

Decomposing the molecular complexity and transformation of dissolved organic matter for innovative anaerobic bioprocessing

Received: 11 June 2024

Accepted: 20 May 2025

Published online: 25 May 2025



Jun Hu¹, Chuan-Guo Liu¹, Wen-Kai Zhang¹, Xue-Wen Liu¹, Bin Dong², Zhan-Dong Wang^{2,3}, Yuan-Guo Xie¹, Zheng-Shuang Hua¹ & Xian-Wei Liu¹✉

The sustainable transformation and management of dissolved organic matter (DOM) are crucial for advancing organic waste treatment towards resource-oriented processes. However, the intricate molecular complexity of DOM poses significant challenges, impeding a comprehensive understanding of the underlying biochemical processes. Here, we focus on the chemical “dark matter” mining using ultra-high resolution mass spectrometry technologies to elucidate the molecular diversity and transformation in anaerobic bioprocessing of food waste. We developed an analytical framework that reveals the persistence of DOM in the final effluent is mainly determined by its molecular properties, such as carbon chain length, aromaticity, unsaturation, and redox states. Our in-depth characterization and quantitative analysis of key biochemical reactions unveils the evolution of DOM composition, providing valuable insights into the targeted conversion of persistent molecules toward full utilization. Additionally, we establish a correlation between the redox state and energy density of a broad range of DOM molecules, enabling us to comprehend and evaluate their biodegradability. These insights enhance the mechanistic understanding of DOM transformation, guiding the rational design and regulation of sustainable organic waste treatment strategies.

Recycling organic waste from households, agriculture, and industries is essential for mitigating environmental impacts, maximizing resource utilization, and protecting human health^{1–3}. Annually, approximately 1.5 billion tons of municipal organic waste⁴, 5.9 billion tons of agricultural biomass⁵, and 380 billion cubic meters of municipal wastewater^{6,7} are produced worldwide, offering significant potential for renewable energy production. Anaerobic bioprocessing of organic waste is highly effective for carbon recovery and greenhouse gas (GHG) emission reduction^{8,9}, while also offering a promising approach for recovering phosphorus and sulfur from wastewater¹⁰.

Despite its widespread application and varying degrees of success, the persistent challenge remains to align cost-effectiveness, discharge standards, and sustainability in treatment plants. For more than a century, traditional treatment methods have relied on adjusting variables, such as organic loadings, pH and residence time within the complex “black box” of biochemical processes. These methods, characterized by their complexity and inefficiency, have often resulted in losses due to an inadequate understanding of organic matter transformation. This transformation primarily involves converting biosolids into dissolved organic matter (DOM), followed by its further

¹Chinese Academy of Sciences Key Laboratory of Urban Pollutant Conversion, Department of Environmental Science and Engineering, University of Science and Technology of China, Hefei 230026, China. ²State Key Laboratory of Fire Science, University of Science and Technology of China, Hefei 230026, China. ³National Synchrotron Radiation Laboratory, University of Science and Technology of China, Hefei, Anhui 230029, PR China.

✉ e-mail: xianweiliu@ustc.edu.cn

conversion for pollutant removal and resource recovery¹¹. Consequently, there is a critical need for in-depth exploration of DOM's biochemical roles at the molecular scale to innovate the treatment process.

The chemical complexity and variability of DOM, especially in complex environmental systems, pose significant challenges for its identification and analysis¹². Ultra-resolution mass spectrometry, particularly Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR MS), emerges as a powerful tool for uncovering the elusive aspects of DOM due to its superior accuracy and resolution. Current research is largely centered on DOM in natural aquatic ecosystems, such as lakes^{13–15}, groundwater^{16,17}, rivers¹⁸ and seas¹⁹, with a specific focus on its relationship with climatic and water properties. The decomposition and stabilization of organic matter are key in environmental studies, especially in engineered systems that demand greater insights into DOM transformation processes rather than just molecular composition analysis. However, most existing studies conducted qualitative analyses of DOM composition based on the Van Krevelen diagrams in lab settings, failing to capture detailed characteristics and biochemical reactions of DOM in real-world waste treatment systems^{11,12}. Our understanding of DOM's dynamic transformation within biological waste treatment systems remains limited due to a lack of innovative analytical approaches.

In this study, we explore the molecular complexities inherent in anaerobic digestion (AD) from seven full-scale food waste treatment plants (FWTPs) across China (Supplementary Fig. 1). To address the biochemical black box challenge, we propose a DOM analysis framework that allows us to evaluate DOM chemical diversity, analyze its molecular characteristics, identify the core molecular community, quantify key biochemical reactions, and correlate redox states with biodegradability. This approach provides a comprehensive

understanding of DOM evolution beyond mere molecular compositions. The knowledge obtained from DOM analysis in this study sheds light on the molecular intricacies of organic matter metabolism, paving the way for refining and advancing anaerobic bioprocessing strategies and solutions.

Results

Sludge characteristics and DOM chemical diversity evaluation

Samples from typical seven two-stage AD systems were analyzed for organic matter conversion. In primary digestion (PD), the organic waste is hydrolyzed and fermented to volatile fatty acids (VFAs). Subsequently, the VFAs and other residual organics are further fermented to produce CH₄ in secondary digestion (SD). The pre-hydrolysis and acidification of organic substrates in the PD not only provide advantageous substrates for methanogens, but also help prevent acidification-induced fermentation collapse in the SD through the consumption of easily degradable fractions of substrates²⁰. The anaerobic acidogenesis with acetate as the main product resulted in a pH drop to about 3.87, which is believed to limit the metabolic activities of most microbes. The SD, operated a pH of around 7.77, significantly contributes to further substrate degradation and conversion, with removal rates of 72.4% for total solids (TS), 85.4% for volatile solids (VS), 87.2% for chemical oxygen demand (COD), and 82.0% for total dissolved phosphorus (TDP) (Fig. 1a). In contrast, there is a notable 27.1% increase in total dissolved nitrogen (TDN), likely driven by the solubilization of solid nitrogenous compounds and therefore elevating the nitrogen content of DOM²¹.

To explore the compositional diversity and molecular transformation process, DOM samples were analyzed by ESI (-) FT-ICR MS, following an analytical workflow illustrated in Supplementary Fig. 2. The initial FT-MS data set contained 14 samples with a total of

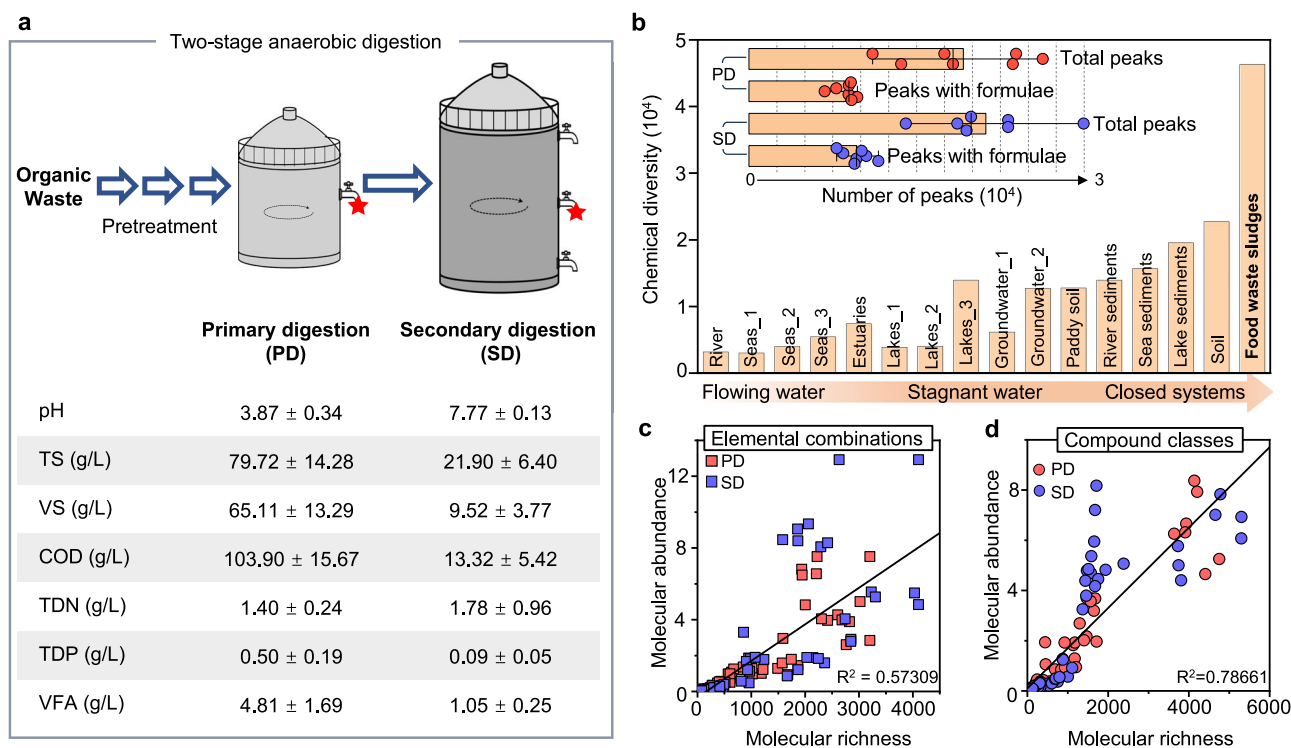


Fig. 1 | Characteristics of anaerobic sludge and chemical diversity of DOM in the two-stage anaerobic digestion. **a** Physical-chemical properties of sludges in food waste treatment plants (FWTPs). Each parameter represents the mean ± standard deviation (SD) from seven biologically independent replicates. Red dots indicate sampling locations, and VFAs include C₂–C₈ fatty acids, as detailed in Supplementary Table 2. **b** Comparative analysis of chemical diversity (molecular richness)

of DOM across various environments. More detailed information on molecular characteristics is provided in Supplementary Table 1. Relationship between molecular richness and abundance for **(c)** element combination and **(d)** compound classes. Here, molecular richness refers to the number of unique formulae within a defined molecular group, while molecular abundance represents the weighted contribution of these formulae based on their relative intensity.

