

Below copied comments from R. Bundy. In between her text (**Reviewer**) our answers are given (**Answer**)

Comments on “Comparing CLE-AdCSV applications using SA and TAC to determine the Fe binding characteristics of model ligands in seawater”

Reviewer:

The authors present a very careful and interesting study that compares some of the most common CLE-AdCSV techniques used to measure the organic speciation of dissolved iron (dFe) in seawater. These experiments are very painstaking to do, so I commend the authors on such an undertaking.

Answer:

Dear Randelle Bundy.

Thank you for your reaction on our manuscript. Your comments, below, are interesting and help to improve the discussion on the voltammetric applications.

Reviewer:

Overall, I found the comparison between techniques using model ligands an important contribution to the field, but unfortunately it appears that there might have been some major issues with the application of the SA methods particularly in the kinetic experiments (e.g. Figure 6). On a broader note, I also believe that this paper is demanding more than can be expected out of these electrochemical methods. These voltammetric approaches have always been operationally defined (as all seawater iron analysis methods are) and it has been known for some time that applying these methods at different analytical windows will give you a methodspecific perspective on the continuum of iron ligands in seawater.

Answer:

We agree fully that the methods are operationally defined and we also prove this in our manuscript. Indeed, the detection window is one of the criteria with which we can explain differences between different applications. However, therefor we must know the detection window, and in order to calculate the detection window, we must know which Fe-AL species are formed and which are electro active.

We wish to understand factors that control Fe biogeochemistry in the ocean, and in order to do so it seems reasonable to expect that the methods used to do this 1) obey assumptions required for the data transformation and 2) produce comparable results with minimum uncertainty and 3) resolve quantitatively the simple case, of a single ligand present. If the methods we use for our investigations do not comply with these requirements we have to ask whether what we are doing is worth the effort and time required. In our study we analyzed model ligands, in an effort to see if the methods applied comply with these three criteria. Ultimately our results suggest we obtain numbers that can neither be compared between techniques and also provide dubious quantitative information about the research question. Since results were sometimes higher or sometimes lower than expected, we have to highlight at least that the uncertainty in results is probably much higher than the uncertainty obtained just from curve fits. Where does this lead interpretation of what are typically quite minor changes in L and K in global ocean data sets?

We started the experiments for two reasons. Firstly, already for some time there has been a discussion about standardizing the ligand analysis methods, and in fact this is also an important component of the GEOTRACES program. Our manuscript could be a start. The second reason

was coming from the work of Slagter et al, (2019). Slagter et al., compared TAC and SA in the Arctic. We knew TAC was not performing very well when humics are present, so we wanted to use SA. However, we also wanted to connect the future with SA to our past with TAC. So we applied both to a part of the Arctic samples. We then saw a more or less constant ratio between the obtained [L], without any relation to the presence of the humic rich transpolar drift. That was surprising. Our co-author Indah Ardiningsih extended the comparison by comparing TAC SA and NN. Again the results were puzzling.

Reviewer:

Bruland et al. (2000)

highlighted this very eloquently, and although that intercomparison applied to Cu speciation, its findings are applicable to Fe speciation as well. The main goal of these CLE-AdCSV methods has never been to quantify a specific ligand's concentration and binding strength, but instead to get a broad picture of the multitude of ligands that are capable of binding Fe in seawater, and to qualitatively evaluate this in seawater where a continuum of ligands exists.

Answer:

We do not agree with the above. There are many publications that split the ligands into more than one fraction or describe ligands as “weak” or “strong”, even assigning the nature of the natural ligand based on the comparison of K from natural ligands and model ligands as we did in this work. All these interpretations are dependent on the idea that L and K can be reliably quantified with low uncertainty. Indeed, this kind of splitting was recommended in the review of Gledhill and Buck (2012), but our study suggests such splitting could result in over interpretation of data. The issue is likely less significant for Cu than Fe since the excess ligand concentration is typically larger for Cu, Cu is quite soluble (and shows much less affinity for cell walls, mercury drops etc.) allowing for much greater confidence that equilibrium can be achieved.

