Assessment of substrate biodegradability improvement in anaerobic Co-digestion using a chemometrics-based metabolomic approach

Francesc Puig-Castellví, Laëtitia Cardona, Delphine Jouan-Rimbaud Bouveresse, Christophe Cordella, Laurent Mazéas, Douglas N Rutledge, Olivier Chapleur

Assessment of Substrate Biodegradability Improvement in Anaerobic Co-digestion using a chemometrics-based Metabolomic Approach

Francesc Puig-Castellví,[1,2] Laëtitia Cardona,[2] Delphine Jouan-Rimbaud Bouveresse,[3] Christophe B. Y. Cordella,[3] Laurent Mazéas,[2] Douglas N. Rutledge,[1,4] and Olivier Chapleur[2]

[1] Université Paris-Saclay, INRAE, AgroParisTech, UMR SayFood, 75005 Paris, France
[2] Université Paris-Saclay, INRAE, PRocédés biOtechnologiques au Service de l'Environnement, 92761 Antony, France
[3] Physiologie de la Nutrition et du Comportement Alimentaire, AgroParisTech, INRAE, Université Paris-Saclay, 75005 Paris, France
[4] National Wine and Grape Industry Centre, Charles Sturt University, Wagga Wagga, Australia

Corresponding author: Olivier.chapleur@inrae.fr

Declarations of interest: none

Abstract:

Anaerobic co-digestion (AcoD) can increase methane production of anaerobic digesters in plants treating wastewater sludge by improving the nutrient balance needed for the microorganisms to grow in the digesters, resulting in a faster process stabilization. Substrate mixture proportions are usually optimized in terms of biogas production, while the metabolic biodegradability of the whole mixture is neglected in this optimisation. In this aim, we developed a strategy to assess AcoD using metabolomics data. This strategy was explored in two different systems. Specifically, we investigated the co-digestion of wastewater sludge with different proportions of either grass or fish waste using untargeted High Performance Liquid Chromatography coupled to Mass Spectrometry (HPLC-MS) metabolomics and chemometrics methods. The analysis of these data revealed that adding grass waste did not improve the metabolic biodegradability of wastewater sludge. Conversely, a synergistic effect in the

metabolic biodegradability was observed when fish waste was used, this effect being the highest for 25% of fish waste. In conclusion, metabolomics can be regarded as a promising tool both for characterizing the biochemical processes occurring during anaerobic digestion, and for providing a better understanding of the anaerobic digestion processes.

Keywords: metabolomics, anaerobic digester, co-digestion, methanization, HPLC-MS, CCA

1. Introduction:

Anaerobic digestion (AD) is a sustainable multistep process for the treatment of organic waste used to reduce the amount of solid organic matter resulting from the management of wastewater, and in consequence, the costs for waste handling. Moreover, this process generates renewable energy in the form of biogas by the action of the microorganisms digesting the organic matter (Madigou et al., 2019).

Biogas production in anaerobic digesters relies strongly on the stability of the microbial community growing in the anaerobic digesters (Calusinska et al., 2018). One key parameter is the carbon-to-nitrogen (C/N) ratio. Low C/N ratios, found for example in wastewater sludge, can result in low digestion rates and low biogas production (Li et al., 2011). To increase the C/N ratio and stimulate sludge digestion, the waste is digested in combination with other substrates richer in carbon. This approach is known as anaerobic co-digestion (AcoD), and is considered to be very ecologically efficient since several types of waste can be processed simultaneously (Borowski and Kubacki, 2015). In this work, we have explored the effect of mixing wastewater sludge (WAS) with either garden grass (GG) or fish waste (FW) as co-substrates.

In the digestion of the organic waste, the complex polymers (carbohydrates, proteins, nucleic acids and fats) are broken down into simpler compounds (sugars, amino acids, nucleic

55 bases, glycerol and fatty acids) that are ultimately converted to methane and CO_2 by the
56 microbial community living in the ADs (Chatterjee and Mazumder, 2019; Meegoda et al.,
57 2018). Despite the metabolic course of the AD, the evolution of the digestion is typically
58 assessed with indirect measurements of the digesters performance (i.e., pH and biochemical
59 oxygen demand (Meegoda et al., 2018)), or by studying the microbial dynamics by RNA and
60 DNA sequencing (De Vrieze et al., 2018). Alternatively, specific metabolites or regulatory
61 compounds can also be analyzed to look for underlying metabolic mechanisms
62 (Vanwonterghem et al., 2014). Lastly, to our knowledge, only a few studies have employed
63 metabolomic approaches to explore the global metabolic dynamics in ADs (Beale et al., 2016;
64 Murovec et al., 2018; Yang et al., 2014).

65 Metabolomics is the study of small molecules (metabolites) within cell extracts, tissues
66 and living organisms (Patti et al., 2012), and has been used extensively to evaluate the dynamic
67 metabolic response of living systems to physiopathological stimuli and genetic modifications
68 (Klassen et al., 2017). Metabolomics has also been used to study the metabolites from microbial
69 communities (the so-called "community metabolomics") (Llewellyn et al., 2015). Community
70 metabolomics studies concerning AD are still very scarce (Jones et al., 2016), since its
71 application has been challenging due to the very large range of metabolites for which there is
72 limited *a priori* knowledge (Vanwonterghem et al., 2014). To reduce the complexity of this
73 analysis, the community metabolomics studies of ADs are focused on either the solid fraction
74 (sludge) (Beale et al., 2016) or on the liquid fraction (digestate) (Murovec et al., 2018).
75 Metabolomics data analyses in these two studies were exploratory and despite they showed the
76 suitability of metabolomics to characterize AD samples, the application interest of these studies
77 was not demonstrated.

In this line, the two objectives of the present work are 1) to identify the metabolites that can be used to assess the GG and FW co-digestion of sludge, and 2) to identify the optimal proportion of GG or FW co-substrate that maximizes WAS digestion.

To achieve these goals, some considerations were taken into account beforehand. First, we have only analyzed the digestate to focus on the extracellular digestion products, as the metabolomic characterization of sludge would also include the intracellular metabolites from the microorganisms growing in the digesters. Second, a time-course experimental design was used to capture the main metabolomic dynamics over time. And third, samples were measured using HPLC-MS and further explored with untargeted chemometrics approaches (Principal Components Analysis (PCA) and Common Components Analysis (CCA)) (Bouhlel et al., 2018; Martin et al., 2015) to unravel the underlying metabolomic dynamics occurring in the studied ADs. In sum, this study provides a novel methodological approach to characterize the metabolomic processes in ADs that can be used to assess digester efficiency while at the same time expanding knowledge in this new field of community metabolomics.

2. Methods

2.1. Feedstock preparation and experimental set-up

Wastewater sludge (WAS) was collected from an industrial wastewater treatment plant (Valenton, France), GG was from the IRSTEA Institute's lawn, and FW was obtained from a fish market. GG and FW were crushed separately and the resulting minced solids were stored at 4°C for two days. The chemical characteristics of WAS, FW and GG are given in **Table A.1**.

