Water Research 308 (2027) 126789

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

# Differentiating molecular fingerprints of dissolved organic nitrogen between natural soil and iron-based sorption media for stormwater treatment

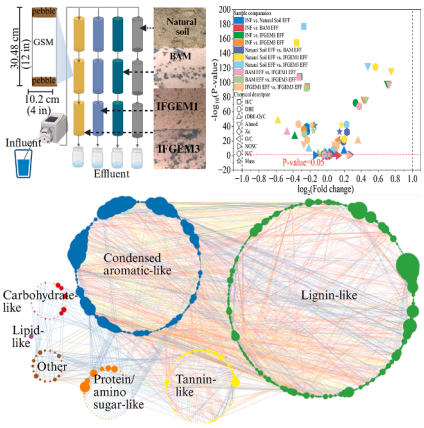
Jinxiang Cheng [a], Andrea Valencia [a], Diana Ordonez [a], Amy M. McKenna [b,c], Ni-Bin Chang [a,*]

[a] Department of Civil, Environmental, and Construction Engineering, University of Central Florida, Orlando, FL, USA
[b] National High Magnetic Field Laboratory, Florida State University, Tallahassee, FL, USA
[c] Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO, USA

## HIGHLIGHTS

- Natural soil exhibits distinctive DON decomposition pathways in stormwater treatment.
- Iron-based sorption media mix has better DON fractional removal than natural soil.
- Chemical descriptors and energetic properties help distinguish the lability of DON.

## GRAPHICAL ABSTRACT




## ARTICLE INFO

*Keywords:*
Stormwater treatment
Porous media
Dissolved organic nitrogen
Nitrogen cycle
Systems chemistry

## ABSTRACT

Dissolved organic nitrogen (DON) remains a significantly under-investigated component in the disruptive nitrogen cycle especially in different soils or specialty adsorbents that can efficiently remove inorganic nitrogen after pollution when simultaneously retaining, removing and remaking DON. The complexity of DON fate, transport, and transformation processes in low impact development infrastructure design is still a pending question for stormwater treatment. To compare different DON degradation pathways and clarify inherent complexity, we integrated paired influent–effluent column experiments with Fourier transform ion cyclotron resonance mass spectrometry, molecular descriptor and zonation analyses, fate-resolved reaction inference, Kendrick mass defect-based network analysis, and depth-resolved nitrogen-cycling gene profiling to examine DON transformation in natural soil, biosorption activated media (BAM), and two iron-filings-based green environmental media (IFGEM1 and IFGEM3) under three nitrate–phosphate loading conditions. Across the three nutrient-loading conditions, IFGEM1 and IFGEM3 achieved NOx-N ($NO_3^-$-N + $NO_2^-$-N) removal efficiencies of

* Corresponding author.
*E-mail address:* nchang@ucf.edu (N.-B. Chang).

https://doi.org/10.1016/j.watres.2026.126789
Received 31 May 2026; Received in revised form 19 August 2026; Accepted 24 August 2026
Available online 24 August 2026
0043-1354/© 2026 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

98.6–99.2% and 98.9–99.9%, respectively, while their effluents showed systematic shifts in the assigned CHON formula pool. Relative to natural soil, IFGEM effluents showed lower relative contributions of condensed aromatic-like formulas and higher contributions of protein/amino sugar-like formulas, together with greater formula richness and stronger separation in molecular descriptor space, most notably for IFGEM3. Natural soil increasingly retained a lignin-rich, oxidizing–saturated DON pool as loading increased, whereas IFGEM3 promoted greater removed-to-produced turnover and concentrated newly produced DON in reducing–saturated chemical space under higher loading. Increasing N and P loading also redistributed inferred transformation portfolios: natural soil and BAM retained broader, more mixed reaction spectra, whereas IFGEM1 and IFGEM3 shifted towards more selective pathway combinations. KMD-derived formula-association networks showed broader cross-class connectivity in IFGEM3 than in natural soil, with the network edges interpreted as putative molecular linkages rather than confirmed transformations. Pre-spiking qPCR profiles showed depth-dependent differences in baseline nitrogen-cycling gene abundances. Overall, iron-based engineered media were associated with systematic restructuring of the detectable DON molecular composition across the tested nutrient-loading conditions.

## 1. Introduction

Urban stormwater runoff is an important pathway for transporting nitrogen (N) and phosphorus (P) from developed landscapes to receiving waters, where nutrient enrichment can contribute to eutrophication, algal proliferation, and broader aquatic ecosystem impairment (Yang and Lusk, 2018; Jani et al., 2020). Stormwater control measures, including infiltration basins, bioretention systems, retention ponds, and engineered reactive-media filters, are therefore widely used to attenuate runoff volume and reduce nutrient export from urban catchments (Li and Davis, 2014; O'Reilly et al., 2012; Osman et al., 2019). Their performance, however, is still commonly evaluated using bulk water-quality endpoints, such as total nitrogen (TN), nitrate, ammonium, total phosphorus (TP), and orthophosphate, although nitrogen in stormwater occurs as multiple inorganics, particulate organic, and dissolved organic forms with different mobility and reactivity (Li and Davis, 2014; Jani et al., 2020). This distinction is critical because dissolved organic nitrogen (DON) can constitute a substantial fraction of stormwater nitrogen and can respond differently from nitrate or ammonium during treatment (Lusk and Toor, 2016a; Jani et al., 2020).

DON is not simply an inert residual nitrogen pool. Bioavailable fractions can support microbial and phytoplankton growth, whereas less reactive fractions contribute to longer-term nitrogen storage and transport (Berman and Bronk, 2003; Bronk et al., 2007). Treatment may therefore alter both DON quantity and quality; biological nutrient-removal systems, for example, can generate low-molecular-weight DON that remains ecologically relevant after discharge (Pehlivanoglu-Mantas and Sedlak, 2006; Eom et al., 2017). Stormwater studies have addressed related but distinct aspects of this issue. Li and Davis (2014) documented DON and $NO_3^-$-N release from a conventional bioretention cell, but did not characterize the exported DON at the molecular level or establish whether the mechanisms responsible for that release apply to the media examined here. Lusk and Toor (2016a) showed that urban streams with similar DON concentrations differed in biodegradability and molecular composition, while a separate retention-pond study found corresponding differences between runoff and pond outflow (Lusk and Toor, 2016b). Thus, bulk DON concentration alone does not fully describe DON quality in urban water systems. However, these studies did not isolate the effects of specific filter media under matched influent and hydraulic conditions. It remains unclear whether natural soil, biosorption activated media (BAM), IFGEM1, and IFGEM3 exhibit distinct effluent molecular fingerprints, how these patterns vary with nutrient loading, and how they relate to bulk nitrogen performance. We addressed these questions by comparing paired influent and effluent samples across three loading conditions using bulk nitrogen measurements, formula occurrence, compound-class distributions, and molecular descriptors. KMD/SOM (Self-Organizing Map) associations and pre-spiking gene profiles were used as hypothesis-generating context rather than as confirmation of biochemical pathways. Engineered sorption media provide an opportunity to improve passive nutrient control in stormwater treatment while also creating distinct physicochemical and microbial environments for DON transformation. Biosorption activated media (BAM), composed primarily of tire crumb, clay/silt, and sand, was developed as a functionalized amendment to enhance sorption capacity, soil moisture retention, infiltration, and nutrient attenuation in stormwater infiltration systems (O'Reilly et al., 2012). Iron-filings-based green environmental media (IFGEM) extends this design by incorporating recycled iron filings, thereby introducing reactive iron surfaces and redox-active microsites that can enhance nitrate removal, phosphorus retention, and coupled biogeochemical processing (Wen et al., 2020; Valencia et al., 2020). Previous studies of related green sorption media, mixed with recycled zero valent iron (ZVI), clay, sand etc., have shown that media composition, redox conditions, influent chemistry, and microbial structure can influence DON removal and transformation (Valencia et al., 2020; Robles-Lecompte et al., 2024; Cheng et al., 2026). Recent studies linked DON fractional removal to microbial species richness in a cascade upflow biofiltration process and have shown that seasonal hydrologic forcing can regulate DON lability and molecular fate in a field-scale filtration system (Robles-Lecompte et al., 2024;Cheng et al., 2026 ). Several related studies provide the immediate foundation for the present work. Valencia et al. (2020) reported the underlying fixed-bed column experiment, comparing bulk nutrient and DON behavior and selected nitrogen-cycling gene abundances in natural soil, BAM, IFGEM1, and IFGEM3 under three influent conditions. Its primary emphasis was treatment performance and microbial ecology, rather than a systematic cross-media comparison of formula-level descriptor space, chemical-trait diversity, and mass-difference network structure. Robles-Lecompte et al. (2024) subsequently integrated FT-ICR MS and qPCR in a cascade upflow system containing BIPGEM and ZIPGEM, whereas Cheng et al. (2026) examined DON lability and formula associations in a field-scale CPS/ZIPGEM system across seasons. These studies resolved cascade-process and field-seasonal behavior, respectively, but did not provide a four-media comparison spanning natural soil, conventional BAM, and two iron-amended formulations under a common influent-loading and column-operation protocol.

The present study is therefore intended as a complementary molecular extension rather than a correction of the earlier work. We examine whether the four media exhibit distinguishable influent-to-effluent changes in the PPL-isolated, negative-ESI-detectable CHON formula pool across three nutrient-loading conditions, and which formula distributions and molecular descriptors most clearly differentiate those responses. Formula occurrence, compound-class composition, and chemical-trait diversity are used to characterize media-associated molecular profiles. KMD/SOM-derived associations are interpreted as putative formula linkages, and pre-spiking qPCR profiles provide baseline microbial context, rather than confirmation of biochemical pathways. By relating these molecular patterns to bulk nitrogen performance, the study provides complementary molecular information for characterizing media-associated changes in the composition and potential reactivity of

residual detectable DON.

In this study, we investigated DON molecular transformation across four stormwater filtration media—natural soil, BAM, IFGEM1, and IFGEM3—operated under three nitrate–phosphate spiking conditions. Paired influent and effluent samples were characterized by negative-ion ESI FT-ICR MS, and the assigned CHON formulas were treated as an operationally defined subset of DON-related molecular signals rather than as a comprehensive proxy for bulk DON. An assigned elemental formula does not establish a unique compound identity or molecular structure, and DON constituents that are poorly retained by styrene divinyl benzene polymer cartridges (PPL)-solid phase extraction (SPE) (PPL-SPE) or weakly ionized under negative-ion electrospray ionization (ESI) may be underrepresented. Molecular fingerprints were evaluated through molecular-class assignment, formula-based chemical descriptors, oxidation–saturation zonation, operational formula-occurrence classification, and KMD/SOM-based putative formula-association analysis, while pre-spiking depth-resolved qPCR profiles provided baseline microbial context. Specifically, this study addressed three research questions: (i) how do natural soil, BAM, and IFGEM media differ in bulk nutrient removal and DON molecular fingerprints under varying nutrient loads; (ii) which DON classes and molecular regimes are preferentially removed, retained, or produced during media passage; and (iii) how are media-specific DON transformation patterns associated with microbial nitrogen-cycling potential across column depth? By integrating treatment performance with fate-resolved molecular fingerprints, this study advances a mechanistic framework for evaluating stormwater nutrient-control media based not only on inorganic nutrient attenuation, but also on the molecular reorganization and potential residual reactivity of DON.

## 2. Methods

### 2.1. Experiment setting up, water sampling and bulk chemical analysis

Four identical fixed-bed columns were constructed using PVC pipes with an internal diameter of 10 cm (4 in). Each column was divided into three 30 cm (12 in) sections to allow media sampling at different depths. A filter layer containing pebbles was installed at the bottom of each section to minimize clogging, and all column sections were sealed to prevent external interference. The four columns were packed with natural soil, BAM, IFGEM-1, and IFGEM-3, respectively. Natural soil was collected from a stormwater retention pond (dry pond) located at Ocala, Florida where the iron-based sorption media were installed as a filtration layer in the vadose zone. The physicochemical properties of the natural soil and engineered iron-based sorption media are provided in Table S1, and the fixed-bed column configuration is shown in Fig. 1. Natural soil was collected from Basin 9b, a basin adjacent to Silver Springs State Park, Florida (USA). The engineered media were prepared as volumetric mixtures: BAM (85% sand, 10% tire crumb, 5% clay), IFGEM1 (96.2% sand, 3.8% iron filings), and IFGEM3 (83% sand, 10% tire crumb, 5% iron filings, 2% clay). The recipes were adopted from previous IFGEM development studies rather than designed as an iron-dose series. IFGEM1 was retained as an established binary sand–iron reference formulation, whereas IFGEM3 was the optimized multicomponent formulation selected by Valencia et al. (2019) after testing seven clay–iron recipes. Thus, differences between IFGEM1 and IFGEM3 reflect their complete media compositions rather than the 3.8% versus 5% iron content alone.

