Contribution of structural recalcitrance to the formation of the deep oceanic dissolved organic carbon reservoir

Nannan Wang\textsuperscript{1,2}, Ya-Wei Luo\textsuperscript{2}, Luca Polimene\textsuperscript{3}, Rui Zhang\textsuperscript{2}, Qiang Zheng\textsuperscript{2}, Ruanhong Cai\textsuperscript{2}, Nianzhi Jiao\textsuperscript{2*}

\textsuperscript{1} College of Environment and Ecology, Xiamen University, Xiamen 361102, China.
\textsuperscript{2} State Key Laboratory of Marine Environmental Sciences, Xiamen University, Xiamen 361102, China.
\textsuperscript{3} Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth, UK.

Running title: Structural recalcitrance dominates DOM persistence

Correspondence: *Nianzhi Jiao; Address: Zhoulongquan BLDG, Xiangan South Road, Xiamen, Fujian 361102, China; E-mail: jiao@xmu.edu.cn; Tel.: 86-592-2880199; Fax: 86-592-2880150.
Originality-Significance Statement

There is a lively debate over why the large recalcitrant dissolved organic carbon (DOC) reservoir persists in the ocean for extended time periods (up to millennia). Here, we evaluate the relative contributions of structurally recalcitrant DOC and recalcitrant DOC due to dilution to deep oceanic DOC pool by modeling transformations of the different DOC components in incubation experiments with different DOC concentration levels. We conclude that the majority of DOC in the deep ocean is structurally recalcitrant, which supports the hypothesis that the recalcitrance of RDOC is largely related to its chemical properties rather than its low concentration. This also implies that recalcitrant DOC is independent of from RDOC concentration threshold for bacterial uptake.
Summary

The origin of the recalcitrant dissolved organic carbon (RDOC) reservoir in the deep ocean remains enigmatic. The structural recalcitrance hypothesis suggests that RDOC is formed by molecules that are chemically resistant to bacterial degradation. The dilution hypothesis claims that RDOC is formed from a large diversity of labile molecules that escape bacterial utilization due to their low concentrations, termed as RDOC$_c$. To evaluate the relative contributions of these two mechanisms in determining the long-term persistence of RDOC, we model the dynamics of both structurally recalcitrant DOC and RDOC$_c$ based on previously published data that describes deep oceanic DOC degradation experiments. Our results demonstrate that the majority DOC (84.5±2.2%) in the deep ocean is structurally recalcitrant. The intrinsically labile DOC (i.e., labile DOC that rapidly consumed and RDOC$_c$) accounts for a relatively small proportion and is consumed rapidly in the incubation experiments, in which 47.8±3.2% of labile DOC and 21.9±4.6% of RDOC$_c$ is consumed in 40 days. Our results suggest that the recalcitrance of RDOC is largely related to its chemical properties, whereas dilution plays a minor role in determining the persistence of deep-ocean DOC.
Introduction

The ocean contains a large amount of recalcitrant dissolved organic carbon (RDOC). This reservoir plays a critical role in global carbon sequestration and potentially affects global climate (Hansell et al., 2009; Jiao et al., 2014). Although we know that RDOC persists in the water column for thousands of years (Brophy and Carlson, 1989; Lancelot et al., 1993; Carlson and Ducklow, 1995; Jiao et al., 2010; Hansell, 2013), the mechanisms underpinning this extraordinary stability remain objects of controversy (Arrieta et al., 2015b, a; Jiao et al., 2015).

Two primary but not mutually exclusive mechanisms have been proposed to explain the persistence of RDOC. The dilution hypothesis postulates that the deep oceanic dissolved organic carbon (DOC) consists of a large diversity of extremely diluted molecules, termed RDOC_c. These molecules, although chemically labile, would not be consumed by bacteria because their concentrations could be below the thresholds for direct assimilatory uptake and the metabolic requirements for their degradation cannot be satisfied (Jannasch, 1967; Barber, 1968; Dittmar and Paeng, 2009; Kujawinski, 2011; Arrieta et al., 2015b, a). This hypothesis is supported by experimental observations showing that the growth of microbes ceased if their substrates fell below a defined concentration threshold (Jannasch, 1967, 1994; Martens-Habbena et al., 2009; Arrieta et al., 2015b, a).

Another explanation for the long-term persistence of RDOC is the structural recalcitrance of the molecules in a specific environmental context, which corresponds to the previously proposed concept of biologically inert RDOC (RDOC_i) (Jiao et al., 2014). This idea is supported by the evidence that the chemical composition of the DOC in the deep ocean, where the RDOC dominates, is different from that of
decomposable DOC (Jannasch, 1967; Lancelot et al., 1993; Kaiser and Benner, 2009; Jiao et al., 2010; Benner and Amon, 2015; Walker et al., 2016b). Indeed, solid phase extracted (SPE) DOC from deep ocean generally features higher double-bond equivalent values, lower O/C and H/C ratios, as well as a higher degree of unsaturation and rings in molecules, than that from surface waters (Koch et al., 2005; Flerus et al., 2012; Koch et al., 2014; Hansman et al., 2015; Arakawa et al., 2017; Jiao et al., 2018). Although a complete chemical characterization of marine DOM is currently lacking, it was reported that RDOC produced from bacterial consumption of simple substrates contains carboxyl and fused alicyclic functional groups, which determines its resistance to biodegradation and refractory nature (Ogawa et al., 2001; Hertkorn et al., 2006; Lechtenfeld et al., 2014). For instance, the carboxyl-rich alicyclic molecules (CRAM), which account for ~8% of the DOC, are one of the most abundant organic components ever identified in the deep ocean (Hertkorn et al., 2006; Hertkorn et al., 2012; Lechtenfeld et al., 2015; Rossel et al., 2015).