>280,000 identified peaks observed across a broad mass range. We used the software Formultitude (formerly known as Formularity²²) and its Compound Identification Algorithm (CIA)²³ function to assign and filter elemental formulae to a list of masses. The resulting data shows that 45.5% of peaks identified in PD samples, and 45.0% of peaks in SD samples, were successfully assigned with molecular formulae. Chemical diversity in anaerobic sludge DOM (46327 unique formulae) are obviously higher than that in various aquatic environments^{13–15,18,19,24–33} (Fig. 1b and Supplementary Table 1). The molecular diversity of DOM appears to increase as the sources transition from flowing water to stagnant water, and eventually to closed systems. In contrast to natural aquatic environments, the AD-based waste treatment plant is a well-controlled closed system, where the diverse chemical composition of DOM is largely shaped by the variety of Chinese dietary culture, specific food waste treatment processes, and sludge microbiomes (Supplementary Notes 1–2 and Supplementary Figs. 3–4). The DOM in anaerobic sludges features diverse molecular compositions, categorized by elemental combinations (CHO, CHON, etc.) and classes (amino sugar, carbohydrate, etc.)³⁴. To quantify chemical diversity, we defined the number of formulae as richness and intensity-weighted formulae as abundance in a defined molecular group. A positive correlation between richness and abundance across elemental and compound categories suggests that molecular diversity correlates with abundance of DOM molecular groups in the two-stage AD systems (Fig. 1c, d). The molecular richness-abundance relationship offers a more scientific criterion for assessing DOM chemical diversity, providing a comprehensive view of DOM complexity in various environmental systems.

DOM molecular composition, distribution, and properties

To assess DOM chemical fingerprint in the PD-SD two-stage anaerobic digestion, we calculated intensity-weighted elements (C, H, O, N, S, and P), elemental combinations (CHO, CHON, CHONP, etc.), elemental ratios (H/C, O/C, etc.) and compound classes (lipid, lignin, protein, etc.). We observed a shift in molecular composition toward lower carbon and higher hydrogen content from primary to secondary digestion, indicating an increase in high H/C molecules, such as long-chain fatty acids (LCFA). This shift is likely driven by the degradation of CHONP, CHOP, and CHOSP molecules, leading to an increase in hydrogen-rich CHO molecules and a significant reduction in phosphorus content (Fig. 2a and Supplementary Fig. 5). The removed phosphorus is likely assimilated into microbial biomass or precipitated as inorganic minerals, such as phosphate salts, under anaerobic conditions. Lignin-, lipid-, and protein-like molecules are the dominant compound classes in food waste sludge, which aligns well with prior studies on DOM in sewage sludges^{35,36}, and food waste³⁷. This indicates that the persistence of these more recalcitrant compounds may be a common characteristic of AD systems. Molecular richness (number of formulae) of lignin, which is over twice that of other classes, remained largely unchanged from PD to SD (Supplementary Fig. 6). However, its molecular abundance (intensity-weighted formulae) decreased from 48.4% to 34.9% (Fig. 2b). This implies that a specific subset of lignin molecules was more susceptible to microbial degradation and conversion, yet not completely mineralized or converted to FT-MS undetectable DOM, such as molecules with masses <100 m/z. Lipid-like molecules showed a significant increase in both richness and abundance, primarily due to their inherent resistance to microbial degradation³⁸. Additionally, biological hydrolysis and conversion processes likely contributed to the further accumulation of soluble lipid molecules³⁹. In contrast, protein-like molecules remained largely unchanged in both molecular richness and abundance during the SD process (Fig. 2a and Supplementary Fig. 6).

Mass distribution analysis of DOM revealed a significant increase in small-molecular-weight molecules (0–400 m/z), primarily due to the buildup of lignin- and lipid-like compounds, while the increase

observed in the 500–800 m/z range was predominantly attributed to lipid-like compounds. The degradation of large-molecular-weight molecules, including lignin (600–1000 m/z), lipids (750–1000 m/z), proteins (800–1000 m/z), and other compounds, contributed to the significant drop in the 800–1000 m/z range of the overall DOM molecular community (Fig. 2c–f and Supplementary Figs. 7–8). This decrease is likely a result of the hydrolysis of dissolved organic polymers, a process that occurs more readily than the further breakdown of their monomeric components⁴⁰. During the SD process, lipids were the only compound class to show a significant increase, with high-molecular-weight molecules (400–800 m/z) exhibiting greater molecular richness but lower intensities. In contrast, smaller molecules (100–400 m/z) displayed a more modest increase in molecular richness, with their intensity distribution being more scattered than concentrated in specific regions (Fig. 2e and Supplementary Fig. 9). This pattern suggests that the SD process not only enhances molecular diversity but also leads to a higher prevalence of smaller, more recalcitrant lipid molecules in the effluent.

The molecular compositional evolution of DOM during the AD process is closely linked to its inherent structural features, such as carbon chain length, degree of unsaturation, and aromaticity (Fig. 2g). Notably, the distribution of C–C single bonds within the long carbon-chain dominated molecular community likely plays a key role in the biodegradation of various DOM compound classes (Fig. 2h and Supplementary Figs. 10–11). For instance, lipids, characterized by a higher proportion of C–C single bonds, accumulated significantly during the SD process. In contrast, lignin, which contains fewer C–C single bonds, was more readily degraded (Fig. 2b, h). However, lignin molecules that contain more aromatic rings (higher AI_{mod}), tended to remain in the final effluent (Fig. 2i). Similarly, the high C–C single bond content and aromatic structures in proteins contributed to their limited degradation (Fig. 2h, j). Ammonia-induced inhibition, particularly at high organic loadings, likely further hindered their breakdown⁴¹. In summary, these findings highlight how the structural characteristics of lignin, lipids, and proteins influence their degradation and persistence in the AD process. This deeper understanding is crucial for developing more effective strategies for the targeted removal and recycling of persistent DOM compounds.

Core molecular communities of DOM

We further expanded the compositional analysis of DOM beyond individual samples to a global view across seven two-stage digesters in China, revealing common molecular patterns. The DOM molecular composition between PD and SD is effectively distinguished using principal component analysis (PCA) (Fig. 3a). Our analysis identified a core molecular community, with 7.7% of molecular formulae shared among PD samples and 5.1% shared among SD samples (Fig. 3b). In the core DOM community, lignin-like compounds were predominant, accounting for 57.21% of in PD and 56.20% in SD (Fig. 3c and Supplementary Fig. 12). Both lignin and proteins, which together accounted for over 80% of the core molecules, remained relatively stable in molecular richness from PD to SD, suggesting their higher recalcitrance and persistence throughout AD. Consistent with the individual sample results in Fig. 2, the proportion of lipid-like compounds in the core community increased significantly from 7.67% in PD to 11.38% in SD. Most of these compounds were composed of CHON and CHO elements. The redox states, indicated by the NOSC, shifted to a more reduced state (Fig. 3c and Supplementary Fig. 13). This shift was primarily driven by an increase in reduced compounds resulting from anaerobic degradation, lacking a sufficient oxidation process (Supplementary Notes 2–3).

In the SD stage, the distribution of DOM at larger molecular sizes (450–600 m/z) showed a significant decrease (Fig. 3d). This pattern was also visible in the distribution of homologous compounds (Fig. 3e), suggesting that the breakdown of larger molecules into smaller ones occurs within the same compound classes, such as lignin,