The inoculum was obtained from a mesophilic full-scale anaerobic digester treating primary sludge at the Valenton (France) wastewater treatment plant. To prepare the inoculum before use, it was left under anaerobic conditions at 35°C for two weeks to digest the residual organic matter remaining in the inoculum. For the experimental set-up, 27 anaerobic batch

bioreactors consisting of 1 L glass bottles were used. Nine different mixtures were prepared by blending WAS substrate with FW or GG co-substrates in different proportions (100% WAS, 75% WAS/25% GG, 50% WAS/50% GG, 25% WAS/75% GG, 0% WAS/100% GG, 75% WAS/25% FW, 50% WAS/50% FW, 25% WAS/75% FW, and 0% WAS/100% FW) of grams of Chemical Oxygen Demand (gCOD). Each type of substrate mixture was used in three reactors (triplicates). In all bioreactors, the amount of added biomass was fixed at 12 gCOD. Then, all 27 bioreactors were inoculated with 1.2 gCOD of anaerobic sludge. All the digesters were complemented with a biochemical potential buffer (International Standard ISO 11734 (1995) (International Organization for Standardization, Geneva, 1995)) up to a final volume of 700 mL. Bioreactors were sealed with a screw cap and a rubber septum and headspaces were flushed with N_2 (purity > 99.99 %, Linde gas SA), and incubated for 4 weeks at 35°C in the dark without agitation.

2.2. Sampling

A particular sampling experimental design was used to focus on the metabolic fingerprint associated with the biogas production. For every digester containing either FW or GG, three points were monitored: one at the start of the co-digestion (day 0), and the two samples closest in time (before and after) to the largest methane production (days 14 and 21 for GG digesters, and days 21 and 28 for FW digesters) (Cardona et al., 2019). Finally, for digesters containing only sludge, all 4 time-points were monitored (days 0, 14, 21, and 28). Considering the triplicates, in total, 84 samples were collected**.** Sampling was performed by collecting 6 mL of liquid phase from the digester through the septum using a syringe. Then, samples were centrifuged at 10,000 g for 10 minutes to collect the supernatants, which were then snap frozen in liquid nitrogen and kept at -20ºC prior to metabolomic analysis.

2.3. Metabolomic analysis

Metabolomic analysis was performed on all collected supernatants. Prior to injection, samples were diluted to 1/10 in water. Instrumentation consisted in an Accela 1250 pump system connected to a LTQ-Orbitrap XL mass spectrometer (Thermo Scientific, MA, US) operated in positive electrospray ionization mode (ESI+). The detection was performed in full scan over an m/z range from 50 to 500 at a resolution of 100,000. The analytical column was a 50 × 2.1 mm inner diameter, 1.7 µm Syncronis C18 (Thermo Scientific, MA, US). The two mobile phases were acetonitrile 0.05% formic acid (phase A) and H_2O 0.05% formic acid (phase B). For each sample, 10 µL were injected into the analytical system. The flow rate was set at 0.4 mL min^{-1}, and the chromatographic method consisted of a linear gradient of A/B solvents changing from 10:90 to 80:20 over 23 minutes, followed by a stabilization phase of 5 minutes to return to the initial condition.

To remove possible batch effects, samples were injected in random order. Moreover, Quality Controls (QCs) composed of a pool of all samples were injected every 5 samples, and blank samples were injected every 10 samples. In total, 116 samples were injected (84 experimental samples, 21 QC samples, and 11 blank samples).

2.4 HPLC-MS data preprocessing

Raw HPLC-MS data were converted into mzXML-format files using MSConvert (ProteoWizard 3.0). Then, the list of chromatographic features (regions of interest, or ROIs) from each sample were extracted using the *XCMS* R-package (version 1.52.0) (Smith et al., 2006). ROIs were generated with the method *centWave*, using a *m/z* error threshold of 10 ppm and a peakwidth between 20 and 50 seconds. ROIs found in different samples were grouped using the *group* method with a bandwidth of 30. ROI retention times from the same ROI groups were unified across samples using the *obiwarp* method (Prince and Marcotte, 2006). A second grouping was carried out using a bandwidth of 25. Then, samples with missing ROIs were filled

using the *fillPeaks* method. From this analysis, a table of peak intensities composed of 116 rows (one per sample) and 476 columns (one per ROI) was obtained. This table was then imported as a matrix into Matlab R2009b (Mathworks, Inc., MA, US) for intensity drift and batch corrections. Drifts in signal intensity were corrected using LOESS (Rusilowicz et al., 2016; Zelena et al., 2009). On the other hand, differences within batches (between triplicates) and between batches (different substrate mixtures and time-points) were minimized with the PQN normalization method (Frank Dieterle et al., 2006), using the same approach as in Puig-Castellví (Puig-Castellví et al., 2016).

2.5. Chemometric data analysis

First, the matrix composed of the metabolomics data from the 84 experimental samples was auto-scaled (van den Berg et al., 2006) and investigated by Principal Components Analysis (PCA (Bro and Smilde, 2014)). With this method, a preliminary overview of the data was obtained.

With the aim of identifying the underlying metabolomic processes that govern the different biological dynamics occurring in the GG- and in the FW-containing digesters, the data were arranged into two matrices (one per co-substrate). Samples included in the first matrix were obtained from the two mono-digestions (WAS and GG), and from the co-digestion of WAS and GG, and collected at days 0, 14 or 21. Samples in the second matrix were from the analogous mono- and co-digestions of WAS and FW. In this second case, samples were collected at days 0, 21 and 28. Thus, each matrix is composed of 45 samples (3 samples per condition and per time, with 5 substrate conditions and 3 time-points). These two matrices will be referred to as ‘GG dataset’ and ‘FW dataset’, respectively.

Each dataset was centred and norm-scaled (Bylesjö et al., 2009) and investigated by Common Components Analysis (CCA) (Bouhlel et al., 2018; Rutledge, 2018). CCA is an

unsupervised chemometric method that estimates a series of orthogonal common components (CCs) that are linear combinations of the original variables, with strong weightings for strongly correlated variables that present the same dispersion of the observations. Analogously to PCA, for every CC, scores and loadings vector are calculated. Loadings give information about the metabolic profile (or metabolite composition) of the analyzed samples, while scores show the relative importance of these metabolic profiles for every sample. Besides, in CCA, a vector of weights (called saliences) corresponding to the importance of the variables for each CC is also iteratively calculated. The optimal number of CCs can be assessed by plotting the sum of saliences for each component, in the same manner as the assessment of the optimal number of components in a PCA with the Scree test (Ledesma et al., 2015) or by applying the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) to the residuals matrices to detect the moment when their lines and/or columns are no longer sufficiently correlated to merit further factorial analysis (Rencher, 2002).

For every CC, variables (*e.g.*, ROIs) associated with loading values beyond 2 standard deviations ($\pm 2 \times SD$) were selected (Bouhlel et al., 2018) and tentatively assigned if possible.

2.6 Metabolite identification

Tentative metabolite identification of the selected ROIs was performed based on the comparison of the accurate molecular mass measured by HPLC-MS with the calculated exact mass. Molecular formula of candidates were first estimated by using the *Rdisop* R-package (Böcker et al., 2009). This package allows the calculation of all the possible molecular formulae for a given accurate mass, within a delta error window in ppm (difference between the experimental mass and the adduct mass), and using a limited number of elements. In this analysis, we sought all the molecular formulae within 10 ppm that matched with $C_cH_hN_nO_oP_pS_sNa_{na}K_kCa_{ca}$. Then, for every accurate molecular mass, proposed molecular formulae were filtered using heuristics (Kind and Fiehn, 2007), and finally, the molecular

formula associated to the lowest delta value was selected. Finally, compound names were given to the selected molecular formulae after inspection of the possible structures in HMDB (Wishart et al., 2018), LipidMaps (Fahy et al., 2007), and PubChem (Kim et al., 2019) compound libraries.