Both $RDOC_c$ and $RDOC_t$ are products of the microbial carbon pump (MCP) since they are gradually generated by the successive and repetitive actions of bacteria degrading relatively more labile substrates (Jiao et al., 2014). However, the implications that these two pools have for the capacity of the MCP to store carbon by forming RDOC differ dramatically. If the majority of RDOC is $RDOC_c$, the capacity of the ocean to store carbon as RDOC would be limited by the DOC concentration threshold for bacterial uptake (i.e., RDOC cannot exceed that threshold otherwise it is consumed by microbes). In contrast, if the majority of RDOC is $RDOC_t$, the capacity of the ocean to store carbon via the MCP would be independent from RDOC concentrations. Therefore, the trade-off between these two forms of recalcitrance
would profoundly influence the capacity of the ocean to store carbon as RDOC and thus influence the impacts of ocean carbon dynamics on climate change.

In this study, we assess the relative contributions of the dilution and structural recalcitrance hypotheses in determining the long-term persistence of DOC by reanalyzing published data from a DOC degradation experiment (Arrieta et al., 2015b), in which natural bacterial communities from the deep Pacific and Atlantic oceans were incubated with deep-water SPE DOC at different concentrations in about 40 days. The dataset includes measurements of SPE DOC utilization and Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) data of hydrophobic SPE DOM (~40% of total seawater DOC). Although the FT-ICR-MS is not totally quantitative due to the matrix effect (i.e., matrix-dependent ion suppression), appropriate standardization and experimental designs allow us to identify molecules that are consumed or produced (Arrieta et al., 2015a) and investigate the influence of the matrix effect on results (Osterholz et al., 2015). Here, we model the dynamics of DOC components under the framework that i) the initial deep oceanic DOC is composed of structurally recalcitrant DOC (RDOC$_c$), biodegradable DOC with extremely low individual concentrations that are below the corresponding microbial uptake thresholds (RDOC$_t$) and labile DOC ready for microbial utilization (LDOC); and ii) LDOC is utilized in both controls and concentrated treatments, while RDOC$_c$ can only be partially utilized in concentrated treatments (Figure 1). By leveraging the FT-ICR-MS profiling, which characterizes SPE DOM molecular composition changes, we define a utilization index for each compound by comparing the relative intensity of normalized FT-ICR-MS signal at the start and at the end of the experiments. Then, we calculate the proportion of intrinsically labile DOC (LDOC and RDOC$_c$) consumed by averaging the utilization
index of compounds that are significantly utilized. Ultimately, the contributions of RDOC<sub>c</sub> and RDOC<sub>t</sub> to the deep oceanic DOC pool are quantified by comparing the observed SPE DOC consumption to the consumption of the intrinsically labile DOC with constraints on the DOC production process.

**Results and Discussion**

The intrinsically labile DOC was rapidly utilized

We first estimate the percentage of consumed intrinsically labile DOC in relation to initial intrinsically labile DOC (c<sub>i</sub> for LDOC and c<sub>c</sub> for RDOC<sub>c</sub> as shown in Table 1) based on significantly utilized molecules. The percentage of LDOC consumed are estimated to be 54.9±3.3% and 47.8±3.2% for experiments O and P respectively (Table 2), meaning that approximately half of the LDOC was consumed in the 40-day incubation. This indicates that the intrinsically labile DOC was quickly consumed, which is consistent with the observations from recent bioassay experiments that approximately 75% of labile DOC was consumed in less than 30 days (Lechtenfeld et al., 2015) and supports labile DOC cycles with timescales of days to weeks (Benner and Amon, 2015; Hansman et al., 2015). The propagation uncertainty caused by measurement variability is approximately one order of magnitude lower than the estimates, allowing us to rule out measurement uncertainty caused by the FT-ICR-MS data.

The additional removal of deep oceanic DOC, i.e., RDOC<sub>c</sub> could be observed within concentrated treatments by comparing the normalized FT-ICR-MS signals at the end of controls and 5-fold concentrated treatments. It shows that approximately 25% of compounds in the RDOC<sub>c</sub> pool has been utilized, i.e., 25.2±3.4% for station O and
21.9±4.6% for station P (Table 2). Together with the evidence of microbial growth
data shown in Arrieta et al. (2015a), it supports the existence of RDOC\(_c\) in the deep
ocean. The percentage of RDOC\(_c\) consumed (21.9% to 25.2%) is smaller than that of
LDOC (47.8% to 54.9%), which can be attributed to the fact that molecules from the
RDOC\(_c\) pool are too diluted and have less chance to be encountered by microbes for
consumption (Hansell and Carlson, 2014). Analogous to the estimates of \(c_l\) with
narrow ranges of propagation uncertainty, the propagation of measurement
uncertainty on \(c_c\) is lower than 4.6% as well (Table 2). Notably, our method provides
the lower limits of \(c_l\) and \(c_c\) rather than their true values (see Supplementary
Methods), implying that there would be more intrinsically labile DOC consumed in
reality than in our estimates.

The majority of deep oceanic DOC is RDOC\(_r\)

The observed percentage of SPE DOC consumed (i.e., \(c_{1X}\) and \(c_{5X}\)) in the incubation
periods are 2.1% and 4.1% for controls and 5-fold concentrated treatments,
respectively (Table 3; Supplementary Methods). The percentage of RDOC\(_r\) in SPE
DOC is estimated from equation (8) in Methods. Taking results from experiment O as
an example (Table 2), LDOC and RDOC\(_c\) account for less than 4.4% and 9.1% of
SPE DOC, respectively, which implies that RDOC\(_r\) represents up to 86.6±1.2% of
SPE DOC. The extensive sensitivity tests show that even using a very wide range of
parameter values (Table 3), the estimate of the percentage of RDOC\(_r\) in SPE DOC
changes no more than 2.3%, except for the effect of \(c_{5X}\) that yields no more than 9.0%
of the change. Therefore, our results are robust with respect to the choice of
parameters. Repeating the same model process on the dataset from experiment P
shows similar results in which the percentage of RDOC\(_r\) in SPE DOC is estimated to
be $84.5\pm2.2\%$. Based on these results, we propose that the majority of the SPE DOC is recalcitrant due to its structurally recalcitrant property rather than the dilution.