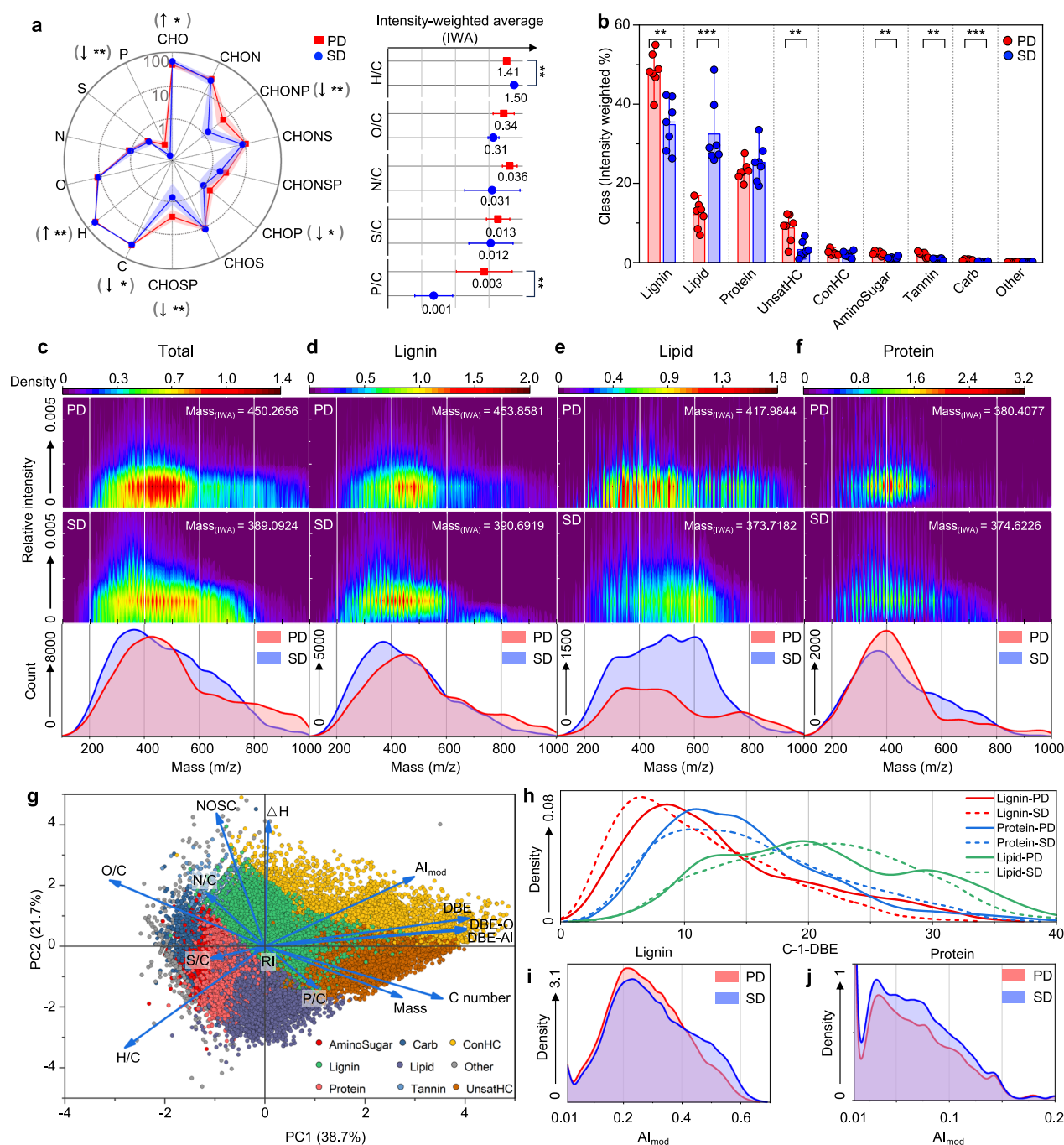


Fig. 2 | Comparing molecular characteristics of DOM between PD and SD sludge samples. Molecular composition is shown based on (a) elements and elemental combinations and (b) compound classes, calculated based on relative intensity weights. Elemental ratios (e.g., H/C, O/C) and mass averages were obtained using intensity-weighted averages (IWA). Each value represents the mean of seven biologically independent replicates. In panel (a), error bars represent standard deviation (SD). In panel (b), the bar + data overlap plot shows the mean (bar), median (center line), whiskers representing $1.5 \times$ the interquartile range (IQR), and outliers as points beyond the whiskers. Significance was evaluated by pair-sample *t*-test: *** represents $p \leq 0.001$, ** represents $p \leq 0.01$, * represents $p \leq 0.05$, and no asterisk indicates $p > 0.05$; exact *p*-values are provided in the Source Data File. Mass and relative intensity (RI) distribution are shown for (c) total, (d) lignin, (e) lipid and (f) protein

compounds. The density plots of mass-RI were generated using a 2D Kernel density plot, with the mass range from 100 to 1000 m/z and RI from 0 to 1, using a 1000 1000 grid. The mass-RI distribution of additional compound classes including unsathC, ConHC, AminoSugar, Tannin and Carb was provided in Supplementary Fig. 7. **g** Principal component analysis (PCA) of the relationship between molecular composition and structural factors: H/C ratio, double bond equivalent (DBE), aromaticity index (AI_{mod}) and other relevant parameters. **h** C-C single-bond distribution of lignin, lipid and protein compounds. The C-1-DBE was used to represent C-C single bonds, while non-C=C bonds (e.g., C=O, C=N) were excluded from the calculations; distributions of other compound classes was provided in Supplementary Fig. 11. AI_{mod} distribution of (i) lipid and (j) protein molecules of PD and SD DOM samples.

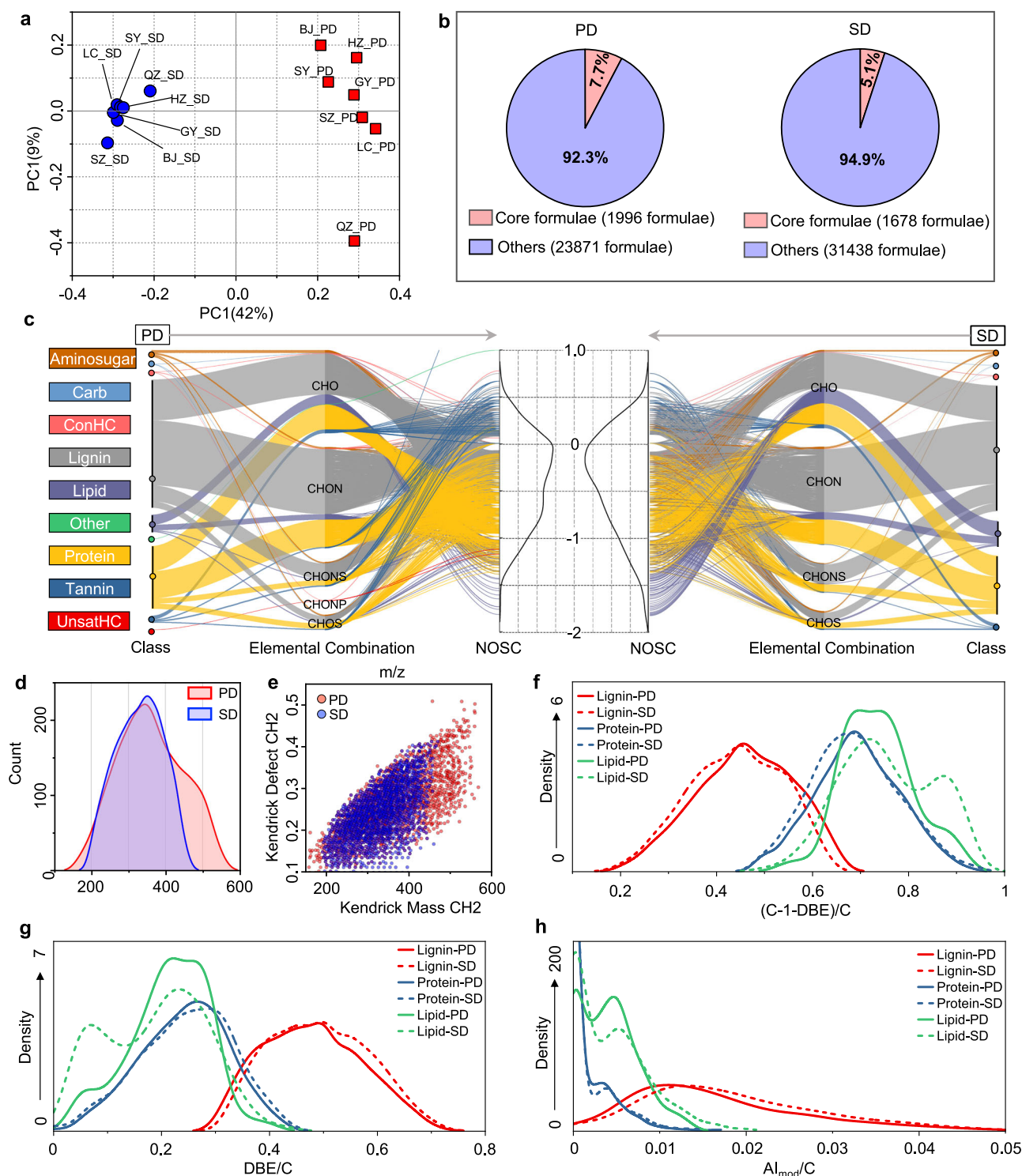


Fig. 3 | Characteristics of the core molecular community in AD. a Principal component analysis (PCA) of DOM molecules in the PD and SD. **b** Percentage and relative abundance of the core molecular community. **c** Parallel plots for comparing compound classes, element combinations and NOSC between PD

and SD samples. **d** Mass distribution, **e** Kendrick defect CH₂ for the core molecular community. **f** (C-1-DBE)/C, **g** DBE/C and **h** AI_{mod}/C distribution for the core molecular community of PD and SD.

proteins, and lipids (Supplementary Fig. 14). These core molecular communities persist during anaerobic fermentation, largely due to their structural characteristics, including the presence of single and double carbon bonds, as well as their higher aromaticity (Fig. 3f–h). These features are significantly more pronounced than those found in smaller, more bioavailable molecules like cellobiose, glucose, and acetate. The carbon chain length (mainly C–C single bonds) plays a

more critical role in the degradation of lipid compounds (Fig. 3f and Supplementary Fig. 15), whereas double bonds (including C=C and aromatic double bonds) have a higher influence on lignin and protein molecules (Fig. 3g and Supplementary Fig. 16). The high aromatic ring content in lignin may be one of the main reasons for its high richness and abundance in the final effluent (Fig. 3h and Supplementary Fig. 16). This observation highlights the challenge in degrading complex