Following a 4-week stormwater-conditioning period, the columns were subjected sequentially to three 3-h nitrate–phosphate pulses (SC1–SC3) at a constant flow rate of 8 mL min. Detailed conditioning, flushing, sampling (Table S2), and bulk water-quality analysis procedures (Table S3, S4 and S5) are provided in Supplementary Method S1.2.2 Baseline DNA extraction and qPCR quantification of nitrogen-cycling gene markers

The depth-resolved qPCR dataset was generated during the fixed-bed column experiment previously reported by Valencia et al. (2020). Media samples were collected during weeks 2 and 4 of the four-week stormwater-conditioning period, before application of SC1–SC3. Samples were obtained from the column top, port 1 located at the top of the middle section, and port 2 located at the top of the bottom section and were stored at −80°C until DNA extraction. Because the subsequent

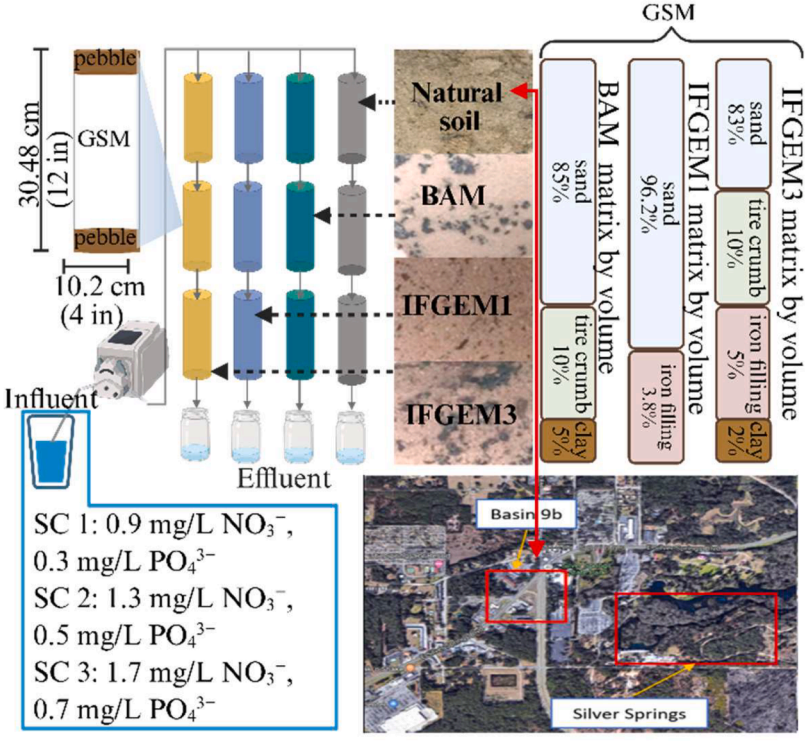



**Fig. 1.** Column experiment framework.

nutrient-loading conditions were each applied for only 3 h, these qPCR measurements were intended to characterize the pre-existing, depth-resolved genetic structure established during column conditioning rather than pulse-specific microbial responses.

Total DNA was extracted using the DNeasy PowerSoil Kit (Qiagen) according to the manufacturer's protocol. The kit workflow included cell disruption, inhibitor removal, and silica-membrane DNA purification; no separate post-extraction cleanup procedure was documented in the recoverable experimental records. DNA-level abundances were evaluated for ammonia-oxidizing bacteria (AOB), complete ammonia oxidation (comammox), nitrite oxidizing bacteria (NOB), denitrifiers (nirS), dissimilatory nitrate reducers to ammonium (nrfA), anaerobic ammonium oxidation (anammox), and iron-reducing bacteria (Gmet0909).

SYBR Green qPCR was performed in 20-$\mu$L reactions containing 10 $\mu$L of SYBR Green master mix, 0.8 $\mu$L each of the forward and reverse primers, 5 $\mu$L of DNA template, and 3.4 $\mu$L of nuclease-free water. The target-specific primer sequences, annealing temperatures, cycle numbers, and complete thermal programs are provided in Table S6. Quantification standards were used in the original assays, as reported by Valencia et al. (2020).

Neither the previously published workflow nor the recoverable project records document the use of a universal bacterial 16S rRNA gene as a normalization reference. The 16S rRNA primer sets listed in Table S6 targeted specific microbial groups, including anammox bacteria and Nitrospira, and should not be interpreted as universal reference-gene assays. Target-gene abundances were reported as copies $g^{-1}$ dry medium rather than as target-to-16S ratios.

The original instrument files and calibration worksheets required to recover assay-specific standard-curve slopes, intercepts, $R^2$values, amplification efficiencies, linear dynamic ranges, and detection limits were not available in the archived project records. These quantities were therefore not reconstructed or replaced with values from the primer-source publications because amplification performance depends on the standard material, reaction chemistry, instrument, template matrix, and inhibition conditions. Although all results were reported as copies g dry medium, absolute copy-number values obtained from different target-specific assays were not treated as directly comparable. For example, abundance was not quantitatively ranked against AOB, comammox, NOB, Anammox, or Gmet0909. Accordingly, the qPCR profiles are used in the present study only for descriptive, within-target comparisons among media and column depths. They are not used to quantitatively rank different gene-target assays or to infer gene expression, microbial activity, or process rates during SC1–SC3.

### 2.2. DON isolation and FT-ICR MS measurement

Individual influent and effluent water samples of 500 mL were collected at the completion of each of the three influent-condition runs and stored for subsequent molecular analysis. Prior to molecular characterization, samples were prepared using solid-phase extraction to reduce impurities and concentrate dissolved organic compounds. Briefly, each sample was passed through a Bond Elut™ PPL cartridge, 200 mg, 3 mL, and the retained organic matter was eluted from the sorbent bed using methanol, following the procedure described by Dittmar et al. (2008). The resulting extracts were preserved at −20 °C until further analysis. A detailed description of the solid-phase extraction procedure is provided in Supplementary Method S2. Sample-specific DOC or DON recoveries were not measured before and after SPE. Accordingly, the molecular data represents the operationally defined PPL-extractable, methanol-elutable fraction rather than the complete DON pool. Because PPL retention is dependent on molecular composition and sample matrix, highly polar or hydrophilic N-containing compounds may be underrepresented, and differential recovery between influent and effluent samples cannot be excluded (Raeke et al., 2016; Stucheli et al., 2018). Each sample state was represented by one 500-mL water sample, one PPL extract, and one accepted FT-ICR MS molecular profile. The archived dataset comprised three influent profiles and twelve effluent profiles, yielding 15 profiles in total. No independent column, SPE-extraction, or instrumental molecular replicates were available for a given medium–condition combination. Bulk water-quality triplicates were analyzed separately and were not treated as biological or technical replicates. The extracts were analyzed by negative-ion 9.4-T ESI FT-ICR MS. Peaks with signal magnitudes greater than four times the baseline root-mean-square noise at ($m/z$) 500 (S/N (>4)) were exported; molecular-formula assignments with absolute mass errors exceeding 0.5 ppm were discarded, and formula assignment was restricted to ($m/z$) 200–1200. Complete instrumental, acquisition, calibration, and formula-processing procedures are provided in Supplementary Method S2.

### 2.3. Molecular formula assignment and compositional analysis

The accepted molecular-formula inventory comprised CHO, CHON, and CHOS classes. Because CHO and CHOS formulas contain no nitrogen, formula-level analyses of N-containing organic matter were restricted to CHON, the only assigned N-bearing class in the dataset. Here, the CHON formula set is treated as an operationally defined molecular subset rather than a comprehensive representation of bulk DON. Specifically, it represents formulas recovered by the PPL-SPE procedure, detected by negative-ion ESI FT-ICR MS, and retained after application of the predefined peak-filtering and formula-assignment criteria. To reveal the intricate breakdown of DON, a suite of chemical descriptors (i.e. H/C, N/C, O/C, DBE, NOSC, (DBE-O)/C, AImod (Xu et al., 2022), Xc (Zhang et al., 2022), $MLB_L$ (Zhou et al., 2024) were employed and isolated according to each specific DON pool. Additionally, no CHONS formulas met the adopted assignment criteria. P-containing formulas were also not included in the original formula-assignment search space. Consequently, the present analysis does not characterize S- or P-containing DON. Candidate CHON formulas were retained within $H/C \leq 2.2$ and $O/C \leq 1.2$as formula-validation boundaries; these limits are analytical filters and should not be interpreted as defining the full range of naturally occurring DON. Accepted CHON formulas were assigned to seven van Krevelen molecular classes, with formulas outside the defined class boundaries grouped as Other (Lin et al., 2021); complete descriptor equations and molecular-class boundaries are provided in Supplementary Method S3. For each medium and influent-pulse scenario, paired formula occurrence was compared within the accepted CHON inventory after application of the same peak-processing and formula-assignment criteria to all samples. Three directly observed occurrence categories were defined: influent-only formulas, shared formulas, and effluent-only formulas. For continuity with established before–after FT-ICR MS terminology, these categories are referred to in subsequent analyses as operationally "removed," operationally "retained," and operationally "produced," respectively (Stubbins and Dittmar 2015; Riedel et al., 2016; Ward et al., 2017; Hu et al., 2023). All paired samples in the archived dataset were prepared using the same PPL-SPE procedure and analyzed under the same negative-ion 9.4-T ESI FT-ICR MS and formula-assignment workflow, thereby reducing avoidable procedural variability and supporting internally consistent paired comparisons. Nevertheless, standardized processing does not eliminate matrix-dependent extraction recovery, ionization competition, signal suppression or enhancement, and threshold-related non-detection. Accordingly, an influent-only formula cannot be interpreted as unequivocally removed from the water, nor can an effluent-only formula be interpreted as unequivocally formed de novo. Throughout this manuscript, the operational fate labels describe formula occurrence only within the defined PPL-SPE–ESI–FT-ICR MS analytical window and do not constitute direct evidence of absolute molecular removal, persistence, production, or a unique biochemical transformation mechanism in the bulk DON pool. Chemical-trait diversity was calculated using

assigned CHON formulas as the analytical entities and H/C, O/C, N/C, DBE, (DBE − O)/C, $AI_{mod}$, $X_c$, NOSC, and theoretical molecular mass as continuous traits. The complete trait-space construction, abundance weighting, diversity-index calculations, and redundancy assessment are provided in Supplementary Method S4.

### 2.4. Molecular putative formula-association network analysis

A Kendrick mass defect (KMD) analysis was used to provide an overview of the composition of the identified formulae in samples (Patel and Erickson, 1981; Kim et al., 2003; Merel, 2023). The KMD calculations are shown in Eqs. (1) and (2).

$$\text{Kendrick mass (F)} = \text{(observed mass)} \times \frac{\text{nominal mass (F)}}{\text{exact mass(F)}} \quad (1)$$

$$\text{KMD} = \text{nominal mass} - \text{Kendrick mass(F)} \quad (2)$$

Based on KMD's definition, organic materials sharing identical KMD values exhibit homology and only vary in molecular weight, with Δm representing integral multiples of the chosen mass of the functional group. Formula pairs were screened against eight predefined exact mass-difference categories. A pair was retained as a candidate formula association when its elemental and exact-mass difference matched one of the predefined categories within the adopted tolerance. The associated terms—such as methylation/demethylation, hydrogenation/dehydrogenation, and carbonylation/decarbonylation—are used only as conventional shorthand for net elemental differences and do not establish that a specific biochemical or enzymatic process occurred. Candidate formula associations were subsequently projected onto a PathDetect-SOM descriptor-space scaffold for comparative network visualization. The SOM did not generate or validate reaction assignments, and the resulting committor values were interpreted as graph-based relative accessibility rather than empirical reaction probabilities, rates, or transformation fluxes. The complete mass-difference library, matching tolerances, SOM construction, network aggregation, and graph calculations are provided in Supplementary Method S5 and Table S7.

## 3. Results

### 3.1. Engineered sorption media couple high inorganic-N removal with systematic restructuring of DON

Under the highest spiking condition (SC3), nitrogen removal differed strongly among sorption media (Fig. 2a). Both IFGEM formulations achieved near-complete NOx removal and the highest TN removal, whereas BAM and natural soil showed lower TN/NOx attenuation. In contrast to the consistently strong NOx removal, ammonium ($NH_4^+$) and DON removal were more formulation-sensitive, with IFGEM3 showing the strongest $NH_4^+$ removal and engineered media generally yielding higher DON removal than natural soil (Fig. 2a). Across all media and loading conditions, ΔDONranged from 169.3 to 658.7 $\mu g\ L^{-1}$,

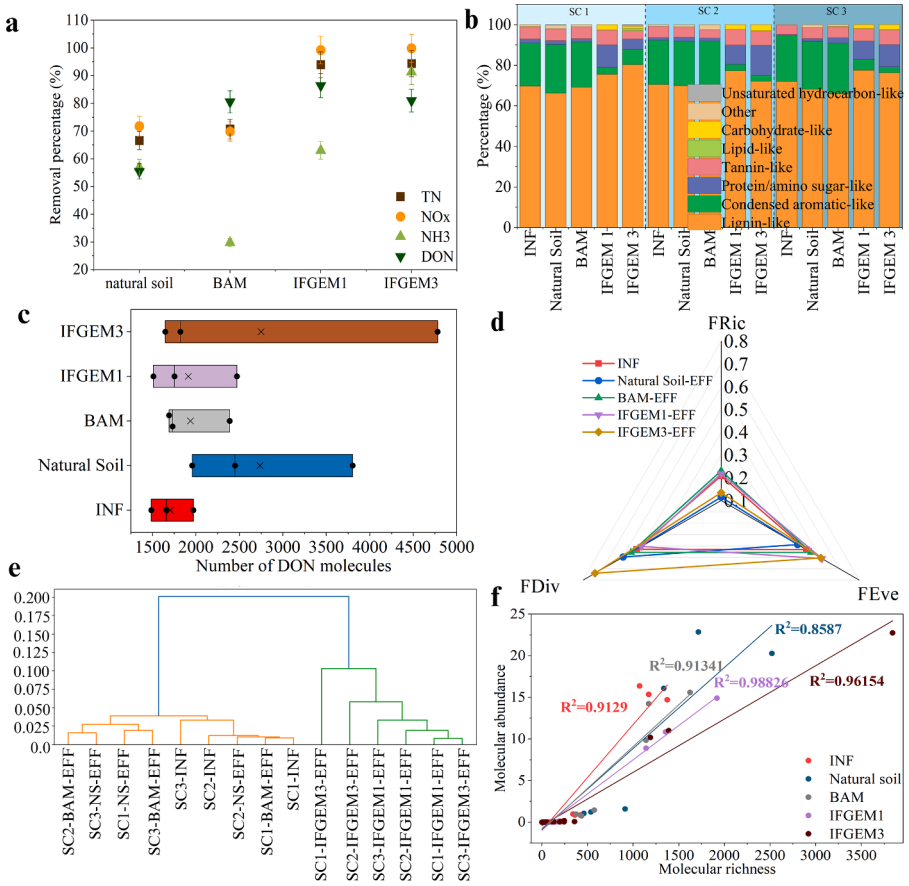



**Fig. 2.** Treatment-driven shifts in nitrogen removal and dissolved organic nitrogen (DON) molecular characteristics. (a) Removal percentage of TN, NOx, $NH_3$ and DON under SC3 (SC1–SC2 shown in Fig. S1). (b) Formula-count-based relative distribution (%) of DON molecular classes (lignin-like, condensed aromatic-like, protein/amino sugar-like, tannin-like, lipid-like, carbohydrate-like, other, and unsaturated hydrocarbon-like) in INF and EFF samples across SC1–SC3. (c) DON molecular richness (number of assigned DON molecules) in INF and EFF samples across treatments. (d) Chemical-trait diversity indices (FRic, FEve, FDiv) for INF and EFF under SC1 (SC2–SC3 in Fig. S3). (e) Dendrogram based on Bray–Curtis distances computed from the DON class-percentage matrix shown in c (ordination of the same matrix in Fig. S2). (f) Relationship between molecular richness and molecular abundance across treatments; lines indicate linear fits with $R^2$ values.

corresponding to reductions of 24.8–97.1%. All replicate-based 95% confidence intervals excluded zero (Table S4), indicating that the combined variability associated with the difference-derived DON calculation did not alter the direction of the observed influent-to-effluent decreases. Removal performance under the lower loading conditions (SC1–SC2) is provided in Fig. S1. Because bulk removal alone cannot reveal whether treatment merely reduces nitrogen concentrations or also alters the chemical identity of organic nitrogen, we next interrogated DON composition at the molecular-class level.