Reviewer: From studies where

these methods have been applied on the basin scale, we have gained unparalleled insights into Fe and ligands dynamics in the ocean, that are robust and oceanographically consistent and are able to be captured in global biogeochemical models (Tagliabue et al. 2017).

Answer: Of course we accomplished a lot (Here we mean with ‘we’ and below ‘our’ not the authors and the reviewer but the whole community working on the subject.). But that does not mean we cannot improve. Moreover, as far as we know Tagliabue et al. do not use our data because it does not fit their models. Moreover, even if a specific model converges to a solution does not mean that the data are robust. This must be ascertained with experiments as the one shown here, where a simple system must be perfectly resolved. Modellers are still struggling to fit DFe in the global models.

As an illustration we quote Ye et al., 2020:

“Global ocean biogeochemical models consider the organic complexation of iron with different degrees of complexities (Tagliabue et al., 2016). While some still assume constant ligand concentrations, others have been trying to capture the spatial variability of organic ligands with different assumptions of their relationship to biological activities. For example, Misumi et al. (2013) linked ligand concentration to oxygen consumption, and Tagliabue and Völker (2011) considered an empirical relationship between total ligand concentration and DOC. In common with other global iron models, we previously described the role of iron-binding ligands **in the iron cycle with a constant ligand** (Ye & Völker, 2017) and with a

prognostic ligand assuming a source of ligands from DOC (Völker & Tagliabue, 2015). Although surface DFe distributions in these models encompassed a similar range to observed DFe concentrations, there were some systematic discrepancies: DFe concentrations are generally too low in the Pacific Ocean, and the models could not reproduce the observed inter-basin pattern in the deep ocean when keeping the range of concentrations close to measured data (for details, see section 3.1.4). These discrepancies indicate that some processes controlling the spatial variability of ligands and DFe are still likely missing in biogeochemical models of iron.”

Reviewer: Thus far, these are the only methods we currently have to learn about the Fe-binding ligand pool as a whole. Other methods exist to measure specific ligands and functional groups, and what we have learned from those techniques is thus far unable to help us explain global distributions of dFe. It would be a shame in my opinion, to not comment on the valuable insights CLE-AdCSV techniques have brought us over the years, and to reduce the findings of this paper to a take home message that these methods are fatally flawed and we need new methods for measuring organic Fe speciation. Innovation in techniques is always a good idea, but I worry about how this paper might impact the direction of the field. Given the stature and expertise of the authors, I would have appreciated some discussion of the strengths of CLE-AdCSV methods and the instances of where and when they do work, and where and when they should be applied with caution and pause. I have outlined some of my other major comments and thoughts below, and have noted some additional specific comments at the end of the document.

Answer: We agree that the valuable insight the method brought us is a good addition to the introduction.

In the introduction we added at lines 64-68:

“The application of this method enlarged our knowledge on the marine chemistry of Fe and the results formed the base of the explanation why DFe depth profiles deviated from those of other trace metals (Van den Berg, 1995; Rue Bruland, 1995; Hutchins et al., 1999; Croot et al., 2004; Laglera and van den Berg, 2009; Gledhill and Buck, 2012; Bundy et al., 2014; Buck et al., 2015; Hassler et al., 2019; Lauderdale et al., 2020).”

We added in section 4.2-Overall, lines 641-643:

“In future use of CLE-AdCSV careful consideration is needed for the interpretation of the obtained [L] in relation to the application and environmental variations in ligand groups, especially the humic substances.”

We did suggest how to continue with CLE-AdCSV, we end in the conclusion with the text:

At line 781: “In case voltammetric methods are used, we advise the SA5-application but to be careful to draw conclusions from obtained K^{cond} -values, since apart of the constraint forced by D, these estimates have a much larger error than apparent from those obtained from the fits to the Langmuir isotherm.”

Still one cannot turn a blind eye to results that show the technique limitations.

Reviewer:

General comments

Kinetic experiments and the loss of signal with the SA method: The authors mention that all vials were conditioned overnight prior to the start of experiments, but in my lab, we have found that

thorough conditioning for usually more than one week is absolutely required for accurate titrations and optimal sensitivity with the SA method (Abualhaija and van den Berg, 2014).