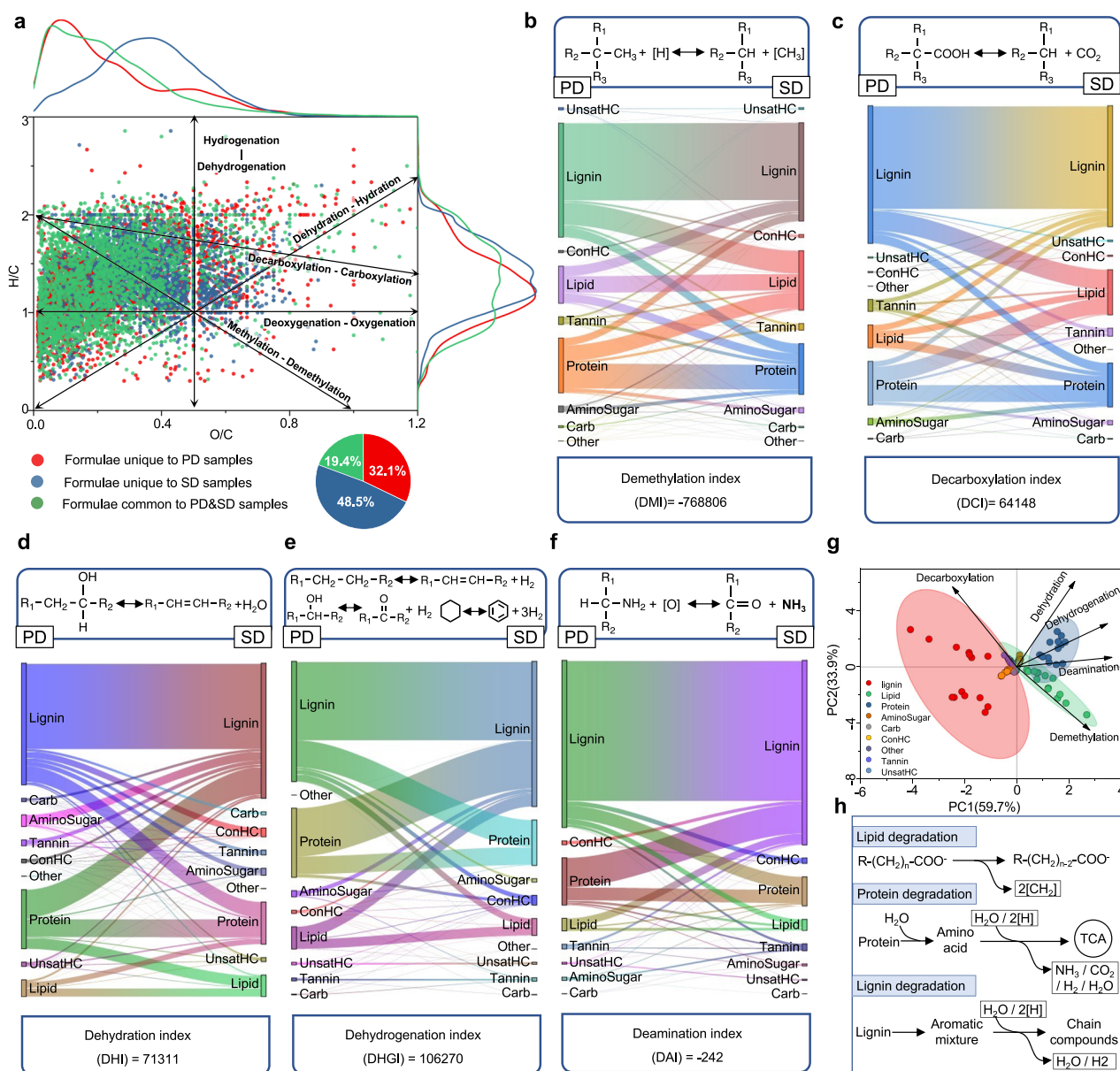


Fig. 4 | Quantitative analysis of possible biochemical reactions in the DOM transformation between PD and SD. **a** Van Krevelen Diagram (VKD) with marginal density plots of molecular formulae unique to PD samples, SD samples, and common to the two samples. Possible chemical reactions occurring between different compound classes: **b** demethylation, **c** decarboxylation, **d** dehydration, **e** dehydrogenation, and **f** deamination between PD and SD molecules. All the

molecules in PD and SD were included in these reaction analyses, and quantitative data for chemical reactions occurred among various compound classes is available in Supplementary Tables 3–7. **g** The relationship between compound classes and chemical reactions based on Principal Component Analysis (PCA). **h** Main chemical reactions in the process of lipids, proteins, and lignin degradation.

molecular structures in DOM of sludge and underscores the importance of the core molecular community as a target for future research to better understand and utilize these molecules in resource recovery and treatment processes.

Biochemical reaction prediction

A grand challenge in DOM research is to visualize the potential biochemical processes that DOM undergoes throughout its lifetime. We plotted all the assigned formulae in the van Krevelen diagram, categorizing them into (i) formulae unique to PD samples, (ii) formulae unique to SD samples, and (iii) formulae common to both PD and SD samples. This allows us to visualize the distribution of DOM molecules based on H/C and O/C ratios, as well as the pathways of key chemical reactions⁴² (Fig. 4a). We found that 19.4% of total DOM molecules originated from

the PD were disappearing in the SD. Additionally, newly detected molecules in the SD accounted for 48.5% of the total. The observed shift towards higher hydrogen-to-carbon (H/C) and oxygen-to-carbon (O/C) ratios in the SD stage suggests that both reduction and oxidation reactions occur simultaneously during the digestion process (Fig. 4a and Supplementary Fig. 17). However, considering the intensity of DOM molecules, where those with lower O/C ratios exhibit higher intensity (Supplementary Fig. 18 and Fig. 2a), it becomes evident that reduction reactions predominate over oxidation reactions, in line with anaerobic conditions of the digestion process (Supplementary Note 3). To quantify these potential biochemical reactions, we calculated the functional group difference in molecular formulae representing key reactions that occur in AD process, such as demethylation/methylation, or decarboxylation/carboxylation (Fig. 4b–g and Supplementary Fig. 19). Based on

these calculations, we introduce demethylation index (DMI), decarboxylation index (DCI), dehydration index (DHI), dehydrogenation index (DHGI) and deamination index (DAI), together allowing to quantify each reaction in the digestion of lipid, lignin, protein and other compound classes.

Lipid degradation in AD exemplifies a demethylation-like process ($-\text{CH}_3 + \text{H}$), where long-chain fatty acids are shortened with a loss of two carbons (-2CH_2) by β -oxidation (Fig. 4b, h). The negative DMI suggests that a greater number of long-chain compounds are generated than consumed, which seems to be contrary to the fact that AD of organic waste is a process of decomposing large molecules into small molecules. In addition to accumulation of long-chain lipids, the estimated 64.4% conversion rate of lignin to lipid-like compounds might partially account for the negative DMI (Fig. 4b and Supplementary Table 3). Furthermore, decarboxylation, another prevalent reaction, is associated with compositional shifts among lignin-, protein- and lipid-like molecules. Our analysis demonstrates a notable lignin-to-lipid transition with DCI (lignin \rightarrow lipid) = 26688, suggesting a likely shift from lignin-like to lipid-like compounds (Fig. 4c and Supplementary Table 4). A closer view of dehydration reactions revealed strong metabolic links between proteins and both lignin (DHI (protein \rightarrow lignin) = 41933) and ConHC molecules (DHI (protein \rightarrow ConHC) = 13519) (Fig. 4d and Supplementary Table 5). Dehydrogenation reactions, which are essential for forming double bonds and producing hydrogen, are particularly prominent in ConHC, lignin, lipid, and protein compounds, as indicated by the high dehydrogenation index (DHGI) (Fig. 4e and Supplementary Table 6). Oxidative deamination reactions are notably prominent in protein-like compounds (Fig. 4f and Supplementary Table 7), leading to shifts in DOM composition toward higher oxygen and lower hydrogen and nitrogen contents. For

example, activated CHON compounds, particularly those containing only one nitrogen atom, undergo oxidative deamination, resulting in an increase in CHO compounds (Fig. 2a).

The DOM transformation in AD involves distinct chemical reactions specific to various compound classes. (Fig. 4g, h). Lignin depolymerization showed a slight preference for decarboxylation reactions, but no single pathway dominated, likely because of its diverse functional groups and inherently hard-to-degrade nature⁴³. In contrast, lipid degradation was primarily driven by a demethylation-like pathway through β -oxidation, followed by decarboxylation that removes carboxyl groups from long-chain fatty acids (LCFAs). Meanwhile, the breakdown of large protein molecules into amino acids and other bioavailable products typically involves dehydrogenation, dehydration, and deamination reactions⁴⁰. This quantitative analysis of reaction types enhances our understanding of molecular metabolism in AD and highlights the significant interplay between different compound classes, contributing to DOM's chemical diversity.