2.7 Assessment of metabolic degradability during AcoD

In this paper, we propose to quantify the metabolic biodegradability (MB) of the different studied mixtures to assess the improvement in the AcoD performance compared to the corresponding mono-digestion experiment.

$MB_{0\rightarrow i}$ is defined as the observed change in the overall metabolite composition during AD between the initial time-point (t_0) and t_i. Since CC scores represent the relative importance of the distinct metabolic profiles, MB can be approximated as the distance between the two sample points in a scores plot obtained from a CCA analysis of a metabolomics dataset. For a CCA model of two components (CC1 and CC2), this expression can be generalized as eq. (1):

$$MB_{s,0\rightarrow i} = \sqrt{(u_{CC1,s,ti} - u_{CC1,s,t0})^2 + (u_{CC2,s,ti} - u_{CC2,s,t0})^2} \quad \text{eq. (1)}$$

In eq. (1), $u_{CC1,s,t0}$ and $u_{CC1,s,t0}$ represent the CC1 scores values associated to a given substrate composition *s* at the two time-points, t_0 and t_i; and $u_{CC2,s,t0}$ and $u_{CC2,s,t0}$ represent the corresponding CC2 scores values for the same substrate composition and times. As graphical examples, the $MB_{0\rightarrow i}$ for 50% GG, 100% GG, 50% FW, and 100% FW are represented with arrows in **Fig. 2.**.

Then, the metabolic biodegradability rate (MBR) can be calculated using eq. (2) as follows:

$$MBR_{substrate\_mixture}\ (\%) = \frac{MB_{substrate\_mixture}}{MB_{pure\ co-substrate}} * 100, \quad \text{eq. (2)}$$

where $MB_{substrate\_mixture}$ is the MB value obtained for every tested substrate, and $MB_{pure\ co\text{-}substrate}$ is the MB value for the pure co-substrates (in this study, 100%GG or 100%FW depending on

the studied AcoD system) at the latest time-point, as the latter is presumed to have the highest observed MB.

Finally, the factor of synergy in the co-digestion (FSC) can be estimated from the MBR. The other hand, FSC is the ratio between the observed and the theoretical MBR for a particular co-digestion mixture:

$$\mathrm{FSC}_{\text{substrate mixture}} = \frac{\mathrm{MBR}_{\text{substrate\_mixture}}}{\mathrm{MBR}_{\text{theoretical}}}, \qquad \text{eq. (3)}$$

where the theoretical MBR is the weighted sum of the MBR from the substrate and the co-substrate by their corresponding ω_s relative proportions:

$$\mathrm{MBR}_{\text{theoretical}} = \sum_{s=1}^{s=S} \omega_s \mathrm{MBR}_s, \qquad \text{eq. (5)}$$

For instance, if 100%WAS has a MBR of 20% and 100%GG has a MBR of 100%, then the theoretical MBR of 50% WAS/50% GG will be of 60%.

A FSC of 1 indicates that the MBR of the studied CoD is equivalent to the MBR obtained by digesting the two substrates separately. A FSC above 1 means that a synergistic effect in the MBR occurred due to the CoD, while a FSC below 1 reflects a loss of efficiency in the metabolic degradation due to the CoD.

3. Results and Discussion

Detection and identification of metabolites from bio-waste matter is a challenging task, since knowledge about their chemical composition is limited (and especially for WAS (Alves Filho et al., 2015; Vanwonterghem et al., 2014)). Moreover, the metabolic range of detected metabolites can be wide, as it may include not only the metabolites intrinsic to the three substrates (WAS, GG, and FW), but also those metabolites resulting from the degradation of these substrates. With the aim of addressing this challenge, we have taken advantage of both

247 the resolution and sensitivity of HPLC-MS, and of advanced chemometric methods capable of
248 untangling the complex mixtures being analyzed.

249 3.1 Unsupervised analysis of the metabolomics dataset

250 After the analysis of the ADs samples with HPLC-MS, the acquired data were
251 preprocessed and analyzed with *XCMS* R-package (see sections 2.3 and 2.4 in methods). With
252 this methodology, a total of 476 features per sample were obtained.

253 The resulting data were arranged into a matrix consisting of 84 rows (samples) and 476
254 columns (features). The data matrix was first analyzed with PCA to investigate its data variance.
255 The first three PCs only captured 32.02% of the dataset variance, revealing a complex
256 underlying data structure. The first component (14.14% of the variance) separated samples
257 collected at the first time-point from the rest, PC2 (10.67%) separated samples from the two
258 co-substrates (grass and fish waste), and PC3 (7.21%) separated samples according to WAS
259 relative composition (**Fig. 1**). The observed scores distribution for the other components could
260 not be satisfactorily linked to any studied factor (proportion of co-substrate or time).

261 From this initial analysis, it is observed that the major difference between the samples
262 is due to time, specifically between the first (square) and second (triangle or circle, depending
263 on the co-substrate) time-points (PC1 in **Fig. 1A**). This can be interpreted as follows: there
264 exists a set of metabolites, present in all samples regardless of the nature of the substrate, which
265 are rapidly consumed by the microbial communities before the second time-point. After these
266 metabolites are consumed, the metabolism of the microbial communities shifts to consume the
267 less assimilable metabolites. This process is less efficient and much slower, and it continues
268 until the end of the experiment.

269 As a second observation, during the course of the experiment, the most important
270 metabolic changes are produced in the reactors containing a greater amount of co-substrate

(**Fig. 1A**). This results in the clusters of samples containing co-substrate being much larger than the cluster of 100% WAS samples (the orange cluster in **Fig. 1A**). This suggests that the most biodegradable compounds are found in GG and FW substrate rather than in WAS.

3.2 Common Components Analysis (CCA)

In a previous work (Cardona et al., 2019), it was concluded that the use of FW caused a delay in the performance compared to GG-containing digesters. In order to cope with this difference in delays, different time-points were monitored depending on the type of co-substrate used. Thus, after observing that both GG- and FW-containing digesters presented independent trends in terms of performance and metabolomic profiles (as seen in **Fig. 1B**), to simplify the interpretation of the metabolomics data, these two sets of digesters were analyzed as two independent experiments.

In order to unravel the metabolic processes occurring in GG- and FW-containing anaerobic digesters, GG and FW datasets were further explored with CCA analysis. With CCA, metabolites presenting the same dispersion of the set of samples (and therefore descriptive of the same underlying biological process) are grouped together in a series of CC components that are a linear combination of the original metabolites, and where the most significant metabolites (see *Methods*) are considered to be representative of the metabolomic fingerprint of the digesters.

The Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) showed that there were only two components that meaningfully describe each dataset, the other components extracting only noise; as they could not be linked to substrate composition nor to time. In other words, the changes in the metabolite concentrations observed over time across reactors were associated with two main metabolic processes (one per CC). CC scores are presented in **Fig. 2**.

Regarding the scores obtained for the analysis of the two datasets, CC1 and CC2 (**Fig. 2**) are descriptive of metabolites from the co-substrate. This can be seen since 100% WAS samples remained clustered together regardless of the collection time, and the most prominent changes over time occurred in the digesters containing the highest amount of co-substrate.