Answer:

We did not condition only overnight, we repeated overnight equilibration five times, hereafter a copy of the text from methods '3.1.2 Conditioning and equilibration', at lines 214-216:

“Before use, all materials such as vials, bottles and cells, were conditioned overnight with the prepared combinations of seawater and ligand. The conditioning procedure was performed at least three times for the analysis with TAC and at least five times for analysis with SA.”

We are aware of the stricter conditioning before using SA. Our materials, equipment, all was divided into SA and TAC use. We are familiar with the SA methods and applied SA25 and SA5 and published the results (Slagter et al., 2019, Ardiningsih et al., 2021)

Reviewer: We are not sure why this is the case with SA, but we have observed it repeatedly as we purchase new Teflon vials and condition them.

Answer: Teflon is very reactive, aggressive cleaning before use, will mean a thorough conditioning is required. Only then there is no net flux of ligands and metals from/to the material surface.

Reviewer: As most titrations that we perform in the lab take several hours to complete and no SA signal loss is observed over that time frame, the dramatic loss of SA signal in the kinetic experiments after only minutes in some cases, is alarming. We also routinely perform overnight equilibrations with SA at 5, 10 and 25 μM in my lab with no loss of signal. Was the effect of conditioning ever tested with SA in this study? It is possible that the formation of a non-electroactive $\text{Fe}(\text{SA})_2$ complex at higher SA concentrations (and also higher Fe concentrations) is related to this conditioning issue. For example, we have found that adding high Fe concentrations to our Teflon vials along with buffer and the appropriate SA concentration yields the best conditioning results, and no resultant loss in signal. Perhaps this might be because $\text{Fe}(\text{SA})_2$ is formed under the conditions of the conditioning, and this complex has different adsorption properties to Teflon than the $\text{Fe}(\text{SA})$ species. The ultimate reason for the conditioning is unclear to me, however it is clear that with proper conditioning no loss of SA signal should be observed over time in the vials. I am worried that the results of the kinetic studies and even the model ligand results are overshadowed by these potential conditioning issues, rather than a lack of equilibrium of Fe with SA.

Answer:

We only succeeded in doing overnight equilibration at low [SA] of 5 μM . At 25 μM we had to follow the procedure of Buck with 15 minutes waiting time between the addition of SA and the measurement. A procedure described in the literature, also in your work (2014, 2015, 2016). If $\text{Fe}(\text{SA})_2$ formation and adsorption on the teflon walls is indeed the reason of the need of conditioning, this indicates that research is needed, since the formation of $\text{Fe}(\text{SA})_x$ species will shift the center of the analytical window (D).. Without proper knowledge of D and of which species are electroactive it is unclear how to interpret results. However, adding more Fe will not

favor Fe-SA species with a coordination $SA > Fe$, but the reverse, it will favor species with a coordination $SA < Fe$.

Reviewer:

The authors discuss the impact of mercury drops in the titration cell on SA measurements and first say that the mercury should have no effect (since they did this experiment also on a Metrohm with smaller mercury drops) and that the drop in signal is instead due to a disequilibrium because of the formation of $Fe(SA)_2$. They then seem to contradict this idea later in the manuscript. Accumulation of mercury in the titration cell is a big problem for SA measurements, because the adsorption potential for $Fe(SA)$ is 0V. Thus, uncharged mercury in the bottom of the cell can adsorb mercury and thus competes with the “active” drop for the binding of $Fe(SA)$. The other issue is that the accumulation of mercury in the bottom of a cell, for a BASi instrument in particular (where a stir bar is used instead of a suspended rod), can also physically impede the stirring process, which then dramatically decreases the sensitivity (peak height) of the measurement. The bottle versus in-cell kinetic experiments perfectly illustrates this, and yet the discussion of these results is largely presented in the context of the disequilibrium argument for SA. I think the authors should make this result very clear in the manuscript, and highlight that this observation is not a definitive support of disequilibrium. I also do not really understand the goal of the experiments where the mercury “puddle” is placed into fresh seawater and then no additional $Fe(SA)$ signal is observed. I may have misunderstood how this experiment was performed, but based on how it is written in the text I do not think a lack of signal in this experiment signifies irreversible formation of $Fe(SA)$, but instead reflects the fact that $Fe(SA)$ adsorbs at 0V and is not reversibly removed until a negative potential is applied to the mercury drop(s).

