Real-time fermentation tracking of p-coumaric acid using Waveguide-Enhanced Raman Spectroscopy (WERS)

Lieutaud Christopher, Perez Florent, Ompala Chardel, Emidio Rita, Godon Blandine, Fojcik Clémentine, Imatoukene Nabila

Real-time fermentation tracking of *p*-coumaric acid using Waveguide-Enhanced Raman Spectroscopy (WERS)

Runing title: Raman monitoring of *p*-coumaric acid

Lieutaud Christopher[a], Perez Florent[a], Ompala Chardel[a], Emidio Rita[a], Godon Blandine[b], Fojcik Clémentine[c], Imatoukene Nabila[b*]

[a] InSpek, 73 rue Leon Bourgeois, 91120 Palaiseau, France

[b] URD Agro-Biotechnologies Industrielles (ABI), CEBB, AgroParisTech, 51110 Pomacle, France

[c] Abolis Biotechnologies, Genopole Campus 1, 91030 Evry, France

*** Corresponding author:** nabila.imatoukene@agroparistech.fr; postal address: 3 Rue des Rouges Terres, 51110 Pomacle

Abstract

The efficient bioproduction of *p*-coumaric acid (*p*-CA) in *Saccharomyces cerevisiae* remains limited by the compound's low solubility and inhibitory effects at elevated concentrations, highlighting the need for real-time process monitoring to support advanced control strategies. This study evaluates the use of waveguide-enhanced Raman spectroscopy (WERS) as an *in situ*, real-time analytical tool for tracking *p*-CA fermentation dynamics. A recently developed Raman-on-a-chip sensor successfully quantified key metabolites directly within turbid, complex fermentation media. Chemometric models built from Raman spectra accurately predicted *p*-CA, ethanol, glycerol, glucose and acetic acid concentrations, with Root Mean Squared Error of Prediction of 0.37 $g.L^{-1}$, 2.9 $g.L^{-1}$, 1.3 $g.L^{-1}$, 6.3 $g.L^{-1}$ and 0.54 $g.L^{-1}$, and corresponding correlation coefficients of 0.63, 0.86, 0.80, 0.43 and 0.87. These performances support reliable monitoring and early detection of metabolic shift. When combined with an optimized semi-defined medium and fed-batch strategy, WERS provided a reproducible monitoring framework that supported improved *p*-CA production performance, representing a 2.3-fold increase compared to the batch process. Beyond demonstrating analytical feasibility, these results position WERS as a scalable Process Analytical Technology (PAT) suitable for integration into feedback-controlled or continuous extraction systems aimed at mitigating *p*-CA product inhibition to enhance yield, maintain culture viability, and streamline industrial *p*-CA biomanufacturing.

Keywords: waveguide-enhanced Raman spectroscopy; real time process monitoring; *p*-coumaric acid; *S. cerevisiae*; fed-batch process optimization; Process Analytical Technology

Introduction

Microbial production of aromatic compounds derived from the phenylpropanoid pathway has attracted significant industrial interest due to their applications in food, cosmetics, pharmaceuticals, and biomaterial synthesis [1,2]. Among these, *p*-coumaric acid (*p*-CA) is a phenolic compound of considerable industrial importance, used as the entry point in the phenylpropanoid pathway and is widely present in the plant cell wall [3]. It is a precursor of a wide range of other valuable plant secondary metabolites such as other *p*-hydroxycinnamic acids, flavonoids, and lignin [4]. It also serves as a precursor for pharmaceuticals, food additives, cosmetic industries, and biopolymers [5–8]. Engineering *Saccharomyces cerevisiae* for enhanced *p*-CA production has therefore become an important biotechnological objective, with recent efforts focusing on both metabolic rewiring and bioprocess optimization [2,9,10].

While medium design is a central parameter influencing flux distribution, precursor availability, and product yields, it is only one component of a broader optimization strategy. In our previous work, we developed a cost-effective semi-defined medium comprising four components, which enabled efficient *p*-CA synthesis during batch fermentation [11]. Such minimal formulations reduce complexity, enhance reproducibility, and facilitate scalability, attributes that are especially valuable for industrial process implementation [12].

Nevertheless, bioproduction of *p*-CA at scale is hindered by the compound's physicochemical and biological properties. Its hydrophobicity limits solubility in aqueous fermentation media, and its antimicrobial activity constrains culture robustness at high titers [13]. These limitations have motivated the development of *in situ* product removal and continuous extraction strategies, also known as in-stream product recovery (ISPR) [14,15]. Yet, their successful implementation requires real-time monitoring of metabolite formation to rapidly detect deviations, adjust operating conditions, and prevent inhibitory accumulation [16]. To date, high-performance liquid chromatography (HPLC) is the standard method for quantifying aromatic compounds such as *p*-CA and for monitoring fermentation progress. However, HPLC analysis is inherently offline, labor-intensive, and time-consuming. It relies on costly instrumentation and trained personnel, creating bottlenecks that impede rapid decision-making in dynamic fermentation environments [17]. These limitations have driven interest in optical process analytical technologies (PAT), particularly Raman spectroscopy, which enables non-invasive, label-free, and real-time monitoring of metabolites directly within bioreactors [18–21].

Integrated photonics have shown in recent years a significant development for multiple sensing devices with advantages regarding cost and signal efficiency [22]. In this study, we aimed to demonstrate the use of waveguide-enhanced Raman spectroscopy (WERS, "Raman-on-a-chip") sensors for in-line, real-time monitoring of *Saccharomyces cerevisiae* fermentations producing *p*-CA. Conventional process analytical technologies, such as chromatography or off-line spectroscopic analyses, do not provide the temporal resolution or robustness required for continuous or semi-

continuous bioprocess control, particularly when dealing with antimicrobial or hydrophobic products like *p*-CA. Raman spectroscopy, on the other hand, enables label-free, non-invasive, and preparation-free analysis of complex media, but its widespread adoption in bioprocess monitoring has been limited by the high cost and fragility of conventional free-space Raman probes. WERS implements the Raman interaction on a photonic chip, yielding probes that can be orders of magnitude cheaper and smaller than free-space Raman probes, while maintaining sensitivity in turbid media such as fermentation broths [23,24].

Within this work, we targeted the simultaneous monitoring of five key molecules, glucose; *p*-CA; acetic acid; glycerol; and ethanol; to evaluate the ability of the Raman-on-a-chip sensor to quantitatively track both substrate consumption and product formation during fed-batch fermentations. These analytes are key indicators of carbon flux, cellular stress, and byproduct formation, and their real-time quantification could enable more responsive process control and higher product yields. By comparing Raman-based spectral predictions with reference HPLC measurements, this study assesses the predictive accuracy and robustness of the Raman-on-a-chip platform and demonstrates its potential as a cost-efficient, scalable PAT solution for future biomanufacturing applications.

Materials and Methods

1. Strain and culture media

1.1 ABG010: Engineered *S. cerevisiae* strain

An engineered *S. cerevisiae* strain (ABG010) was provided by Abolis Biotechnologies, France. This strain was designed for *de novo* production of *p*-CA. Further details on the relevant characteristics of ABG010 are available in previous work [14].