Interestingly, the score distributions for both AcoD systems are triangle-shaped (drawn using dashed grey lines in **Fig. 2**). In these two triangles, the three vertices correspond to (1) the 100% WAS samples, (2) the 100% co-substrate samples at the initial time-point, and (3) the same samples at the last time-point. This particular score distribution is observed due to two reasons. First, because we used an experimental design to screen the degradability of the mixtures at different proportions. And second, because the largest variability in the studied system is related to the degradation of the co-substrate and the lowest one to WAS. Hence, due to the particularities of AcoD, this score distribution should be also expected for any other proportion-based experimental design of a AcoD system using organic waste as co-substrate.

In both GG and FW AcoD systems, CC1 components scores sign change (from positive to negative) between the first and the second time-point, implying that either metabolites have been consumed or produced after the first time-point. In addition, the small differences observed between the scores from the second and the third time-point indicate that CC1 only describes the metabolic alterations from the earlier stage. So, this component is representative of the most rapidly consumed metabolites from the co-substrate, and also of the products resulting from this fast metabolic process. On another hand, CC2 separates digesters rich in co-substrate (positive scores) from those with a lower amount of co-substrate (negative scores). For both GG and FW AcoD systems, the metabolic changes over time in CC2 were more important for digesters rich in co-substrate.

3.3 Feature selection

For the GG dataset, 35 features were selected for CC1 and 41 for CC2 while for the FW dataset, 18 features were found to be representative for CC1 and 17 for CC2 (**Figure 3**). The lower number of relevant features for the FW dataset could be the result of the smaller number of metabolites in fish than in grass (due to the existence of the plant secondary metabolite in the latter (Fiehn, 2002)).

Spectral features assigned to metabolites are shown in **Table. A.2** (for GG dataset) and **Table. A.3** (for FW dataset). All metabolites represented by CC1 (**Fig3A** and **Fig3B**) are associated with positive loadings, and therefore correspond to metabolites being consumed over time. On the other hand, metabolites represented by CC2 (**Fig3C** and **Fig3D**) are associated with either positive or negative loadings, indicating that these components are descriptive of a more complex metabolomic dynamic that is less influenced by time and more dependent on the co-substrate mixture, as commented in section 3.2.

3.4 Metabolomic fingerprint of the GG anaerobic digesters

In GG digesters, CC1 is mainly descriptive of the breakdown of the proteins from grass and the consumption of the products obtained in this process. Related to this process, CC1 includes 7 dipeptides, 4 aminoacids, one acetylated amino acid, and 7 compounds from amino acid catabolism (**Table. S1)**. This result is in agreement with the first phase of the anaerobic digestion, also known as the hydrolytic phase. In a previous metabolic study of the ADs of vegetal matter, it was observed that the different phases of the anaerobic digestion give distinct metabolomic signatures (Yang et al., 2014).

Apart from the protein breakdown and assimilation, the CC1 component is also descriptive of the degradation of nucleic bases and nucleotides. Detected degradation products from this pahtway are xanthine, beta-alanine, and nicotinic acid (**Table. A.2)**.

Some plant-specific metabolites were also included in CC1, such as D-mannitol and an acetamide compound. . Acetamides are produced during fermentation processes of plant matter (Linskens and Jackson, 1988). Finally, the detection of two sulfide compounds (dipropyl sulfide and ethylpropyl disulfide) can be regarded as a signature of the hydrogen sulfide commonly generated during microbial degradation processes (Zerrouki et al., 2015).

The metabolomic profile for CC2 from the GG dataset is mainly descriptive of oxidized free fatty acids (11 features). The reduced forms of these lipids may be grass constituents (as they are associated with positive loadings), and they may have been oxidized after cell walls were disrupted during the digestion process. In the same line, a fatty aldehyde was detected and linked to grass substrate in this component. In addition, a polyphenol was found specific for this component.

Conversely, there are only 4 features in CC2 that are characteristic of WAS. 3 of these features (X257, X273, and X306 in **Table A.2**) were assigned to small compounds containing heteroatoms (such as phosphorous (Alves Filho et al., 2015) and sulfur (Du and Parker, 2013)) and to unsaturations. The relative structural simplicity of these compounds may indicate that they have been generated in the degradation of more complex organic compounds originally found in the WAS in earlier digestion stages (i.e., during the microbial aerobic digestion). The other compound was assigned to a secosteroid. Secosteroids are formed after ring cleavage of sterol lipids, which may have occurred by enzymatic reaction during the anaerobic digestion. The presence of secosteroids in the WAS substrate may also be the result of the partial digestion of steroids from sewage waste. These compounds may come from the dissolved organic matter (DOM) of water, which is known to contain as major components oxidized sterols (Woods et al., 2012), terpenoids (Lam et al., 2007), and other lipids (Edith Kaiser et al., 2003).

3.5 Metabolomic fingerprint of the FW anaerobic digesters

The CC1 component resolved from the FW dataset shows a similar (although less extensive) metabolomic profile to that obtained for the GG dataset. That is, this component is rich in amino acids (5 metabolites, one of them acetylated) and amino acid catabolism products (2 metabolites), as well as nucleic base degradation products (2 metabolites). Hence, CC1 is representative of the hydrolytic stage in anaerobic digesters, where macromolecular compounds such as proteins and DNA from the substrate are broken down into simpler molecules. This CC also includes 2 sulfide compounds, which can be related to hydrogen sulfide formation occurring in the degradation processes of organic matter (Ghaly et al., 2010). Specific to FW, cadaverine and histamine were also detected in CC1 (**Table A.3**). These molecules are known as biogenic amines, or nitrogenous compounds mainly formed by the decarboxylation of amino acids (Kuley et al., 2017). Biogenic amines are one of the many compounds resulting from fish spoilage by microbial growth (Ghaly et al., 2010). All highlighted metabolites in CC1 are associated with positive loadings, denoting a decrease of these compounds over time, as seen in the same component for the GG dataset.

Finally, CC2 only includes 4 assigned features. From those, one is a secosteroid, two correspond to alkylated glycerol compounds, and the last one is triethanolamine (**Table S2**). In all cases, these metabolites can be regarded as modified lipids, found predominantly in WAS substrate (**Fig3D**). Since CC2 scores do not change significantly over time (**Fig. 2B**), it may indicate that these transformations took place prior to the anaerobic digestion.

3.6 Assessment of the metabolic degradability in AcoD

In order to assess from a metabolomics point of view whether the blending of WAS with a co-substrate (GG or FW) improves the degradability of the former, we have calculated the metabolic biodegradability rate (MBR) and the Factor of Synergy in Co-digestion (FSC) for every studied mixture. The MBR indicates how much a mixture is degraded referenced to the 100% co-substrate mono-digestion. On other hand, the FSC can be used to identify whether or

not the MB of the tested co-digestions are equivalent to the ones obtained by digesting the two substrates separately. More information about these terms is given in section 2.7.