Since FT-ICR-MS data only provide molecular formulas and no chemical structures, it is impossible to unambiguously designate a specific compound as being refractory due to dilution or chemical structure without any experimental treatments. Here, we have reanalyzed FT-ICR-MS datasets from substrate utilization experiments with concentrated treatments and explored the contributions of both $\text{RDOC}_c$ and $\text{RDOC}_t$ to long-term persistence of RDOC in the deep ocean. The key rationale of our method is that a fraction of $\text{RDOC}_c$ becomes bioavailable when increasing the DOC concentrations, the process of which allows the differentiation between compounds with low concentration (i.e., $\text{RDOC}_c$) and structurally recalcitrant compounds (i.e., $\text{RDOC}_t$) in each mass peak and thus the quantification of their respective contributions to the peak magnitude. One caveat to use the FT-ICR-MS data is that the efficiency of the SPE method (Bond-Elut-PPL, 1g) is ~40\% as shown in Arrieta et al. (2015a) and possibly introduces compositional bias towards hydrophobic compounds while losing other DOC compounds (Coppola et al., 2015; Broek et al., 2017). However, it is shown that the radiocarbon ($\Delta^{14}$C) values and C/N ratios of PPL SPE DOC are statistically indistinguishable from those of bulk DOC at depth, which suggests that PPL SPE DOC is a close representative of the deep oceanic DOC dominated by refractory compounds (Broek et al., 2017). Considering the extracted DOC in experiments O and P are from deep ocean, we could ignore the influence of compositional bias caused by the PPL extraction and generalize our conclusion that $\text{RDOC}_t$ dominates the deep oceanic DOC. Matrix-dependent ion suppression is another issue related to the use of FT-ICR-MS data (Trufelli et al., 2011; Osterholz et al., 2015). In the enrichment experiments, incubation and DOC concentrated
treatments can result in the changes of both DOC composition and concentration, and therefore lead to the matrix effect. Our model considers matrix-dependent ion suppression to be an inherent assumption, and we evaluate the matrix effect caused by DOC composition and concentration changes in incubation experiments (Figure S2).

When matrix effect is considered, RDOC\textsubscript{t} accounts for 82.1~89.1% and 77.8~87.9% of SPE DOC for experiments O and P, respectively (Table S1), showing that the matrix effect is unlikely to affect our estimates. However, considering that the matrix effect is a complex issue with poorly understood mechanisms, we should pay a great attention to sample preparation and the use of calibration approach for the FT-ICR-MS community in publishing future datasets (Trufelli, et al., 2010).

Our results are consistent with the size-reactivity model and other studies that link the radiocarbon age of organic matter to its chemical composition (Loh et al., 2004; Repeta and Aluwihare, 2006; Walker et al., 2014; Benner and Amon, 2015; Walker et al., 2016a; Walker et al., 2016b). The size-age-composition relationship that organic matter size is negatively correlated with radiocarbon age and carbon:nitrogen ratios supports the dominant role of chemical composition in determining the recalcitrance of the RDOC pool (Walker et al., 2016b). In addition, if the majority of deep oceanic DOC is RDOC\textsubscript{c}, deep ocean $\Delta^{14}$C calculated from the mass balance model would be difficult to reconcile with its observation (Wilson and Arndt, 2017). Most recently, an experimental study using large volume water column (>100 tons) shows that the microbial transformation of LDOC to RDOC in a standing ecosystem can be completed in a very short time (a few months) (Jiao et al., 2018), providing further evidence for the dominance of RDOC\textsubscript{t}.
It is worth noting that the study of Arrieta et al. (2015a) also refers to the RDOC\textsubscript{c} fraction by constructing a utilization index of the relative abundance of molecules, which showed that there are more than 70% and 40% of DOC is RDOC\textsubscript{c} in the deep ocean for experiments O and P, respectively, i.e., no more than 30% and 60% of DOC is RDOC\textsubscript{t}. This discrepancy from our study can be attributed to different assumptions in the estimation. Their estimation assumes that each molecule is either intrinsically labile or structurally recalcitrant. Therefore the significantly utilized molecules detected by utilization index are only intrinsically labile, i.e., LDOC and RDOC\textsubscript{c}. However, our method assumes that each observed peak likely contains both intrinsically labile DOC and RDOC\textsubscript{t} components, since that the structural isomers with different bioavailability cannot be distinguished with the limit of current FT-ICR MS technology (Stubbins et al., 2014; Osterholz et al., 2015).

**Explanations of low DOC consumption in concentrated experiments**

Arrieta and coworkers have stated that the ‘dilution hypothesis is the primary mechanism controlling on the biogeochemical cycling of DOC in the deep ocean’ (Arrieta et al., 2015b). It has been pointed out that there is no significant increase in DOC consumption (<6%) when increasing DOC concentrations (Jiao et al., 2015). Arrieta and coworkers explained that it would take a longer time for complete consumption of labile substrates and illustrated this process by a simulation study that assumed that DOC utilization depends only on concentration. Here, we carry out a similar simulation but under a more general case that deep oceanic DOC is a mixture of structurally recalcitrant and intrinsically labile DOC. If no RDOC\textsubscript{t} is available in the original SPE DOC (Figure 2A), the simulation results are equal to those reported in Arrieta et al. (Arrieta et al., 2015a). The SPE DOC utilized after 40 days is 2.1±0.7%, 2.6±0.9%, 3.7±1.2% and 5.2±1.7% of the initial concentrations for controls,
2-, 5-, and 10-fold concentrated treatments, respectively. However, if the percentage of RDOC in total DOC varies from 10% to 90% (Figure 2B-D), there is no apparent difference in DOC consumption rate among the four scenarios at the 40-day time scale. Based on our conservative estimate of the percentage of RDOC in total DOC (84.5%), the DOC utilized after 40 days would be 2.1±0.7%, 2.5±0.8%, 3.5±1.1% and 4.8±1.5% of the initial concentration for controls and concentrated treatments, which are still well within the range of the observations (<6%) and have slight difference from those estimated under the assumption that no RDOC exists.