Strain stocks were maintained at –80 °C in YPD medium containing 1% yeast extract (Fisher Scientific, Belgium), 2% peptone (Fisher Scientific, Belgium), 2% glucose (Thermo Fisher (Kandel) GmbH, Germany), supplemented with 3% glycerol. Pre-cultures were grown overnight using the same medium in shake flasks at 30 °C and 180 rpm before inoculation into bioreactors to achieve an initial optical density at 600 nm ($OD_{600\ nm}$) of 0.5, and fermentations were conducted for 72h.

1.2 Media

1.2.1 *p*-CA production in synthetic medium

Fermentations were conducted in a semi-defined synthetic medium for *p*-CA production. This medium consisted of 50 $g.L^{-1}$ of D-glucose, 11.36 $g.L^{-1}$ of urea, a solution of vitamins (0.00025 $g.L^{-1}$ biotine, 0.005 $g.L^{-1}$ calcium pantothenate, 0.005 $g.L^{-1}$ nicotinic acid, 0.125 $g.L^{-1}$ myo-inositol, 0.005 $g.L^{-1}$ thiamine HCl, 0.001 $g.L^{-1}$ pyridoxal HCl, 0.001 $g.L^{-1}$ *p*-aminobenzoic acid), 0.075 $g.L^{-1}$ EDTA, trace elements (0.015 $g.L^{-1}$ $FeSO_4$, 0.023 $g.L^{-1}$ $ZnSO_4$, 0.005 $g.L^{-1}$ $MnCl_2$, 0.0015 $g.L^{-1}$ $CoCl_2$, 0.002 $g.L^{-1}$ Na_2MoO_4, 0.0015 $g.L^{-1}$ $CuSO_4$, 2.5 $g.L^{-1}$ $MgSO_4$, 0.023 $g.L^{-1}$ $CaCl_2$, 0.005 $g.L^{-1}$ H_3BO_3, 0.0005

105 g.L^{-1} KI, 15 g.L^{-1} KH_2PO_4). A stock solution of each component was prepared and autoclaved separately at 121 °C for
106 20 min or filtered using 0.2 µm filters.

107 **1.2.2 Medium simplification through plackett Braman screening**

108 A Plackett–Burman experimental design (DoE) was employed to identify the most influential components affecting *p*-
109 CA titer. Sixteen medium components were screened with (+) and without (-) compounds. DoE was constructed and
110 statistical analysis for the development of the medium was carried out using MODDE ® v12.0 software (Sartorius Stedim
111 Data Analytics, Sweden). Variables tested (sixteen) are presented in Table 1 which included twenty-three different
112 experimental runs and four central points. Cultures were performed in 250 mL baffled flasks containing 25 mL of each
113 medium. Based on the screening results, a simplified 8-component semi-defined medium was adopted for all subsequent
114 Raman-monitored fermentations, as it provided equivalent *p*-CA titers with reduced formulation complexity and cost.

115 **Table 1:** Plackett Burman design matrix, with X1 – $ZnSO_4$, X2 - $MnCl_2$, X3 - $CoCl_2$, X4 - $CuSO_4$, X5 - Na_2MoO_4, X6 -
116 $CaCl_2$, X7 - $FeSO_4$, X8 - H_3BO_3, X9 – KI, X10 - Calcium pantothenate, X11 - nicotinic acid, X12 - myo-inositol, X13 -
117 thiamine HCl, X14 - pyridoxal HCl, X15 - *p*-aminobenzoic acid, X16 – EDTA.

N°	**X1**	**X2**	**X3**	**X4**	**X5**	**X6**	**X7**	**X8**	**X9**	**X10**	**X11**	**X12**	**X13**	**X14**	**X15**	**X16**
1	+	-	+	+	-	-	-	-	+	-	+	-	+	+	+	+
2	+	+	-	+	+	-	-	-	-	+	-	+	-	+	+	+
3	-	+	+	-	+	+	-	-	-	-	+	-	+	-	+	+
4	-	-	+	+	-	+	+	-	-	-	-	+	-	+	-	+
5	+	-	-	+	+	-	+	+	-	-	-	-	+	-	+	-
6	+	+	-	-	+	+	-	+	+	-	-	-	-	+	-	+
7	+	+	+	-	-	+	+	-	+	+	-	-	-	-	+	-
8	+	+	+	+	-	-	+	+	-	+	+	-	-	-	-	+
9	-	+	+	+	+	-	-	+	+	-	+	+	-	-	-	-
10	+	-	+	+	+	+	-	-	+	+	-	+	+	-	-	-
11	-	+	-	+	+	+	+	-	-	+	+	-	+	+	-	-
12	+	-	+	-	+	+	+	+	-	-	+	+	-	+	+	-
13	-	+	-	+	-	+	+	+	+	-	-	+	+	-	+	+
14	-	-	+	-	+	-	+	+	+	+	-	-	+	+	-	+
15	-	-	-	+	-	+	-	+	+	+	+	-	-	+	+	-
16	-	-	-	-	+	-	+	-	+	+	+	+	-	-	+	+
17	+	-	-	-	-	+	-	+	-	+	+	+	+	-	-	+
18	+	+	-	-	-	-	+	-	+	-	+	+	+	+	-	-
19	-	+	+	-	-	-	-	+	-	+	-	+	+	+	+	-
20	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
21	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

118 **2. Batch and fed-batch fermentation strategies**

119 Fermentation experiments were performed at URD ABI AgroParisTech (Reims, France) using 1.5 L glass-jacketed
120 bioreactors equipped with a PRO-LAB™ controller unit, C-BIO2™ operator and control software from GPC (GPC, La

Rochelle, France). Bioreactors were equipped with standard PG13.5 sensor ports, pH and dissolved oxygen (DO) probes from Hamilton Company, as well as automated control of temperature, agitation, and aeration. The working volume was set to 500 mL.

All fermentations were conducted at 30 °C under aerobic conditions, with a constant air flow of 0.5 vvm and a mixing rate of 800 rpm. The pH was controlled at 6.0 ± 0.05 using KOH 1 M or H_2SO_4 1 M solutions. Antifoam PMC Ouvrie was added as needed. Each bioreactor was inoculated at an initial optical density (OD_{600} = 0.5) from shake-flask pre-cultures.

2.1 Batch fermentation

In batch experiments, the ABG010 strain was grown in a medium composed of 30 $g \cdot L^{-1}$ D-glucose, 11.36 $g \cdot L^{-1}$ urea, 0.00025 $g \cdot L^{-1}$ biotin, 0.015 $g \cdot L^{-1}$ $FeSO_4$, 0.023 $g \cdot L^{-1}$ $ZnSO_4$, 0.0015 $g \cdot L^{-1}$ $CuSO_4$, 2.5 $g \cdot L^{-1}$ $MgSO_4$, 0.0005 $g \cdot L^{-1}$ KI, and 15 $g \cdot L^{-1}$ KH_2PO_4, with a working fermentation volume of 1 L under identical physical conditions.