The metabolic biodegradability rate (MBR) of GG mono-digestion did not improve in a statistically significant amount from day 14 (89%±14%, **Fig4A**) to day 21 (100%±9%, **Fig4A**). For WAS mono-digestion, the MBR was of 18%±8% at day 14, and of 20%±1% at day 21 (**Fig4A**). When the two substrates were used in mixture, the best improvement in the MBR was found for 50% WAS/50% GG AcoD after day 14, with almost 11% of observed synergy in the degradation (FSC = 1.11, **Fig4C**). 25% WAS/75% GG AcoD also showed a synergy effect in the degradability (FSC = 1.08, **Fig4C**). However, at day 21, all tested mixtures presented a FSC below 1, indicating that the degradation pathways triggered by mixing together the two substrates are less efficient that those employed during WAS and GG mono-digestions (**Fig4C**).

On another hand, in terms of MBR, neither the WAS mono-digestion nor the FW mono-digestion were improved in a statistically significant proportion between day 21 and day 28 (from 31%±1% to 24%±15%, and from 103%±3% to 100%±2%, respectively (**Fig4B**)). This can be interpreted as the degradation of organic molecules being completed before day 21, as observed in the heatmap representation of **Fig3**. When the two substrates were mixed and digested, a higher MBR than the $MBR_{theoretical}$ was obtained for all tested conditions (**Fig4B**). In graphical terms, in **Fig. 2B**, it can be observed that scores from 25% (orange), 50% (brown) and 75% FW (purple) at day 0 are located outside of the theoretical triangle that describes the metabolic variability of this AcoD system, causing that the calculated MB are larger in those conditions than in 100% FW mono-digestion (blue). Our hypothesis is that the mixing of the two co-substrates produced an additional breakdown of WAS macromolecules before the start of the digestion process. As a consequence, at the initial time-point, samples from co-digestion experiments were found to be metabolically more concentrated than 100% FW samples (shown

415 in **Fig3B**). Due to this additional process, the best AcoD was found for 75% WAS/25% FW
416 AcoD after 28 days with an increase of more than 100% in the MBR from the theoretical value
417 (FSC = 2.13, **Fig4D**). 50% WAS/50% FW and 25% WAS/75% FW AcoD also showed an
418 improved MBR (FSC of 1.58 and 1.43, respectively (**Fig4D**)). The synergy effect observed in
419 these 3 mixtures after 28 days was slightly higher than that after 21 days (**Fig4D**).

420 In summary, we demonstrated that metabolomics can be used to assess the anaerobic
421 co-digestion of sludge with a co-substrate using two different AcoD systems. This methodology
422 can be therefore applied to investigate the improvement in the co-digestion associated to other
423 co-substrates. Finally, it must be noted that this methodology is potentially easy to extrapolate
424 to the waste management industry, since it is not restricted to only HPLC-MS data, and any
425 metabolomics-based high-throughput technique, including NMR and GC-MS, can be used for
426 the same purpose.

427 ## 4. Conclusion

428 A chemometrics-based metabolomics strategy to determine the metabolic fingerprint of
429 AcoD and to assess substrate degradability during AcoD was developed.

430 On one hand, regarding the two AcoD studied systems, a comprehensive insight of the
431 metabolic processes occurring was achieved. We observed that most of the metabolic
432 alterations over time that occurred during AcoD were linked to the co-substrate, the compounds
433 in the co-substrate being more biodegradable than those in WAS. For GG AcoD, these
434 alterations were mainly linked to protein degradation processes. For FW AcoD, the two most
435 important groups of compounds being biodegraded over time were the amino acids and the
436 biogenic amines.

437 On the other hand, regarding the assessment of AcoD, the addition of green waste in the
438 AD did not improve the MBR of waterwaste sludge, whereas a synergistic effect in the MBR

was observed when fish waste was used. The most improved MB in fish co-digestion was found when WAS was mixed with 25% FW.

5. Funding

This work is part of the DIGESTOMIC project funded by the National Research Agency (ANR-16-CE05-0014).

6. Bibliography

Alves Filho, E.G., Alexandre e Silva, L.M., Ferreira, A.G., 2015. Advancements in waste water characterization through NMR spectroscopy: review. Magn. Reson. Chem. 53, 648–657. https://doi.org/10.1002/mrc.4158

Beale, D.J., Karpe, A.V., McLeod, J.D., Gondalia, S.V., Muster, T.H., Othman, M.Z., Palombo, E.A., Joshi, D., 2016. An 'omics' approach towards the characterisation of laboratory scale anaerobic digesters treating municipal sewage sludge. Water Res. 88, 346–357. https://doi.org/10.1016/J.WATRES.2015.10.029

Böcker, S., Letzel, M.C., Lipták, Z., Pervukhin, A., 2009. SIRIUS: decomposing isotope patterns for metabolite identification. Bioinformatics 25, 218–24. https://doi.org/10.1093/bioinformatics/btn603

Borowski, S., Kubacki, P., 2015. Co-digestion of pig slaughterhouse waste with sewage sludge. Waste Manag. 40, 119–126. https://doi.org/10.1016/j.wasman.2015.03.021

Bouhlel, J., Jouan-Rimbaud Bouveresse, D., Abouelkaram, S., Baéza, E., Jondreville, C., Travel, A., Ratel, J., Engel, E., Rutledge, D.N., 2018. Comparison of common components analysis with principal components analysis and independent components analysis: Application to SPME-GC-MS volatolomic signatures. Talanta 178, 854–863. https://doi.org/10.1016/J.TALANTA.2017.10.025

Bro, R., Smilde, A.K., 2014. Principal component analysis. Anal. Methods 6, 2812–2831. https://doi.org/10.1039/C3AY41907J

Bylesjö, M., Cloarec, O., Rantalainen, M., 2009. Normalization and Closure. Compr. Chemom. 109–127. https://doi.org/10.1016/B978-044452701-1.00109-5

Calusinska, M., Goux, X., Fossépré, M., Muller, E.E.L., Wilmes, P., Delfosse, P., 2018. A year of monitoring 20 mesophilic full-scale bioreactors reveals the existence of stable but different core microbiomes in bio-waste and wastewater anaerobic digestion systems. Biotechnol. Biofuels 11, 196. https://doi.org/10.1186/s13068-018-1195-8

Cardona, L., Levrard, C., Guenne, A., Chapleur, O., Mazéas, L., 2019. Co-digestion of wastewater sludge: Choosing the optimal blend. Waste Manag. 87, 772–781. https://doi.org/10.1016/J.WASMAN.2019.03.016

Chatterjee, B., Mazumder, D., 2019. Role of stage-separation in the ubiquitous development of

476 Anaerobic Digestion of Organic Fraction of Municipal Solid Waste: A critical review. Renew.
477 Sustain. Energy Rev. 104, 439–469. https://doi.org/10.1016/J.RSER.2019.01.026

478 De Vrieze, J., Pinto, A.J., Sloan, W.T., Ijaz, U.Z., 2018. The active microbial community more accurately
479 reflects the anaerobic digestion process: 16S rRNA (gene) sequencing as a predictive tool.
480 Microbiome 6, 63. https://doi.org/10.1186/s40168-018-0449-9

481 Du, W., Parker, W., 2013. Characterization of Sulfur in Raw and Anaerobically Digested Municipal
482 Wastewater Treatment Sludges. Water Environ. Res. 85, 124–132.
483 https://doi.org/10.2175/106143012X13407275694671

484 Edith Kaiser, *,†, Andre J. Simpson, ‡, Karl J. Dria, ‡, Barbara Sulzberger, † and, Hatcher‡, P.G., 2003.
485 Solid-State and Multidimensional Solution-State NMR of Solid Phase Extracted and Ultrafiltered
486 Riverine Dissolved Organic Matter. https://doi.org/10.1021/ES020174B