Till now, there are two plausible interpretations about the low DOC consumption in enrichment experiments. The first one, illustrated in Arrieta et al. (2015a), hypothesizes that a large fraction of DOC in the deep ocean is intrinsically labile compounds and attributes the low DOC consumption even under concentrated treatments to the short-term incubation. In other words, the concentrated DOC could be completely consumed if given longer time for incubation (Figure 3A). In contrast, our results show that the percentage of intrinsically labile DOC consumed is high in less than 40-days of incubation (>47.8% for LDOC and >21.9% for RDOCc), but the small fraction of intrinsically labile DOC results in the small bulk DOC consumption (Figure 3B). Interestingly, this discussion is also consistent with the results obtained by a different modeling approach that explored the radiocarbon signature of DOC in relation to the dilution theory (Wilson and Arndt, 2017). While both the models suggest that DOC recalcitrance is largely related to its chemical composition, they also imply that RDOCc plays a secondary role in explaining the long-term persistence of deep oceanic DOC.

Conclusions
In summary, our results support the hypothesis that RDOC in the deep ocean is dominated by structurally recalcitrant \( \text{RDOC}_t \), i.e. at least 84.5% of RDOC in this case. The LDOC and \( \text{RDOC}_c \) account for a relatively minor fraction of the bulk RDOC (<5.0% for LDOC and <10.5% for \( \text{RDOC}_c \)) and can be rapidly used. The high proportion of \( \text{RDOC}_t \) implies that the capacity of the ocean to store carbon via MCP is not constrained by a dilution threshold. Further investigations are required to understand if the current pool of oceanic RDOC could increase, potentially counteracting the anthropogenically induced increase in atmospheric CO\(_2\).

**Acknowledgements**

This study is supported by the National Key R&D Program of China2016YFA0601404, the NSFC grants 91428308 and 41476093, and the National Key Research Programs 2013CB955700. This study is a contribution to the international IMBER project. L.P. was funded by the (UK) NERC National Capability in marine modeling and by the NERC funded project 'Understanding and modeling the Microbial Carbon Pump under changing nutrient concentrations and temperature' (NE/R011087/1).

**Conflict of interest**

All authors have seen and approved the final version submitted, and declared no conflict of interest.

**Supporting Information**

Additional Supporting Information may be found in the online version of this article.

**Supplementary Methods.** Descriptive information regarding the methods used within this study.
Supplementary Tables and Figures

Table S1. Matrix effect on estimating the percentage of RDOC\(_i\) in SPE DOC \((f_i)\). Here, we consider the matrix effect caused by the incubation process (i.e., change of signal intensity from DOC component during incubation experiments) and concentrated treatments (i.e., change of signal intensity due to DOC concentrated treatments). The measurement uncertainty caused by matrix effect (i.e., \(\sigma_{ME}\)) is calculated as the 95\(^\text{th}\) percentile of signal intensity difference by comparing 1X initial and 1X final FT-ICR-MS, 1X initial and 5X initial FT-ICR-MS, 5X initial and 5X final FT-ICR-MS, respectively. We then investigate the matrix effect on the estimation of \(f_i\) according to equation (17) in Supplementary Methods.

Figure S1. Histogram distribution of measurement variability of FT-ICR-MS fingerprints of DOC by assembling all molecules together. The measurement variability for each molecule is calculated as the standard deviation of its relative intensities from replicates. The average measurement variability FT-ICR-MS (i.e., \(\sigma_m\)) is chosen as the 95\(^\text{th}\) percentile of this distribution for the downstream propagation uncertainty analysis.

Figure S2. Comparison of signal intensity of FT-ICR-MS to assess DOC composition change during the incubation experiments (A, B, D, E) and the matrix effect due to DOC concentrated treatments (C, F). The diagonal line is shown in the red color in each figure.

References


Table 1. Model parameters with their indicators. All are unitless.