2.2 Fed-batch fermentation

The experiments were conducted using the same physical parameters as batch experiments. The composition of the initial medium was the same as in batch, except the initial glucose concentration which is 20 $g.L^{-1}$ of D-glucose. The working volume was 500 mL of fermentation medium. Feeding with the fresh medium was started only after residual ethanol produced from D-glucose consumption phase was totally consumed, maintaining a target concentration of approximately 2 $g.L^{-1}$ throughout the production phase. Fresh medium was composed of concentrated glucose solution (200 $g.L^{-1}$), 5.62 $g.L^{-1}$ of urea, and the same other components as in the initial medium. Fed-batch fermentation was used to acquire Raman spectroscopy readings and verified by off-line analysis every two hours.

2.3 Selection of the optimal operating mode

The comparative evaluation of batch and fed-batch modes clearly indicated that fed-batch fermentation provided superior performance in terms of *p*-CA titer (Figure 2). Moreover, the fed-batch strategy enables longer fermentation runs and exposes the culture to a broader range of operating conditions. This increased variability across key process parameters provides a richer and more informative dataset, which is advantageous for developing and validating Raman-based modelling approaches. Consequently, all subsequent Raman-monitored fermentation campaigns were conducted under fed-batch operation using reactors equipped with Raman-on-a-chip probes.

The combination of reproducible fed-batch performance and reliable in-line Raman data acquisition formed the basis for developing and validating predictive chemometric models correlating Raman spectra with metabolite concentrations. Five fermentations were carried out in fed-batch to ensure robust data collection, enough for model building.

3. Monitoring of metabolites bioproduction using Raman–on-a-chip

Real-time monitoring of the fermentation process was performed using WERS sensors whose design is described in (Kita et al., 2020 The probes were specifically engineered for bioprocess integration and manufactured from stainless

steel 316L to ensure chemical inertness, biocompatibility, and resistance to sterilization procedures. Each immersion probe was interfaced with the bioreactor through a PG 13.5 standard port, enabling fully sterile insertion and removal. Prior to each run, probes were sterilized using ethanol.

Each probe contained a silicon nitride photonic chip that guided both the excitation and scattered Raman light through a waveguide whose design enhanced the optical interaction with the fermentation broth, allowing robust signal collection even in turbid media [25]. The probe was connected via optical fiber to a control unit that housed a 785 nm diode laser (≈ 100 mW at the probe tip), collimation optics, and a thermoelectrically cooled CCD spectrometer.

During fermentation, Raman spectra were automatically acquired every 30 min throughout the 72-hour fed-batch campaigns. For each time point, 30 individual spectra (10 s integration each) were collected and averaged to account for intra-sample variability and to improve the signal-to-noise ratio. Five metabolites were targeted for quantification: D-glucose, *p*-CA, ethanol, glycerol, and acetic acid, representing key indicators of carbon flux, redox balance, and production efficiency. Reference samples were withdrawn every 2 hours during daytime and analyzed by HPLC (see Section 4.2). For time points between HPLC analyses, mainly during the night, reference concentrations were interpolated linearly between adjacent values. Over the five monitored fermentations, the combined dataset comprised ≈ 17 000 spectra and 2 370 HPLC reference values (Table 2).

Raman spectroscopy is sensitive to experimental conditions, including laser power fluctuations, optical alignment, probe fouling, temperature changes, hardware changes, and matrix effects [26]. These sources of variability can significantly alter baseline intensity and peak shapes. To take these phenomena in consideration, the strategy involved for data acquisition included a selection of run designs to include all sources of hardware variability and have more representative models [27]. Each run was carried out with different probes, different control units, and different conditions of acquisition detailed in Table 2. Run 1 contained fewer spectra since the first night was not monitored. Run 2 contained substantially fewer spectra due to switching to off-line Raman measurements. In this case, spectra were only acquired concurrently with HPLC sampling.

For spectra acquired without corresponding reference values (i.e., between HPLC sampling points and during the night), reference concentrations were obtained by linear interpolation between adjacent HPLC measurements. Runs 3, 4 and 5 were used with full spectra collection during day and night times.

The combined dataset from all five runs was used to train Partial Least Squares (PLS) regression models for quantification of the five target molecules. For each molecule, 80% of the data were used for model training and cross-validation, with the constraint that all spectra from a given sample were kept together within the same split and cross-validation folds. The remaining 20% of the data was held out for independent testing.

Table 2: Spectral database composition

Experiment	Number of Raman spectra	Number of reference values	Notes

Run 1	2 970	480	Partial night missing
Run 2	420	420	Off-line Raman only (hardware issue)
Run 3	4 510	480	Full fed-batch
Run 4	4 840	480	Full fed-batch
Run 5	4 413	510	Full fed-batch

For the fermentation dataset, Raman spectra were split into 80/20 train/test sets, ensuring that all spectra from a given sample remained grouped to avoid data leakage. Spectral preprocessing and modeling strategies were customized for each metabolite and are detailed in the following sections.

3.1 Spectral preprocessing

Spectral preprocessing was conducted in two main stages. First, all spectra were visually inspected and corrected for cosmic ray artefacts. Such events, caused by high-energy particles from space striking the detector during acquisition, are common in Raman spectroscopy [28]. They produce sharp, localized spikes in the spectrum that occur randomly across measurements. Following this correction, the initial portion of the Raman spectra (below 580 cm^{-1}) was excluded from further analysis due to the limited informational content in this range, which was predominantly influenced by the WERS device signal from the fiber core and cladding [29]. This procedure yielded a clean baseline dataset, subsequently processed according to each target molecule. Indeed, the optimal preprocessing pipeline for one compound (e.g., glucose) is not necessarily suitable for another (e.g., *p*-CA). Therefore, an automated optimization pipeline was implemented to determine, for each analyte, the preprocessing sequence that minimized the modeling error. The performance criterion used to select the best preprocessing configuration was the Root Mean Squared Error (RMSE) obtained during cross-validation (RMSEcv).

12 preprocessing methods were considered, including smoothing, various normalizations, baseline correction, and derivatives. A preprocessing string generator systematically produced all meaningful combinations of these methods up to a maximum sequence length N. “Meaningful combinations” were defined as those avoiding redundant or incompatible operations (for example, two consecutive first-order derivatives, since the second derivative is already included in the processing library, or sequential normalizations that do not add information). In this study, the maximum preprocessing string length was set to N=3, which represented a practical balance between empirical performance and computational cost. This resulted in a total of 233 distinct preprocessing pipelines evaluated for each of the five target molecules. At the end of each preprocessing chain, a Principal Component Analysis (PCA) with 40 components was applied. The raw spectra, originally of dimensionality 1886, were thus projected onto a lower-dimensional space that preserved the relevant variability of the signal while substantially improving computational efficiency. Using 40 components is generally sufficient to retain the dominant variance structures in Raman datasets, ensuring that subsequent modeling steps benefit from both reduced noise and faster processing times.