487 Fahy, E., Sud, M., Cotter, D., Subramaniam, S., 2007. LIPID MAPS online tools for lipid research.
488 Nucleic Acids Res. 35, W606-12. https://doi.org/10.1093/nar/gkm324

489 Fiehn, O., 2002. Metabolomics – the link between genotypes and phenotypes. Plant Mol. Biol. 48,
490 155–171. https://doi.org/10.1023/A:1013713905833

491 Frank Dieterle, Alfred Ross, Götz Schlotterbeck, and, Senn*, H., 2006. Probabilistic Quotient
492 Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures.
493 Application in 1H NMR Metabonomics. https://doi.org/10.1021/AC051632C

494 Ghaly, A.E., Dave, D., Budge, S., Brooks, M.S., Ghaly, A.E., Dave, D., Budge, S., Brooks, M.S., 2010. Fish
495 Spoilage Mechanisms and Preservation Techniques: Review. Am. J. Appl. Sci. 7, 859–877.
496 https://doi.org/10.3844/ajassp.2010.859.877

497 International Organization for Standardization, Geneva, S., 1995. ISO 11734:1995 - Water quality --
498 Evaluation of the anaerobic biodegradability of organic compounds in digested sludge --
499 Method by measurement of the biogas production [WWW Document]. Exam. Biol. Prop. water.
500 URL https://www.iso.org/standard/19656.html (accessed 1.29.19).

501 Jones, O.A.H., Lear, G., Welji, A.M., Collins, G., Quince, C., 2016. Community Metabolomics in
502 Environmental Microbiology, in: Beale, D.J., Kouremenos, K.A., Palombo, E.A. (Eds.), Microbial
503 Metabolomics: Applications in Clinical, Environmental, and Industrial Microbiology. Springer
504 International Publishing, Cham, pp. 199–224. https://doi.org/10.1007/978-3-319-46326-1_7

505 Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B.A., Thiessen, P.A., Yu, B.,
506 Zaslavsky, L., Zhang, J., Bolton, E.E., 2019. PubChem 2019 update: improved access to chemical
507 data. Nucleic Acids Res. 47, D1102–D1109. https://doi.org/10.1093/nar/gky1033

508 Kind, T., Fiehn, O., 2007. Seven Golden Rules for heuristic filtering of molecular formulas obtained by
509 accurate mass spectrometry. BMC Bioinformatics 8, 105. https://doi.org/10.1186/1471-2105-8-
510 105

511 Klassen, A., Faccio, A.T., Canuto, G.A.B., da Cruz, P.L.R., Ribeiro, H.C., Tavares, M.F.M., Sussulini, A.,
512 2017. Metabolomics: Definitions and Significance in Systems Biology. Springer, Cham, pp. 3–17.
513 https://doi.org/10.1007/978-3-319-47656-8_1

514 Kuley, E., Durmus, M., Balikci, E., Ucar, Y., Regenstein, J.M., Özoğul, F., 2017. Fish spoilage bacterial
515 growth and their biogenic amine accumulation: Inhibitory effects of olive by-products. Int. J.
516 Food Prop. 20, 1029–1043. https://doi.org/10.1080/10942912.2016.1193516

517 Lam, B., Baer, A., Alaee, M., Lefebvre, B., Moser, A., Williams, A., Simpson, A.J., 2007. Major
518 Structural Components in Freshwater Dissolved Organic Matter. Environ. Sci. Technol. 41,

519 8240–8247. https://doi.org/10.1021/es0713072

520 Ledesma, R.D., Valero-Mora, P., Macbeth, G., 2015. The Scree Test and the Number of Factors: a
521 Dynamic Graphics Approach. Span. J. Psychol. 18, E11. https://doi.org/10.1017/sjp.2015.13

522 Li, Y., Park, S.Y., Zhu, J., 2011. Solid-state anaerobic digestion for methane production from organic
523 waste. Renew. Sustain. Energy Rev. 15, 821–826. https://doi.org/10.1016/J.RSER.2010.07.042

524 Linskens, H.-F., Jackson, J.F., 1988. Wine Analysis. Springer Berlin Heidelberg.

525 Llewellyn, C.A., Sommer, U., Dupont, C.L., Allen, A.E., Viant, M.R., 2015. Using community
526 metabolomics as a new approach to discriminate marine microbial particulate organic matter in
527 the western English Channel. Prog. Oceanogr. 137, 421–433.
528 https://doi.org/10.1016/J.POCEAN.2015.04.022

529 Madigou, C., Lê Cao, K.-A., Bureau, C., Mazéas, L., Déjean, S., Chapleur, O., 2019. Ecological
530 consequences of abrupt temperature changes in anaerobic digesters. Chem. Eng. J. 361, 266–
531 277. https://doi.org/10.1016/J.CEJ.2018.12.003

532 Martin, J.-C., Maillot, M., Mazerolles, G., Verdu, A., Lyan, B., Migné, C., Defoort, C., Canlet, C., Junot,
533 C., Guillou, C., Manach, C., Jabob, D., Jouan-Rimbaud Bouveresse, D., Paris, E., Pujos-Guillot, E.,
534 Jourdan, F., Giacomoni, F., Courant, F., Favé, G., Le Gall, G., Chassaigne, H., Tabet, J.-C., Martin,
535 J.-F., Antignac, J.-P., Shintu, L., Defernez, M., Philo, M., Alexandre-Gouaubau, M.-C., Amiot-
536 Carlin, M.-J., Bossis, M., Triba, M.N., Stojilkovic, N., Banzet, N., Molinié, R., Bott, R., Goulitquer,
537 S., Caldarelli, S., Rutledge, D.N., 2015. Can we trust untargeted metabolomics? Results of the
538 metabo-ring initiative, a large-scale, multi-instrument inter-laboratory study. Metabolomics 11,
539 807–821. https://doi.org/10.1007/s11306-014-0740-0

540 Meegoda, J., Li, B., Patel, K., Wang, L., Meegoda, J.N., Li, B., Patel, K., Wang, L.B., 2018. A Review of
541 the Processes, Parameters, and Optimization of Anaerobic Digestion. Int. J. Environ. Res. Public
542 Health 15, 2224. https://doi.org/10.3390/ijerph15102224

543 Murovec, B., Makuc, D., Kolbl Repinc, S., Prevoršek, Z., Zavec, D., Šket, R., Pečnik, K., Plavec, J., Stres,
544 B., 2018. 1H NMR metabolomics of microbial metabolites in the four MW agricultural biogas
545 plant reactors: A case study of inhibition mirroring the acute rumen acidosis symptoms. J.
546 Environ. Manage. 222, 428–435. https://doi.org/10.1016/J.JENVMAN.2018.05.068

547 Patti, G.J., Yanes, O., Siuzdak, G., 2012. Metabolomics: the apogee of the omics trilogy. Nat. Rev. Mol.
548 Cell Biol. 13, 263–269. https://doi.org/10.1038/nrm3314

549 Prince, J.T., Marcotte, E.M., 2006. Chromatographic Alignment of ESI-LC-MS Proteomics Data Sets by
550 Ordered Bijective Interpolated Warping. https://doi.org/10.1021/AC0605344