Answer:

The idea that a mercury electrode at 0V is equivalent to a mercury drop at the bottom of the cell is not correct. A cell held at 0V (vs the reference) holds a potential and a surface charge. The potential of zero electrode charge for a polarographic instrument with an Ag/AgCl reference electrode is in between -0.4 and -0.5 V in seawater. The continuous decrease of a peak cannot be assigned to the effect of the Hg puddle or the stirrer, otherwise, it would be impossible to measure duplicates and triplicates of a sample.

About the mercury puddle, we agree that our experiments did not give a clear proof of what happens with Fe-SA species at mercury drops at the bottom of the cell. Adsorption on the mercury at the bottom of the cell was indicated by Buck et al. (2007) to be the reason of a decrease or instability of the signal. However, this was contradicted by Abualhaija and Van den Berg (2014) who argued that the changing conditions in the oxygen concentration was the reason of the instability, since samples were not purged only stirred with the SA method using BASi equipment. That is why we checked, our results are inconclusive as said at the end of point 3 page 26, line 713, section in-cell kinetics: “It remains hard to explain the strong reduction of the peak height without a relationship with the mercury volume at SA5”

But it is obvious that discrepancies among authors justify our investigation,

We concluded the in-cell kinetics section with:

In-cell kinetic experiments with SA did not reach equilibrium and showed a continuous decay of the peak height. This can be explained by a combination of processes like adsorption of $Fe(SA)$ -complexes on

dispensed mercury at the cell bottom, and formation of the irreversible ($\text{FeSA} + \text{SA}$ forms FeSA_2) and according to Abualhaija and Van den Berg (2014) to the non electroactive $\text{Fe}(\text{SA})_2$.

Here it is important to understand that FeSA adsorption onto Hg (electrode or puddle) may also be driven by hydrophobicity.

Reviewer:

Comparison of the TAC and SA methods with model ligands: I found this section very interesting, and it is an important aspect of this work to report to the community. However, the model ligand section is hard to follow because it is not always clear what the measured values are being compared to in terms of the expected or “true” value. The expected values of each model ligand based on what has been seen previously in the literature would be immensely helpful to include in Table 2. Also, please label each $\log K_{\text{cond}}$ with either a Fe^{3+} or Fe' subscript, because I was often getting confused which one you were reporting in some figures, tables and in the text, particularly in Table 2. For example, you report the model A ligand $\log K_{\text{cond}}$ with respect to Fe^{3+} and the model B ligands with respect to Fe' . Yet, in Table 2 you report the results for model B ligands with respect to Fe^{3+} . Please keep these consistent so that comparison across methods and to previous studies are possible. I also think it would be helpful to report the “true” or expected values for each model ligand because that might also make Table 3 more meaningful. Table 3 is useful in terms of how the different methods perform relative to one another, but how about how they perform relative to what is the expected value? This would be much more insightful to consider. I actually had to make my own table while I was reading the manuscript in order to see for myself how the measured values compare to previous literature results (and thus the “expected” value for each model ligand).

Answer:

About the K^{cond} ,

Indeed we should use consequently one way of expressing the conditional stability constant and we did, all values given are with respect to Fe^{3+} , with the exception of Table 1. In this table we gave both values, with respect to Fe^{3+} and with respect to Fe' , since the applications use a different pH and thus different inorganic alpha values were used. It is clearly indicated in the headings of the columns.

However, in Table 2 and 3 we used $\log K$ without any specification, here we now added in the table captions: “ $\log K$ is used for $\log K^{\text{cond}}$ with respect to Fe^{3+} .”

We are sorry that the table was cut off. We repaired this now. The editing team pointed it out.