Across SC1–SC3, the influent molecular profiles were dominated by lignin-like formulas, which accounted for approximately 69.7–72.1% of the assigned CHON formula pool, followed by condensed aromatic-like formulas at approximately 21.4–22.9% (Fig. 2b). Natural-soil and BAM effluents retained broadly similar relative class distributions, with condensed aromatic-like formulas accounting for approximately 21.9–25.1%. (Table S8) In comparison, IFGEM effluents showed lower relative contributions of condensed aromatic-like formulas (approximately 2.8–7.5%) and higher contributions of formulas occupying the protein/amino sugar-like (approximately 5.1–14.9%) and carbohydrate-like (approximately 1.1–3.0%) regions (Table S8). Across the 12 effluent profiles, DON removal efficiency was descriptively associated with a higher protein/amino sugar-like formula-count proportion ($\rho$=0.867) and a lower condensed aromatic-like proportion ($\rho$=−0.657), whereas formula richness showed no positive relationship with removal ($\rho$=−0.559) (Table S9 and S10). These coefficients describe cross-profile tendencies rather than independent treatment effects. Because SC1–SC3 were applied sequentially to the same set of four columns, the three profiles within each medium represent repeated observations from a single column rather than independent experimental units; accordingly, the correlations are interpreted descriptively and not as inferential or causal evidence. Importantly, the class percentages are based on formula counts and cannot be converted into class-specific DON concentrations. The increased relative representation of protein/amino sugar-like formulas therefore occurred within a smaller residual DON pool and does not, by itself, establish absolute production of this molecular class. These directionally consistent patterns across the three tested loading conditions indicate that IFGEM effluents occupied a different relative class-composition space within the PPL-extractable, negative-ESI-detectable CHON fraction. However, because these values are based on formula-count proportions, the higher relative contribution of protein/amino sugar-like formulas cannot be attributed uniquely to de novo formation or an increase in their absolute concentration; it may also reflect preferential persistence of these formulas or selective loss of condensed aromatic-like and other formulas from the detected pool. Subsequent analyses therefore evaluated whether these class-level patterns were accompanied by changes in chemical-trait diversity and molecular-descriptor space.

Consistent with compositional reassembly, DON molecular richness (total number of assigned DON molecules) increased from influent to effluent, with treatment-dependent differences. IFGEM3 media mix exhibits the broadest richness range and the highest observed richness (Fig. 2c). To test whether treatment effects were systematic rather than condition-specific, we quantified dissimilarity using Bray–Curtis distances computed from the DON class-percentage matrix and evaluated sample grouping and dissimilarity. The dendrogram separated samples into coherent branches, distinguishing an IFGEM-associated group from a natural soil/BAM/influent-associated group across spiking conditions (Fig. 2e); a Non-metric Multidimensional Scaling (NMDS) visualization based on the same input matrix is shown in Fig. S2. Having established that IFGEM produces a consistent DON compositional regime, we next assessed how these shifts translate into changes in the structure of DON "trait space."

Chemical-trait diversity was summarized in the main text using FRic, FEve, and FDiv, which respectively describe the breadth of occupied formula-trait space, the evenness of the relative-signal distribution within that space, and the relative weighting of formulas toward its margins (Fig. 2d). These indices were calculated from H/C, O/C, N/C, DBE, (DBE − O)/C, $AI_{mod}$, $X_c$, NOSC, and theoretical molecular mass. Under SC1, the observed FRic values were lower for natural soil and IFGEM3 than for the influent, BAM, and IFGEM1 profiles, whereas IFGEM1 and IFGEM3 showed higher FEve values. IFGEM3 exhibited the highest observed FDiv (0.740), indicating that formulas with greater relative ion-signal weights were distributed closer to the margins of its profile-specific chemical-trait space. Corresponding FRic, FEve, and FDiv patterns under SC2 and SC3 are shown in Fig. S3. FDis and RaoQ are reported in Table S11 but are not interpreted separately because they showed substantial redundancy. The detailed calculation of chemical-trait can be in the supplementary method S4. These indices characterize relative organization within the detected CHON formula pool and do not independently establish absolute molecular abundance, biological function, or a specific DON transformation mechanism.

The 9.4-T FT-ICR-MS was used in this study to help realize molecular abundance of DON by measuring the cyclotron frequency of ions, which correlates directly to their mass-to-charge ratio ($m/z$), while the intensity of the induced image current in the detection electrodes is proportional to the ion abundance. It is indicative that across the four sorption media in this study, molecular abundance scaled linearly with molecular richness ($R^2$ = 0.86–0.99, Fig. 2f), providing an integrative constraint linking treatment-induced diversification to overall molecular signals. Comparison between reproducible DON molecules that are consistently produced by microbial processes versus refractory or recalcitrant DON molecules that are resistant to degradation can help realize the structure changes in DON family. Together, Fig. 2a–f shows that IFGEM media mixes simultaneously maximize inorganic-N removal under high loading and produce reproducible molecules with functional reconfiguration of DON.

### 3.2. *IFGEM restructures the dominant lignin-like DON pool across mass–abundance and descriptor space*

Because lignin-like molecules remained the dominant DON class across influent and effluent waters in Fig. 2, we next asked whether this apparent class-level was persistently masked deeper within class reorganization after filtration through different sorption media. Two-dimensional kernel density maps of relative abundance versus mass ($m/z$) showed that lignin-like molecules in both natural soil and IFGEM3 were concentrated within a shared low-relative-abundance in the mid-to-high-mass domain across SC1–SC3 (Fig. 3a–c). Because FT-ICR MS spectra plot signal intensity against accurate mass-to-charge ratio and can resolve large numbers of molecular formulas in complex DOM mixtures (Hur et al., 2010; Koch et al., 2007), these maps provide a high-resolution view of how the lignin-like DON population is distributed in mass–intensity space (He et al., 2020; Qi et al., 2022). Here, the x-axis represents the accurate mass-to-charge ratio ($m/z$) of assigned lignin-like formulas, and the y-axis represents their normalized relative abundance (relative peak intensity within each sample rather than absolute concentration). The color scale indicates the two-dimensional kernel density estimated from the ($m/z$, relative abundance) point cloud, such that hotspots identify regions where lignin-like formulas are most densely distributed in mass–abundance space. The two-dimensional kernel-density maps were treated as descriptive visualizations, whereas the corresponding one-dimensional m/zdistributions were compared statistically between natural soil and IFGEM3. The lignin-like formula-mass distributions differed under SC1 ($D = 0.09112$, raw $P < 0.0001$), SC2 ($D = 0.08003$, raw $P = 2.23893 \times 10^{-4}$), and SC3 ($D = 0.10304$, raw $P < 0.0001$); all three comparisons remained significant after Holm adjustment ($P_{adj} < 0.0003$; Table S12). The nature of the difference varied among loading conditions. Under SC1, IFGEM3 showed a broader central mass distribution than natural soil, with an interquartile range of 256.03 compared with 196.01 $m/z$, whereas their median masses were similar (552.13 and 547.19 $m/z$,

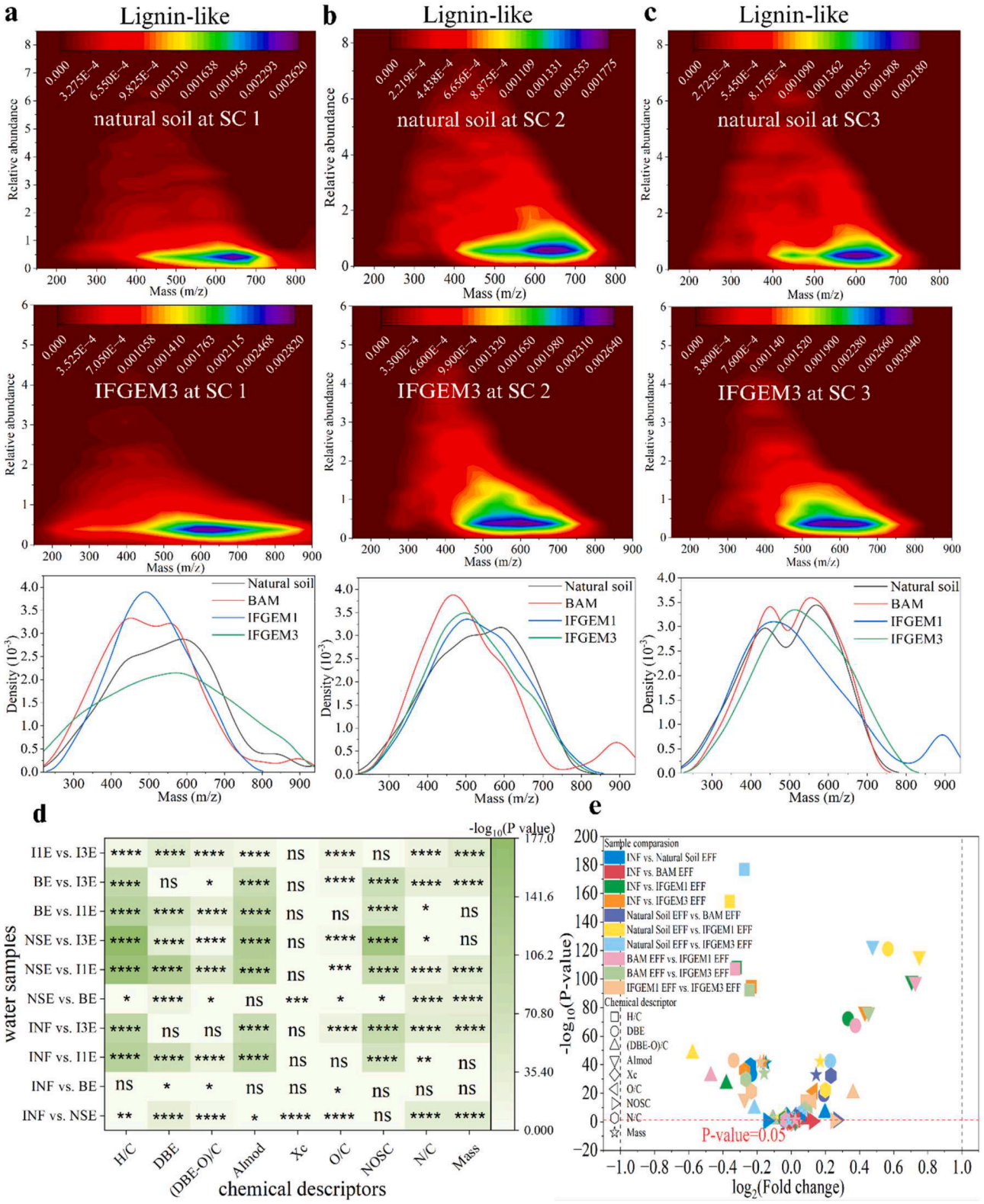



**Fig. 3.** IFGEM restructures the dominant lignin-like DON pool across mass–abundance and descriptor space. (a)–(c), Two-dimensional kernel density plots of relative abundance versus mass (*m/z*) for lignin-like molecules in natural soil effluent (top row) and IFGEM3 effluent (middle row) under SC1 (a), SC2 (b) and SC3 (c). The bottom row shows the corresponding one-dimensional mass-density distributions for natural soil, BAM, IFGEM1 and IFGEM3 under each spiking condition. The corresponding two-dimensional kernel density plots for BAM and IFGEM1 are shown in Fig. S4. (d) One-way ANOVA heat map of chemical descriptors across pairwise sample comparisons under SC1. Sample abbreviations: I1E, IFGEM1 effluent; I3E, IFGEM3 effluent; BE, BAM effluent; NSE, natural soil effluent; INF, influent. Shading indicates $-\log_{10}$(P value), and significance levels are indicated within cells. Corresponding analyses for SC2 and SC3 are shown in Fig. S5. (e) Volcano plot summarizing pairwise differences in chemical descriptors under SC1, plotted as $\log_2$(fold change) versus $-\log_{10}$(P value). Corresponding volcano plots for SC2 and SC3 are shown in Fig. S6.

respectively). Under SC2, IFGEM3 showed a lower median mass and a slightly narrower interquartile range, while under SC3 it showed a higher median mass but a comparable overall range. Thus, the principal differences involved condition-dependent changes in distribution shape and location rather than a consistently broader mass distribution in IFGEM3. When all four media were viewed together, BAM and IFGEM1 occupied intermediate descriptive positions in the one-dimensional mass-density curves (Fig. 3a–c, bottom row; corresponding kernel-density maps are shown in Fig. S4). Lignin-like formulas therefore remained numerically dominant across media, while the observed shapes of their formula-mass distributions varied among the available profiles. The mean mass values in Table S13 likewise did not shift monotonically, indicating that the observed differences were not adequately represented by the centroid alone.