<table>
<thead>
<tr>
<th>Meaning</th>
<th>Symbol</th>
<th>Standard value</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>The percentage of RDOC\textsubscript{t} in SPE DOC pool</td>
<td>$f_t$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>The percentage of RDOC\textsubscript{c} in SPE DOC pool</td>
<td>$f_c$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>The percentage of LDOC in SPE DOC pool</td>
<td>$f_l$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>The percentage of consumed RDOC\textsubscript{c} in relation to initial RDOC\textsubscript{c}</td>
<td>$c_c$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>The percentage of consumed LDOC in relation to initial LDOC</td>
<td>$c_l$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>The percentage of consumed DOC in relation to SPE DOC in controls</td>
<td>$c_{1X}$</td>
<td>2.1%</td>
<td>Arrieta et al. (2015a)</td>
</tr>
<tr>
<td>The percentage of consumed DOC in relation to SPE DOC in 5-fold concentrated treatments</td>
<td>$c_{5X}$</td>
<td>4.1%</td>
<td>Arrieta et al. (2015a)</td>
</tr>
<tr>
<td>The new DOC produced per unit of LDOC consumed</td>
<td>$p_l$</td>
<td>12.5%</td>
<td>(Stoderegger and Herndl, 1998)</td>
</tr>
<tr>
<td>The new DOC produced per unit of RDOC\textsubscript{c} consumed</td>
<td>$p_c$</td>
<td>12.5%</td>
<td>(Stoderegger and Herndl, 1998)</td>
</tr>
</tbody>
</table>
Table 2. Propagation of measurement uncertainty of FT-ICR-MS on estimating the percentage of RDOC, in SPE DOC ($f_t$). To derive the propagation of measurement uncertainty from equation (17) in Supplementary Methods, we set $c_{1X} = 2.1\%$, $c_{5X} = 4.1\%$, $p_l = 12.5\%$, $p_c = 12.5\%$ as standard values and calculate the measurement uncertainty of FT-ICR-MS ($\sigma_m$) using the replicates (see Supplementary Methods).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimation (%)</th>
<th>Propagation uncertainty (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp. O</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c_l$</td>
<td>54.9</td>
<td>3.3</td>
</tr>
<tr>
<td>$c_c$</td>
<td>25.2</td>
<td>3.4</td>
</tr>
<tr>
<td>$f_l$</td>
<td>4.4</td>
<td>0.3</td>
</tr>
<tr>
<td>$f_c$</td>
<td>9.1</td>
<td>1.2</td>
</tr>
<tr>
<td>$f_t$</td>
<td>86.6</td>
<td>1.2</td>
</tr>
<tr>
<td>Exp. P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c_l$</td>
<td>47.8</td>
<td>3.2</td>
</tr>
<tr>
<td>$c_c$</td>
<td>21.9</td>
<td>4.6</td>
</tr>
<tr>
<td>$f_l$</td>
<td>5.0</td>
<td>0.3</td>
</tr>
<tr>
<td>$f_c$</td>
<td>10.5</td>
<td>2.2</td>
</tr>
<tr>
<td>$f_t$</td>
<td>84.5</td>
<td>2.2</td>
</tr>
</tbody>
</table>
Table 3. Sensitivity analysis for each parameter in estimating the percentage of RDOC$_t$ in SPE DOC ($f_t$).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Standard Value (%)</th>
<th>Parameter range (%)</th>
<th>$f_t$ range (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp. O</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c_l$</td>
<td>54.9</td>
<td>51.6~58.1$^a$</td>
<td>86.3~86.9</td>
</tr>
<tr>
<td>$c_c$</td>
<td>25.2</td>
<td>21.9~28.6</td>
<td>85.2~87.6</td>
</tr>
<tr>
<td>$c_{1X}$</td>
<td>2.1</td>
<td>1.4~2.8</td>
<td>84.6~88.5</td>
</tr>
<tr>
<td>$c_{5X}$</td>
<td>4.1</td>
<td>2.6~5.6</td>
<td>78.8~93.2</td>
</tr>
<tr>
<td>$p_l$</td>
<td>12.5</td>
<td>0.0~25.0</td>
<td>85.9~87.1</td>
</tr>
<tr>
<td>$p_c$</td>
<td>12.5</td>
<td>0.0~25.0</td>
<td>85.0~87.7</td>
</tr>
<tr>
<td>Exp. P</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c_l$</td>
<td>47.8</td>
<td>44.6~51.0</td>
<td>84.2~84.8</td>
</tr>
<tr>
<td>$c_c$</td>
<td>21.9</td>
<td>17.3~26.5</td>
<td>81.7~86.3</td>
</tr>
<tr>
<td>$c_{1X}$</td>
<td>2.1</td>
<td>1.4~2.8</td>
<td>82.6~86.5</td>
</tr>
<tr>
<td>$c_{5X}$</td>
<td>4.1</td>
<td>2.6~5.6</td>
<td>76.3~91.9</td>
</tr>
<tr>
<td>$p_l$</td>
<td>12.5</td>
<td>0.0~25.0</td>
<td>83.7~85.1</td>
</tr>
<tr>
<td>$p_c$</td>
<td>12.5</td>
<td>0.0~25.0</td>
<td>82.7~85.8</td>
</tr>
</tbody>
</table>

$^a$: the ranges of parameter $c_l$ and $c_c$ are set to be one standard deviation (calculated from the propagation uncertainty analysis) away from their mean values.
**Figure Legends**

**Figure 1.** Schematic representation of the transformation of deep oceanic DOC components in incubation experiments with different DOC concentration treatments. The SPE DOC in the deep ocean is composed of structurally recalcitrant RDOC_i, recalcitrant DOC due to dilution (RDOC_c) and labile DOC ready for microbial utilization (LDOC). The natural bacterial communities collected in the deep ocean are exposed to ambient and 5-fold concentrations of natural DOC collected from their original locations. (A) In controls, a fraction of LDOC is consumed by bacteria to generate new DOC and respire back to CO_2 or assimilate to POC. Both RDOC_c and RDOC_i are not consumed during this process. (B) In the 5-fold concentrated treatments, except the LDOC, a fraction of RDOC_c components whose concentrations become sufficient for microbial utilization are also utilized and contribute to the SPE DOC decrease. Detailed description of symbols can be seen in Table1.

**Figure 2.** Expected DOC utilization as a function of DOC enrichment when the RDOC is composed of intrinsically labile DOC (LDOC and RDOC_c) and RDOC_i. The simulation is the same as in Arrieta et al. (Arrieta et al., 2015a) but assumes that 0% (A), 10% (B), 50% (C), and 90% (D) of RDOC is RDOC_i. The 40-day time is denoted as a dotted line. The result of (A) is the same as Figure 1B in Arrieta et al. (Arrieta et al., 2015a) assuming that the DOC utilization is solely limited by concentration. Detailed simulation can be seen in Supplementary Methods.

**Figure 3.** Two possible interpretations for the small proportion of SPE DOC consumption in a 40-day incubation. (A) A large fraction of the SPE DOC is intrinsically labile and complete consumption of it would take a long time. This is what has been proposed in Arrieta et al. (2015a). (B) According to our estimates, there is a small amount of intrinsically labile DOC in the SPE DOC and it is rapidly utilized.
Figure 1.