About the true value:

We do know for the section A model ligands how much we added, so indeed there is a true value for $[\text{L}]$. This is not the case for the B model ligands, since $[\text{L}]$ is determined in a method-specific way, and above all the molecules consist of mixtures of unknown composition. For the conditional stability constant we assumed two true values, for DTPA and DFOB (indicated in the log log plots as a dashed black line). To show how little we know of Fe-L stability constants in seawater is illustrated in this review process by the comments submitted by Peter Croot: Witter et al., 2000 published for phytic acid $\log K^{\text{cond}} = 22.3$, whereas Schlosser and Croot, (2008) published a value of 18.6. These values both with respect to Fe^{3+} differ a factor 5000.

Since we compared applications, we prefer to show how the different methods perform relative to one another. Still figure 2 shows for two model ligands how the applications compare to the true values.

Reviewer

I think it is also important to thoroughly discuss the detection window being used in each method, and how that compares to the analytical window of each model ligand experiment. A quick calculation of the analytical window for each model ligand case shows that several of the titrations are likely outside of the analytical window for some of the methods. For example, based on the expected logK for each model ligand from previous literature values, the TAC method was best suited for measuring most of the model siderophores and it often performed better than SA for these model ligands. In the opposite case, the SA 5 μM method performed better with the weaker humic and fulvic ligands. Making the connection between analytical window and the model ligand being examined clear in light of the results obtained is critical.

Answer:

Indeed the detection window or analytical window (D) is important to interpret the quality of the results. In Table 1 we show D of the applied methods as we calculated them with our calibration. Note that both SA applications have a comparable D, even though the [SA] differed a factor 5. This is explained by Abualhija and van den Berg (2014), due to the formation of the two different Fe-SA species. The D of TAC is higher, and thus indeed should be better equipped to detect the K^{cond} of DFOB. Both SA applications should be better equipped to detect the K^{cond} of the humic substances, DTPA and phytic acid (if for the latter we assume Schlosser and Croot (2008) have published the correct value).

We added as introduction to section 4.2 Titrations-Overall Line 593-onwards : “The log-log plots for DTPA and desferrioxamine B, between known and observed conditional stability constants show that the data points obtained by TAC are closest to the theoretical curve of DTPA, and by SA5 are closest to desferrioxamine B (Figures 2A,C). However, the TAC application has the highest D and should therefore be better equipped to detect desferrioxamine B, and least equipped to detect DTPA.”

Reviewer

Recommendations for future work and insights from past work: Given the implications of this work to the field and the extensive knowledge and background of the authors, I was hoping there would be a final section of the manuscript with recommendations going forward. The authors mention that we need to find new ways to measure the speciation of Fe in seawater, but make no suggestions. The authors should also comment on how past results might be interpreted. When I made my own table where I compared the measured ligand concentrations and logKs from each model ligand study to past results seen in the literature, to my eyes there was no systematic “best” method. It was often dependent on the model ligand and the analytical window where that model ligand falls, relative to the analytical window of the method used. Some discussion that brings all of the insights from this paper together and gives recommendations going forward beyond, “we need a better way” would be very powerful. As both an electrochemist and a mass spectrometerist, I can say with certainty that no method is perfect, and each has its benefits and pitfalls.

Answer:

Any compilation of published methods for model ligands shows strong deviations of values as a function of the added ligand, making this work more necessary. Another important conclusion of this work is that we find that errors in the calculation of log K are higher than the usual difference of D among methods. This increases extraordinarily our uncertainty and limits extraordinarily our ability to compare values obtained with different methods.

The suggestion, of recommendations for future work is indeed a good suggestion, although a full review of the pros and cons of available alternatives is beyond the scope of this manuscript. We extended the conclusion section Lines 780-810 as follows:

