system, placed on a stainless steel WT7070G Mobile Worktable, can work in cycling mode thanks to a pressure feed reservoir (AC4504 Senco air compressor included).

### *Summary of Key Data Formats output by this instrument:*

Operational parameters such as pressure, temperature, and cycle count can only be logged manually for basic documentation and process traceability. While the instrument does not generate analytical data, these logs support reproducibility and batch record-keeping when linked to sample IDs and experimental protocols.

## 3. STRUCTURAL/FUNCTIONAL CHARACTERIZATION PLATFORM

We have successfully established a comprehensive platform for the design, production, characterization, and structural analysis of recombinant proteins and peptides. The platform supports a broad spectrum of biological systems and analytical techniques, enabling both applied biotechnology and fundamental research in protein science. The services offered for macromolecular functional and structural characterization are:

### **1. Recombinant Protein Expression in Plant Hosts**

We have implemented a dedicated module for the planning, optimization, and functional characterization of recombinant protein expression in plant hosts.

#### **1.1. Design and *In Silico* Modelling**

The structural features of the protein of interest (POI) are used to guide *in silico* modelling. Based on this, we define the optimal vector system, codon usage, expression platform, and purification strategy. This enables tailoring to the desired properties of the POI (e.g., stability, purity, or minimization of allergenic/toxic features).

#### **1.2. Transient Expression and In Vivo Functional Testing**

A pipeline for rapid POI testing has been implemented using plant cell protoplasts, leaf agro infiltration, or heterologous systems (yeast or bacteria). These assays allow monitoring of POI half-life, subcellular localization, and co-/post-translational modifications using biochemical and microscopy-based techniques.

#### **1.3. Optimization of Expression Parameters**

Systematic testing enables the selection of optimal POI sequences and expression conditions. Host physiology and/or transcriptomic responses during recombinant expression are evaluated, especially in plant tissues.

This module is supported by facilities for plant transformation, growth, microscopy, and biochemical analysis.

### **2. Protein and Peptide Production in Multiple Expression Systems**

The platform offers flexible protein production services across bacterial, mammalian, and plant expression systems.

*D6.21 Harmonized, distributed IBISBA-IT platforms for metagenomic analysis, biomolecule production and structural/functional characterization in support of the accelerated generation of customized bioprocesses.*

### **2.1. Bacterial Expression Systems**

- Utilization of *E. coli* and other hosts for cost-effective, high-throughput protein production.
- Purification using affinity, ion exchange, and size exclusion chromatography.

### **2.2. Mammalian Expression Systems**

- Use of mammalian cell lines for the production of complex proteins requiring native folding and post-translational modifications.
- Protein A/G affinity, ion exchange, and hydrophobic interaction chromatography applied for precise purification.

### **2.3. Plant-Based Expression Systems**

- Expression in plant cells, tissues, and whole plants for sustainable protein production.
- Extraction protocols adapted to overcome plant-specific matrix effects (e.g., secondary metabolites).
- Protein engineering tools enable modification of amino acid sequences to improve yield, activity, stability, or facilitate detection/purification.

We provide guidance on protein engineering strategies, considering proteostasis mechanisms and subcellular compartmentalization. Specific tags and fluorescent proteins can be added to improve in vivo tracking and downstream purification.

## **3. Measurement of molecular mass distribution in solution using mass photometry.**

**Mass photometry (MP)** is a cutting-edge technique based on light-scattering enabling detection of individual, unlabeled molecules in dilute solutions. Its broad measurement range makes it an essential technique for researchers studying complex biological molecules. The technique offers precise insights into the molecular mass distribution of various biomolecules. The core of the MP instrument is a mass photometer, a sophisticated device that measures the light scattering of individual molecules as they absorb onto a glass microscope slide. This allows for the detection of single molecules without the need for labels, providing a direct and accurate analysis of molecular masses. The MP instrument boasts an impressive measurement range, accurately determining molecular masses from 40 kDa to 5 MDa. This wide range makes it an invaluable tool for studying a variety of biomolecules, including proteins, nucleic acids, membrane proteins, and aggregates, as well as large protein complexes. Its versatility in application underscores its significance in molecular biology and biochemistry research.

