A molecular perspective on the ageing of marine dissolved organic matter

R. Flerus[1], B.P. Koch[1,2], O.J. Lechtenfeld[1], S.L. McCallister[3], P. Schmitt-Kopplin[4], R. Benner[5], K. Kaiser[5], G. Kattner[1]

[1]{Alfred Wegener Institute for Polar and Marine Research, Ecological Chemistry, Bremerhaven, Germany}

[2]{University of Applied Sciences, Bremerhaven, Germany}

[3]{Virginia Commonwealth University, Department of Biology, Center for Environmental Studies, Richmond, VA, USA}

[4]{Helmholtz Zentrum München, German Research Center for Environmental Health, Analytical BioGeoChemistry, Neuherberg, Germany}

[5]{University of South Carolina, Marine Science Program, Columbia, SC, USA}

Correspondence to: B. P. Koch (Boris.Koch@awi.de)

We thank the anonymous reviewers for their valuable comments. Both reviewers consider parts of our normalization and data evaluation procedure to be too complicated. The reviewers were of different opinions regarding the focus of our manuscript. Whereas reviewer 1 judges our results and interpretation to be valuable for the DOM community, reviewer 2 suggested to focus rather on chemical characteristics than combining 14C values with SPE-DOM extracts.

We think that our data interpretation, especially the degradation continuum, which is the main focus of our manuscript, is very important for future research on the sources and fate of marine DOM. In addition, it is important to combine FT-ICR MS data with other parameters of bulk DOM, because FT-ICR MS is a valuable and more and more used tool for marine DOM analysis. Combining these data provides a much broader insight in processes rather than using every parameter isolated.

We decided to keep the focus of our paper. We reworked the chapter of the normalization procedure with respect to clarify the procedure. We also reworked the comparison of young vs. aged samples, added semi-quantitative information and improved the chapters “degradation continuum” and “microbial carbon pump” to clarify our interpretation and hypotheses.

Specific response to the comments:

(In the case of very long reviewer comments, we extracted the questions of interest)

R1C-1: Page 2, Lines 21-22: "All other compounds should persist partly thermohaline circulation." This sentence reads quite strangely to me, and I think a preposition is missing somewhere. Perhaps adding 'during' or 'throughout' is better? "All other compounds should persist partly during/throughout thermohaline circulation." This sentence is also in the caption to Figure 10 and needs updating there as well.

AC: *We followed the suggestion and rewrote this sentence accordingly.*

R1C-2: Page 3, Line 15: "modern" refractory DOM seems to be an oxymoron. Do you mean younger, refractory DOM? An additional explanation to this sentence and the Jiao et al. citation would be useful here.

AC: *We followed the suggestion and added the Jiao et al. citation.*

R1C-3: Page 5, Lines 8-11: The reference given for the water sampling details is a manuscript that is in preparation. A few more sentences describing the sample stations would be useful for readers.

AC: *We added a map of the sampling stations to provide better information to the reader.*

R1C-4: Page 6, Lines 4-5: How much ultra-pure water was used to re-dissolve the 50 uL of dried extract? Add this to the experimental information.

AC: *We used 6.5 mL of ultra-pure water to re-dissolve the 50 µL of dried extract. We added this information to the manuscript.*

R1C-5: Page 7, Lines 12-14: Section 2.4 states that the mass spectra are evaluated at 200-500 m/z, but the spectra are internally calibrated at 340-540 m/z, with only 5 peaks...Why is the calibration list used so short? How can the authors be confident that 0.5 ppm is a sufficient limit for formula assignments?

R1C-6: Page 7, Line 18: Why do you only use a m/z range of 200-500? There is quite a lot of data at 500-600 m/z (according to Fig. 1)…

AC: *In order to achieve maximum comparability to our previous datasets, our mass calibration was carried out according to the procedures in earlier published studies (Koch et al. 2008, Flerus et al. 2011). The mass accuracy for the internal calibration was generally better than 0.2 ppm. We expect the (higher) threshold of +/- 0.5 ppm to be adequate for correct formula assignment. The mass limit of 500 m/z was incorporated as an additional level of quality control. However, to comply with the reviewer's suggestion and after thorough verification we expanded the data evaluation to m/z 600 and checked the validity of the trends also for this expanded mass range. All trends described in the manuscript were still the same as observed before. We added this result as well as the citations for mass calibration in our previous work.*

R1C-7: Page 8, Line 3: I understand that the peak magnitudes were summed for each mass spectrum, but I do not understand the intensity threshold described as 100 ppm. …What S/N threshold is used in the Bruker software to construct the mass lists, prior to summing the peak magnitudes? …

R2C-1: For example, the authors speak of a 100 ppm threshold – what exactly is this?