“All methods have drawbacks, and the characterization of Fe-binding organic ligands at nM concentrations in seawater with a high ionic strength and at 40-80 $\mu\text{mol/kg}$ DOC (Hansell et al. 2009) is and remains challenging. In case voltammetric methods are used, we advise the SA5-application, but should not be overinterpreted and issues relating to the technique fully acknowledge. In particular, it is hard to justify determination of equilibrium constants using an adsorption isotherm that assume equilibrium, if a signal is not stable and the experimental system is therefore not at equilibrium (for whatever reason), especially bearing in mind the method specifically uses the term “equilibrium” in its title. Furthermore, it appears that apart from the constraint forced by D, these estimates have a much larger error than apparent from those obtained from the fits to the Langmuir isotherm, which may explain why direct links between speciation predicted from conditional parameters do not relate strongly to bioavailability (Shaked et al., 2020, 2021) although this could also be caused by the limited role assigned to iron reduction prior to bio-assimilation. Given that the research questions that are typically addressed when determining conditional parameters from the Langmuir isotherm actually relate to how variations in the parameters impact on Fe speciation (e.g. abundance of Fe not bound to organic matter and thus assumed to be bioavailable), it would seem appropriate to broaden the methodology applied to these questions to other methods capable of e.g. estimating concentrations and distributions of the dominant groups of binding sites and/or the lability and solubility of Fe. For example, Whitby et al. (2020) and Laglera et al., (2019) have used alternative methods based on voltammetry to suggest that humic substances are more important than thought before. Other recent studies have focused on the influence of the pH on metal organic complexation and vary the pH of titrations (Ye et al., 2020, Gledhill et al., 2015; Zhu et al., 2021). Here we note that the acid-base chemistry of organic matter – which underpins metal binding - is severely understudied, but can nevertheless quantify the distribution and concentration of the total cation binding sites present, at least in the portion of dissolved organic matter that can be isolated from seawater (Lodeiro et al., 2020). With further knowledge of acid-base chemistry of dissolved organic matter and how it changes with inputs of fresh material from biological activity, the total binding site concentration could be independently constrained and only binding affinities derived. Alternatively models can be applied that allow for the estimation of metal speciation for ambient conditions (e.g. Hiemstra and van Riemsdijk, 2006; Stockdale et al., 2015), with predictions of how binding to organic matter could influence e.g. iron solubility (Zhu et al., 2021). The wider application of methods employing cation chelating resins, such as diffusive thin film gradients (Zhang and Davison, 2015; Town et al., 2009; Bayens et al., 2018) might also offer alternative insights into the lability of metals in seawater. Another way to apply voltammetric methods is to characterize metal-binding ligands by pseudovoltammetry (Luther et al, (2021), although unfortunately this method is not currently suitable for DFe. Whilst continued application of the CLE-AdCSV approach will no doubt further develop our knowledge on how operationally defined ligand concentrations and stability constants vary in the ocean under a restrictive set of conditions apparently specific to one added ligand, new approaches to both the

determination and the interpretation of metal binding to organic matter will surely stimulate discussions in the field of the organic metal complexation and furthermore be likely lead to new insights.

Reviewer:

Specific comments:

There are several small typos, only some of which I have detailed here.

Section 2: You list your assumptions and refer to them by number, but they are not numbered. Numbering them might be helpful, since this whole section reads like a list and you refer to specific assumptions later in the manuscript.

Answer: Thank you this is repaired.

Line 104: Add a space between “knowledge” and “In”

Answer: Thank you this is repaired.

Line 108: Do you mean 20 minutes or 15 minutes? You use 15 minutes throughout the manuscript, and say that you also used a timer.

Answer: Thank you this is repaired into 15.

Line 129: Add a space between “[L]” and “and”

Answer: Thank you, this is now adapted.

Line 181: Were the samples filtered prior to being stored frozen? If so, how were they filtered? Where they frozen at -20°C?

Answer: we changed the first sentence into (line 183): “The natural seawater used in the experiments consisted of mixed leftover filtered samples of the northern Western Atlantic GEOTRACES cruise GA02 (stored frozen at -20 °C) (Rijkenberg et al., 2014).”

Line 182: Was the UVSW aged prior to use or used immediately?

Answer: Used without aging, as shown by” UV irradiated sea water was stored for 3 days at most.”

Line 215: Was all kinetic conditioning done in UVSW?

Answer: We added: in UV-irradiated seawater

Line 225: Remove the second “in” after “placed”

Answer: Thank you, this is now adapted.

Line 227: Do you have a reference for this?

Answer: Yes we changed the text already due to remarks of other reviewers and added Rapp et al., 2017; Wuttig et al., 2019).