551 Puig-Castellví, F., Alfonso, I., Pinã, B., Tauler, R., 2016. ^{1}H NMR metabolomic study of auxotrophic
552 starvation in yeast using Multivariate Curve Resolution-Alternating Least Squares for Pathway
553 Analysis. Sci. Rep. 6. https://doi.org/10.1038/srep30982

554 Rencher, A.C., 2002. Methods of multivariate analysis. J. Wiley, p. 708.

555 Rusilowicz, M., Dickinson, M., Charlton, A., O'Keefe, S., Wilson, J., 2016. A batch correction method
556 for liquid chromatography–mass spectrometry data that does not depend on quality control
557 samples. Metabolomics 12, 56. https://doi.org/10.1007/s11306-016-0972-2

558 Rutledge, D.N., 2018. Comparison of Principal Components Analysis, Independent Components
559 Analysis and Common Components Analysis. J. Anal. Test. 2, 235–248.
560 https://doi.org/10.1007/s41664-018-0065-5

561 Smith, C.A., Want, E.J., Maille, G.O., Abagyan, R., Siuzdak, G., 2006. XCMS : Processing Mass

562 Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment , Matching , and
563 Identification. Anal. chemestry 78, 779–787.

564 van den Berg, R.A., Hoefsloot, H.C., Westerhuis, J.A., Smilde, A.K., van der Werf, M.J., 2006.
565 Centering, scaling, and transformations: improving the biological information content of
566 metabolomics data. BMC Genomics 7, 142. https://doi.org/10.1186/1471-2164-7-142

567 Vanwonterghem, I., Jensen, P.D., Ho, D.P., Batstone, D.J., Tyson, G.W., 2014. Linking microbial
568 community structure, interactions and function in anaerobic digesters using new molecular
569 techniques. Curr. Opin. Biotechnol. 27, 55–64. https://doi.org/10.1016/j.copbio.2013.11.004

570 Wishart, D.S., Feunang, Y.D., Marcu, A., Guo, A.C., Liang, K., Vázquez-Fresno, R., Sajed, T., Johnson,
571 D., Li, C., Karu, N., Sayeeda, Z., Lo, E., Assempour, N., Berjanskii, M., Singhal, S., Arndt, D., Liang,
572 Y., Badran, H., Grant, J., Serra-Cayuela, A., Liu, Y., Mandal, R., Neveu, V., Pon, A., Knox, C.,
573 Wilson, M., Manach, C., Scalbert, A., 2018. HMDB 4.0: The human metabolome database for
574 2018. Nucleic Acids Res. 46, D608–D617. https://doi.org/10.1093/nar/gkx1089

575 Woods, G.C., Simpson, M.J., Simpson, A.J., 2012. Oxidized sterols as a significant component of
576 dissolved organic matter: Evidence from 2D HPLC in combination with 2D and 3D NMR
577 spectroscopy. Water Res. 46, 3398–3408. https://doi.org/10.1016/J.WATRES.2012.03.040

578 Yang, D., Fan, X., Shi, X., Lian, S., Qiao, J., Guo, R., 2014. Metabolomics reveals stage-specific
579 metabolic pathways of microbial communities in two-stage anaerobic fermentation of corn-
580 stalk. Biotechnol. Lett. 36, 1461–1468. https://doi.org/10.1007/s10529-014-1508-3

581 Zelena, E., Dunn, W.B., Broadhurst, D., Francis-McIntyre, S., Carroll, K.M., Begley, P., O'Hagan, S.,
582 Knowles, J.D., Halsall, A., Wilson, I.D., Kell, D.B., Kell, D.B., 2009. Development of a Robust and
583 Repeatable UPLC−MS Method for the Long-Term Metabolomic Study of Human Serum. Anal.
584 Chem. 81, 1357–1364. https://doi.org/10.1021/ac8019366

585 Zerrouki, S., Rihani, R., Bentahar, F., Belkacemi, K., 2015. Anaerobic digestion of wastewater from the
586 fruit juice industry: experiments and modeling. Water Sci. Technol. 72, 123.
587 https://doi.org/10.2166/wst.2015.193

588

Figures:

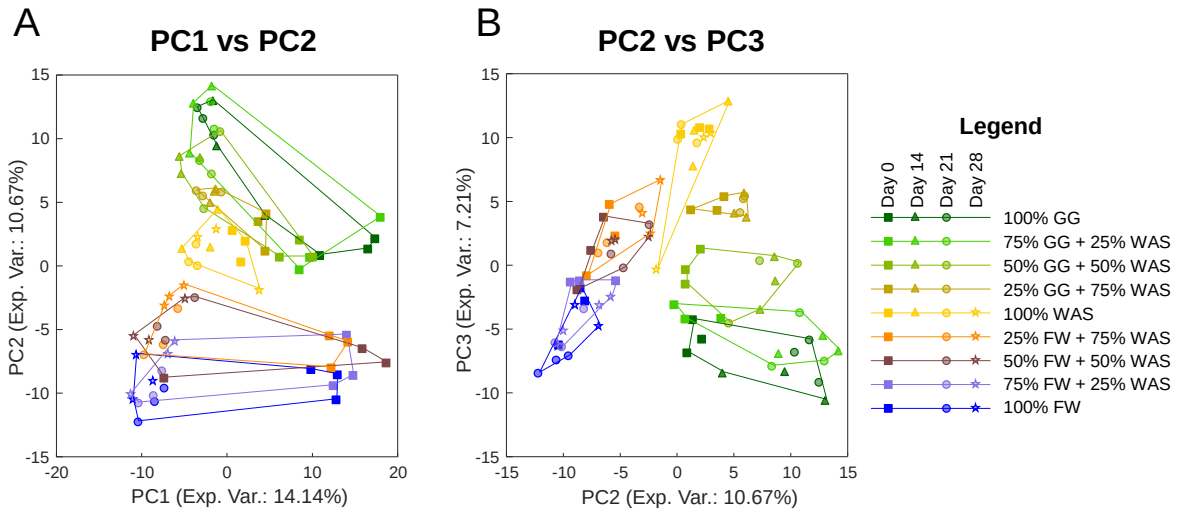


Fig. 1. Principal Components Analysis: **A.** Scores plot of PC1 vs PC2. **B.** Scores plot of PC2 vs PC3. The substrate composition is given by the color and the sample collection time is represented with different symbols.