A) Controls

Start \[ f_i \]

\[ \text{RDOC}_t \]

\[ f_c \]

\[ \text{RDOC}_c \]

\[ f_i \]

End

\[ \text{RDOC}_t \]

\[ \text{RDOC}_c \]

\[ \text{LDOC} \]

\[ \text{new DOC} \]

\[ \text{CO}_2 \text{ or POC} \]

B) 5-fold concentration experiments

Start

\[ \text{RDOC}_t \]

\[ \text{RDOC}_c \]

\[ \text{LDOC} \]

End

\[ \text{RDOC}_t \]

\[ \text{RDOC}_c \]

\[ \text{new DOC} \]

\[ \text{CO}_2 \text{ or POC} \]

\[ \text{LDOC} \]

\[ \text{new DOC} \]

\[ \text{CO}_2 \text{ or POC} \]
Figure 2.
Figure 3.

A  Dilution hypothesis

B  Structural recalcitrance hypothesis

Bulk DOC (t₀)  Bulk DOC (t₁)  Bulk DOC (t₂)

= intrinsically labile DOC (LDOC+RDOCᵢ)
= intrinsically recalcitrant RDOC (RDOCᵢ)
= consumed DOC

Time
Supplementary Methods

Data description

The dataset (http://digital.csic.es/handle/10261/111563) used in this study is taken from DOC enrichment experiments (Arrieta et al., 2015a), in which the dilution hypothesis was tested against whether bacterial growth increased with increasing DOC concentrations. The FT-ICR-MS fingerprints of SPE DOC at the beginning and at the end of the experiments in controls (1X DOC) and in concentrated samples (5X DOC) were also collected, which allow us to investigate the degradation of thousands of different molecules simultaneously. The FT-ICR-MS datasets, together with the change of SPE DOC concentrations, are used in our study.

Model description

As shown in the conceptual model (Figure 1), DOC is assumed to be composed of structurally recalcitrant DOC (RDOC), recalcitrant DOC due to dilution (RDOCc) and labile DOC ready for microbial utilization (LDOC), with their percentage in DOC pool being \( f_t \), \( f_c \), and \( f_l \), respectively (see Table 1 for details of variable indicators).

Here, LDOC compounds are labile, whose concentrations are above the corresponding microbial uptake thresholds in both controls and 5-fold concentrated treatments. RDOCc compounds are also intrinsically labile, but their concentrations are below the corresponding uptake thresholds in controls and above those in 5-fold concentrated treatments. We have

\[
f_t + f_c + f_l = 1. \quad (1)
\]
In controls, only LDOC is utilized, while RDOC\(_c\) is not utilized because its concentration is below the microbial uptake thresholds. A fraction of LDOC consumed by bacteria is regenerated to new DOC with the ratio of the regenerated DOC to consumed LDOC being \(p_l\). The rest, i.e., 1 - \(p_l\) of consumed LDOC, is respired to CO\(_2\) or assimilated to POC. The remaining components from RDOC\(_c\) and RDOC\(_t\) pools are not utilized either because of the dilution limit (i.e., RDOC\(_c\)) or structurally recalcitrance (i.e., RDOC\(_t\)). The process describing transformation of DOC components in controls can be represented as follows:

\[
f_t + f_c + f_1(1 - c_1) + f_l c_l p_l = 1 - c_{1X}, \tag{2}
\]

where \(c_{1X}\) denotes the percentage of consumed DOC in relation to SPE DOC in controls and \(c_l\) represents the ratio of consumed LDOC to bulk LDOC. By subtracting equation (2) from (1), we have following equation for controls:

\[
f_l c_l (1 - p_l) = c_{1X}. \tag{3}
\]

In 5-fold concentrated treatments, a fraction of RDOC\(_c\) compounds become bioavailable because their concentrations are high enough to support the bacterial metabolism. Therefore, except for the consumption of LDOC, which is supposed to be the same as that in controls, RDOC\(_c\) is also utilized. Therefore, for 5-fold concentrated treatments,

\[
f_t + f_c (1 - c_c) + f_c c_c p_c + f_1 (1 - c_l) + f_l c_l p_l = 1 - c_{5X}, \tag{4}
\]

where \(c_{5X}\) denotes the percentage of consumed DOC in relation to SPE DOC in 5-fold concentrated treatments, \(c_c\) represents the ratio of consumed RDOC\(_c\) to bulk RDOC\(_c\), and \(p_c\) is the new DOC produced per unit of RDOC\(_c\) consumed. By subtracting equation (4) from (2), we have
Equations (3) and (5) explicitly link the degradation of LDOC and RDOC components with the SPE DOC utilization. Then \( f_i \) and \( f_c \) can be derived from these two equations:

\[
f_i = \frac{c_{1X}}{c_i(1 - p_i)},
\]

\[
f_c = \frac{c_{S5X} - c_{1X}}{c_c(1 - p_c)}.
\]

Finally, the percentage of RDOC in SPE DOC can be estimated from equation (1) as:

\[
f_t = 1 - \frac{c_{1X}}{c_i(1 - p_i)} - \frac{c_{S5X} - c_{1X}}{c_c(1 - p_c)}.
\]

The required parameters, i.e., \( c_{1X}, c_{S5X}, p_c, p_i, c_e \), and \( c_i \), are estimated in the following section.