Redox features of DOM transformation

Redox conditions play a crucial role in DOM degradation, influenced by DOM turnover processes, such as DOM inflow, solid solubilization, as well as outflow and gas production. Here, we used the standard molar combustion enthalpy ($\text{kJ mol}^{-1} \text{C}$) as an approximation for the Gibbs free energy associated with the complete oxidation of a given organic compound. This allowed us to assess the dependence of molecular energy potential on nominal oxidation state of carbon (NOSC)^{44,45}. Notably, a strong linear relationship was observed between the average combustion enthalpy (ΔH) and NOSC values of each compound classes ($R^2 = 0.817$) (Fig. 5a). This relationship was

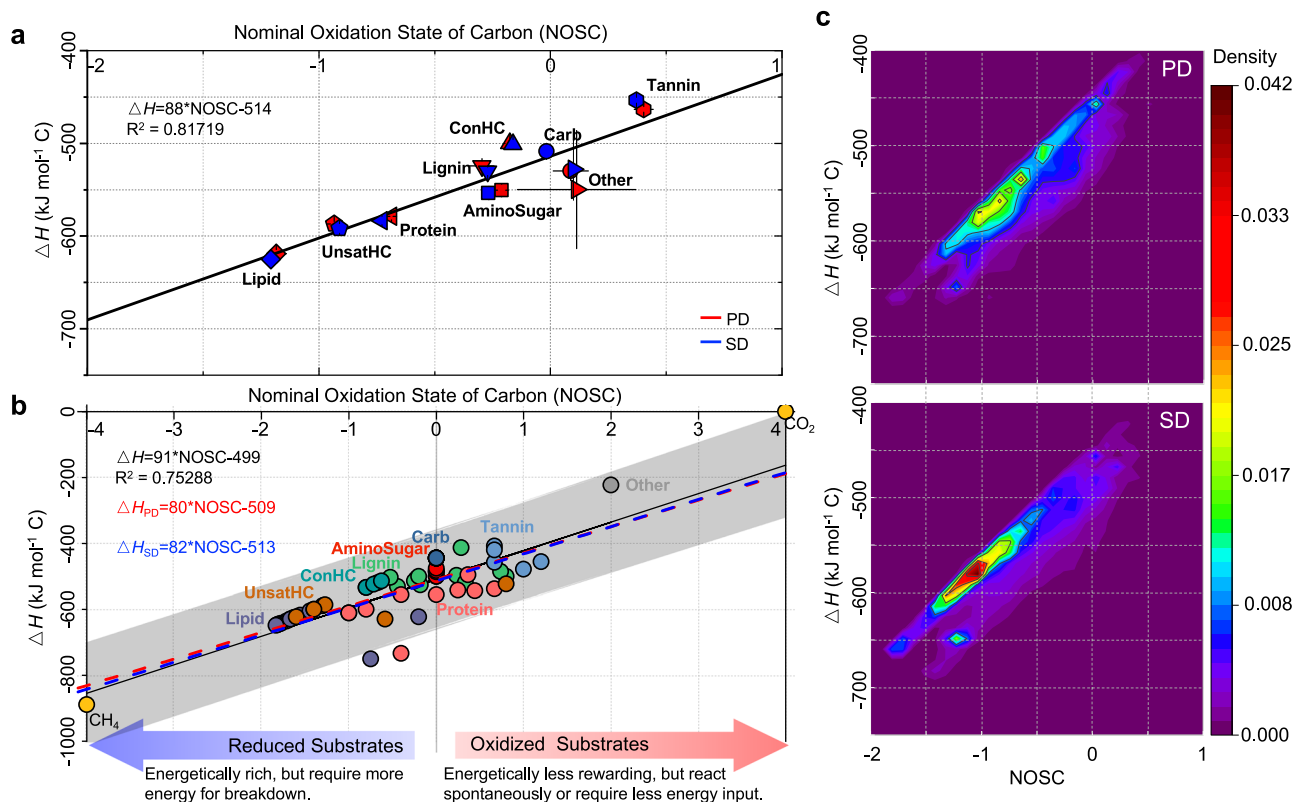


Fig. 5 | Dependence of energy density on NOSC values for compound classes presented in the DOM of sludge samples. a The relationship between average combustion enthalpy (ΔH) ($\text{kJ mol}^{-1} \text{C}$) and NOSC values of compound classes of DOM in PD and SD with error bars indicating standard deviation (SD) for both ΔH and NOSC ($n = 7$ biologically independent replicates). **b** Linear relationship validated using known compounds in waste sludges, with the shaded area representing

the 95% prediction interval of the fitted linear regression model. **c** Intensity-weighted ΔH -NOSC relationships, with relative intensity weights applied as the accounting unit for each molecule. The calculation methods for ΔH -NOSC follow the studies by LaRowe et al. (ref. 44) and Gunina et al. (ref. 45), which are detailed in Methods section. All formulae in PD and SD were included in the ΔH -NOSC analyses, which are available in the Source Data file.

validated using a range of known chemicals including all types of compound classes (Fig. 5b and Supplementary Table 8). The linear decrease in energy density of a broad range of organic substances by $91 \text{ kJ mol}^{-1} \text{ C}$ per NOSC unit upon DOM oxidation is supported by experimental data from both PD and SD samples, which all show linear ΔH -NOSC relationships within the prediction range of simulated data (Supplementary Fig. 20).

Most compound classes showed a shift toward lower values both for ΔH and NOSC from PD to SD (Fig. 5a), indicating that AD modifies the thermodynamic and redox properties of DOM, leaving more recalcitrant molecules in the final effluent. This observation is further supported by the intensity-weighted ΔH -NOSC relationships, where the relative intensity weight serves as the unit of account for each molecule. Notably, the entire molecular community exhibits a pronounced shift toward a more reducing state and higher energy potential during the AD process (Fig. 5c). Microorganisms in anaerobic sludges preferentially utilize readily degradable substrates, such as low molecular weight carbohydrates, carboxylic acids, and a portion of proteins (NOSC ≥ 0), which usually require less energy investment^{45,46}. However, the primary DOM compounds in the sludges, including lignin, lipid and unsaturated hydrocarbons pose greater thermodynamic challenges for microbial breakdown due to their higher energy demands and less favorable redox conditions.

Discussion

Understanding the biochemical process, especially at a molecular level in anaerobic bioprocessing, is the key to develop innovative strategies for organic waste treatment and resource recovery. However, the complexity of these systems presents significant challenges. We address these challenges by developing a comprehensive framework for analyzing the molecular diversity and transformation of DOM in AD system. Unlike many studies that focused on samples from a single site or a limited number of samples (typically two: pre- and post-fermentation) in controlled laboratory settings, our study collected samples from widely distributed FWTPs across diverse geographic regions. This approach allowed us to identify more “global” trends and derive more reliable common patterns based on average or core molecular data. Based on this analysis, we revealed the compositional evolution of DOM from PD to SD. We found that lignin-, lipid-, and protein-like compounds, dominated the DOM community in the final effluent. This composition is a result of multi-phase biochemical reactions, involving biological solubilization of organic solids and water-insoluble liquids (e.g., lipid, etc.), transformation between different forms of DOM, biogas production (e.g., CH_4 , CO_2 , etc.), and cell biosynthesis (Fig. 6a). Organic solids are a primary source of DOM molecules, especially for high-solid AD. The DOM pool experiences a smaller reduction compared to COD and SCOD during the SD process (Fig. 1a and Fig. 2b), suggesting that the inflow of DOM molecules from organic solid solubilization outweighs their breakdown. Another possible reason is that FT-ICR MS-undetected molecules are notably higher in PD than in SD, given that smaller molecules such as VFAs are considerably more abundant in PD than in SD (Fig. 1a and Supplementary Table 2). The molecules unique to SD samples are more diverse than those of PD samples, as well as those common to PD and SD samples (Fig. 4a). This indicates that persistent DOM from biological solubilization is the main component of the final effluent, which is evidenced by the increasing dominance of lipid-like molecules and the persistence of lignin-like molecules (Fig. 2a and Supplementary Figs. 21–22), despite the removal of large amounts of DOM through the biogas production and cell growth. Another important DOM source is the upstream DOM from PD. Nearly 20% of the DOM molecules in PD are also detected in the SD effluents, most of which belong to lignin and lipid molecules and become an important part of the final effluent DOM (Fig. 4a and Supplementary Fig. 23).

Our research underscores the critical role of intrinsic molecular characteristics, such as mass, carbon chain length, degree of

unsaturation, and redox properties, in shaping the molecular diversity and compositional evolution in AD (Figs. 2–5 and Supplementary Figs. 24–26). The persistence of more recalcitrant DOM is largely governed by its structural properties. Carbon chain length and C-C single bonds are particularly influential in determining molecular composition, especially for molecules derived from the complete hydrolysis of larger compounds. Additionally, higher aromaticity, particularly in lignin and protein molecules, contributes to their retention in the final effluent. These insights are essential for understanding the persistence of these molecules and identifying strategies to convert them into more bioavailable forms. Specifically, lipid-like compounds in the effluent, characterized by long-chain molecules with high C-C single bonds, could be converted into acetyl-CoA through β -oxidation⁴⁷. Lignin-like compounds, rich in aromatic rings, could be broken down through oxidative processes including dioxygenases-catalyzed ring-opening^{48,49}, O-demethylation, hydroxylation, and decarboxylation¹³, enhancing their bioavailability.

Unraveling the complexities of molecular transformations in mixtures is crucial but challenging⁵⁰. We propose an approach to dissect biochemical reactions by quantifying functional group changes in various compound classes. This methodology allows for a quantitative evaluation of key reactions like demethylation, dehydrogenation, and decarboxylation, which are fundamental to the metabolism of complex organic matters (Fig. 4 and Supplementary Fig. 19). These chemical interactions contribute significantly to the evolution of DOM composition and may provide insights into the persistence of more recalcitrant molecules in different environments. Future efforts could be directed to regulating these reactions to produce more oxidized, short-chain, and bioavailable compounds, potentially through manipulating the key enzymes and microorganisms^{51–53}. Sludge microbiota regulates the balance of carbon and energy through key metabolic processes: hydrolysis, acidogenesis, acetogenesis, and methanogenesis. These microbial-driven reactions mediate shifts in DOM composition, with community dynamics shaping the transformation of DOM compound classes (e.g., amino sugars, carbohydrates, and lipids) and elemental compositions (e.g., CHO and CHONS) (Supplementary Note 1 and Supplementary Fig. 4). As a result, the chemical composition shifts toward a more uniform redox state, with NOSC values transitioning from a broad range to values concentrated near zero (Supplementary Fig. 27). However, compounds under reduced states dominate the DOM community in the final effluent, which has a higher energy content but lower energy availability for microbial acidogenesis. From a DOM degradation kinetics perspective, acidogenesis of persistent DOM is an efficiency-limiting process for anaerobic digestion, which in turn affects the subsequent wastewater treatment.