This redistribution was accompanied by a chemically directional shift in descriptor space. Under SC1, IFGEM effluent showed higher H/C values than INF, natural soil, and BAM (1.179 in IFGEM1 and 1.111 in IFGEM3, versus 0.944 in INF, 0.918 in natural soil EFF, and 0.937 in BAM EFF; Table S13), together with markedly lower AImod (0.267 and 0.323, versus 0.435, 0.449, and 0.442, respectively), consistent with reduced aromaticity. The same IFGEM samples also exhibited lower or negative NOSC (−0.103 and −0.087, versus 0.132, 0.146, and 0.121) and sharply elevated $MLB_L$ (13.709 and 7.301, versus 2.030, 2.209, and 1.774), indicating a more reduced and more labile lignin-like assemblage after IFGEM treatment. This pattern persisted under SC2–SC3, where IFGEM effluents consistently maintained higher H/C (1.136–1.224 versus 0.906–0.947 in INF/natural soil/BAM), lower DBE (10.736–12.876 versus 14.264–16.845), lower AImod (0.231–0.303 versus 0.434–0.469), negative NOSC (−0.138 to −0.091 versus 0.095–0.169), and much higher MLBL (11.048–17.972 versus 0.202–2.717; Table S13). By contrast, mean Xc remained narrowly clustered at approximately 2.63–2.69 across types of media and spiking conditions, helping explain why Xc was among the least discriminating descriptors in the ANOVA (Analysis of variance) heat maps (Fig. 3d) in which the threshold of heatmap can be seen from the right edge of Fig. 3d Mean O/C also stayed within a relatively narrow range (0.422–0.477), suggesting that the principal IFGEM effect was not a simple oxygenation shift, but a coordinated change in saturation, aromaticity, redox state, and lability.

The statistical analyses in Fig. 3d–e support this interpretation. Under SC1, one-way ANOVA identified widespread pairwise differences across H/C, DBE, (DBE–O)/C, AImod, O/C, N/C, and mass, whereas Xc was frequently non-significant and NOSC differentiated only a subset of comparisons (Fig. 3d). However, the two IFGEM effluents were themselves significantly different across multiple descriptors, indicating that the formulations of the two types of sorption media did not generate an identical lignin-like molecular space. The volcano plot further showed that the largest descriptor effect sizes and strongest statistical separation occurred in comparisons involving IFGEM effluents, whereas contrasts among INF, natural soil, and BAM generally remained closer to the origin or exhibited smaller effects (Fig. 3e). Chemically, these IFGEM-related contrasts coincided with the broader class- and descriptor-level pattern of higher saturation, lower aromaticity and oxidation state, lower relative representation of condensed aromatic-like formulas, and higher representation of protein/amino sugar-like formulas. Their significance therefore reflects a coordinated change in the composition of the detectable CHON pool rather than statistical separation alone. These coordinated descriptor changes are consistent with a shift toward greater potential lability and bioavailability of the residual detectable DON fraction, but they do not directly demonstrate biodegradability or biological uptake. The corresponding ANOVA and volcano analyses under SC2 and SC3 preserved the same overall pattern of grouping, although the magnitude of individual contrasts varied with spiking conditions (Fig. S5 and S6). Taken together, Fig. 3 shows that the IFGEM signature first identified in Fig. 1 is rooted not only in altered DON class proportions, but also in a deeper reorganization of the dominant lignin-like pool toward a less aromatic, more reduced, and more labile molecular ensemble.

### 3.3. Fate–zone–class coupling reveals divergent DON processing regimes in natural soil versus IFGEM3

Figs. 2 and 3 in the above section collectively show that IFGEM-based green sorption media alter DON class composition and shift molecular descriptors toward a less aromatic, more reduced and more labile space. To identify how these differences arise inside each medium, we resolved DON transformation from influent (INF) to effluent (EFF) jointly by fate (removed, retained, produced), zonation (Zones 1–4), and DON class, with an enthalpy gauge as an additional thermodynamic axis (Fig. 4a; Supplementary Figs. S7–S8). Here, zonation refers to the classification of DON molecules into four regions in the NOSC–(DBE−O)/C space, which jointly describes molecular oxidation state and saturation level (LaRowe and Van Cappellen, 2011). Specifically, NOSC was used to indicate oxidation/reduction status, with positive values representing oxidized molecules and negative values representing reduced molecules, whereas (DBE−O)/C was used to indicate saturation/unsaturation, with positive values reflecting more unsaturated structures and negative values reflecting more saturated structures (Liu et al., 2025; Zhang et al., 2026). The detailed definition can be seen in the supplementary method 3.

Under SC1, natural soil routes the dominant lignin/condensed-aromatic backbone primarily into retention and local recycling, with a pronounced Zone-4 (oxidizing–saturated) imprint in the retained branch (Fig. 4a; Table S14). As loading increases, the natural-soil Sankey diagrams (SC2–SC3; Fig. S7; Table S14) show an increasingly retention-dominated routing, with most of the effluent mass continuing to be organized through Zone 4 and remaining lignin-rich; production becomes a secondary pathway and removal stays minor. In contrast, IFGEM3 exhibits a different internal routing architecture across all conditions (Fig. S8). Compared with natural soil, IFGEM3 sends a much larger fraction of influent molecules through active turnover (production and removal) and funnels conspicuous “produced” streams into reducing zones (Zones 2–3) that terminate in protein/amino sugar-like and carbohydrate-like classes in the effluent, while the “removed” stream is visually enriched in aromatic-rich pathways where notably those associated with condensed-aromatic structures. This fate–zone coupling provides a mechanistic explanation for why IFGEM3 effluents differ from natural soil even when lignin-like formulas remain abundant overall. Thus, Sankey routing indicates two distinct regimes: natural soil increasingly preserves a Zone-4, lignin-dominated framework with rising loading, whereas IFGEM3 sustains strong reducing-zone turnover that generates N-containing products while selectively removing aromatic-rich components.

If IFGEM3 preferentially redirects molecules into reducing zones and produces N-containing classes, the resulting effluent should show systematic shifts in bond-type descriptors within major classes. However, the density curves of DBE/C and (C − 1−DBE)/C may further help explain the structural origin of the changes in (DBE−O)/C. Whereas C − 1−DBE)/C represents carbon chain length or mainly C–C single-bond structures, DBE/C represents double-bond structures (Hu et al., 2025). Under SC1, density curves show that, relative to natural soil, IFGEM3 shifts lignin-like and protein/amino sugar-like molecules toward lower DBE/C (fewer C═C and aromatic double bonds), higher (C − 1−DBE)/C (greater C–C single-bond contribution), and lower AImod/C (lower aromatic-ring character) (Fig. 4b–d). The same direction of separation persists under SC2 and SC3 (Fig. S9), consistent with the Sankey observation that IFGEM3 routes more molecules into reducing-zone transformation pathways rather than simple oxidizing-zone retention.

The descriptor densities therefore link the fate–zone patterns to concrete structural reassembly: IFGEM3 produces an effluent pool that is less aromatic and more single-bond enriched than natural soil, in line with a stronger Zone 2/Zone 3 imprint.

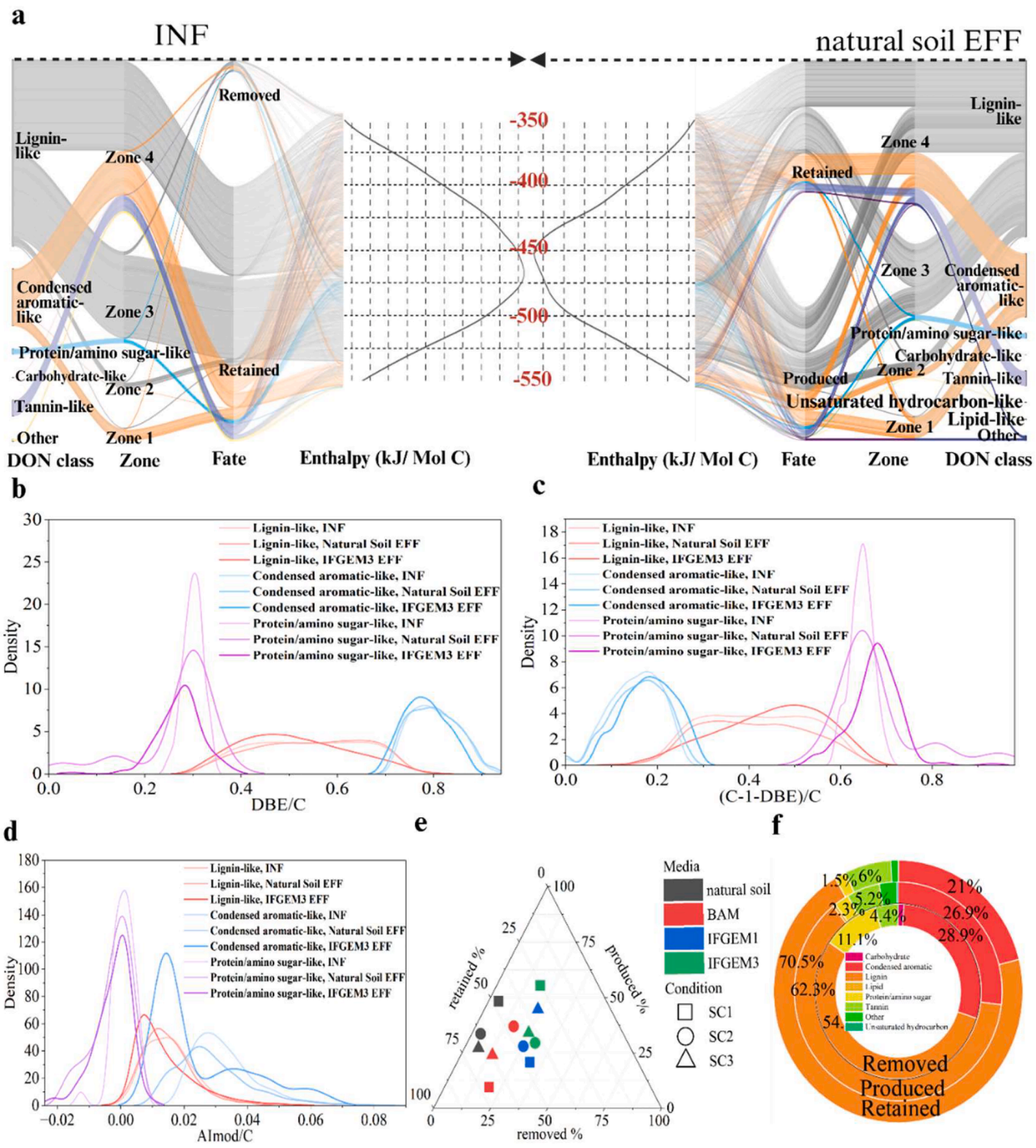



**Fig. 4.** Fate–zone–class coupling distinguishes DON transformation in natural soil and engineered media. (a) Sankey diagram illustrating transformation of DON classes from INF to natural soil EFF under SC1, resolved by zonation (Zones 1–4), fate (removed, retained, produced), and enthalpy gauge (kJ mol C). Natural soil under SC2 and SC3 is shown in Fig. S7, and IFGEM3 under SC1–SC3 is shown in Fig. S8. b–d. (b) Density curves under SC1 for DBE/C, (c) Density curves under SC1 (C − 1−DBE)/C. (d) Density curves under SC1 AImod/C for lignin-like, condensed aromatic-like, and protein/amino sugar-like molecules in INF, natural soil EFF, and IFGEM3 EFF (SC2–SC3 in Supplementary Fig. S9). (e) Ternary plot summarizing the fraction of molecules assigned to removed, retained, and produced pools across media and conditions. (f) DON class composition of removed, produced, and retained pools for natural soil under SC1.

To generalize fate behavior beyond individual Sankey diagrams, we summarized the fraction of molecules assigned to removed, produced, and retained pools across media and spiking conditions. In the ternary space (Fig. 4e), natural soil shifts toward the retention vertex with increasing loading (retained 47.3% → 62.6% → 66.7% from SC1→SC2→SC3; produced 48.4% → 33.6% → 27.2%; removed remains low at 3.8–6.1%). IFGEM3 occupies a contrasting region characterized by sustained turnover: under SC1 it is strongly production-oriented (produced 55.4%, retained 25.4%, removed 19.2%), and under SC2–SC3 it remains balanced but still turnover-rich (removed 29.8%/24.7% and produced 29.5%/34.0%, with retention only ~41%). Other media fall between these endmembers, reinforcing that engineered media—especially IFGEM3—shift DON processing away from passive retention and toward active transformation.

Therefore, the ternary plot consolidates the Sankey message, with natural soil behaving increasingly as a retention system under higher loading, whereas IFGEM3 maintaining high DON turnover across conditions.

Finally, we asked which DON classes dominate each fate pool, and how that class–fate coupling changes with loading. For natural soil under SC1, the fate-resolved doughnut plot (Fig. 4f; inner=removed, mid=produced, outer=retained) shows that the retained pool is most lignin-enriched (70.5% lignin-like), while the produced pool remains lignin-dominated (62.3%) but carries a higher condensed aromatic fraction (26.9%); the small removed pool is relatively enriched in non-lignin components, including protein/amino sugar-like molecules (11.1%) and condensed aromatics (28.9%). Extending across conditions (Fig. S10), natural soil retains a stable pattern in which retention remains lignin-dominated (~71–72% lignin-like; ~21–23% condensed aromatics), whereas the removed pool shifts with loading (e.g., condensed aromatics rise to 39.4% in the removed fraction under SC2, while SC3 removed becomes more lignin-heavy at 70.1% lignin-like), indicating that removal remains small but compositionally sensitive to spiking.

In sharp contrast, IFGEM3 exhibits a pronounced class selectivity by fate (Fig. S11). Across SC1–SC3, the retained fraction is overwhelmingly lignin-like (85.4–89.9%), consistent with a lignin backbone that is preferentially preserved even under high turnover. Meanwhile, the removed fraction is strongly enriched in condensed aromatic-like molecules (e.g., 71.7% under SC1; 45.2% under SC2; 58.8% under SC3), supporting the interpretation that IFGEM3 preferentially strips aromatic-rich components from the flow path. Critically, the produced pool in IFGEM3 becomes markedly enriched in protein/amino sugar-like DON under higher loading (e.g., 32.4% under SC2 and 23.7% under SC3), consistent with the Sankey routing of produced streams through reducing zones and into N-containing classes.

Fate-specific class patterns close the a–f evidence chain: natural soil primarily retains a lignin-rich pool (Zone-4 imprint) with limited, condition-sensitive removal, whereas IFGEM3 media mix couples strong turnover to selective condensed-aromatic removal and enhanced production of protein/amino sugar-like DON, providing a mechanistic basis for the effluent signatures observed in Figs. 2 and 3.