AC: *We reworked the description of our normalization procedure with the focus on better explaining and verifying the single steps which were applied. The basic mass list was obtained from the Bruker software applying a threshold of S/N 3, which was given by the software. The normalization to the sum of all magnitudes in the spectra was directly performed on the basic mass list.*

R1C-8: Page 8, Lines 24-25: Regarding Figure 3, why are there no values for the y-axes shown on the plots? The authors state how much magnitude the summed peaks account for in all the spectra (66 + 1.4%), so why aren't values shown on the y-axis?

R2C-2: ...you can certainly put a number on the y-axis in Figure 3. Without an axis, I have no way of evaluating the slope in Figure 3.

AC: *We added the values for the y-axes in ‰.*

R1C-9: Page 9, Line 4: I believe Table 1a should just be Table 1.

AC: *Right and changed accordingly.*

R1C-10: Page 10, Line 10: For Figure 6, it may be useful to put a line on the van Krevelen plot at H/C=1.0. The y-axis only shows 0 and 2.5, and I think having a line at 1.0 would be beneficial to readers.

AC: *We followed the suggestion.*

R1C-11: Page 10, Lines 26-28: The authors give H/C ratios for young DOM as 1.27 and 1.25 for aged DOM. Are these 2 values statistically different from each other? If you compare the magnitude-averaged H/C ratio for all 137 samples, is a difference of 0.02 actually statistically different?

R2C-3: Provide examples of compounds that have an H/C of 1.27 versus 1.25. Are these values significantly different?

AC: *We added a figure showing H/C ratios vs age. The figure shows perfectly the correlation between the H/C ratios and the DOM age.*

R1C-12: Page 10, Line 29: For Figure 7, how are the dotted lines actually calculated? It is not clear from the figure caption, and these lines are not discussed in the text. I think they could easily be deleted from the figure to avoid confusion, since the authors don't actually discuss them in the text.

AC: *We added the lines to simplify a comparison with the data presented in Hertkorn et al 2006.*

R1C-13: Page 11, Lines 10-11: The authors state that CRAM is obviously a significant contributor to SPE-DOM. This is not obvious. This statement is based only on the location of points in the van Krevelen diagram, which is an area of overlapping functionality that is likely due to any number of structural isomers of formulas that could be assigned to a variety of biomolecular compound classes.

AC: *A recent study by Witt et al. (2009, cited in the manuscript) shows that the structural variability of a single elemental formula in the FT-ICR mass spectra is probably not as high as expected for natural organic matter. From that point of view we think that there is a link between the trends presented in the manuscript and CRAM, which is presented in Hertkorn et al. 2006. We added some critical thoughts to this section of the discussion part, including also the possibility of structural isomers.*

R1C-14: Page 11, Line 20: For Table 2, I think the caption should be 'Formulas utilized for magnitude summation in order to calculate the degradation index' or something more along those lines. The wording of the current caption is confusing.

AC: *We followed the suggestion.*

R1C-15: Page 11, Degradation Index: The authors give the equation but do not justify it. How was the calculation developed? Because this index is an important part of the manuscript, the authors should do a better job of describing the index and rationalizing the calculation.

R2C-3: I haven't quite figured out why the index has 100 and +1 in the denominator – seems more complicated than necessary. Also seems that it should be normalized to a particular depth.

AC: *We agree with the reviewers and reworked both with focus on simplifying, the degradation index as well as the discussion about the development to make it a tool which can be easily used by other scientists.*
We don´t normalize the degradation index to a particular depth, since we think, that the degradation state e.g. in the deep Pacific Ocean compared to the Atlantic Ocean could be different at the same depth.

R1C-16: Page 11, Line 22: The authors state that the index can differ due to the instrument. Does that mean which specific FTICR-MS was used? What are the differences between the instruments utilized by the authors (magnetic field, manufacturer, operator, installation time-frame, etc.)