**Parameter estimation**

**\( c_i \) and \( c_e \):** Following the idea proposed by Arrieta and coworkers, a utilization index is derived from FT-ICR-MS data, which is defined by subtracting the remaining relative contribution at the end of controls from the initial relative contribution for each mass peak, as follows

\[
I^l = \frac{H^0 - (1 - c_{1X})H^{1X}}{H^0},
\]

where \( H^0 \) and \( H^{1X} \) are the normalized FT-ICR-MS signal of SPE DOC at the beginning and at the end of controls, respectively. The term \((1 - c_{1X})H^{1X}\) can be interpreted as the DOC normalized relative peak magnitude (Lechtenfeld et al., 2014).
A positive $I^l$ indicates the net utilization (bacterial usage minus regeneration) of the corresponding intrinsically labile compound. When $I^l$ is negative, it means that the regeneration of this compound is more than its consumption. As a result, $I^l$ provides the lower boundary of the compound production but no information about the decrease of this compound. Therefore, $c_l$ is estimated by averaging the utilization index $I^l$ on molecules whose index values are positive, i.e.,

$$c_l > c_l^* = \frac{1}{N_l} \sum_{I^l > 0} I^l,$$  \hspace{1cm} (10)

where $N_l$ denotes the number of molecules whose index values is positive.

Compared with controls, a fraction of RDOC$_c$ compounds are utilized in 5-fold concentrated treatments and the consumption of LDOC is supported to be the same. Thus, we can also identify the utilization index of RDOC$_c$ by comparing the relative contribution of FT-ICR-MS at the end of controls and 5-fold concentrated treatments:

$$I^c = \frac{(1 - c_{1X})H^{1X} - (1 - c_{5X})H^{5X}}{(1 - c_{1X})H^{1X}},$$ \hspace{1cm} (11)

where $H^{5X}$ is the normalized FT-ICR-MS signal of SPE DOC at the end of 5-fold concentrated treatments. Let $N_c$ be the number of mass peaks with positive values of $I^c$, we calculate $c_c$ as

$$c_c > c_c^* = \frac{1}{N_c} \sum_{I^c > 0} I^c,$$ \hspace{1cm} (12)

The terms $c_l^*$ and $c_c^*$ in equations (10) and (12) could possibly provide the lower limits of $c_l$ and $c_c$. This is because 1) the defined utilization index (i.e., $I^l$ and $I^c$) calculates the proportion of consumed labile DOC to SPE DOC, which would be smaller than the ratio of consumed labile DOC to labile DOC (i.e., $c_l$ and $c_c$); 2) due to
the dynamical transformation process of DOC, the observed utilization index actually represents a net decrease of intrinsically labile compounds that subtract the DOC regeneration from its consumption.

$c_{1X}$ and $c_{5X}$: Because no DOC consumption data are reported for Stations O and P (Arrieta et al., 2015a), the DOC consumption information from Stations K, L, and N are used as the proxies of that for both Stations O and P. The DOC consumption data from Station M is excluded because they are anomalous due to the observation of a second phase of intense growth (Arrieta et al., 2015a). The percentage of consumed SPE DOC is approximated to be $2.1\pm0.7\%$ for $c_{1X}$ (average ± SE) and $4.1\pm1.5\%$ for $c_{5X}$ (average ± SE).

$p_c$ and $p_l$: A portion of intrinsically labile DOC consumed by bacteria is converted into bacterial-derived DOC, representing approximately $25\%$ of the respired carbon (Stoderegger and Herndl, 1998). Considering that the bacterial respired carbon amounts to approximately $35.0\% \sim 99.0\%$ of the assimilated organic carbon (Del Giorgio and Cole, 1998), we have the ranges of the newly produced DOC per unit of LDOC or RDOC$ _c$ consumed from $0.0\%$ to $25.0\%$, and set the mean value of their range (i.e., $12.5\%$) as their standard value. The bacterial-derived DOC can be progressively used during the incubation, which leads to an even lower amount of newly produced DOC per unit of consumed intrinsic DOC in reality.

**Error propagation analysis**

Uncertainty propagation for $f_i$, $f_c$ and $f_l$ derived from the measurement variability of FT-ICR-MS fingerprints of SPE DOC is estimated as follows. Except for the controls at the beginning of the experiments, each scenario has two or three replicates for the FT-ICR-MS fingerprints of SPE DOC. We first obtain the distribution of measurement
variability by assembling the standard deviation of intensity for each molecule across all scenarios and determine its 95% confidence level as the measurement variability \( \sigma_m \). The confidence level is chosen to allow us to investigate an exaggerated uncertainty of measurement variability on the estimation of \( f_i \). Applying the error propagation formula, which is the function of measurement variability (Glover et al., 2011), to equation (10):

\[
\sigma^2(c_i) = \frac{1}{N_i^2} \sum_{l \geq 0} (1 - I^l)^2 \left\{ \frac{\sigma^2(H_1^X)}{(H_1^X)^2} + \frac{\sigma^2(H_0^0)}{(H_0^0)^2} \right\} \tag{13}
\]

104 Based on equation (6), the error for \( f_i \) propagated from the \( c_i \) is estimated as

\[
\sigma^2(f_i) = \left( \frac{c_{1X}}{1 - p_l} \right)^2 \frac{1}{c_i^2} \sigma^2(c_i). \tag{14}
\]

105 Assuming that all molecules share the same measurement variability, i.e., \( \sigma_m^2 = \sigma^2(H_1^X) = \sigma^2(H_0^0) \), by plugging equation (13) into equation (14), we have

\[
\sigma^2(f_i) = \left( \frac{c_{1X}}{1 - p_l} \right)^2 \frac{\sigma_m^2}{N_i^2 c_i^4} \sum_{l \geq 0} (1 - I^l)^2 \left\{ \frac{1}{(H_1^X)^2} + \frac{1}{(H_0^0)^2} \right\} \tag{15}
\]