### 3.4. *Reaction fingerprints link fate–zone partitioning to selective DON turnover*

Building on the fate–zone–class framework (Fig. 4), we next asked which putative biochemical transformations most plausibly connect molecules assigned to the removed pool with those newly appearing in the produced pool, and whether these transformations are consistent with the observed fate-dependent class enrichment and thermodynamic zonation.

Across all sorption media and spiking conditions, hydrogenation/dehydrogenation overwhelmingly dominated the inferred removed-to-produced linkages (Supplementary Fig. S12). In natural soil, hydrogenation/dehydrogenation accounted for ~77–83% of candidate transformations (SC1–SC3), similar to BAM (~80–86%), but was even more pronounced in engineered sorption media (IFGEM1: ~89–91%; IFGEM3: ~89–93%). All other reaction categories contributed only minor fractions individually. It is indicative that oxidation/reduction and methylation/demethylation were typically at the few-percent level (generally ≤7% in natural soil/BAM and ≤4% in IFGEMs), while carbonylation/decarbonylation and carboxylation/decarboxylation together comprised a small residual share; aromatization/dearomatization was consistently negligible (≤~1%). This quantitative ranking shows that changes in hydrogen saturation state are the primary driver of molecular turnover between removed and produced pools, with heteroatom functionalization acting as a secondary, condition- and media-dependent modifier.

At the molecular-class level, the dominant putative reaction types preferentially reshuffled the lignin-like and condensed aromatic-like fractions. In natural soil under SC1 (Fig. 5a), hydrogenation/dehydrogenation and carbonylation/decarbonylation generated the bulk of cross-class connections between condensed aromatic-like and lignin-like compounds, whereas oxidation/reduction largely preserved class identity while still contributing to within-class turnover. The same qualitative routing—dominance of lignin-like/condensed aromatic-like interconversion with relatively minor participation of other classes—persisted under SC2 and SC3 (Supplementary Figs. S13–S14). In contrast, IFGEM3 exhibited a clearer routing into a protein/amino sugar-like produced pool, most evident under higher spiking (SC2–SC3), where carbonylation/decarbonylation and hydrogenation/dehydrogenation supported substantial conversion from lignin-like/condensed aromatic-like precursors into protein/amino sugar-like products (Supplementary Figs. S15–S17). Together, these reaction maps suggest that engineered sorption media do not merely “retain” influent DON, but actively restructure it along distinct biochemical channels, particularly by generating N-rich, protein/amino sugar-like products from aromatic/lignin-like precursors.

These putative pathway-level inferences are mirrored by the fate-specific class compositions. Lignin-like compounds remained the dominant class across fates and media (Fig. 5b), but were especially enriched in the retained pool — ~70% in natural soil and BAM versus ~85–90% in IFGEM1/IFGEM3 across conditions—consistent with lignin-like DON forming a relatively persistent background. In contrast, the produced pool diverged sharply in its N-rich component (Fig. 5c). Protein/amino sugar-like molecules stayed low in natural soil and BAM (generally a few percent, with only modest increases under some conditions), whereas IFGEM media showed strong enrichment specifically in the produced pool, reaching ~15–29% in IFGEM1 and up to ~24–33% in IFGEM3 (most prominently under SC2–SC3). Complementarily, condensed aromatic-like compounds were preferentially concentrated in the removed pool in engineered sorption media (Supplementary Fig. S13e), while natural soil and BAM exhibited more balanced condensed aromatic fractions across removed/produced/retained pools. This fate-dependent class partitioning therefore supports a mechanistic picture in which engineered sorption media selectively capture condensed aromatics while producing protein/amino sugar-like DON, rather than producing a compositionally neutral turnover.

Finally, mapping molecules into NOSC–(DBE−O)/C space connects these transformations to shifts in redox and saturation state (Fig. 5d–e; Supplementary Figs. S13f and S18). In natural soil, the retained pool was consistently dominated by Zone 4 (oxidizing–saturated), increasing from 58.51% (SC1) to 61.33% (SC2) and 65.15% (SC3), with a stable secondary contribution from Zone 3 (reducing–saturated; 28.43–33.99%). In contrast, the produced pool in natural soil shifted across influent regimes: it was Zone-4–biased under SC1 (50.34%) but became Zone-3–enriched under SC2–SC3 (42.52% and 41.98%), alongside increased representation of unsaturated zones (Zones 1 and 2 rising to 31.91% under SC3). The removed pool remained split between Zones 3 and 4 (e.g., SC1: 41.11% vs 35.56%; SC3: 41.24% vs ~46.40%), consistent with partial turnover spanning both reducing- and oxidizing-saturated domains. This zonation pattern is consistent with the reaction-level analysis in Fig. 5a–e, where removed-to-produced turnover in natural soil under SC1 is resolved largely within the lignin-like and condensed aromatic-like backbone through hydrogenation/dehydrogenation, methylation/demethylation, oxidation/reduction, carbonylation/decarbonylation and hydration/dehydration links, thereby providing a reaction-level basis for the mixed Zone 3/Zone 4 distribution observed here.

IFGEM3 exhibited a markedly stronger fate–zone decoupling (Fig. S18; Table S13). The removed pool was overwhelmingly Zone-4 enriched (72.02–78.07%), with an additional Zone-1 contribution (11.22–17.55%), indicating preferential removal of formulas in oxidizing space. Conversely, the produced pool was concentrated in Zone 3, increasing from 50.63% (SC1; with 12.76% in Zone 2) to >81%

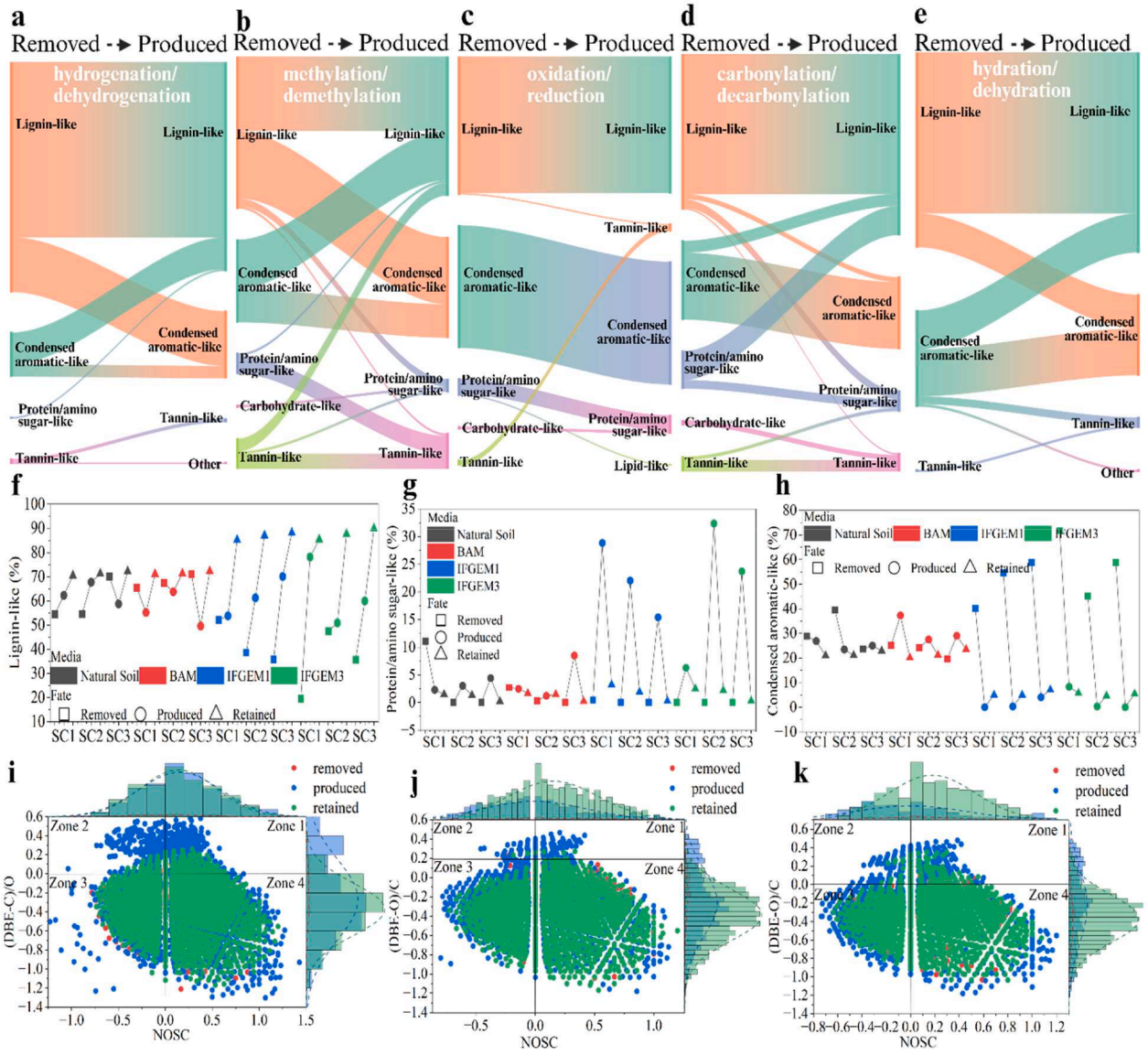



**Fig. 5.** Quantitative analysis of putative bioreaction reactions in the DON transformation within natural soil (from removed to produced) under SC1. (a) hydrogenation/dehydrogenation, (b) methylation/demethylation, (c) oxidation/reduction, (d) carbonylation/decarbonylation and (e) hydration/dehydration. (f) the percentage of lignin-like compounds, (g) the percentage of protein/amino sugar-like compounds and (h) the percentage of condensed aromatic-like compounds in the removed, produced, retained pool of natural soil, BAM, IFGEM1 and IFGEM3. (i) the zonation of removed, produced and retained of natural soil under SC1. (j) the zonation of removed, produced and retained of natural soil under SC2. (k) the zonation of removed, produced and retained of natural soil under SC2.

Zone 3 dominance under SC2–SC3 (82.37% and 81.51%). The retained pool in IFGEM3 consistently occupied both Zone 4 (52.36–57.01%) and Zone 3 (41.18–45.86%), suggesting that—even within the persistent fraction—IFGEM3 maintains a larger reducing-saturated component than natural soil. These fate-dependent redistributions in thermodynamic space align with the reaction fingerprint in Supplementary Fig. S12 (dominance of hydrogenation/dehydrogenation) and provide a baseline ecological context linking fate assignment (Fig. 3) to systematic, media-specific restructuring of DON.

## 4. Discussion

### *4.1. Contrasting DON molecular patterns and baseline microbial genetic contexts in IFGEM3 and natural soil*

To move beyond class percentages and fate partitioning and identify how DON formulas were internally reorganized within each medium, we compared the Kendrick mass defect (KMD)-based Path Detect-DON putative formula-association networks of natural soil and IFGEM3 under SC1 (Fig. 6a, b). The class clustering in Fig. 6c provides the descriptor-space scaffold needed to interpret these putative formula-association networks. The annotated formulas resolve into chemically coherent SOM neighborhoods. The formula-association edges connect operationally defined influent-only and effluent-only formulas through selected exact-mass differences and the SOM provides the chemical state-space scaffold onto which those edges are projected. It shows that putative formula-association networks were dominated by lignin-like and condensed aromatic-like molecules, with smaller but distinct protein/amino sugar-like, carbohydrate-like, tannin-like, lipid-like, and other modules, so that the links in Fig. 6a, b identifying putative mass-difference relationships within or among molecular neighborhoods rather than as isolated pairwise mass differences. The SOM lets you see whether the inferred links stay within the same chemical neighborhood, move between adjacent related neighborhoods, or jump across distinct class domains indicating possible pathways. Against that framework, the natural-soil network in Fig. 6a appears comparatively conservative.