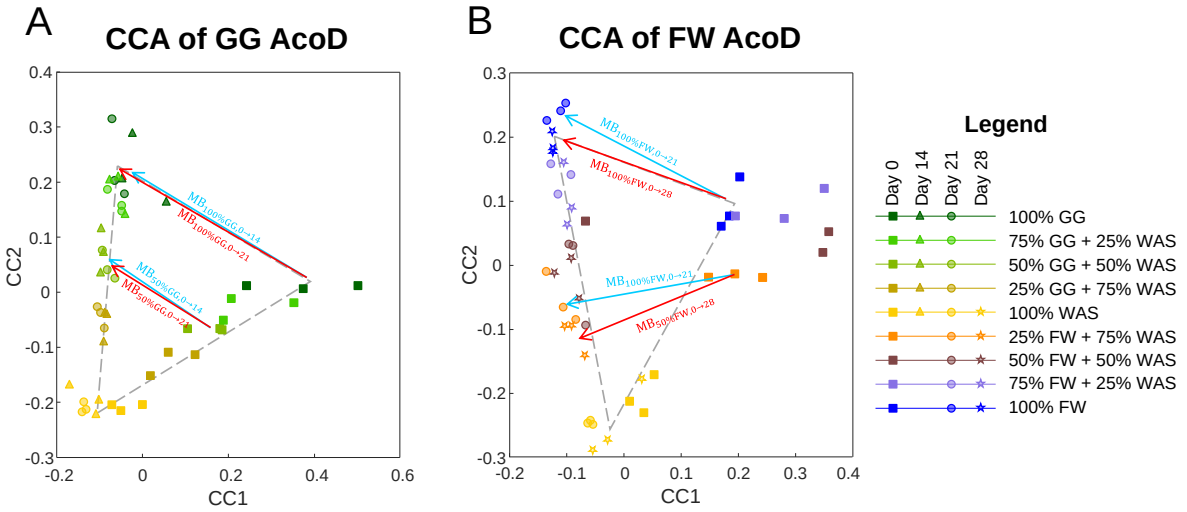


Fig. 2. CCA scores. **A)** CC1 vs CC2 scores on GG dataset. **B)** CC1 vs CC2 scores on FW dataset.. In both plots, scores distribution shape is similar to a triangle (drawn using dashed grey lines). In **A** and **B**, the metabolic biodegradabilities (MB) of 50% GG and 100% GG (**A**), and 50% and 100% FW (**B**) linked to the anaerobic digestion at the two screened time-points days are illustrated with arrows. Cyan arrows denote the MB achieved at the second time-point, while read arrows denote the corresponding MB at the third time-point.

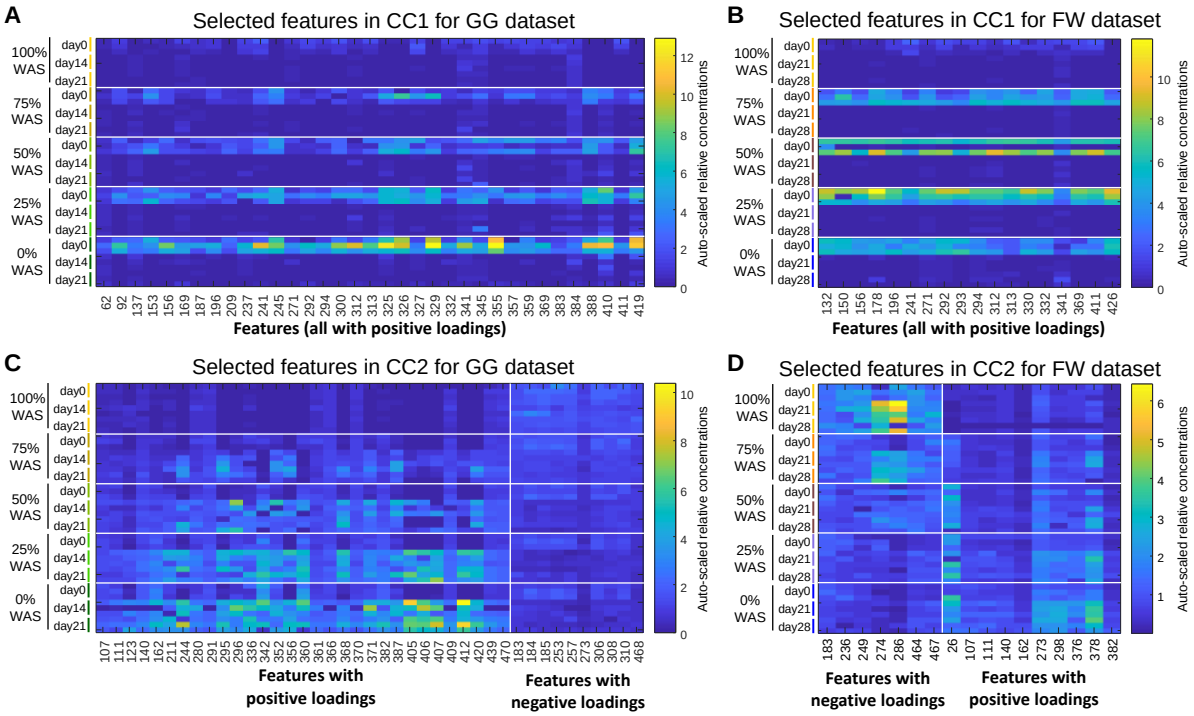
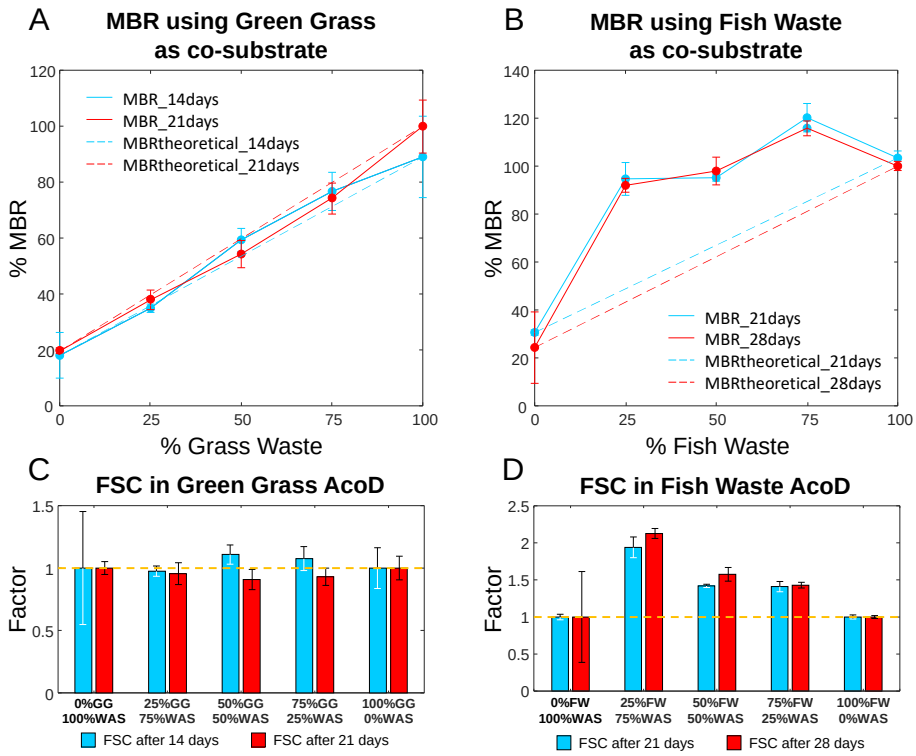


Figure 3. Heatmap representation of temporal evolution of the relative concentration for the selected features. **A)** Selected features in CC1 for GG dataset. **B)** Selected features in CC1 for FW dataset. **C)** Selected features in CC2 for GG dataset. **D)** Selected features in CC2 for FW dataset. Putative metabolite names for the shown features can be consulted in **Table A.2** and **Table A.3**.



607

608 **Fig4.** Metabolic biodegradability performance. **A-B)** Metabolic Biodegradability Rate (MBR) for the
609 different conditions screened. A %MBR of 100 corresponds to the anaerobic digestion of only co-
610 substrate during the longest period (21 days for GG **(A)**, 28 days for FW **(B)**). **C-D)** Factor of Synergy
611 in CoD (FSC) for the studied mixtures and the two CoD durations.