In the current waste treatment paradigm, non-targeted removal often leads to underutilized DOM, particularly for substances like lignin and lipids. To enable effective microbial utilization of highly reduced substrates, further upgrading is imperative to shift the molecular compositions towards relatively high oxygen and low hydrogen content, necessitating enhanced oxidation processes (Fig. 6b). Reconstructing the microbial community is a feasible solution for the augmentation of biological oxidation, which could be achieved through targeted enrichment and redesigning sludge microbiomes by incorporating key bacteria responsible for the oxidation of lignin⁵⁴ and long-chain fatty acids^{39,47}. Additionally, introducing easily metabolizable co-substrates for microbes targeting persistent DOM, coupled with effluent liquid backflow processes, could further promote microbial digestion and stimulate co-metabolization. Moreover, combining chemical oxidation with anaerobic digestion could significantly improve the breakdown of fats, oils, and grease through advanced oxidation, resulting in low-molecular-weight substrates accessible for microbial treatment^{55,56}. Techniques such as oxidative depolymerization have shown promise in lignin valorization⁵⁷, suggesting a potential role for tandem chemical and

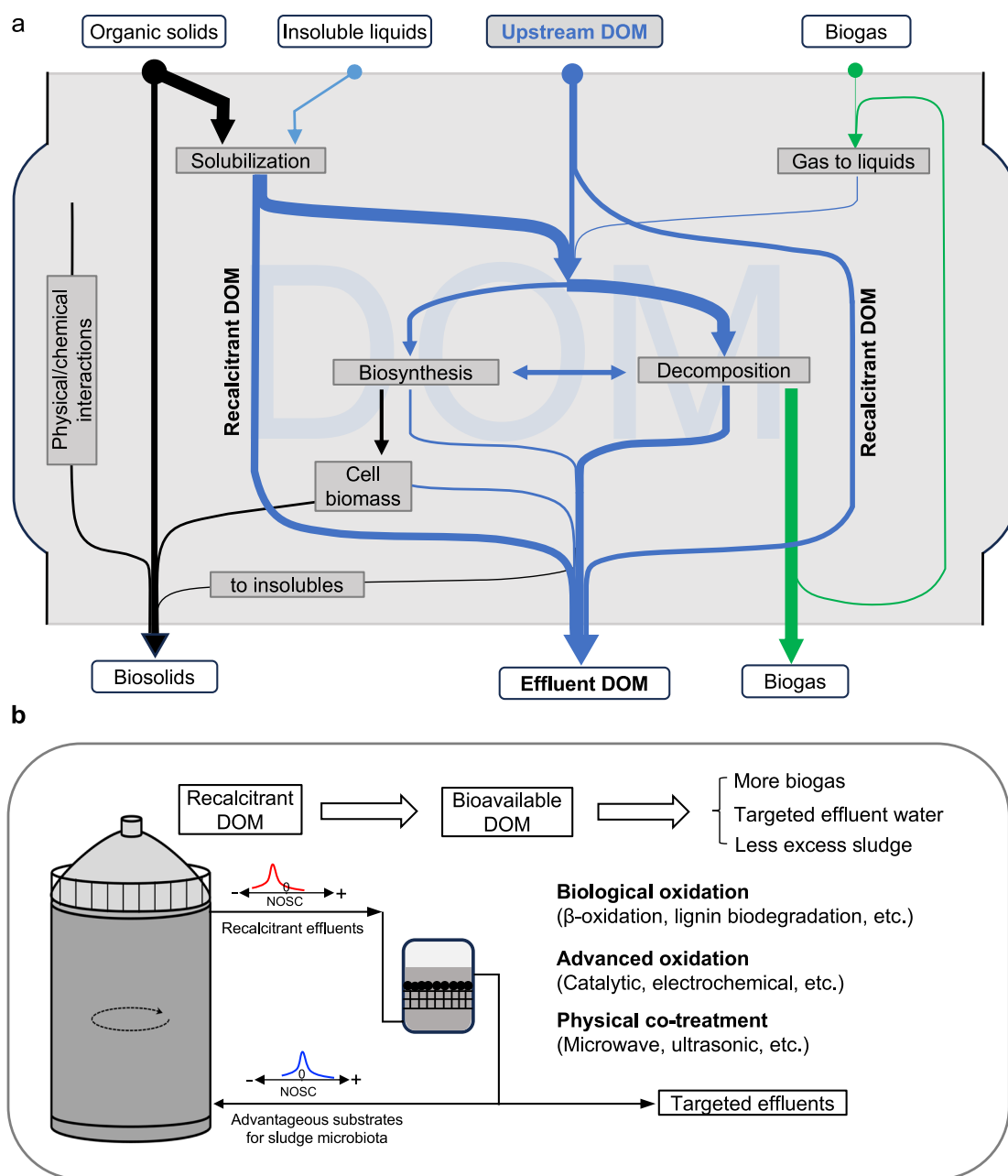


Fig. 6 | The global evolution of DOM metabolic flow and DOM transformation regulation for innovative anaerobic bioprocessing systems. a The evolution of DOM molecules during the AD. The solubilization here includes hydrolysis and

other biochemical processes to solubilize the solids. **b** Potential solutions for overcoming the recalcitrance of core molecular community for targeted effluents in the biological waste treatment.

biological processes. However, it is crucial to carefully balance chemical oxidation to prevent the inhibition of anaerobic microbes. Physical treatment is a supplementary way to increase the accessibility of organic materials^{58,59}. The integration of anaerobic digestion with these augmentation methods will be a promising strategy, facilitating the deconstruction of poorly degradable compounds into small and oxidized molecules accessible to microbial utilization.

In summary, our study comprehensively elucidated the molecular complexity and transformation of DOM in AD. The framework we developed here for DOM analysis offers a versatile tool for deciphering biochemical processes in various complex organic systems. Through this approach, we have unveiled intricate details about the composition and evolution of DOM, enhancing our understanding of biological waste treatment processes. The identified persistent yet prevalent

compounds in the sludge, primarily long-chain, aromatic, and unsaturated molecules, present valuable opportunities for energy recovery. Targeting these compounds for conversion into bioavailable substrates could significantly improve energy efficiency and clean water production, moving beyond the current paradigm of high-cost treatment techniques. This approach holds potential for wide application in diverse anaerobic bioprocessing systems, ranging from municipal organic waste treatment, lignocellulosic biorefinery, and biological plastics recycling.

Methods

Sludge sampling and characterization

Samples were collected from 7 representative full-scale, two-stage digesters located in Beijing (BJ), Guangyuan (GY), Hangzhou (HZ),

Liaocheng (LC), Quanzhou (QZ), Shenyang (SY), Shenzhen (SZ) across China. The cities BJ, LC, and SY belong to North China, while GY, HZ, QZ and SZ are in South China, regions traditionally distinguished by their culture, diet, habits, and customs. These differences could potentially influence the DOM molecular composition in AD process. Additional details regarding sample geographical features are provided in Supplementary Table 9. These samples were collected from the effluent of PD and SD at the end-stage of a feed cycle. At least 250 mL sample was taken from each digester and then shipped on dry ice to the lab. Upon arrival, the samples were separated into aliquots, specifically prepared for analysis of physicochemical parameters and DOM. For liquid sample analysis, the samples were centrifuged at 7000 rpm (5752 rcf) for 10 min and the supernatant was filtered with 0.22 µm Millipore membrane. These samples were stored short-term in fridge (<14 days). Owing to high organic contents in the DOM samples, all the DOM samples were diluted 10 times before filtering and other treatment. The sludge properties such as total solid (TS), volatile solid (VS), chemical oxygen demand (COD), total dissolved nitrogen (TDN), total dissolved phosphorus (TDP) were analyzed using standard methods as shown in Supplementary Table 10. The VFAs, including C2–C8 fatty acids were determined by gas chromatograph (Model 6890NT, Agilent Inc., USA) equipped with a flame ionization detector and a 30 m x 0.25 mm x 0.25 µm fused-silica capillary column (DB-FFAP), following the method described by Raposo et al.⁶⁰.

DOM sample preparation and FT-ICR MS experiment

Solid phase extraction (SPE) was performed using solid phase extraction columns (HyperSep™ retain PEP 60 mg, Thermo Scientific) with balanced retention of polar and non-polar analytes. The detailed SPE procedure is provided in Supplementary Method 1. The analysis of DOM was conducted using a 7 T Bruker solarix FT-ICR MS in negative ion mode with the ESI source and a krypton lamp, following a previous protocol⁶¹ with minor adjustments. The nebulizer gas (N₂) was adjusted to 2.0 bar and maintained at 400 °C to ensure complete vaporization of the samples injected at a rate of 20 µL/h. The operating capillary voltage was -1500 V, and the end plate voltage was -500 V. Ion accumulation and flight times to the analyzer were optimized to 0.02 s and 0.8 s, respectively. We collected 200 coadded scans of broadband spectra in the 54–20000 m/z range and the mass resolution set to >320000 at 400 m/z. The sodium trifluoroacetate (NaTFA) was used as a quality control sample within the mass range of 112–2968 m/z, achieving a mass accuracy of ≤ 0.5 ppm through calibration. The mass spectra of the DOM of PD and SD samples from Shenzhen (SZ) were provided to as an example to underline the differences between samples and the resolution power (Supplementary Fig. 28).