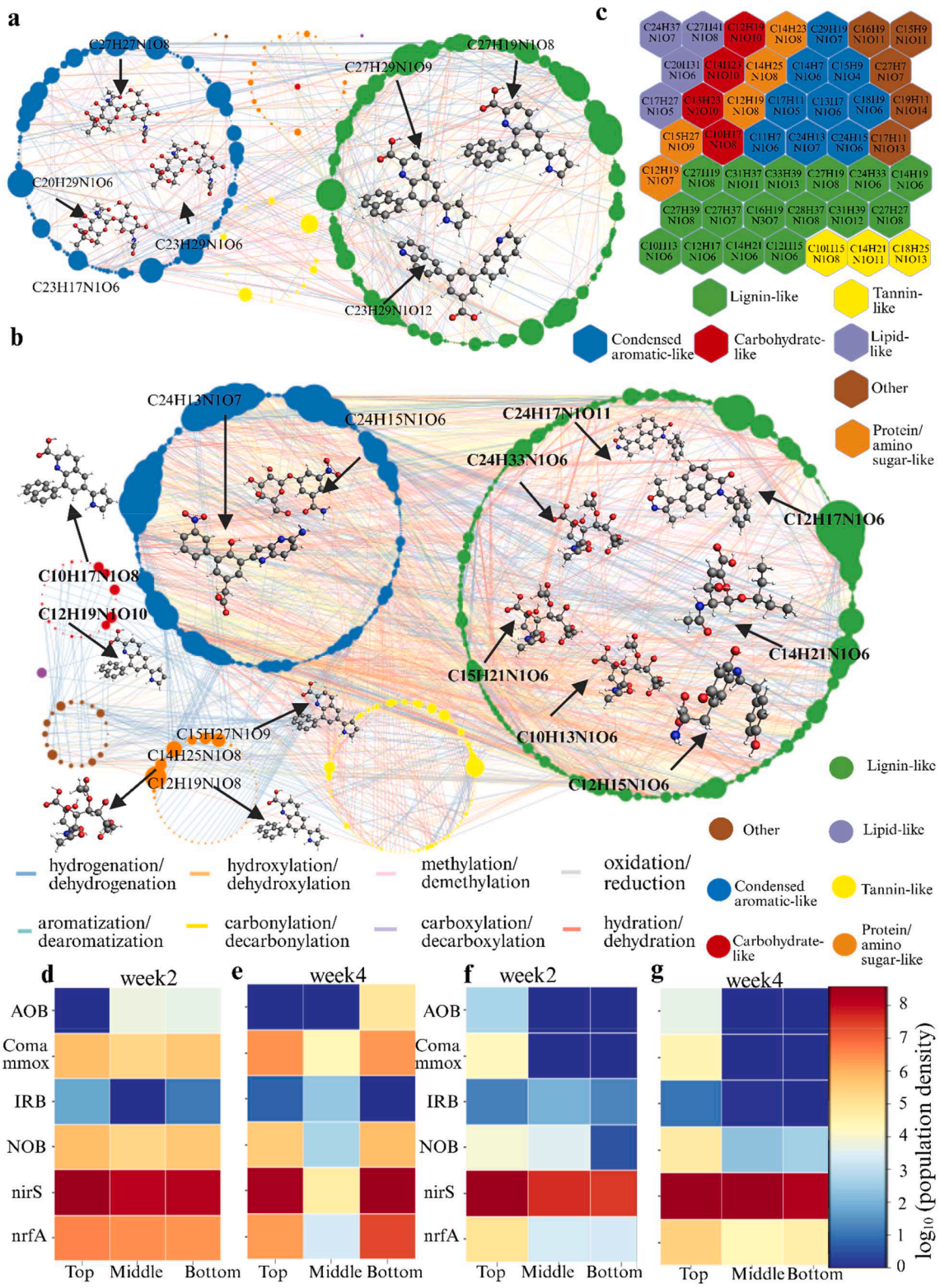



**Fig. 6.** Network-resolved DON transformation and baseline nitrogen-cycling stratification distinguish IFGEM3 from natural soil. a Putative formula-association networks within natural soil under SC1, b Putative formula-association network within IFGEM3 under SC1. c Self-organizing map used as the descriptor-space scaffold for KMD-based Path Detect-DON analysis. (lignin-like (green), other (brown), condensed aromatic-likes (blue), carbohydrate-likes (red), protein/amino sugar-likes (orange), lipid-likes (purple), and tannin-likes (yellow)). d the heatmap of genes in week 2 of incubation (before any spiking) for natural soil. e the heatmap of genes in week 4 of incubation (before any spiking) for natural soil. f the heatmap of genes in week 2 of incubation (before any spiking) for IFGEM3. g the heatmap of genes in week 4 of incubation (before any spiking) for IFGEM3. Panels d–g show pre-spiking, depth-resolved DNA-level abundances of selected nitrogen-cycling and iron-reduction-associated gene markers during weeks 2 and 4 of stormwater conditioning. These profiles characterize baseline microbial genetic potential and do not represent gene expression or microbial activity during the subsequent 3-h SC1–SC3 nutrient pulses. Because each target was quantified using a distinct primer and calibration assay, values should be interpreted within each target rather than quantitatively compared across different gene targets.

Most nodes and reaction paths remain concentrated within the two dominant condensed aromatic-like and lignin-like neighborhoods, and the smaller peripheral classes contribute only limited connectivity. By contrast, the IFGEM3 network in Fig. 6b is more compositionally distributed, with clearer linkages extending from the dominant lignin-like and condensed aromatic-like cores into protein/amino sugar-like, tannin-like, and carbohydrate-like neighborhoods. Taken together, these SC1 putative formula-association networks indicate that natural soil showed formula associations concentrated primarily within an existing aromatic/lignin-rich backbone, whereas IFGEM3 supports a broader and more cross-class reorganization of DON.

This contrast is not confined to a single influent regime. Under SC2 and SC3, the supplementary natural-soil putative formula-association networks (Figs. S19 and S20) remain visually anchored by the same large lignin-like and condensed aromatic-like clusters, with only modest peripheral participation from tannin-like or protein/amino sugar-like molecules. In other words, increasing nutrient loading changes the internal density of the network but does not fundamentally change its architecture. Natural soil continues to process DON largely within a restricted aromatic-dominated reaction space. However, the IFGEM3 putative formula-association networks behave differently under SC2 and SC3 (Figs. S21 and S22). Although lignin-like and condensed aromatic-like clusters again remain the largest communities, the surrounding protein/amino sugar-like and tannin-like modules are larger and more integrated than in natural soil, and the overall network retains a more distributed topology across classes. Thus, the difference between natural soil and IFGEM3 is not simply a condition-specific response to SC1, but a recurring medium-associated formula-association pattern across the three tested conditions: natural soil remains comparatively conservative, whereas IFGEM3 repeatedly showed broader cross-class formula associations.

To place the media-associated DON molecular patterns in the context of the microbial communities established during column conditioning, we examined the depth-resolved qPCR profiles collected at weeks 2 and 4 before nutrient spiking (Fig. 6d–g; Valencia et al., 2020). In natural soil, the nirS was broadly distributed across the three column depths at both time points, while AOB, comammox, NOB, nrfA, and the Gmet0909 iron-reduction-associated marker exhibited target-specific depth variation. By week 4, several markers showed relatively stronger lower-column signals within their respective assays, indicating increased vertical differentiation of the baseline gene-marker profiles.

In IFGEM3, AOB and comammox showed relatively stronger upper-section signals within their respective assays, whereas nirS was distributed across the three depths and nrfA and Gmet0909 exhibited additional depth-dependent patterns (Fig. 6f and 6g). The corresponding BAM and IFGEM1 profiles are provided in Fig. S23. Collectively, these observations indicate that the four media developed different spatial distributions of nitrogen-cycling and iron-reduction-associated gene markers during the four-week stormwater-conditioning period. Natural soil showed broader depth-wise co-occurrence of several marker genes, whereas IFGEM3 showed greater depth differentiation for selected targets.

Viewed together, Fig. 6 juxtaposes two distinct evidence streams: KMD-derived formula-association networks from water samples collected following the nutrient pulses and depth-resolved qPCR profiles obtained during the pre-spiking conditioning period. Natural soil showed formula associations concentrated mainly within lignin-like and condensed aromatic-like regions, whereas IFGEM3 showed broader cross-class associations; the baseline qPCR profiles additionally showed different depth-wise distributions of selected gene markers between the two media. Their side-by-side presentation provides hypothesis-generating ecological context only. Because the datasets were collected at different experimental stages and do not share matched observational units, the qPCR profiles do not establish gene expression, process activity, in-column redox stratification, statistical agreement, or causal linkage with the KMD-derived patterns. Detailed analytical limitations are discussed in Section 4.4.

### 4.2. Increasing N and P loading shifts DON fingerprints and putative formula-association

Building on Fig. 6, which established that natural soil and IFGEM3 followed contrasting DON transformation regimes under SC1, we next examined whether these regimes remained stable or were further reorganized as influent N and P loading increased from SC1 to SC3 simultaneously instead of single spiking condition. Fig. 7a–d shows that loading effects were not uniform across media, but instead reshaped DON fingerprints in a medium-specific manner. After Holm adjustment, the strongest differentiation in natural soil occurred between SC1 and SC2, where H/C, DBE, NOSC, N/C, and mass differed significantly, while the contrast between SC2 and SC3 was comparatively narrower and was concentrated in N/C and mass. BAM showed a similarly selective response: SC1 versus SC2 and SC2 versus SC3 differed significantly for many descriptors, but SC1 versus SC3 was largely nonsignificant except for N/C. In contrast, the engineered media were more sensitive to increasing loading. IFGEM1 showed only limited separation between SC1 and SC2, but much clearer divergence once SC3 was reached, whereas IFGEM3 exhibited the broadest and most persistent condition-dependent separation, with significant differences across most pairwise comparisons for H/C, DBE, (DBE–O)/C, AImod, O/C, N/C, and mass. Across all four sorption media, Xc remained comparatively insensitive, suggesting that loading primarily altered saturation/aromaticity, oxidation state, nitrogen enrichment, and molecular size rather than the most condensed core structures. Taken together, Fig. 7a–d indicates that increasing N and P loading does not simply scale DON transformation intensity, but reorganizes the molecular fingerprint space, with the strongest reorganization occurring in the iron-engineered systems, especially IFGEM3.

These fingerprint-level shifts are matched by corresponding changes in the inferred transformation spectrum (Fig. 7e). Rather than maintaining a fixed reaction profile across loading levels, each medium redistributed the relative importance of transformation types as nutrient loading increased. Natural soil and BAM retained comparatively broad and mixed transformation portfolios, in which hydrogenation/dehydrogenation, carbonylation/decarbonylation, oxidation/reduction, methylation/demethylation, and hydration/dehydration all contributed appreciably. This more even partitioning is consistent with the more buffered and transitional behavior seen in Fig. 7a–b. By contrast, IFGEM1 and IFGEM3 showed a more directional reweighting of putative mass-difference categories, with hydrogenation/dehydrogenation emerging as a major component across all conditions, accompanied by substantial contributions from carbonylation/decarbonylation and condition-dependent shifts in oxidation/reduction and methylation/demethylation. In other words, as N and P loading increased, the engineered media did not simply produce different DON molecules; they also appeared to channel DON transformation through a narrower and more selective set of reaction routes. This makes Fig. 7e a baseline ecological context between the statistical descriptor changes in Fig. 7a–d and the zonation redistribution of the produced pool in Fig. 7f.

As defined above, the zonation analysis of the operationally defined effluent-only formula set further clarifies the chemical direction of these loading-associated shifts. In natural soil, this formula set changed from being Zone 4-dominated under SC1 (50.34% in Zone 4 vs. 23.45% in Zone 3) to being Zone 3-dominated under SC2 (42.52% in Zone 3 vs. 33.88% in Zone 4), and remained chemically mixed under SC3, where Zone 3 was still the largest fraction (41.98%) but Zones 1 and 2 together also accounted for 31.91% (Table S15). This redistribution indicates a greater relative representation of lower-NOSC, yet still comparatively saturated, formulas under the higher loading conditions. The Zone 3 designation describes formula-derived reducing character rather than a directly measured decrease in-column redox potential. Plausible, non-exclusive contributors include preferential loss of Zone 4 formulas,

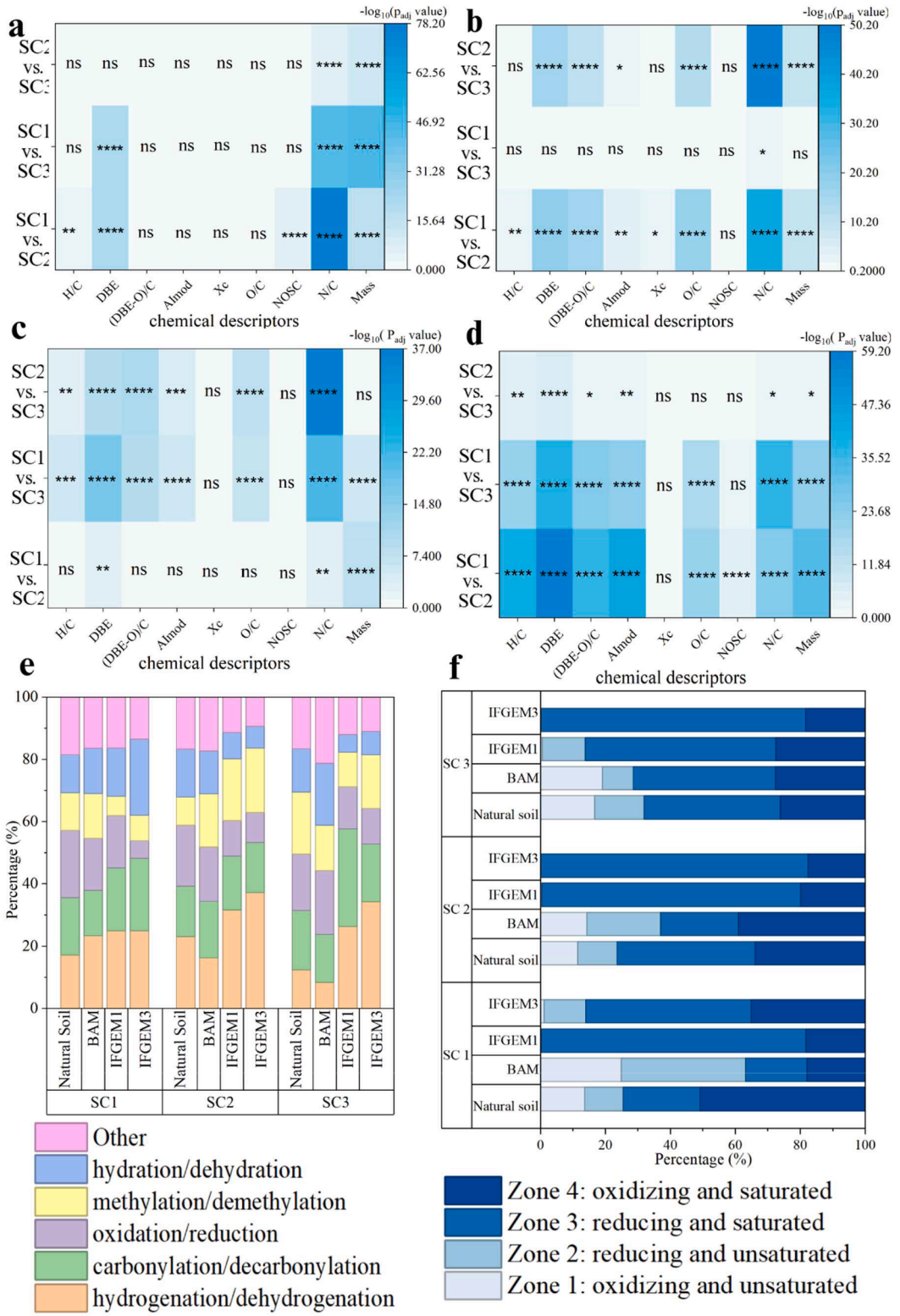



**Fig. 7.** Comparative impact of varying N and P loading on the shift of DON fingerprints, inferred transformation spectra and produced-pool zonation across media. (a)–(d), Heat maps showing pairwise comparisons of effluent DON chemical descriptors among spiking conditions for natural soil (a), BAM (b), IFGEM1 (c) and IFGEM3 (d). Comparisons include SC1 versus SC2, SC1 versus SC3, and SC2 versus SC3. P was adjusted independently within each medium using the Holm procedure. Shading indicates $-\log_{10}(P_{adj}$ value) and symbols denote Holm-adjusted significance levels (ns, not significant; $P_{adj} < 0.05$, $P_{adj} < 0.01$, $P_{adj} < 0.001$, $P_{adj} < 0.0001$). Chemical descriptors include H/C, DBE, (DBE−O)/C, AImod, Xc, O/C, NOSC, N/C, and mass. (e) Stacked bars summarizing the relative contribution of major inferred removed→produced DON transformation classes across SC1–SC3 for all four media. Major pathway classes are shown separately, whereas less abundant pathways are grouped as Other. (f) Horizontal stacked bars showing the zonation distribution of the operationally defined effluent-only formula set across spiking conditions and media. Zone 1, oxidizing and unsaturated; Zone 2, reducing and unsaturated; Zone 3, reducing and saturated; Zone 4, oxidizing and saturated.

greater persistence or appearance of Zone 3 formulas, and short-term abiotic or biofilm-associated processing. Because pulse-resolved depth-specific redox conditions and microbial activity were not measured during SC1–SC3, these alternatives cannot be distinguished by the present dataset. This indicates that higher loading progressively moved newly formed DON in natural soil away from a primarily oxidized/saturated signature toward a more reduced and compositionally diverse state. BAM showed an intermediate pattern: the produced pool under SC1 was concentrated in the more unsaturated Zones 1 and 2 (63.19% combined), became more balanced under SC2 with an increase in Zone 4 (38.49%), and then shifted toward Zone 3 dominance under SC3 (43.78%). In IFGEM1, produced DON was already strongly concentrated in Zone 3 across all three conditions (81.56%, 79.64%, and 58.29% for SC1, SC2, and SC3, respectively), with Zone 4 as a secondary compartment. IFGEM3 showed the clearest loading response of all media: while the produced pool under SC1 was distributed mainly between Zone 3 (50.63%) and Zone 4 (34.88%), it became overwhelmingly concentrated in Zone 3 under SC2 and SC3 (82.37% and 81.51%, respectively), with virtually no contribution from Zones 1 and 2. Thus, Fig. 7f demonstrates that increasing N and P loading was associated with greater representation of effluent-only formulas in the reducing-and-saturated Zone 3, whereas natural soil retains a broader, less constrained chemical response.