3.2 Chemometric modelling

For each target molecule, an automated modeling pipeline was executed to identify the model yielding the lowest RMSEcv. Four regression approaches were evaluated: PLS [30], Support Vector Regression (SVR), Relevance Vector Machine (RVM) [31], and Ridge regression.

The complete dataset (≈17,000 spectra and 2,370 reference values) was first aggregated and then split into training and test sets using an 80/20 stratified split. Stratification ensured that all spectra corresponding to a given experimental timepoint were kept together, thereby preserving the underlying sample structure and preventing information leakage. For each of the 233 preprocessing pipelines, all four models went through Bayesian hyperparameter optimization. This procedure sequentially explores the hyperparameter space by building a probabilistic surrogate model of the objective function (here, RMSEcv). Instead of performing an exhaustive or grid-based search, the algorithm estimates which regions of the hyperparameter space are most promising and samples them more intensively. This strategy enables efficient convergence toward well-performing configurations while substantially reducing computational cost. In total, 4 × 233 = 932 optimized models were generated for each molecule. All models were subsequently ranked according to their RMSEcv, and the best-performing model was selected.

4. Reference analytical methods

4.1 Quantification of dry matter

Biomass was quantified *via* determination of the cell dry weight (CDW). Aliquots of 5 mL culture were centrifuged at 4500 × *g* for 10 min, and the resulting pellets were washed once with an equivalent volume of distilled water to remove traces of medium components. Pellets were then dried overnight at 105 °C, and their cell dry weight was determined gravimetrically. Yeast growth was additionally monitored by measuring optical density at 600 nm (OD_{600}) using a Cary 60 UV-Vis spectrophotometer (Agilent, Marcy l'Etoile, France).

4.2 Reference method HPLC

2 mL of samples were collected from the bioreactor every 2 hours for 72 hours, excluding nighttime. Quantification of *p*-CA, D-glucose, ethanol, glycerol, and acetic acid was carried out using HPLC (Thermo scientific, ultimate 3000) equipped with both an ultraviolet-diode array detector (UV-DAD) using an accucore aQ C18 column (100x3 mm, Thermo scientific) for *p*-CA quantification and a Shodex refractive index detector and a Thermo scientific ultraviolet detector at 210 nm using an Aminex HPX-87H column (300x7.8 mm, Bio-rad Laboratories S.A., Marnes La Coquette, France) for D-glucose, ethanol, glycerol and acetic acid quantifications. For more information on HPLC analysis, it can be useful to see our previous work [14].

Results

1. ***p*-CA production in synthetic medium**

The engineered *S. cerevisiae* strain ABG010 exhibited significantly enhanced *p*-CA production when cultivated in synthetic medium, reaching 1.08 ± 0.03 g.L^{-1}. This corresponds to more than a fourfold increase relative to the 0.25 g.L^{-1} obtained in complex medium (data not shown). In our previous work, *p*-CA titers of 0.50 g·L^{-1} were achieved in YNB medium and 0.25 g·L^{-1} in an optimized semi-defined medium [32]. Collectively, these results suggest that synthetic medium provides more favorable conditions for *p*-CA biosynthesis, potentially due to improved precursor availability or reduced metabolic competition associated with complex nutrients. However, both the synthetic medium used in this study and YNB are relatively costly, owing to their content of vitamins and other high-value components. This consideration motivated our efforts to minimize the medium complexity and cost while maintaining comparable production levels.

2. Plackett Burman screening for medium simplification

First, a screening of DoE was performed to select compounds with a high impact on *p*-CA production to use the simplified medium in batch and fed-batch fermentation strategy.

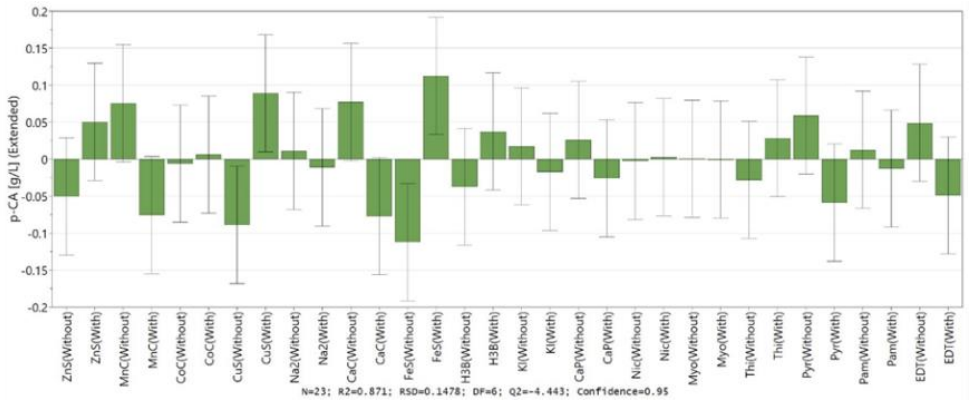

Figure 1: Regression coefficients of the plackett Burman design for *p*-CA production. N = 23; R^2 = 0.865; RDS = 0.1012; Q^2 = 0.665.

As shown in Figure 1, $MnCl_2$, $CoCl_2$, Na_2MoO_4, $CaCl_2$, H_3BO_3, KI, calcium pantothenate, nicotinic acid, myo-inositol, thiamine-HCl, pyridoxal-HCl, *p*-aminobenzoic acid, and EDTA exhibited no significant effect on *p*-CA production. In contrast, $CuSO_4$ and $FeSO_4$ showed a positive and statistically significant impact. Although $ZnSO_4$ did not reach significance, its positive trend justified retaining it in the formulation. Overall, the experimental design enabled the removal of thirteen components, many of which were among the most expensive additives, including several vitamins, thereby simplifying and reducing the cost of the medium.

An almost equivalent level of *p*-CA production (0.825 g.L^{-1}) was achieved after removing the non-essential components identified in the experimental design, compared to the initial medium containing more components. This simplification provides several advantages, including a substantial reduction in medium cost, easier preparation, and significant time savings, thereby improving the overall practicality and scalability of the process.