**4. Structural characterization of macromolecules using X-ray crystallography or Cryo-EM.** Our platform offers comprehensive support for structure determination of macromolecules, utilizing advanced X-ray crystallography and/or Cryo-EM tools and protocols.

**4.1.** The platform comprises two refrigerated incubators for the growth of microorganisms for plasmid amplification and recombinant protein production; a growth chamber for plants; an ultracentrifuge for subcellular fractionations and an inverted microscope platform for fluorescence microscopy analysis. Protein purification is supported by an Äkta Micro FPLC for the purification of macromolecular samples, and protein characterization by a mass photometer for the measurement of the distribution of masses in solution.

**4.2.** Recombinant protein is purified by affinity, ion exchange, hydrophobic, and/or size exclusion chromatography. Protein structures are determined by X-ray crystallography and/or Cryo-EM single particle analysis, supported by an Oryx Nano crystallization robot and access to synchrotrons and Cryo-EM microscopes.

## **Research Domains Enabled by the Platform**

**This platform enables applied and exploratory research across the following areas:**

1. Biogenesis and evolution of eukaryotic compartments, particularly seed storage protein accumulation and protein body formation in cereal seeds.
2. Endoplasmic Reticulum Quality Control (ERQC) and ER-Associated Degradation (ERAD) mechanisms.
3. Unfolded Protein Response (UPR) in seed development, under both normal and stress conditions.
4. Biogenesis of membrane proteins, including transporters and tail-anchored proteins, with a comparative focus on plant tonoplast and animal lysosomal systems.
5. Optimization of recombinant protein production in plant systems using ER chaperones and rational engineering strategies.

## **Core Facilities and Equipment Available**

The platform relies on a suite of fully operational laboratories and equipment, including:

- Protein biochemistry, molecular biology, and class 1 microbiology labs
- Chemical analysis instrumentation
- Microscopy (fluorescence, inverted)
- Mass photometer
- Plant growth chambers and artificial greenhouse
- Bioreactors and incubators for fermentation and expression
- Protein purification systems (ÄKTA Micro FPLC)
- Oryx Nano crystallization robot
- Ultracentrifuge for fractionation
- Access to Cryo-EM and synchrotron infrastructure

### 3.1 STACKABLE REFRIGERATED INCUBATORS- INNOVA S44I (EPPENDORF)

The Innova S44i is a large capacity, temperature controlled (refrigerated) stackable shaker allowing the use of high-volume microbial culture flasks in large numbers and in a compact footprint.

#### *Summary of Key Data Formats output by this instrument*

The INNOVA S44i stackable refrigerated incubator primarily produces time-series operational data related to temperature, shaking speed (RPM), and run duration. This data can be exported via USB as TXT or CSV files.

### 3.2 INVERTED MICROSCOPE – AXIO OBSERVER 7M (ZEISS)

The Axio Observer from Zeiss is an inverted microscope platform designed with multi-modal imaging capabilities. It can switch between wide-field, fluorescence imaging, enabling comprehensive analysis of cellular and molecular structures. The instrument is equipped with a SW Platform Chg AV 4/ZEN 3.7 DLic software with the modules SW ZEN Toolkit Motorized Acquisition DLic, for the automatic multidimensional acquisition in X/Y/Z/T, and SW ZEN Toolkit Deconvolution DLic, for the advanced deconvolution in single or multiple channels.

#### *Summary of Key Data Formats output by this instrument*

- **TIFF, JPEG:** Expression Analysis, Protein or Peptide Purification, Cell Biology, Biochemical and Biophysical Assay, Plant Phenotypic and Growth, Cellular Localization and Interaction
- **PDF:** Statistical and Computational Analysis Outputs, Project Report

For FAIR data management, raw.czi files are retained alongside processed outputs, and metadata can be extracted or converted into interoperable formats using tools such as Bio-Formats or OME standards.