FT-ICR MS data analysis

Raw spectra were converted to a list of m/z values using the Data Analysis software (version 5.2, Bruker Daltonics, CA) with a signal-to-noise (S/N) ratio >7. Internal calibration was performed to adjust experimental mass measurements based on calibrants listed in Supplementary Table 11 before formula assignment. Putative chemical formulae were assigned using the Compound Identification Algorithm (CIA)²³ with a user-friendly software Formularity²². Assignments adhered to a mass accuracy threshold of ≤ 0.5 ppm, within the mass range of 100–1000 m/z. Our formula assignments followed the “golden rules”⁶², considering the elements C, H, O, N, S, and P under the elemental constraints $C_{1-∞}H_{1-∞}O_{1-∞}N_{0-3}S_{0-2}P_{0-1}$, $O/C ≤ 1.2$ and $0.3 ≤ H/C ≤ 3$. Furthermore, we prioritized formulas with the fewest heteroatoms (N, S, P) and the lowest mass error, reducing the likelihood of improbable assignments^{22,63}. These details and formula assignment results are summarized in Supplementary Table 12. The assigned formulae were categorized into eight elemental combinations (CHO, CHON, CHONP, CHONS, CHONSP, CHOP, CHOS and CHOSP). Biochemical compound classes including amino sugar (AminoSugar),

carbohydrate (Carb), condensed hydrocarbon (ConHC), lignin, lipid, protein, tannin, unsaturated hydrocarbon (unsatHC) and all molecules unassigned to the prior categories (“other”), are grouped based on counts of C, H, and O in the assigned formulae with different defined H:C and O:C ranges^{22,34}, which are summarized in Supplementary Table 13. Notably, these classes are a priori assignments that do not rely on any measured proof of identification (structure, e.g., by MS/MS) but are solely used to categorize the data for easier interpretation.

As direct structural characterization of diverse and complex DOM molecules is currently unfeasible (Supplementary Note 4), we utilized well-established computational methods for structural elucidation. Modified aromaticity index (AI_{mod})⁶⁴ was used to quantify the number of benzene rings with the equation: $AI_{mod} = (1 + C - 0.5 * O - S - 0.5 * (N + P + H)) / (C - 0.5 * O - N - S - P)$, where, C, H, O, N, S and P are the number of carbon, hydrogen, oxygen, nitrogen, sulfur, and phosphorus atoms in a given molecular formula. This notation is consistent across all subsequent equations. Double bond equivalent (DBE)⁶⁵ was used to quantify the number of double bonds in the DOM molecules with the equation: $DBE = 1 + C - 0.5 * H + 0.5 * N + 0.5 * P$. To assess carbon–carbon single bonding, we used the C–1–DBE index, which reflects the number of C–C single bonds. Notably, the estimate may be lower than the actual value, as non-C–C double bonds (e.g., C=O, C=N) cannot be accurately distinguished in this calculation. The nominal oxidation state of carbon (NOSC)⁴⁴ of each organic molecule was calculated using the following equation: $NOSC = -H/C + 2 * O/C + 3 * N/C + 2 * S/C - 5 * P/C$. Gibbs free energy of a given organic compound can be computed as the standard molar enthalpy of its complete combustion (ΔH). ΔH values are calculated based on the Patel-Erickson equation^{66,67}, where the standard molar enthalpy of combustion of organic matter is proportional to the number of electrons that it transfers to oxygen during combustion. This proportionality is expressed through the equation⁶⁸: $\Delta H = -114.14 * (4 * C + H - 2 * O - 0 * N + 5 * P + 6 * S) / C$. The quantitative analysis of chemical reactions including demethylation/methylation, decarboxylation/carboxylation, dehydration/hydration, dehydrogenation/hydrogenation, and deamination/amination reactions are calculated based on the functional group differences between the formulae of PD and SD samples. The analysis of biochemical reactions is conducted through the Perl programming language and the detailed calculation process is provided in Supplementary Method 2.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data underlying this study are available from the corresponding author upon request. The data supporting the findings are provided in the main text, Supplementary Information, and the Source Data file. Raw mass spectrometry data are publicly available on Zenodo (<https://doi.org/10.5281/zenodo.15192562>). Source data are provided with this paper.

Code availability

The code for biochemical reaction analysis is available on Github (https://github.com/hzhengsh/DOM_biochemical_reaction) and archived on Zenodo (<https://doi.org/10.5281/zenodo.15172534>).

References

1. Liu, Z. Regulate waste recycling internationally. *Nature* **594**, 333–333 (2021).
2. Guo, Y. et al. Global food loss and waste embodies unrecognized harms to air quality and biodiversity hotspots. *Nat. Food* **4**, 686–698 (2023).
3. Xue, L. et al. China's food loss and waste embodies increasing environmental impacts. *Nat. Food* **2**, 519–528 (2021).

4. Kaza, S., Yao, L., Bhada-Tata, P. & Woerden, F. V. *What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050*. <https://datatopics.worldbank.org/what-a-waste/> (2018).
5. World Bioenergy Association. *Global Bioenergy Statistics Report 2023*. <https://www.worldbioenergy.org/global-bioenergy-statistics/> (2023).
6. UN-Water. *Progress on Wastewater Treatment—2021 Update*. <https://www.unwater.org/publications/progress-wastewater-treatment-2021-update#:~:text=Target%206.3%20is%3A%20E2%80%9CBy%202030> (2021).
7. Jones, E. R., Vliet, M. T. H. van, Qadir, M. & Bierkens, M. F. P. Country-level and gridded estimates of wastewater production, collection, treatment and reuse. *Earth Syst. Sci. Data* **13**, 237–254 (2020).
8. Li, W.-W., Yu, H.-Q. & Rittmann, B. E. Chemistry: reuse water pollutants. *Nature* **528**, 29–31 (2015).
9. Du, W.-J. et al. Spatiotemporal pattern of greenhouse gas emissions in China's wastewater sector and pathways towards carbon neutrality. *Nat. Water* **1**, 166–175 (2023).
10. Heronemus, E., Gamage, K. H. H., Hettiarachchi, G. M. & Parameswaran, P. Efficient recovery of phosphorus and sulfur from anaerobic membrane bioreactor (AnMBR) permeate using chemical addition of iron and evaluation of its nutrient availability for plant uptake. *Sci. Total Environ.* **783**, 146850 (2021).
11. Xiao, K., Abbt-Braun, G. & Horn, H. Changes in the characteristics of dissolved organic matter during sludge treatment: a critical review. *Water Res* **187**, 116441 (2020).
12. Nebbioso, A. & Piccolo, A. Molecular characterization of dissolved organic matter (DOM): a critical review. *Anal. Bioanal. Chem.* **405**, 109–124 (2013).
13. Luo, J. et al. Lake chemodiversity driven by natural and anthropogenic factors. *Environ. Sci. Technol.* **56**, 5910–5919 (2022).
14. Kellerman, A. M., Dittmar, T., Kothawala, D. N. & Tranvik, L. J. Chemodiversity of dissolved organic matter in lakes driven by climate and hydrology. *Nat. Commun.* **5**, 3804 (2014).
15. Hu, A. et al. Ecological networks of dissolved organic matter and microorganisms under global change. *Nat. Commun.* **13**, 3600 (2022).
16. McDonough, L. K. et al. Changes in global groundwater organic carbon driven by climate change and urbanization. *Nat. Commun.* **11**, 1279 (2020).
17. McDonough, L. K. et al. A new conceptual framework for the transformation of groundwater dissolved organic matter. *Nat. Commun.* **13**, 2153 (2022).
18. He, W., Chen, M., Park, J.-E. & Hur, J. Molecular diversity of riverine alkaline-extractable sediment organic matter and its linkages with spectral indicators and molecular size distributions. *Water Res.* **100**, 222–231 (2016).
19. Chen, X. et al. Niche differentiation of microbial community shapes vertical distribution of recalcitrant dissolved organic matter in deep-sea sediments. *Environ. Int.* **178**, 108080 (2023).
20. Wilson, T., Potts, L. & Stallings, R. Review of full scale 2-phase AG anaerobic digester systems. *New England Water Environment Association Annual Conference* (EPA, 2005).
21. Hu, H. et al. Effect of solids retention time on effluent dissolved organic nitrogen in the activated sludge process: studies on bioavailability, fluorescent components, and molecular characteristics. *Environ. Sci. Technol.* **52**, 3449–3455 (2018).
22. Tolić, N. et al. Formularity: software for automated formula assignment of natural and other organic matter from ultrahigh-resolution mass spectra. *Anal. Chem.* **89**, 12659–12665 (2017).
23. Kujawski, E. B. & Behn, M. D. Automated analysis of electrospray ionization Fourier transform ion cyclotron resonance mass spectra of natural organic matter. *Anal. Chem.* **78**, 4363–4373 (2006).
24. Danczak, R. E. et al. Using metacommunity ecology to understand environmental metabolomes. *Nat. Commun.* **11**, 6369 (2020).
25. Lechtenfeld, O. J., Hertkorn, N., Shen, Y., Witt, M. & Benner, R. Marine sequestration of carbon in bacterial metabolites. *Nat. Commun.* **6**, 6711 (2015).
26. Medeiros, P. M. et al. Dissolved organic matter composition and photochemical transformations in the northern North Pacific Ocean. *Geophys. Res. Lett.* **42**, 863–870 (2015).
27. Seidel, M., Vernulapalli, S. P. B., Mathieu, D. & Dittmar, T. Marine dissolved organic matter shares thousands of molecular formulae yet differs structurally across major water masses. *Environ. Sci. Technol.* **56**, 3758–3769 (2022).
28. Sun, X. et al. Linkages between optical and molecular signatures of dissolved organic matter along the Yangtze river estuary-to-east China sea continuum. *Front. Mar. Sci.* **9**, 933561 (2022).
29. Kellerman, A. M., Kothawala, D. N., Dittmar, T. & Tranvik, L. J. Persistence of dissolved organic matter in lakes related to its molecular characteristics. *Nat. Geosci.* **8**, 454–457 (2015).
30. Zark, M. & Dittmar, T. Universal molecular structures in natural dissolved organic matter. *Nat. Commun.* **9**, 3178 (2018).
31. Osterholz, H. et al. Terrigenous dissolved organic matter persists in the energy-limited deep groundwaters of the Fennoscandian Shield. *Nat. Commun.* **13**, 4837 (2022).
32. Li, H.-Y. et al. The chemodiversity of paddy soil dissolved organic matter correlates with microbial community at continental scales. *Microbiome* **6**, 187 (2018).
33. Ding, Y. et al. Chemodiversity of soil dissolved organic matter. *Environ. Sci. Technol.* **54**, 6174–6184 (2020).
34. Bailey, V. L., Smith, A. P., Tfaily, M., Fansler, S. J. & Bond-Lamberty, B. Differences in soluble organic carbon chemistry in pore waters sampled from different pore size domains. *Soil Biol. Biochem.* **107**, 133–143 (2017).
35. Liu, J. et al. Comprehensive understanding of DOM reactivity in anaerobic fermentation of persulfate-pretreated sewage sludge via FT-ICR mass spectrometry and reactivity analysis. *Water Res.* **229**, 119488 (2022).
36. Zhao, P. et al. Molecular composition and chemodiversity of dissolved organic matter in wastewater sludge via Fourier transform ion cyclotron resonance mass spectrometry: effects of extraction methods and electrospray ionization modes. *Water Res.* **232**, 119687 (2023).
37. Yang, X. et al. Unveiling the deterministic dynamics of microbial meta-metabolism: a multi-omics investigation of anaerobic biodegradation. *Microbiome* **12**, 166 (2024).
38. Lalman, J. A. & Bagley, D. M. Anaerobic degradation and methanogenic inhibitory effects of oleic and stearic acids. *Water Res.* **35**, 2975–2983 (2001).
39. Holohan, B. C. et al. Principles, advances, and perspectives of anaerobic digestion of lipids. *Environ. Sci. Technol.* **56**, 4749–4775 (2022).
40. Zhu, S. & Xu, C. *Biochemistry* 4th Edition (Higher Education Press, Beijing, 2017).
41. Yenigün, O. & Demirel, B. Ammonia inhibition in anaerobic digestion: a review. *Process Biochem.* **48**, 901–911 (2013).
42. Kim, S., Kramer, R. W. & Hatcher, P. G. Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the Van Krevelen diagram. *Anal. Chem.* **75**, 5336–5344 (2003).
43. Erickson, E. et al. Critical enzyme reactions in aromatic catabolism for microbial lignin conversion. *Nat. Catal.* **5**, 86–98 (2022).
44. LaRowe, D. E. & Cappellen, P. V. Degradation of natural organic matter: a thermodynamic analysis. *Geochimica Cosmochimica Acta* **75**, 2030–2042 (2011).
45. Gunina, A. & Kuzyakov, Y. From energy to (soil organic) matter. *Glob. Chang. Biol.* **28**, 2169–2182 (2022).
46. Wang, C. & Kuzyakov, Y. Energy use efficiency of soil microorganisms: driven by carbon recycling and reduction. *Glob. Chang. Biol.* **29**, 6170–6187 (2023).