3. Batch and fed-batch fermentation strategies

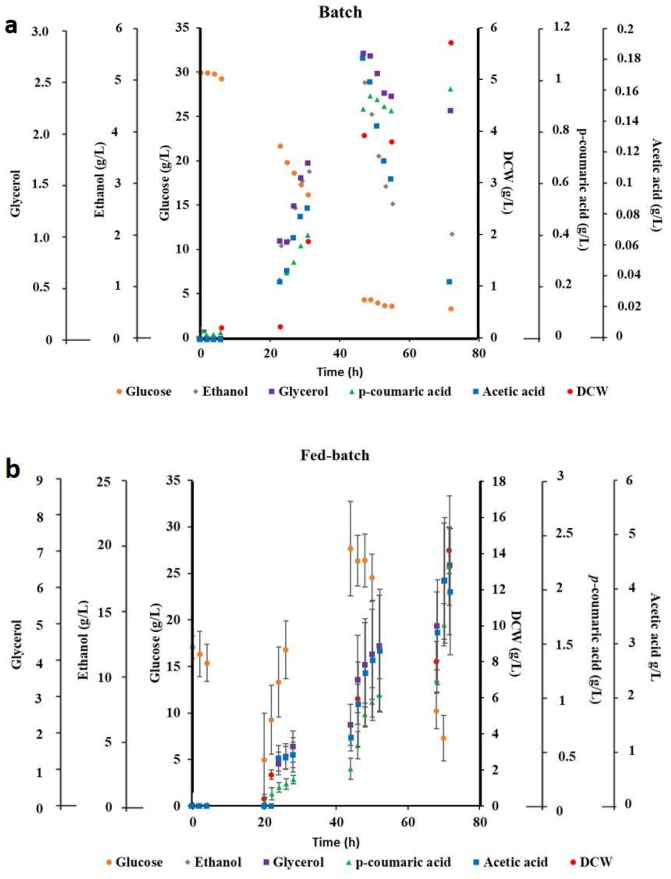

Figure 2: Total production concentrations of *p*-CA, D-glucose consumption, ethanol, glycerol, acetic acid, and cell dry weight productions in batch (a) and fed-batch (b) fermentations. Fed-batch data represent the mean of n = 5 independent replicates, and error bars show standard deviation.

The simplified medium identified through the Plackett Burman design was used in both batch and fed-batch fermentations. In batch mode, glucose was provided at an initial concentration of 30 $g.L^{-1}$. Under these conditions, *p*-CA accumulation occurred during the respiratory phase, approximately between 23 h and 48 h, reaching a final titer of 0.96 ± 0.1 $g.L^{-1}$ after approximately 72 h. This observation aligns with the findings of Frick and Wittmann (2005), who demonstrated through flux analysis that fermentative metabolism in *S. cerevisiae* reduces carbon flux through the pentose phosphate pathway, thereby limiting the availability of erythrose-4-phosphate (E4P), and decreases the intracellular pool of phosphoenolpyruvate (PEP) due to ethanol formation [33]. Because E4P and PEP are the entry substrates of the shikimate pathway leading to L-Phe and L-Tyr, aromatic amino acids that serve as precursors for *p*-CA biosynthesis [34], their reduced availability under fermentative conditions likely accounts for the observed production profile (Figure 2). Glucose was fully consumed in 48 h, after which *p*-CA production plateaued, likely due to substrate exhaustion and *p*-CA toxicity effects. Acetic acid and glycerol accumulated as secondary metabolites, while ethanol was transiently produced during oxygen-limited phases, and then used as the carbon source.

Fed-batch approach successfully extended the productive phase and alleviated both carbon limitation and *p*-CA-associated stress. Under these conditions, the process yielded over 2.21 g.L^{-1} of *p*-CA after 72 h, as determined by HPLC, representing a 2.3-fold increase compared to batch fermentation. The enhanced titer was accompanied by sustained glucose utilization and higher total biomass accumulation, representing 14 g. L^{-1} of CDW. Although the fed-batch strategy enabled a twofold increase in *p*-CA production, substantial levels of by-products persisted. Implementing real-time monitoring via Raman spectroscopy could enhance the redirection of metabolic fluxes and thereby improve process performance.

4. Calibration of WERS system on *p*-CA samples

Before integration of the sensor in the bioreactor, calibration experiments using purified *p*-CA standard. Raman spectra were acquired in distilled water from 50 mg.L^{-1} to 800 mg.L^{-1} (Figure 3).

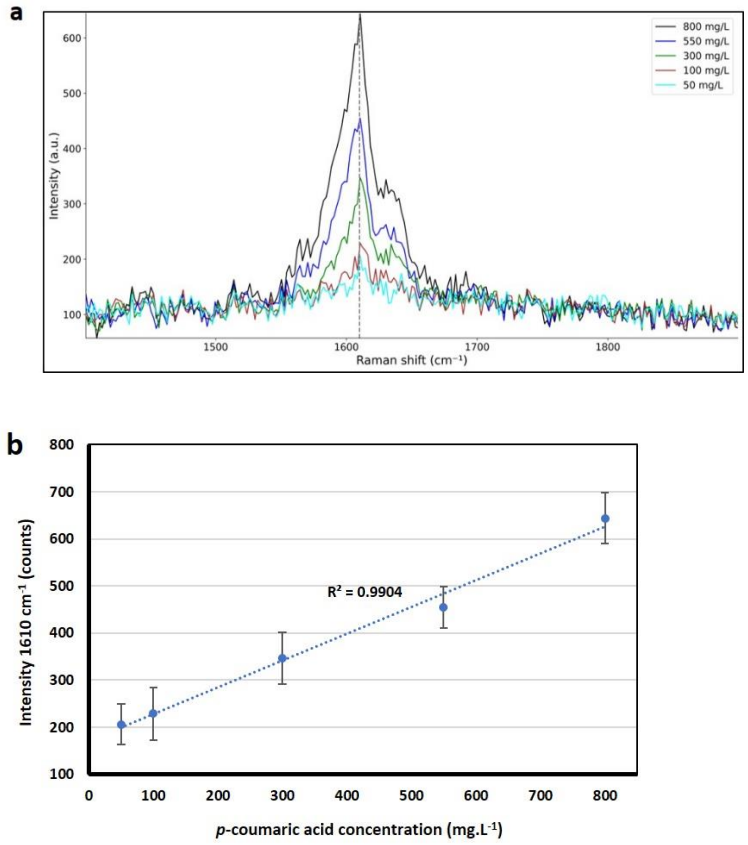

Figure 3: a. Average Raman spectra from *p*-CA diluted with distilled water. b. Correlation between the sample concentration and its Raman intensity at 1610 cm^{-1}

Results revealed a strong linear relationship ($R^2 > 0.99$) between the intensity of the characteristic band at 1610 cm^{-1} and the actual concentration. The detection limit in distilled water was ≈ 50 mg.L^{-1}, validating the suitability of the system for *p*-CA monitoring *in situ*. The specific peak detected is linked to aromatic cycle vibrations [35,36] and is detected between 50 and 800 mg.L^{-1}.

5. On-line monitoring of fed-batch fermentation and WERS modelling

The optimal condition for *p*-CA production which is the fed-batch strategy was reproduced in five independent experiments to construct a model based on Raman probe measurements and to identify correlations with HPLC analyses. The modeling results are summarized in Table 3. Performance was assessed using the coefficient of determination (R^2) and the root mean squared error (RMSE). R^2 measures the proportion of variance in the reference data explained by the model (with values closer to 1 indicating stronger predictive ability), while RMSE quantifies the average prediction error in the same units as the measured concentration.

To ensure an unbiased evaluation of predictive performance, we first set aside 20% of the full dataset as an external test set. These samples were not used at any stage of model development and served exclusively to assess the final generalization performance. The remaining 80% of the data constituted the training pool on which model selection and optimization were conducted. Within this subset, we performed a 15-fold cross-validation, in which the training data were repeatedly partitioned into calibration and evaluation folds. Each fold served once as the evaluation subset while the remaining data were used for model fitting. This repeated resampling procedure enabled robust estimation of model stability and allowed tuning of model hyperparameters while minimizing overfitting.