AC: *The distribution of the exact magnitudes depends e.g. on the instrument, the concentration of the injected sample (ions in the ICR cell) and the settings of an instrument. Hence, the final value of the degradation index will differ slightly from instrument to instrument. But trends observed for a data set analyzed under similar conditions will be the same (Kido Soule et al 2010). Due to that only trends in the degradation index within one sample set should be compared.*

R1C-17: Page 12, Lines 1-2: The authors give degradation index values in the range of 14.6-40.1, yet the coloring scale on Figure 8 appears to be more narrow (18-28). What is the reason for this discrepancy?

AC: *Figure 8 shows only surface water samples, since ODV is not useful for single samples as were analyzed in the deep sea. The range described in the manuscript is referred to all samples. These values were presented in an extra table in the manuscript.*

R2C-4: The quality of this relationship depends strongly on the variability in "normalized" magnitude for a particular peak in a particular sample. How conserved is this for a single sample over different days and instrument conditions? I think this concern is what drives other authors to focus on presence-absence rather than intensity. No information is provided on this important parameter– e.g., were duplicates run to constrain y-axis error bars in Figure 3? The rest of the paper cannot be appreciated without this information.

AC: *The extract of one young sample and one old sample were run at a later time at another instrument (data unpublished). The normalized magnitude trends were still the same and the trends were reproducible. A systematic reproducibility study was carried out recently (Kido Soule et al. 2010) demonstrating the reliability of the method also in terms of peak magnitude.*

R2C-5: This was necessary because samples were not all run at the same DOC concentration (factor of 3-9 difference across samples)?... p.11459. "Samples were adjusted to similar DOC concentrations"
- How were dilution factors (3-9) considered when comparing intensities across spectra; was this achieved by normalizing to total MS area?

AC: *All SPE-DOM extracts were diluted to the same extract DOC concentrations before measurement to avoid influences due to different numbers of ions in the ICR cell. Data evaluation was performed only on the normalized magnitudes of the diluted samples to obtain compositional trends. These trends were not quantitative.*
Inspired by the reviewers comment we decided to even add a semi-quantitative estimation: we multiplied the normalized magnitudes with the DOC concentrations and added an additional discussion to the manuscript. This would be valid since the extraction efficiency is the same for all extracts. The trends were still the same and a semi-quantitative estimation could be performed.

R2C-6: If the data were normalized as I imagine them to be (i.e., relative to the total MS) then aren't the peaks that increase in relative intensity just one component of a two component mass balance?...

AC: *This is of course correct – especially when considering only two peaks. However, we were particularly interested in the slopes of relative magnitude changes for each different peak in the 137 samples. The variety of different slopes, which can be found for the NEG and POS correlating compounds, was in the focus of data interpretation.*

We added a semi-quantitative calculation (see also last question), which showed that probably all compounds decrease with age – the POS correlating compounds most and the NEG correlating compounds least. The negative slope is different for every mass formula.

R2C-7: p.11461 line 4: "we calculated "average" "relative magnitude" "ratios"" - I have no idea what this means (is "relative magnitude" equivalent to their previous normalized ratio?). What does the ratio of 1 represent?
… Basically, what I do understand is the following: the authors are trying to use their data to identify molecules that are removed with depth (100m versus >800 m) versus those that are not…

AC: *We agree that this passage was a bit confusing. We omitted the "relative magnitude ratios" and used instead a semi-quantitative approach by multiplying the normalized magnitudes of each spectrum with DOC concentrations. With this it was possible to estimate a decreasing/increasing rate of SPE-DOM compounds by comparing the spectra of young and old SPE-DOM. Since it was not possible to compare each single young sample with each single aged sample, we compared average spectra of youngest samples (which were found in the upper 100 m) with average spectra of the oldest samples (which were found deeper 800 m).*

Semi-quantitative: *we use here "semi-quantitative" because we estimate only the degradation rate – no absolute quantitative values of the total amount of the compounds.*

R2C-8: There is no way to quantify the fraction of total SPE-DOM that is being ionized and analyzed via this technique. No attempt is even made to do this – for example, how does C:N of the bulk SPE fraction compare to C:N calculated via FTICR-MS? The point here is that the peaks visible by FTICR do not necessarily represent molecules that drive the radiocarbon content of SPE-DOM. For example, all of these peaks could represent compounds with a modern radiocarbon signature throughout the

ocean. Even in that case they could show an intensity relationship with radiocarbon. Of course, this would require all of these peaks to represent only a very small fraction of the PPL-resin, but there is no evidence to the contrary. In summary, the fact that certain peaks show a linear relationship in their intensity to bulk SPE-DOC _14C provides no information on the _14C signature of that compound.