### 3.3 PLANT GROWTH CHAMBER – FITOCLIMA 1200 (ARALAB)

Environmental simulation chamber for plant growth, designed for various applications in biotechnology and plant research. It allows the control of temperature, humidity, airflow, and photoperiod.

### *Summary of Key Data Formats output by this instrument*

The Fitoclima 1200 plant growth chamber (ARALAB) generates environmental control data related to programmed and real-time values of temperature, humidity, light intensity, and photoperiod cycles. This data can be exported via USB key in TXT or CSV format.

## 3.4 ULTRACENTRIFUGE - OPTIMA MAX- XP (BECKMAN COULTER)

The Optima™ MAX-XP tabletop ultracentrifuge allows small sample volumes (from 175 µL up to 32.4 ml) to be analyzed by speeding up the information acquisition process. The instrument is mainly used for subcellular fractionation and for purification and characterization of recombinant biomolecules.

### *Summary of Key Data Formats output by this instrument*

- The OPTIMA MAX-XP ultracentrifuge (Beckman Coulter) produces run data related to rotor type, speed (RPM), run time, temperature, acceleration/deceleration profiles, and program settings. These operational parameters can be logged manually for basic documentation and process traceability and can be exported in CSV/TXT format via a USB drive. While the instrument does not generate analytical data, these logs support reproducibility and batch record-keeping when linked to sample IDs and experimental protocols.

## 3.5 MASS PHOTOMETER TWOMP REFEYN

Refeyn's TwoMP mass photometer measures in a solution the distribution of molecular masses of proteins and nucleic acids based on measurements of the scattering of electromagnetic radiation by each individual molecule or molecular complex.

It can be used to characterize biomolecules, study their function, and optimize conditions for working with them. Additionally, the mass photometer can be used to study protein interactions, oligomerization and macromolecular assembly, and to assess sample integrity and homogeneity.

The TwoMP includes software packages for mass photometry data acquisition (Refeyn AcquireMP) and analysis (Refeyn DiscoverMP). Figures can be directly generated and exported as PNG, SVG, PDF or EPS from within the software.

### *Summary of Key Data Formats output by this instrument*

The instrument produces proprietary raw data files in .rtdc format, containing time-resolved image sequences and embedded acquisition metadata. Processed results — including molecular mass distributions, peak fits, and particle counts — are typically exported in .csv, .xlsx, or .pdf

*D6.21 Harmonized, distributed IBISBA-IT platforms for metagenomic analysis, biomolecule production and structural/functional characterization in support of the accelerated generation of customized bioprocesses.*

formats, with metadata and experimental settings optionally available in .json or csv form depending on software version. To support FAIR data practices, raw and processed files are archived together, and metadata are progressively being harmonized using controlled vocabularies for proteomics and bioimaging (e.g., PSI-MS, EDAM). The datasets are intended to be linked with persistent identifiers and accompanied by clear licensing and protocol documentation, enabling interoperability with structural and functional characterization workflows within the ITINERIS research infrastructure.

### 3.6 FAST PROTEIN LIQUID CHROMATOGRAPHY (FPLC) SYSTEM - ÄKTA PURE MICRO

ÄKTA pure micro is a chromatography system for target protein purification at the microliter scale, providing a complete solution for handling small sample volumes, especially useful for micro-preparative purifications, such as sample preparation for technologies that require processing of small sample volumes, such as Cryo-electron microscopy (Cryo-EM).

It supports sample volumes as low as 50  $\mu\text{L}$  and can collect fractions of volumes as small as 8  $\mu\text{L}$ , depending on the specific column and flow rate used, making it highly suitable for pilot-scale purifications or for optimizing large-scale protocols.

This feature is particularly useful for applications such as affinity chromatography, ion exchange or size exclusion chromatography (SEC) for protein or peptide isolation.

In fact, ÄKTA pure micro can work with a wide range of chromatography columns, generally small in size, from 1 mL to 5 mL in volume.

It is equipped with an automated fraction collector that allows the collection of purified fractions in multi-well plates, based on pre-set parameters such as UV absorbance or conductivity.

ÄKTA pure micro uses high-precision pumps capable of providing low flow rates (down to 0.1 mL/min), which is essential for small-scale purifications where controlled flow is important.