Viewed together, Fig. 7 shows that rising nutrient loading signifies the separation among media-specific DON transformation regimes that was first identified under SC1. Natural soil responds through a broader redistribution of molecular descriptors, reaction types, and produced-pool zonation, indicating a comparatively flexible and heterogeneous processing mode. BAM occupies an intermediate position, showing some loading sensitivity but less directional reorganization than the iron-engineered media. IFGEM1 and especially IFGEM3, in contrast, show a more structured and chemically focused response, in which increasing N and P loading is associated with strong fingerprint divergence, a more selective transformation spectrum, and a pronounced accumulation of produced DON in Zone 3. Therefore, it is indicative that increasing nutrient loading was associated with media-specific shifts in molecular-descriptor space, putative mass-difference spectra, and the zonation of the operationally defined effluent-only formula set. These changes were clearest in IFGEM3 and occurred despite consistently high bulk NOx-N removal. The engineering significance and potential ecological implications of this divergence between bulk treatment performance and residual DON composition are considered below.

### 4.3. *Engineering implications, ecological trade-offs, and relation to previous IFGEM studies*

From an engineering perspective, the present column tests define a tested operating envelope rather than an optimized design. All media were evaluated in 0.90-m downflow beds operated at 8 mL $min^{-1}$, with measured hydraulic retention times of 137 and 124 min for IFGEM1 and IFGEM3, respectively (Valencia et al., 2020). Across the tested influent conditions of 0.9–1.7 mg $L^{-1}$ nitrate and 0.3–0.7 mg $L^{-1}$ phosphate, IFGEM1 and IFGEM3 maintained NOx-N removal efficiencies of 98.6–99.2% and 98.9–99.9%, respectively. Their corresponding bulk DON reductions were 86.4–97.1% and 70.6–93.5%. Similar HRTs of IFGEM1 and IFGEM3, together with their different molecular fingerprints, indicate that media formulation influenced residual DON composition beyond contact time alone.

The tested bed depth and HRT may therefore serve as a starting benchmark for site-specific pilot validation, but they should not be interpreted as optimum or minimum design criteria. Media depth and flow were not independently varied, the molecular samples represented the full-column effluents, and each nutrient condition was applied for only 3 h. Consequently, the present experiment does not define an optimal media depth, a universal HRT, or a long-term breakthrough loading threshold. No deterioration in bulk NOx-N removal was observed through SC3, indicating that the conventional treatment-performance threshold was not reached within the tested range. However, the molecular-descriptor and formula-association patterns changed substantially from SC1 to SC2–SC3, particularly in IFGEM3. This difference suggests that changes in effluent DON quality may precede loss of bulk inorganic-N removal and should be considered alongside concentration-based endpoints during pilot monitoring.

The bulk and molecular results also reveal a potential ecological trade-off. Although IFGEM substantially lowered bulk DON concentrations, the residual detectable CHON fraction showed higher $MLB_L$values and greater formula-count representation of protein/amino sugar-like formulas than the natural-soil and BAM effluents. Higher $MLB_L$is consistent with a larger relative contribution of formulas located on the potentially labile side of the molecular lability boundary; however, it remains a formula-derived screening index rather than a direct measurement of biodegradable or bioavailable DON.

The present results therefore do not demonstrate that IFGEM increased net eutrophication risk in receiving waters. The ecological consequence depends on both the concentration of DON remaining after treatment and the fraction that can be utilized by microorganisms or phytoplankton. A substantial decrease in bulk DON may partly or fully offset a higher potential lability per unit of residual DON. Conversely, if a sufficiently large portion of the residual pool is readily bioavailable, it could contribute to downstream microbial oxygen demand, N remineralization, or primary production. Engineering evaluation of IFGEM should therefore pair bulk TN, NOx-N, and DON removal with direct lability endpoints, such as biodegradable DON assays, algal-growth tests, or oxygen-demand measurements. Where a persistent high-lability residual fraction is confirmed, downstream polishing or media blending could be evaluated rather than relying on bulk nitrogen removal alone.

The present findings are broadly consistent with, but more molecularly resolved than, previous IFGEM-related studies. Valencia et al. (2020) reported the underlying column campaign involving natural soil, BAM, IFGEM1, and IFGEM3, with primary emphasis on bulk nutrient and DON performance, 9.4-T FT-ICR MS characterization, and baseline nitrogen-cycling gene profiles. That study found that IFGEM1 and IFGEM3 maintained stronger DON responses than natural soil across the higher loading conditions. The present study and Valencia et al. (2020) are based on the same column campaign and the same archived 9.4-T FT-ICR MS dataset. Their distinction therefore lies in analytical scope rather than instrumental resolution or experimental replication. Valencia et al. focused primarily on bulk nutrient performance, DON behavior, and microbial abundances, whereas the present study applies an expanded formula-level framework to compare occurrence categories, compound-class distributions, molecular descriptors, chemical-trait diversity, zonation, and putative KMD/SOM formula associations among the four media. Accordingly, the present results should be interpreted as an analytical extension of the earlier work, not as independent confirmation of it.

At field scale, Cheng et al. (2026) found that the iron-containing ZIPGEM retained predominantly refractory formulas while the operationally defined effluent-only formula set showed greater representation of labile, reduced, and saturated formulas across ten seasonal sampling events. The higher $MLB_L$, increased representation of protein/amino sugar-like formulas, and greater Zone-3 contribution observed for IFGEM3 in the present study are directionally consistent with that field-scale tendency. Nevertheless, the studies do not establish identical transformation pathways. IFGEM3 contains tire crumb and clay in addition to iron filings, whereas ZIPGEM contains perlite and clay; the present short-pulse column study also differs from the field system in hydraulic scale, seasonal forcing, biofilm age, nutrient-exposure duration, and influent composition. Analytical resolution provides an additional distinction: 21-T FT-ICR MS can resolve and assign a larger fraction of complex formula space than 9.4-T analysis. These differences

may alter formula coverage and apparent mass-difference connectivity. The convergence among studies should therefore be interpreted as consistent directional molecular responses of iron-amended media, rather than confirmation of a universal or uniquely identified biochemical mechanism.

### 4.4. Analytical scope and study limitations

The media- and loading-associated molecular patterns described above should be interpreted within the analytical scope and evidentiary boundaries of the SPE–ESI–FT-ICR MS and qPCR workflows. The molecular profiles should be interpreted within the selectivity of the PPL-SPE and negative-ion ESI workflow. Although all samples were processed under the same extraction and instrumental conditions, PPL recovery was not measured for the individual influent and effluent samples and may vary with molecular polarity and matrix composition. Highly polar or hydrophilic N-containing compounds may therefore be underrepresented, and some apparent influent–effluent differences in formula occurrence or relative signal contribution may partly reflect differential extraction or ionization rather than treatment-induced compositional change alone. Consequently, the reported patterns describe relative changes within the PPL-extractable, negative-ESI-detectable molecular fraction and should not be interpreted as a quantitative mass balance for the complete DON pool.

The formula-level analysis was restricted to assigned CHON formulas because this was the only N-containing class in the accepted molecular-formula inventory. This analytical scope does not imply that the complete DON pool consisted exclusively of CHON compounds. No CHONS formulas were available in the accepted dataset, and P-containing formulas were not evaluated under the original elemental search space. Therefore, media-associated changes in sulfur- or phosphorus-containing DON could not be assessed. This limitation is particularly relevant to the phosphate-loading experiments because incorporation of inorganic phosphate into organic P-containing molecules during column operation cannot be excluded. The reported patterns should consequently be interpreted as changes within the PPL-extractable, negative-ESI-detectable CHON fraction rather than as a complete molecular inventory of DON.

In addition, influent-only, shared, and effluent-only formulas represent operational detection categories rather than direct molecular mass-balance evidence of removal, persistence, or production. Formula occurrence may be affected by extraction recovery, matrix-dependent ionization, signal suppression, and detection thresholds. Similarly, KMD-derived mass differences identify non-unique, putative formula linkages and cannot independently establish reaction direction, molecular structure, enzymatic activity, or biochemical pathway. The network analyses are therefore used to generate hypotheses regarding possible compositional relationships rather than to confirm specific transformations.

The qPCR profiles represent a different temporal and analytical level of evidence from molecular data. They characterize DNA-level, depth-resolved baseline gene-marker distributions established during the four-week conditioning period, whereas the molecular profiles were obtained from water samples following the subsequent 3-h nutrient pulses. Consequently, the qPCR data cannot resolve gene expression, microbial process activity, or spiking-induced community dynamics during SC1–SC3, and they cannot causally validate the KMD-derived formula associations. Moreover, assay-specific calibration records, including amplification efficiencies, standard-curve statistics, dynamic ranges, and detection limits, were not recoverable from the archived project records. Interpretation was therefore restricted to within-target spatial patterns, and values from different gene-target assays were not quantitatively ranked.

The column configuration also limits direct engineering optimization. Media depth and flow rate were not varied independently, the molecular analysis was based on full-column influent and effluent samples, and each nutrient-loading condition was applied for only 3 h. Long-term breakthrough, clogging, media exhaustion, and regeneration were not evaluated. Consequently, the tested 0.90-m bed, measured HRTs, and SC1–SC3 loading range define an experimental operating envelope rather than optimal or threshold design values. Long-duration tests that independently vary bed depth, hydraulic loading, and nutrient mass loading are required to establish design curves and service-life criteria.

The media-specific molecular differences should therefore be interpreted as emergent outcomes of combined physicochemical and biological contexts rather than as evidence of a single resolved mechanism. Potential contributors include differences in media composition and surface chemistry, iron-associated sorption or redox reactivity, hydraulic contact, and the baseline biofilm community established during conditioning. Because these factors were not independently manipulated, their individual contributions cannot be separated in the present study. The observed molecular and qPCR patterns consequently provide a basis for mechanistic hypotheses, but not direct attribution of formula-level changes to a specific abiotic or microbial process.

Future studies should combine sample-specific SPE recovery assessment, broader heteroatom formula assignment, $^{15}N$ tracing, targeted structural confirmation, pulse-resolved RNA-based microbial analysis, and direct nitrogen-transformation rate measurements to test the proposed molecular and microbial relationships.

## 5. Conclusion

This study shows that changes in DON molecular composition within filtration media were medium-specific and sensitive to nutrient loading rather than being explained by bulk NOx-N removal alone. Across three nitrate–phosphate loading conditions, engineered media—particularly IFGEM3— maintained high NOx-N removal while showing systematic shifts in the detectable DON molecular fingerprint.. Relative to natural soil and BAM, IFGEM media effluents showed a lower relative representation of condensed aromatic-like formulas and a higher relative proportion of protein/amino sugar-like formulas within the detected PPL-extractable fraction, while the dominant lignin-like DON pool exhibited broader mass–abundance distributions, lower aromaticity, lower oxidation state, and higher inferred lability. These results indicate that the media differed not only in bulk DON reduction but also in the relative organization of the detectable CHON formula pool. From an engineering perspective, these findings support evaluating filtration media using residual DON composition and potential lability together with bulk NOx-N removal, rather than relying on concentration-based performance alone. Because media depth, hydraulic retention time, and nutrient loading were not independently optimized, the tested configuration should be viewed as an experimental operating envelope for subsequent pilot validation rather than as a universal design criterion.

Fate–zone–class analyses further resolved these differences into distinct media-associated formula-occurrence and zonation patterns. Natural soil increasingly preserved a lignin-rich, oxidizing–saturated DON framework as nutrient loading increased, with most formula associations remaining within the dominant aromatic regions. In contrast, IFGEM3 showed larger operationally defined influent-only and effluent-only formula fractions and progressively concentrated effluent-only formulas in reducing–saturated chemical space under higher loading. KMD-based network analyses showed broader cross-class formula associations in IFGEM3 and a lower relative representation of aromatic-rich formulas in its effluent, while the identified mass-difference relationships remained putative and non-unique. Pre-spiking gene profiles provided complementary baseline context for this contrast: natural soil showed broader depth-wise co-occurrence of selected nitrogen-cycling gene markers, whereas IFGEM3 showed greater depth differentiation for selected targets. It is noted that the pre-spiking qPCR profiles describe the baseline spatial distribution of selected DNA-level gene

markers established during column conditioning, rather than gene expression or microbial activity during the subsequent 3-h nutrient pulses. The absence of recoverable assay-specific calibration metadata and contemporaneous activity measurements prevent quantitative cross-target comparison or causal attribution of the DON molecular patterns to specific microbial pathways.