47. Usman, M., Zhao, S., Jeon, B.-H., Li, X. & Salama, E.-S. Microbial β -oxidation of synthetic long-chain fatty acids to improve lipid biomethanation. *Water Res.* **213**, 118164 (2022).
48. Bugg, T. D. H. Dioxygenase enzymes: catalytic mechanisms and chemical models. *Tetrahedron* **59**, 7075–7101 (2003).
49. Vaillancourt, F. H., Bolin, J. T. & Eltis, L. D. The ins and outs of ring-cleaving dioxygenases. *Crit. Rev. Biochem. Mol. Biol.* **41**, 241–267 (2006).
50. Leyva, D., Tariq, M. U., Jaffé, R., Saeed, F. & Fernandez-Lima, F. Description of dissolved organic matter transformational networks at the molecular level. *Environ. Sci. Technol.* **57**, 2672–2681 (2023).
51. Napier, J. A., Haslam, R. P., Beaudoin, F. & Cahoon, E. B. Understanding and manipulating plant lipid composition: metabolic engineering leads the way. *Curr. Opin. Plant Biol.* **19**, 68–75 (2014).
52. Ngan, C. Y. et al. Lineage-specific chromatin signatures reveal a regulator of lipid metabolism in microalgae. *Nat. Plants* **1**, 15107 (2015).
53. Bornscheuer, Uwe T., ed. *Lipid Modification by Enzymes and Engineered Microbes*, Vol. 240 (Elsevier, 2018).
54. Khan, M. U. & Ahring, B. K. Lignin degradation under anaerobic digestion: influence of lignin modifications -a review. *Biomass. Bioenergy* **128**, 105325 (2019).
55. Ries, S. et al. Microwave enhanced advanced oxidation treatment of fat, oil and grease (FOG) with organic co-substrates. *Environ. Technol.* **42**, 4500–4510 (2021).
56. Srinivasan, A. et al. Microwave-enhanced advanced oxidation treatment of lipids and food wastes. *Water Air Soil Pollut.* **229**, 227 (2018).
57. Sun, Z., Fridrich, B., Santi, A. de, Elangovan, S. & Barta, K. Bright side of lignin depolymerization: toward new platform chemicals. *Chem. Rev.* **118**, 614–678 (2018).
58. Balch, M. L. et al. Lignocellulose fermentation and residual solids characterization for senescent switchgrass fermentation by *Clostridium thermocellum* in the presence and absence of continuous in situ ball-milling. *Energ. Environ. Sci.* **10**, 1252–1261 (2017).
59. Bozkurt, Y. C. & Apul, O. G. Critical review for microwave pre-treatment of waste-activated sludge prior to anaerobic digestion. *Curr. Opin. Environ. Sci. Heal.* **14**, 1–9 (2020).
60. Raposo, F. et al. Harmonization of the quantitative determination of volatile fatty acids profile in aqueous matrix samples by direct injection using gas chromatography and high-performance liquid chromatography techniques: multi-laboratory validation study. *J. Chromatogr. A* **1413**, 94–106 (2015).
61. He, C. et al. In-house standard method for molecular characterization of dissolved organic matter by FT-ICR mass spectrometry. *ACS Omega* **5**, 11730–11736 (2020).
62. Kind, T. & Fiehn, O. Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinform.* **8**, 105 (2007).
63. Bramer, L. M. et al. ftmsRanalysis: an R package for exploratory data analysis and interactive visualization of FT-MS data. *Plos Comput. Biol.* **16**, e1007654 (2020).
64. Koch, B. P. & Dittmar, T. From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid Commun. Mass Spectrom.* **30**, 250–250 (2016).
65. Koch, B. P. & Dittmar, T. From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid Commun. Mass Spectrom.* **20**, 926–932 (2006).
66. Patel, S. A. & Erickson, L. E. Estimation of heats of combustion of biomass from elemental analysis using available electron concepts. *Biotechnol. Bioeng.* **23**, 2051–2067 (1981).
67. Battley, E. H. The development of direct and indirect methods for the study of the thermodynamics of microbial growth. *Thermochim. Acta* **309**, 17–37 (1998).
68. Popovic, M. Thermodynamic properties of microorganisms: determination and analysis of enthalpy, entropy, and Gibbs free energy of biomass, cells and colonies of 32 microorganism species. *Helvion* **5**, e01950 (2019).

Acknowledgements

We gratefully acknowledge support from the National Key Research and Development Program of China (2019YFC1905002) and the National Natural Science Foundation of China (NSFC 22225602, and 52293441) to X.W.L. We also thank Yun-Tong Xu, Min Feng, and You-Ming Jiang from Xi'an Jiaotong University (XJTU), as well as Ze-Wei Li and Jun-Hao Wan from University of Science and Technology of China (USTC) for their assistance with supplemental experiments and analyses during the manuscript revision.

Author contributions

J.H., X.W.L. designed the study. C.G.L. acquired the samples, J.H., W.K.Z. and X.W.L. (Xue-Wen Liu) conducted the physicochemical analysis of sludge samples. Y.G.X. and Z.S.H. designing the necessary code. J.H., B.D. and Z.D.W. performed FT-MS data analysis. J.H., X.W.L. contributed to the writing of the manuscript.

Competing interests

The authors declare no competing interest.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-60240-3>.

Correspondence and requests for materials should be addressed to Xian-Wei Liu.

Peer review information *Nature Communications* thanks Carsten Simon and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025