AC: *This is an important thought which we addressed in the discussion. It is true that we don't know how large the ionized proportion of our extract is. Even the C/N ratios don't allow better insights on this since the ionization efficiency of N-containing and N-free compounds might be different (magnitude weighted C/N values were around xx and bulk C/N values around xx in the surface).*

However, we think that it is reasonable to assume that peaks which show a very good correlation of relative magnitude and radiocarbon content can reflect molecular changes caused by ageing. In this context it is not ultimately important whether these molecules have the same age or just correlate with radiocarbon content. We think, the reasonable depth-dependent and spatial distribution of these age correlated compounds (as reflected in the age index) still legitimates their use as age indicators.

To comply with the reviewers comprehensible and important comments we added a paragraph to the discussion.

R2C-9: I don't understand the value of the _14C LIM exercise in the context of FTICR-MS data. I would have liked to see all the various relationships that went into generating Figure 10. How different were the slopes of peak intensity versus radiocarbon signature, and wouldn't the point be made more clearly if intensity was plotted against water mass age or depth?

AC: *For the values: see R2C-7. The degradation continuum could also be shown using the water mass age or the depth. However, Figure 10 shows, in addition to the continuum, that most of the compounds would still be present in the oldest water mass in the Pacific Ocean. This could not be shown with depth – because the depth is not increasing from the Atlantic Ocean to the Pacific Ocean, but the average age of DOM and the amount of bulk Dom are changing. Water mass age could be an alternative, but it is hard to find correct literature values for that .Here it is a similar problem as using the depth. E.g. decreases the average age of DOM much faster than the water mass age, especially comparing the deep Atlantic with the deep Pacific Ocean.*

R2C-10: Would peak intensity versus density or salinity give you a linear relationship for many molecular ions? Using that relationship would it be correct to calculate a salinity/density value at which that compound was no longer present? For example, DOC is often linearly related to density or temperature, but extrapolating that relationship to a DOC concentration of 0 is meaningless.

AC: *There is no linear relationship with salinity but there is one with temperature. We used the radiocarbon age as a marker for a potential processing state of our extracts. This would not be possible using the temperature.*

The extrapolation of our compounds to "0" is of interest, because our samples are from the Atlantic Ocean. All compounds represented in the mass spectra decreased with different slopes – it is of interest to find out how many of these compounds would be "0" with further ageing of the DOM (and DOM is getting older on the way from the Atlantic Ocean to the Pacific Ocean).

Our result – that only few compounds will be "0" in the Pacific Ocean – is of interest, because that means, that a lot of other compounds are probably at least old as one Ocean Cycle and will probably

still be present, if they come again to the surface. We cannot verify this at all – but we think, that this hypothesis of compound cycling and remixing in the surface – which is based on our data - is an important thought for future research in marine DOM cycling.

R2C-11: p.11459. Neither blank information nor sample size information is provided for radiocarbon data… Yet, if radiocarbon samples were big enough (>200 ug C or so) then blanks may not be an issue.

AC: *We added additional information. Sample sizes for radiocarbon SPE ranged from a minimum of 150 μg to over 1 mg with the majority of samples with the majority of samples (20) 260 μg C or greater. Given the relatively large samples size any related process blank should have minimal influence on the reported values.*

R2C-12: It is curious that most compounds that are POS related to radiocarbon have essentially the same H/C ratio (differing primarily in O content). This is somewhat true for the negatively correlated peaks as well. Are these compounds structurally related? Could some of them be fragments of larger molecules?

AC: *The ionization source was setup to reduce all possible fragmentation during ionization process - and thus the signals can be expected not being fragments of higher molecular weight compounds.*

R2C-13: Paragraph beginning on line 11. Some examples of compounds that fit into the VK space being discussed would be useful. The paper would also benefit from some simple conclusions (if they are accurate): Do compounds that dominate the deep ocean SPE extractable, ES ionizable DOM reservoir appear to contain a greater number of double bonds? This could be consistent with the conclusions of Lam et al (2007).