After identifying the best-performing model based on cross-validation results, the final model was retrained using the entire 80% training subset and subsequently evaluated on the held-out 20% external test set. Table 3 below reports the performance obtained during the cross-validation phase as well as the results on the independent test set.

Table 3: Modeling results

Molecule	**Best model**	**Cross-validation (80 %)**		**Test (20 %)**	
		R^2	RMSEcv ($g.L^{-1}$)	R^2	RMSE ($g.L^{-1}$)
Acetic acid	Ridge	0.86	0.541	0.87	0.503
Ethanol	Ridge	0.84	2.932	0.86	2.847
Glucose	PLS	0.40	6.279	0.43	6.166
Glycerol	PLS	0.63	1.292	0.80	0.958
p-coumaric acid	PLS	0.66	0.372	0.63	0.367

Two molecules, ethanol and acetic acid, yielded strong models with $R^2 > 0.8$. Glycerol also reached satisfactory predictive performance in the test set ($R^2 > 0.8$). For *p*-CA, the model captured part of the concentration trend ($R^2 > 0.6$), although predictive accuracy remained limited. In contrast, glucose was poorly predicted, with low R^2 values indicating that the Raman signal did not provide sufficient information for robust quantification.

Although this analysis was based on five independent runs using different probes and control units, a separate study employing conventional Raman spectroscopy reported comparable RMSEcv values for glucose and ethanol, with errors of 6.65 and 3.64 $g \cdot L^{-1}$, respectively [21].

Figure 4 describes the behaviours of the models for each molecule during cross-validation from WERS measurements.

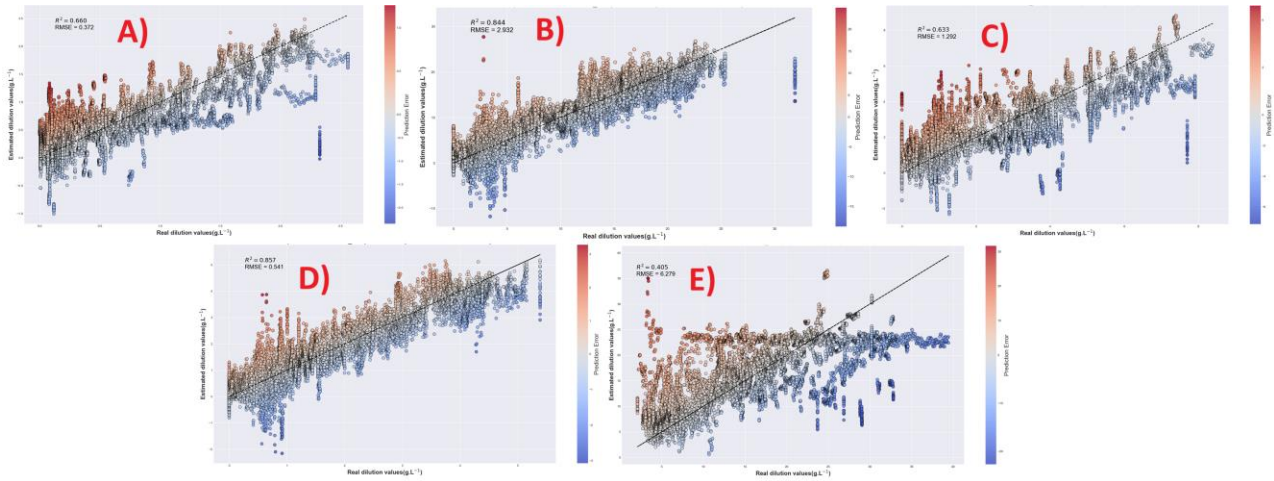

Figure 4: Prediction in cross-validation for each of the different analytes, x-axis is the real concentration value, y-axis is the predicted concentration **A.** *p*-CA **B.** ethanol **C.** glycerol **D.** acetic acid **E.** glucose

Models with the performance levels reported in the above table may already enable meaningful in-line monitoring of the corresponding target molecules. Figure 5 illustrates the predicted concentrations over time for each molecule during Run 4.

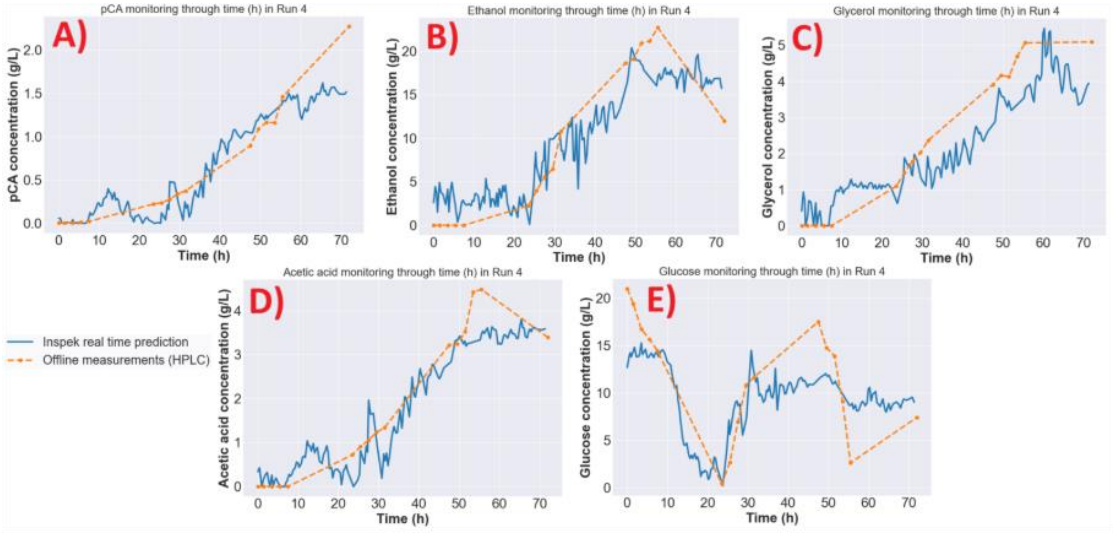

Figure 5: Run 4 molecules cross tracking using Raman cross-validation predictions. Blue is prediction, Orange is HPLC values. **A**. *p*-CA, **B**. ethanol, **C**. glycerol, **D**. acetic acid, **E**. glucose.

These predictions are based on cross-validation, meaning that some spectra from run 4 were included in training in order to predict on other run 4 points. As such, the results provide a first demonstration of what real-time concentration monitoring could look like, even if not yet under fully unknown test conditions. The *p*-CA model achieved a correlation coefficient (R^2) of 0.80 and a root-mean-square error of prediction (RMSEP) below 400 $mg.L^{-1}$. Comparable accuracies were obtained for ethanol, glycerol, acetic acid and glucose ($R^2 > 0.75$). Trends shown in Figure 5 demonstrate that online measurements are well fit with offline reference values provided by HPLC as described in Figure 2. Figure 5.A

with *p*-CA shows a clear curve break in predictions in the late stages with high concentrations, that could be linked to the fact that this molecule is less soluble above 800 mg.L^{-1}.