The pump is designed to work efficiently with small columns and low volumes, providing stable and reproducible gradients for protein separation.

The system features integrated UV absorbance and conductivity detection, allowing one to monitor protein elution, salt gradients and buffer composition during purification and assess protein purity in real time.

The system combines the different components (pump, UV-Vis detector (up to three wavelengths), fraction collector) into a single integrated platform controlled by UNICORN™ software (Cytiva's proprietary software) which allows for precise method creation, data analysis and monitoring of the purification process in real time.

Applications of AKTA pure micro:

- Small-scale protein purification of proteins from bacterial, yeast or mammalian cell cultures;

- Protein characterization, useful for assessing protein purity, molecular weight and functionality, using techniques such as size exclusion chromatography or affinity purification.
- Process development, excellent for developing and optimizing small-scale purification methods before scaling up to larger systems.
- High-throughput screening, to optimize protein purification protocols.

### *Summary of Key Data Formats output by this instrument*

- **CSV:** This format is commonly used for exporting raw data from the system, such as chromatographic data (UV absorbance, conductivity, pressure, flow rate, etc.), fraction collection data, and system parameters.
- **XML:** Used for storing and running configuration data.
- **UDA (UNICORN Data Archive):** This is a proprietary format used by **UNICORN** software to save complete run data, including chromatograms, peak integration, fraction collection, and system logs.
- **PDF (Portable Document Format):** Often used for generating printable reports of purification runs, chromatograms, and results. The .pdf files typically contain graphs, method summaries, and peak analysis information for easy sharing and documentation.
- **TXT:** Used for basic log files or textual data that can be generated during or after a purification run, such as system status, error messages, and alerts.

These formats support data analysis, record-keeping, and sharing, and are compatible with common data analysis software for further processing or archiving. Most data are generated by **UNICORN** software, which allows easy export and analysis across these different file types.

## 4. DATA STORAGE AND HARMONISATION OF THE PLATFORM

The acquisition of instruments for metagenomic analysis, biomolecule production, and structural/functional characterization support platforms for the accelerated generation of customized bioprocesses. As expected, the different instruments and platforms generate heterogeneous data, all of which (including protocols and validated pipelines) will be made compliant with FAIR principles, in keeping with the IBISBA data policy. FAIRification of the data produced will be functional to the integration and harmonization of the platforms within the IBISBA research infrastructure. Through dedicated server storage, data provided by all instruments and platforms will be easily mapped, converted and synchronized for sharing through the ITINERIS Hub.

## 4.1 SERVER

The high-performance computing server deployed at CNR-IBBA, based on a Lenovo ThinkStation P620 with an AMD Ryzen Threadripper PRO CPU (32 cores, 64 threads), 256 GB RAM, 1 TB SSD, and now upgraded with a second NVIDIA RTX A6000 GPU (connected via NVLink), is a critical asset for the computational and data management needs of the ITINERIS project. With an expanded storage capacity (+20 TB HDD), uninterrupted power supply, and dual high-resolution monitors, the server enables the centralized storage, processing, and analysis of experimental data from all workflows — including DNA construct sequences, protein expression and purification records, and structural data from X-ray crystallography and cryo-EM. It is also configured to support advanced *in silico* analyses, such as protein structure and function prediction, modeling of post-translational modifications in eukaryotic systems, and structure determination from raw data. This infrastructure supports the creation of reproducible digital workflows and pipelines, facilitating harmonized data annotation, interoperability, and integration across national and European research infrastructures.

Data types hosted or generated by our instruments and platforms are progressively being prepared for FAIRification using tools and methodologies provided by the GO FAIR Foundation. This includes the generation of a FAIR Implementation Profile (FIP) to guide decisions on the selection and use of FAIR-enabling resources such as semantic artefacts, licenses, repositories, metadata schemas, etc. Emphasis is placed on adopting community standards (e.g. PDB/mmCIF, MRC, CSV, ...), assigning persistent identifiers, and ensuring compatibility with frameworks promoted by ELIXIR and EOSC.