107 Similarly, the error for \( f_c \) propagated from the FT-ICR-MS fingerprints of SPE DOC is estimated as

\[
\sigma^2(f_c) = \left( \frac{c_{5X} - c_{1X}}{1 - p_c} \right)^2 \frac{\sigma_m^2}{N_c^2 c_c^4} \sum_{l \geq 0} (1 - I^c)^2 \left\{ \frac{1}{(H_5^X)^2} + \frac{1}{(H_1^X)^2} \right\}. \tag{16}
\]

109 With the simplistic assumption that the error propagations of \( f_i \) and \( f_c \) are independent, the error propagation of \( f_i \) could be represented as

\[
\sigma^2(f_i) = \sigma^2(f_i) + \sigma^2(f_c). \tag{17}
\]

111 **Sensitivity analysis**

112 Analyses are conducted to investigate the sensitivity of \( f_i \) to each parameter over wide ranges. Six parameters in equation (8) are considered. Specifically, the ranges of \( c_i \)
and $c_c$ are set to be one standard deviation, which is calculated from the error
propagation analysis, away from their base values. The ranges of $c_{1X}$ and $c_{5X}$ are set to
be one standard deviation away from their base values as well. Both $p_l$ and $p_c$ are
varied from 0.0% to 25.0% to cover most likely range.

**Calculate the consumption of SPE DOC as a function of DOC enrichment**

In these simulations, we explore how the utilization rate of SPE DOC changes as a
function of DOC enrichment given that DOC is composed of intrinsically labile DOC
(LDOC and RDOC$_c$) and RDOC$_t$. Our simulation is based on the simplification that
all components in the RDOC$_c$ pool are available for microbial utilization in
concentrated experiments, which could possibly lead to overestimation of the SPE
DOC utilization rate. Consistent with simulations in Arrieta et al. (2015b), we
consider four different DOC concentration levels (controls and the corresponding
two-, five- and ten-fold concentrations of nature DOC). Let $x(t)$ be the concentration
of intrinsically labile compounds at day $t$. Assuming that the utilization rate is a
function of concentration and cell abundance, we have

$$\frac{dx(t)}{dt} = x(t)K_sS,$$

in which $K_s$ is the Monod substrate affinity constant and $S$ the maximum cell
abundances observed in the experiment. The solution of equation (18) can be
represented as:

$$x(t) = x(0)e^{-K_sSt}.$$  

Let $x(0)$ be the initial DOC concentration:
\[ x(t) = z(0)e^{-K_s t}, \]  

in which \( \gamma = f_i + f_c \). \( z(0) \) is set to 30 (in arbitrary units), and the cell abundances \( S \) for controls, two-, five- and ten-fold concentration treatments are 55, 67, 97, and 139 \( \times 10^6 \) prokaryotes \( \text{L}^{-1} \), respectively (Arrieta et al., 2015b). The substrate affinity constant \( S \) is calculated by reconciling the proportion of consumed SPE DOC at \( t = 40 \) in equation (20) to that of the observed values in the controls (i.e., \( c_{1X} \)). Therefore,

\[ K_s = \frac{\ln (1 - c_{1X}/\gamma)}{-40 \times 55 \times 10^6}. \]  

(21)

By plugging (21) into (20), the proportion of consumed SPE DOC at \( t \) day is given as:

\[ c(t) = \frac{x(t) - x(0)}{z(0)} = \gamma(e^{-K_s t} - 1). \]  

(22)

In our simulation, \( \gamma \) ranges from 0 to 1. All SPE DOC is \( \text{RDOC}_t \) when \( \gamma = 0 \) and is \( \text{RDOC}_c/\text{LDOC} \) when \( \gamma = 1 \). We also conducted sensitivity analysis to investigate the change of \( c(t) \) by varying \( c_{1X} \) from 1.4% to 2.8% in equation (21).

**Reference**


Supplementary Tables

**Table S1.** Matrix effect on estimating the percentage of RDOC$_i$ in SPE DOC ($f_i$). Here, we consider the matrix effect caused by the incubation process (i.e., change of signal intensity from DOC component during incubation experiments) and concentrated treatments (i.e., change of signal intensity due to DOC concentrated treatments). The measurement uncertainty caused by matrix effect (i.e., $\sigma_{ME}$) is calculated as the 95$^{th}$ percentile of signal intensity difference by comparing 1X initial and 1X final FT-ICR-MS, 1X initial and 5X initial FT-ICR-MS, 5X initial and 5X final FT-ICR-MS, respectively. We then investigate the matrix effect on the estimation of $f_i$ according to equation (17) in Supplementary Methods.

<table>
<thead>
<tr>
<th></th>
<th>Exp. O</th>
<th>Exp. P</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{ME}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1X initial, 1X final</td>
<td>0.084</td>
<td>0.073</td>
</tr>
<tr>
<td>1X initial, 5X initial</td>
<td>0.074</td>
<td>0.040</td>
</tr>
<tr>
<td>5X initial, 5X final</td>
<td>0.066</td>
<td>0.068</td>
</tr>
<tr>
<td>Max. of $\sigma_{ME}$</td>
<td>0.084</td>
<td>0.073</td>
</tr>
<tr>
<td>Range of $f_i$, %</td>
<td>82.1–89.1</td>
<td>77.8–87.9</td>
</tr>
</tbody>
</table>
Figure S1. Histogram distribution of measurement variability of FT-ICR-MS fingerprints of DOC by assembling all molecules together. The measurement variability for each molecule is calculated as the standard deviation of its relative intensities from replicates. The average measurement variability FT-ICR-MS (i.e., $\sigma_m$) is chosen as the 95th percentile of this distribution for the downstream propagation uncertainty analysis.
Figure S2. Comparison of signal intensity of FT-ICR-MS to assess DOC composition change during the incubation experiments (A, B, D, E) and the matrix effect due to DOC concentrated treatments (C, F). The diagonal line is shown in the red color in each figure. The close to the diagonal line means that the overall FT-ICR-MS signal pattern is similar among different treatments.