Taken together, these findings indicate that iron-based engineered media can alter the detectable DON molecular composition while maintaining high nutrient-removal performance. Increasing N and P loading was associated with shifts in both molecular fingerprints and putative formula-association patterns, with the clearest changes observed for IFGEM3. More broadly, evaluating treatment performance solely by nitrate or total nitrogen removal may overlook changes in the composition and potential lability of residual DON. Combining bulk nitrogen measurements with molecular fingerprinting, operational formula-occurrence analysis, and baseline gene-marker profiling therefore provides a more complete basis for evaluating passive nutrient-treatment systems.

These conclusions should be interpreted within several analytical and experimental limitations. Sample-specific PPL-SPE recovery was not quantified, and matrix-dependent extraction and negative-ESI response may have underrepresented highly polar DON. The formula-level analysis was restricted to the accepted PPL-extractable, negative-ESI-detectable CHON fraction; no CHONS formulas were present in the accepted formula inventory, and P-containing formulas were not evaluated under the original elemental assignment space, which is particularly relevant to the phosphate-spiking conditions. Moreover, KMD mass differences identify non-unique, putative formula linkages rather than confirmed reaction pathways, and the pre-spiking DNA-level qPCR profiles do not resolve microbial activity during the subsequent 3-h nutrient pulses.

Future studies should test whether the operationally defined influent-only and effluent-only formula sets represent true nitrogen removal, incorporation, or formation using $^{15}$N-labeled nitrate and/or representative DON substrates. Sample-specific SPE recovery measurements and expanded N/S/P formula assignment are needed to evaluate whether the observed patterns extend beyond the PPL-extractable CHON fraction. Finally, pulse-resolved molecular sampling combined with post-pulse qPCR, RT-qPCR or metatranscriptomic analysis, and direct nitrogen-transformation rate measurements would test whether the baseline genetic patterns translate into condition-specific microbial activity.

## CRediT authorship contribution statement

**Jinxiang Cheng:** Writing – original draft, Investigation, Formal analysis, Data curation, Conceptualization. **Andrea Valencia:** Investigation, Data curation. **Diana Ordonez:** Investigation, Data curation. **Amy M. McKenna:** Investigation, Data curation. **Ni-Bin Chang:** Writing – review & editing, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization.

## Declaration of competing interest

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

## Acknowledgements

The authors acknowledge financial support from the Florida Department of Transportation (Award ID: BDV24 977–25). A portion of this work was performed at the National High Magnetic Field Laboratory ICR User Facility, which is supported by the National Science Foundation Division of Chemistry and Division of Materials Research through DMR-2128556, (previous grant number DMR-1644779), the United States of America.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2026.126789.

## Data availability

Data will be made available on request.

## References

Berman, T., Bronk, D.A., 2003. Dissolved organic nitrogen: a dynamic participant in aquatic ecosystems. Aquat. Microb. Ecol. 31, 279–305. https://doi.org/10.3354/ame031279.

Bronk, D.A., See, J.H., Bradley, P., Killberg, L., 2007. DON as a source of bioavailable nitrogen for phytoplankton. Biogeosciences. 4 (3), 283–296. https://doi.org/10.5194/bg-4-283-2007.

Cheng, J., McKenna, A.M., Chang, N.-B., 2026. Exploring seasonal lability and fate of dissolved organic nitrogen cycling in a field-scale water filtration system. Water. Res. 292, 125251. https://doi.org/10.1016/j.watres.2025.125251.

Dittmar, T., Koch, B.P., Hertkorn, N., Kattner, G., 2008. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. Limnol. Oceanogr.: Methods 6, 230–235. https://doi.org/10.4319/lom.2008.6.230.

Eom, H., Borgatti, D., Paerl, H.W., Park, C., 2017. Formation of low-molecular-weight dissolved organic nitrogen in predenitrification biological nutrient removal systems and its impact on eutrophication in coastal waters. Environ. Sci. Technol. 51 (7), 3776–3783. https://doi.org/10.1021/acs.est.6b06576.

He, C., Zhang, Y., Li, Y., Zhuo, X., Li, Y., Zhang, C., Shi, Q., 2020. In-house standard method for molecular characterization of dissolved organic matter by FT-ICR mass spectrometry. ACS. Omega 5 (20), 11730–11736. https://doi.org/10.1021/acsomega.0c01055.

Hu, J., Kang, L., Li, Z., Feng, X., Liang, C., Wu, Z., Zhou, W., Liu, X., Yang, Y., Chen, L., 2023. Photo-produced aromatic compounds stimulate microbial degradation of dissolved organic carbon in thermokarst lakes. Nat. Commun. 14, 3681. https://doi.org/10.1038/s41467-023-39432-2.

Hu, J., Liu, C.-G., Zhang, W.-K., Liu, X.-W., Dong, B., Wang, Z.-D., Xie, Y.-G., Hua, Z.-S., Liu, X.-W., 2025. Decomposing the molecular complexity and transformation of dissolved organic matter for innovative anaerobic bioprocessing. Nat. Commun. 16, 4859. https://doi.org/10.1038/s41467-025-60240-3.

Hur, M., Yeo, I., Kim, E., No, M.-H., Koh, J., Cho, Y.J., Lee, J.W., Kim, S., 2010. Correlation of FT-ICR mass spectra with the chemical and physical properties of associated crude oils. Energy Fuels 24 (10), 5524–5532. https://doi.org/10.1021/ef1007165.

Jani, J., Yang, Y.-Y., Lusk, M.G., Toor, G.S., 2020. Composition of nitrogen in urban residential stormwater runoff: concentrations, loads, and source characterization of nitrate and organic nitrogen. PLoS. One 15 (2), e0229715. https://doi.org/10.1371/journal.pone.0229715.

Kim, S., Kramer, R.W., Hatcher, P.G., 2003. Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram. Anal. Chem. 75 (20), 5336–5344. https://doi.org/10.1021/ac034415p.

Koch, B.P., Dittmar, T., Witt, M., Kattner, G., 2007. Fundamentals of molecular formula assignment to ultrahigh resolution mass data of natural organic matter. Anal. Chem. 79 (4), 1758–1763. https://doi.org/10.1021/ac061949s.

LaRowe, D.E., Van Cappellen, P., 2011. Degradation of natural organic matter: a thermodynamic analysis. Geochim. Cosmochim. Acta 75 (8), 2030–2042. https://doi.org/10.1016/j.gca.2011.01.020.

Li, L., Davis, A.P., 2014. Urban stormwater runoff nitrogen composition and fate in bioretention systems. Environ. Sci. Technol. 48 (6), 3403–3410. https://doi.org/10.1021/es4055302.

Lin, Y., Wang, L., Xu, K., Huang, H., Ren, H., 2021. Algae biofilm reduces microbe-derived dissolved organic nitrogen discharges: performance and mechanisms. Environ. Sci. Technol. 55 (9), 6227–6238. https://doi.org/10.1021/acs.est.0c06915.

Liu, Y., Xie, Z., Feng, J., Xie, S., Ma, C., 2025. Thermal regimes during overwintering recovery shape microbial network and dissolved organic matter complexity in microcystis-dominated systems. ISMe J. 19 (1). https://doi.org/10.1093/ismejo/wraf227 wraf227.

Lusk, M.G., Toor, G.S., 2016a. Biodegradability and molecular composition of dissolved organic nitrogen in urban stormwater runoff and outflow water from a stormwater retention pond. Environ. Sci. Technol. 50 (7), 3391–3398. https://doi.org/10.1021/acs.est.5b05714.

Lusk, M.G., Toor, G.S., 2016b. Dissolved organic nitrogen in urban streams: biodegradability and molecular composition studies. Water. Res. 96, 225–235. https://doi.org/10.1016/j.watres.2016.03.060.

Merel, S., 2023. Critical assessment of the Kendrick mass defect analysis as an innovative approach to process high resolution mass spectrometry data for environmental applications. Chemosphere 313, 137443. https://doi.org/10.1016/j.chemosphere.2022.137443.

O'Reilly, A.M., Wanielista, M.P., Chang, N.-B., Xuan, Z., Harris, W.G., 2012. Nutrient removal using biosorption activated media: preliminary biogeochemical assessment

of an innovative stormwater infiltration basin. Sci. Total Environ. 432, 227–242. https://doi.org/10.1016/j.scitotenv.2012.05.083.
Osman, M., Wan, Yusof, K., Takaijudin, H., Goh, H.W., Abdul Malek, M., Azizan, N.A., Ab. Ghani, A., Sa'id Abdurrasheed, A., 2019. A review of nitrogen removal for urban stormwater runoff in bioretention system. Sustainability. 11 (19), 5415. https://doi.org/10.3390/su11195415.
Patel, S.A., Erickson, L.E., 1981. Estimation of heats of combustion of biomass from elemental analysis using available electron concepts. Biotechnol. Bioeng. 23 (9), 2051–2067. https://doi.org/10.1002/bit.260230910.
Pehlivanoglu-Mantas, E., Sedlak, D.L., 2006. Wastewater-derived dissolved organic nitrogen: analytical methods, characterization, and effects—A review. Crit. Rev. Environ. Sci. Technol. 36 (3), 261–285. https://doi.org/10.1080/10643380500542780.
Qi, Y., Xie, Q., Wang, J.-J., He, D., Bao, H., Fu, Q.-L., Su, S., Sheng, M., Li, S.-L., Volmer, D.A., Wu, F., Jiang, G., Liu, C.-Q., Fu, P., 2022. Deciphering dissolved organic matter by Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS): from bulk to fractions and individuals. Carbon Research 1, 3. https://doi.org/10.1007/s44246-022-00002-8.
Raeke, J., Lechtenfeld, O.J., Wagner, M., Herzsprung, P., Reemtsma, T., 2016. Selectivity of solid phase extraction of freshwater dissolved organic matter and its effect on ultrahigh resolution mass spectra. Environ. Sci.: Processes Impacts 18 (7), 918–927. https://doi.org/10.1039/C6EM00200E.
Riedel, T., Zark, M., Vähätalo, A.V., Niggemann, J., Spencer, R.G.M., Hernes, P.J., Dittmar, T., 2016. Molecular signatures of biogeochemical transformations in dissolved organic matter from ten world rivers. Front. Earth. Sci. (Lausanne) 4, 85. https://doi.org/10.3389/feart.2016.00085.
Robles-Lecompte, A., Cheng, J., McKenna, A.M., Chang, N.-B., 2024. Linking pattern shifts of dissolved organic nitrogen fractional removal with microbial species richness in a cascade upflow biofiltration process. Water. Res. 264, 122130. https://doi.org/10.1016/j.watres.2024.122130.
Stubbins, A., Dittmar, T., 2015. Illuminating the deep: molecular signatures of photochemical alteration of dissolved organic matter from North Atlantic Deep Water. Mar. Chem. 177, 318–324. https://doi.org/10.1016/j.marchem.2015.06.020.
Stücheli, P.E., Niggemann, J., Schubert, C.J., 2018. Comparison of different solid phase extraction sorbents for the qualitative assessment of dissolved organic nitrogen in freshwater samples using FT-ICR-MS. J. Limnol. 77 (3), 400–411. https://doi.org/10.4081/jlimnol.2018.1791.
Valencia, A., Chang, N.-B., Wen, D., Ordonez, D., Wanielista, M.P., 2019. Optimal recipe assessment of iron filing-based green environmental media for improving nutrient removal in stormwater runoff. Environ. Eng. Sci. 36 (10), 1323–1336. https://doi.org/10.1089/ees.2019.0094.
Valencia, A., Ordonez, D., Wen, D., McKenna, A.M., Chang, N.-B., Wanielista, M.P., 2020. The interaction of dissolved organic nitrogen removal and microbial abundance in iron-filings based green environmental media for stormwater treatment. Environ. Res. 188, 109815. https://doi.org/10.1016/j.envres.2020.109815.
Ward, C.P., Nalven, S.G., Crump, B.C., Kling, G.W., Cory, R.M., 2017. Photochemical alteration of organic carbon draining permafrost soils shifts microbial metabolic pathways and stimulates respiration. Nat. Commun. 8, 772. https://doi.org/10.1038/s41467-017-00759-2.
Wen, D., Chang, N.-B., Wanielista, M.P., 2020. Assessing nutrient removal in stormwater runoff for urban farming with iron filings-based green environmental media. Sci. Rep. 10, 9379. https://doi.org/10.1038/s41598-020-66159-7.
Xu, J., Zhou, P., Zhang, C., Yuan, L., Xiao, X., Dai, L., Huo, K., 2022. Striding the threshold of photocatalytic lignin-first biorefining via a bottom-up approach: from model compounds to realistic lignin. Green Chem. 24 (14), 5351–5378. https://doi.org/10.1039/D2GC01409B.
Yang, Y.-Y., Lusk, M.G., 2018. Nutrients in urban stormwater runoff: current state of the science and potential mitigation options. Curr. Pollut. Rep. 4 (2), 112–127. https://doi.org/10.1007/s40726-018-0087-7.
Zhang, S., Dong, Z., Shi, J., Yang, C., Fang, Y., Chen, G., Chen, H., Tian, C., 2022. Enzymatic hydrolysis of corn stover lignin by laccase, lignin peroxidase, and manganese peroxidase. Bioresour. Technol. 361, 127699. https://doi.org/10.1016/j.biortech.2022.127699.
Zhang, S., Wang, H., Xiang, Y., Sun, M., Ye, L., Liu, W., Zhu, Y., Wang, C., Wang, L., Lou, Z., 2026. Combined effects of coagulation and ozonation treatment on landfill leachate DOM biodegradability. Sustainability. 18 (3), 1530. https://doi.org/10.3390/su18031530.
Zhou, T., Guo, J., Liu, Q., Liu, Y., Wu, W., Wang, Y., Zhang, S., Peng, Y., 2024. DOM and DON transformation in full-scale wastewater treatment plants: comparison of autotrophic and heterotrophic nitrogen removal units. Chem. Eng. J. 479, 147810. https://doi.org/10.1016/j.cej.2023.147810.