The robustness of the Raman-on-a-chip sensors was demonstrated across multiple probes and control units, confirming consistent signal quality under repeated sterilization and prolonged immersion. These results establish the feasibility of real-time, quantitative monitoring of *p*-CA bioproduction using compact, low-cost photonic sensors and validate the technological readiness of the system for industrial bioprocess applications.

Conclusion

This study demonstrates the feasibility of real-time, quantitative monitoring of *p*-coumaric acid (*p*-CA) bioproduction in *Saccharomyces cerevisiae* using waveguide-enhanced Raman spectroscopy (WERS) sensors. The Raman-on-a-chip approach enabled reliable in-line quantification of key metabolites including glucose, ethanol, glycerol, acetic acid, and *p*-CA directly within turbid fermentation media. The chemometric models achieved prediction accuracies (RMSEP < 0.4 g·L^{-1}, R^2 up to 0.8) suitable for process supervision and early-stage control applications. Together with the optimized semi-defined medium and fed-batch operation, these results establish a reproducible framework for microbial *p*-CA production and confirm the analytical robustness of WERS as a scalable Process Analytical Technology (PAT) tool.

Since *p*-CA has limited solubility and exhibits cellular toxicity at elevated concentrations, maintaining sub-inhibitory levels in the broth is a critical challenge for process intensification. The implementation of WERS as an online monitoring tool provides the foundation for feedback-controlled or continuous extraction strategies aimed at mitigating product inhibition and prolonging culture viability. Future developments will focus on coupling WERS-based process monitoring with continuous product recovery systems to enable fully integrated, self-regulated *p*-CA biomanufacturing.

Author contributions

Lieutaud Christopher: Conceptualization, Methodology, Investigation, Data curation, Writing - review & editing; Perez Florent: Raman Formal analysis, investigation, Writing - review & editing; Ompala Chardel & Emidio Rita: Raman data acquisition and analysis; Godon Blandine: Fermentation Formal analysis, Data curation; Fojcik Clémentine: Investigation; Nabila Imatoukene: Supervision, Conceptualization, Methodology, Formal analysis, Investigation, Writing - review & editing.

Funding

This work was supported by the Région Grand Est (France), Conseil Départemental de la Marne (France), Grand Reims (France), European Union's Single Market Programme via the Open Call BioMan4R2 and from the European Innovation Council and SMEs Executive Agency. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or European Innovation Council and SMEs Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. The Agence Nationale de la Recherche (ANR, Grant ANR-17-CE07-0046) funded strain construction.

Conflict of interest statement

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported here.

References

1. Averesch, N.J.H.; Kayser, O. Editorial: Biotechnological Production and Conversion of Aromatic Compounds and Natural Products. *Front. Bioeng. Biotechnol.* **2020**, *8*, doi:10.3389/fbioe.2020.00646.

2. Rodriguez, A.; Kildegaard, K.R.; Li, M.; Borodina, I.; Nielsen, J. Establishment of a Yeast Platform Strain for Production of P- Coumaric Acid through Metabolic Engineering of Aromatic Amino Acid Biosynthesis. *Metab. Eng.* **2015**, *31*, 181–188, doi:10.1016/j.ymben.2015.08.003.

3. Reverón, I.; de las Rivas, B.; Muñoz, R.; López de Felipe, F. Genome-Wide Transcriptomic Responses of a Human Isolate of Lactobacillus Plantarum Exposed to p-Coumaric Acid Stress. *Mol. Nutr. Food Res.* **2012**, *56*, 1848–1859, doi:10.1002/mnfr.201200384.

4. Mutz, M.; Kösters, D.; Wynands, B.; Wierckx, N.; Marienhagen, J. Microbial Synthesis of the Plant Natural Product Precursor P-Coumaric Acid with Corynebacterium Glutamicum. *Microb. Cell Fact.*

2023, *22*, 1–18, doi:10.1186/s12934-023-02222-y.

5. Boo, Y.C. P-Coumaric Acid as an Active Ingredient in Cosmetics: A Review Focusing on Its Antimelanogenic Effects. *Antioxidants* **2019**, *8*, doi:10.3390/antiox8080275.

6. Pei, K.; Ou, J.; Huang, J.; Ou, S. P-Coumaric Acid and Its Conjugates: Dietary Sources, Pharmacokinetic Properties and Biological Activities. *J. Sci. Food Agric.* **2016**, *96*, 2952–2962, doi:10.1002/jsfa.7578.

7. Zaman, A.; Hasnat, H.; Noman, Z. Al; Islam, M.M.; Nakib, A. Al; Mukherjee, S.; Saha, K.; Ahmed, N.U.; Ashrafi, S.; Saha, T.; et al. Exploring Pharmacological Potentials of P-Coumaric Acid: A Prospective Phytochemical for Drug Discovery. *Bangladesh Pharm. J.* **2023**, *26*, 185–194, doi:10.3329/bpj.v26i2.67808.

8. Zhang, R.; Shah, A.A.; Wang, B.; Gong, C. Current Progress of P-Coumaric Acid Production Using Bioengineering Technologies. *Food Biosci.* **2025**, *71*, 107109, doi:10.1016/j.fbio.2025.107109.

9. Borja, G.M.; Rodriguez, A.; Campbell, K.; Borodina, I.; Chen, Y. Metabolic Engineering and Transcriptomic Analysis of Saccharomyces Cerevisiae Producing p - Coumaric Acid from Xylose. *Microb. Cell Fact.* **2019**, 1–14, doi:10.1186/s12934-019-1244-4.

10. Flourat, A.L.; Combes, J.; Bailly-Maitre-Grand, C.; Magnien, K.; Haudrechy, A.; Renault, J.H.; Allais, F. Accessing P-Hydroxycinnamic Acids: Chemical Synthesis, Biomass Recovery, or Engineered Microbial Production? *ChemSusChem* **2021**, *14*, 118–129, doi:10.1002/cssc.202002141.

11. Combes, J.; Imatoukene, N.; Couvreur, J.; Godon, B.; Fojcik, C.; Allais, F.; Lopez, M. An Optimized Semi-Defined Medium for p-Coumaric Acid Production in Extractive Fermentation. *Process Biochem.* **2022**, *122*, 357–362, doi:10.1016/j.procbio.2022.10.021.

12. Liu, Q.; Yu, T.; Li, X.; Chen, Y.; Campbell, K.; Nielsen, J.; Chen, Y. Rewiring Carbon Metabolism in Yeast for High Level Production of Aromatic Chemicals. *Nat. Commun.* **2019**, *10*, 1–13, doi:10.1038/s41467-019-12961-5.

13. Ultee, A.; Bennik, M.H.J.; Moezelaar, R. The Phenolic Hydroxyl Group of Carvacrol Is Essential for Action against the Food-Borne Pathogen Bacillus Cereus. **2002**, *68*, 1561–1568, doi:10.1128/AEM.68.4.1561.