AC: *We agree, but this is only useful, if the diagram and the example compounds are interpreted in the right way. It might lead to misunderstanding of what can be seen in the van Krevelen Diagram. The elemental formulas obtained from FTMS spectra give no structural information and since only a very small part of DOM can be identified as e.g. carbohydrates or lipids or lignin, we would like to avoid, that compounds plotted in the van Krevelen diagram would be seen as these compounds.*

In the chapter "molecular characteristics of DOM diagenesis" we described an increasing DBE (double bound equivalent) with age (and therefore with depth). A higher DBE means a higher number of double bounds.

R2C-14: It should be explicitly noted here that amino sugars were measured on bulk DOM and not SPE extracts (at least this is what the methods currently imply). How does the relationship in Figure 9a compare to the relationship of amino sugar concentrations (yield) to depth?

AC: *we reworked the figure subscription and added the information.*

R2C-15: The paragraph starting on line 22 is not particularly illuminating. Bacterial activity is high in a region with high DOC concentration and enriched _14C signature – and this somehow provides evidence to indicate bacterial reprocessing of SPE-DOM? In fact, their calculated IDEG is low at this site, suggesting, in their interpretation, the presence of DOM that has not been degraded or altered by bacterial reprocessing. This again strikes me as a misplaced sentence.

AC: *We agree with the reviewer, that this is confusing. We reworked this paragraph with the focus on better explaining.*

R2C-16: p. 11465 section 3.3. Were keeling plots constructed using total DOC concentrations or the calculated concentration of SPE-DOC? The fresh endmember contribution should, by definition, have the same radiocarbon signature in all and any DOM fraction being examined. In Table 3, provide information on the quality of the fit to the data. Line 13-16 - The authors have lost me again here with regard to the relationship between keeling plots, reactivity continuum and relative magnitude ratios. Keeling plots typically assume the presence of only two radiocarbon endmembers.

AC: *We used the bulk DOC concentrations and the SPE-DOM age for this calculation.The intention of this calculation was to check the validity of trends in calculcated SPE-DOM age (+ trends) for bulk DOM by comparingg our results to the results in Beaupre et al (2009, 2010)*
We agree that the Keeling plots typically assume the presence of only 2 endmembers. But if we understood in the right way, the calculation did also fit not perfectly as shown in Beaupre et al. (2010). This discrepancy and our results lead us to the conclusion, that in general there are more than only 2 endmembers.

Our hypothesis: The average age of DOM in the water column is not only driven by mixing of these 2 pools. A high portion of very fresh material produced by primary production (DOM 1) is processed in the surface water masses to a different kind of DOM of unknown compounds (DOM 2). We think that we can see this fraction in the FTMS spectra. The production of DOM2 is done by bacteria. DOM 2 is mixed with water masses, but with time it is reworked (also by bacteria?). DOM 2 consists of thousands of different compounds with different reworking rates. Most of these compounds are not fully reworked during thermohaline circulation, so older parts of these compounds are yet present in the surface. Hence the average age of these compounds is not modern.

Finally, we have 2 processes running parallel: Mixing and reworking of DOM1 (only modern) and DOM2 (modern and old – several different ages) from surface to deeper water layers and with water circulation. Both processes drive the 14C of bulk DOM. This hypothesis would explain, that the 2 component model fits somehow, but not perfectly.

The presence of a very old background component (e.g. black carbon) is not excluded using this model of processes.

R2C-17: Finally, this dataset has the potential to provide some interesting molecular-level information and it frustrates me to see that being abandoned in favor of over-interpreting the radiocarbon data.

AC: *We don´t agree with this. We used the radiocarbon age as a marker for a potential processing state of our extracts. With this, we could also distinguish in the upper layers <200m between water masses with "freshest" SPE-DOM and water masses with not so "fresh" DOM. This was only possible using the 14C age of the extracts. Also if there was a correlation with temperature – this parameter gives no information about the diagenetically DOM-state. The degradation index, which was developed, based on the correlation between SPE-DOM 14C and mass peaks, provides a tool to follow the diagenetic state of SPE-DOM on a relatively high resolution in the upper water columns and maybe also in the deep waters of different world Oceans.*

Literature:

Kido Soule, M. C. K., Longnecker, K., Giovannoni, S. J., and Kujawinski, E. B.: Impact of instrument and experiment parameters on reproducibility of ultrahigh resolution ESI FT-ICR mass spectra of natural organic matter, Organic Geochemistry, 41, 725-733, 10.1016/j.orggeochem.2010.05.017,