14. Combes, J.; Imatoukene, N.; Couvreur, J.; Godon, B.; Brunissen, F.; Fojcik, C.; Allais, F.; Lopez, M. Intensification of P-Coumaric Acid Heterologous Production Using Extractive Biphasic Fermentation. *Bioresour. Technol.* **2021**, *337*, doi:10.1016/j.biortech.2021.125436.

15. Combes, J.; Imatoukene, N.; Moussa, M.; Coquart, N.; Chemarin, F.; Athès, V.; Fojcik, C.; Chadni, M.; Ioannou, I.; Lopez, M.; et al. In-Stream Product Recovery of p-Coumaric Acid Heterologously Produced: Implementation of a Continuous Liquid-Liquid Extraction Assisted by Hollow Fiber Membrane Contactor. *Sep. Purif. Technol.* **2022**, *293*, doi:10.1016/j.seppur.2022.121083.

16. Rathore, A.S.; Bhambure, R.; Ghare, V. Process Analytical Technology (PAT) for Biopharmaceutical Products. *Anal. Bioanal. Chem.* **2010**, *398*, 137–154, doi:10.1007/s00216-010-3781-x.

17. Armstrong, A.; Horry, K.; Cui, T.; Hulley, M.; Turner, R.; Farid, S.S.; Goldrick, S.; Bracewell, D.G. Advanced Control Strategies for Bioprocess Chromatography: Challenges and Opportunities for Intensified Processes and next Generation Products. *J. Chromatogr. A* **2021**, *1639*, doi:10.1016/j.chroma.2021.461914.

18. Abu-Absi, N.R.; Kenty, B.M.; Cuellar, M.E.; Borys, M.C.; Sakhamuri, S.; Strachan, D.J.; Hausladen, M.C.; Li, Z.J. Real Time Monitoring of Multiple Parameters in Mammalian Cell Culture Bioreactors Using an In-Line Raman Spectroscopy Probe. *Biotechnol. Bioeng.* **2010**, doi:10.1002/bit.23023.

19. Karnachoriti, M. Raman Spectroscopy as a Tool for Real- - Time Nutrient Monitoring in Bioreactor Cultivation of Microalgae. **2025**, 817–826, doi:10.1002/jrs.6841.

20. Rubini, M.; Boyer, J.; Poulain, J.; Berger, A.; Saillard, T.; Louet, J.; Soucé, M.; Roussel, S.; Arnould, S.; Verg, M.; et al. Monitoring of Nutrients , Metabolites , IgG Titer , and Cell Densities in 10 L Bioreactors Using Raman Spectroscopy and PLS Regression Models. **2025**.

21. Yang, N.; Guerin, C.; Kokanyan, N.; Perré, P.; Yang, N.; Guerin, C.; Kokanyan, N.; Perré, P. Raman Spectroscopy Applied to Online Monitoring of a Bioreactor : Tackling the Limit of Detection To Cite This Version : HAL Id : Hal-04227441 Raman Spectroscopy Applied to Online Monitoring of a Bioreactor : Tackling the Limit of Detection. **2025**, 0–22.

22. Wang, X.; Li, T.; Hu, J.; Liu, J. Perspectives of Active Si Photonics Devices for Data Communication and Optical Sensing. *J. Appl. Phys.* **2025**, *138*, doi:10.1063/5.0263644.

23. Kita, D.M.; Michon, J.; Hu, J. A Packaged, Fiber-Coupled Waveguide-Enhanced Raman Spectroscopic Sensor. *Opt. Express* **2020**, *28*, 14963, doi:10.1364/oe.392486.

24. Sanchez, D.; Lieutaud, C.; Bonnassies, P.; Ibrahimi, Y.; Imatoukene, N.; Michon, J. Waveguide-Enhanced Raman Sensors for Bioprocess Monitoring. **2023**, 26, doi:10.1117/12.3008430.

25. Jérôme Michon, Derek Kita, J.H. Sensitivity Comparison of Free-Space and Waveguide Raman for Bulk Sensing. *J. Opt. Soc. Am. B* **2020**, *37*, 2012–2020.

26. Rusu, E.; Baia, M. Moving from Raman Spectroscopy Lab towards Analytical Applications : A Review of Interlaboratory Studies. **2023**, 1–16.

27. Picot, F.; Daoust, F.; Sheehy, G.; Dallaire, F.; Chaikho, L.; Bégin, T.; Kadoury, S.; Leblond, F. Data Consistency and Classification Model Transferability across Biomedical Raman Spectroscopy Systems. *Transl. Biophotonics* **2021**, *3*, 1–11, doi:10.1002/tbio.202000019.

28. Coca-lopez, N. Analytica Chimica Acta An Intuitive Approach for Spike Removal in Raman Spectra Based on Peaks ' Prominence and Width. *Anal. Chim. Acta* **2024**, *1295*, 342312, doi:10.1016/j.aca.2024.342312.

29. Wang, P.; Miller, B.L. Waveguide-Enhanced Raman Spectroscopy (WERS): An Emerging Chip-Based Tool for Chemical and Biological Sensing. **2022**.

30. Wold, S.; Sjöström, M.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130, doi:10.1016/S0169-7439(01)00155-1.

31. Tipping, M.E. The Relevance Vector Machine. *Adv. Neural Inf. Process. Syst.* **2000**, 653–658.

32. Combes, J.; Imatoukene, N.; Couvreur, J.; Godon, B.; Fojcik, C.; Allais, F.; Lopez, M. An Optimized Semi-Defined Medium for p-Coumaric Acid Production in Extractive Fermentation. *Process Biochem.* **2022**, *122*, 357–362, doi:10.1016/j.procbio.2022.10.021.

33. Frick, O.; Wittmann, C. Characterization of the Metabolic Shift between Oxidative and Fermentative Growth in Saccharomyces Cerevisiae by Comparative 13C Flux Analysis. *Microb. Cell Fact.* **2005**, *4*, doi:10.1186/1475-2859-4-30.

34. Herrmann, K.M. The Shikimate Pathway : Early Steps in the Biosynthesis of Aromatic Compounds. *Am. Soc. plant Physiol.* **1995**, *7*, 907–919.

35. Ma, J.; Zhou, X.; Ma, J.; Ji, Z.; Zhang, X.; Xu, F. Raman Microspectroscopy Imaging Study on Topochemical Correlation between Lignin and Hydroxycinnamic Acids in Miscanthus Sinensis. *Microsc. Microanal.* **2014**, *20*, 956–963, doi:10.1017/S1431927614000658.

36. Aguilar-Hernández, I.; Afseth, N.K.; López-Luke, T.; Contreras-Torres, F.F.; Wold, J.P.; Ornelas-Soto, N. Surface Enhanced Raman Spectroscopy of Phenolic Antioxidants: A Systematic Evaluation of Ferulic Acid, p-Coumaric Acid, Caffeic Acid and Sinapic Acid. *Vib. Spectrosc.* **2017**, *89*, 113–122, doi:10.1016/j.vibspec.2017.02.002.