Line 260: Bundy et al. (2018) determine the conditional stability constant of ferrioxamine E in seawater (as well as ferrioxamine B). The $\log K_{=14.4}$

$\log K_{=14.4}$ for ferrioxamine E was 14.05 and the $\alpha_{E^{1+}}$

used was 10. These measurements were performed using the SA method at 5 μM .

Answer:

Apologies for missing this. We added this information now in the text at line 271-273:

But Bundy et al. (2018) estimated $\log K_{FeL,Fe}$ values of ferrioxamine B and E, using SA5 in seawater to be close being 14.4 and 14, respectively.

Line 308: Were blanks also in absence of Fe?

Answer: No, the blanks were UV irradiated seawater for the Atlantic Ocean, so these contained approximately 0.64 ± 0.4 nM DFe (N=434) (obtained from all samples <100 m depth described in Gerringa et al., 2015).

Line 313: This sentence is confusing. I think you mean that you added buffer and dFe and not also SA, and then you added SA after equilibration. Why equilibrate the buffer and dFe for one hour? The commonly used method equilibrates the buffer and dFe for two hours before adding the SA (Buck et al. 2007).

Answer:

We changed “the 25 μ M SA buffer” into: “the buffer”. This was indeed confusing.

Line 321: I think you mean “prior to the addition of TAC or SA.”

Answer:

We changed the sentence into: “In the latter case, there is the possibility that Fe-oxide precipitates were formed prior to the addition of TAC or SA and were probably dissolved after the addition of TAC or SA”.

Line 357: Why was this dFe concentration chosen? Was your seawater a deep sample? Why not used the measured dFe concentration?

Answer:

The measured DFe varied per sample. Our mixed Atlantic seawater was >100 m depth. To answer reviewer Peter Croot we calculated the possible average DFe from all samples taken for ligand characterization to be 0.64 nM. However, we do not know which samples were used, they were leftovers. Anyhow it is not important for the calculations which DFe was used, as long as it is the approximately the same magnitude as seawater concentrations.

Line 369-370: In most SA CSV studies the blank is zero, meaning no Fe is added with the buffer or SA addition. Can you include the ordering information for the boric acid and SA that were used? Was distilled or Optima ammonium hydroxide used for the buffer preparation? The fact that there was a blank with these measurements is very disconcerting. On a related note, please note the error associated with the dFe measurements for the model ligand results. In some cases, the standard deviation on the concentration of L and the logK is relatively small, and given that the dFe in each experiment varied quite widely, it would be nice to see the error on these measurements as well.

Answer:

We are not sure whether the reviewer refers to the voltammetric signal or ICPMS analysis. With ICPMS analysis we do not often encounter zero for DFe in solutions or samples. This is a general feature when using reagents that cannot be produced at ultraclean levels: borate, good buffers, artificial ligands and some model ligands. Residual Fe concentrations must be always checked in all reagents, otherwise, small deviations in the concentration of Fe in solution may produce strong deviations in L and K.

We had no information about the Fe content of SA, but we do know that the Fe content of TAC varies and can be too high, and that we have to discard certain batches. That is why we did the analyses. Yes we used high purity ammonium hydroxide (Merck supra pur), added now in Table S1).

See Table S1 for SA, Acros organics.

We added at section 3.7 ICPMS analyses: at line 375

“For the analyses of DFe in UV irradiated seawater with and without added model ligands the limit of detection for DFe was $22 \text{ pM} \pm 8 \text{ pM}$. The DFe of model ligands DTPA, Phytic acid, desferrioxamine B and ferrichrome were in the dilutions used (2 nM and 4nM) below the detection limit. This means the addition of the model ligand did not increase the DFe in the UV irradiated seawater. Analyses of 2 nM vibriobactin resulted once in <detection limit and once in 0.1 nM DFe. However, ferrioxamine E, FA and HA contained measurable amounts of DFe. 2 nM FerrioxamineE= $1.76 \pm 0.04 \text{ nM}$ (N=2), 0.2 mg FA= $1.1 \pm 0.02 \text{ nM}$ (N=3), 0.2 mg HA= $3.39 \pm 0.05 \text{ nM}$ (N=2).”

Line 410:

Change to “ligands like siderophores point to biases”

Answer:

We changed the sentence by removing a confusing and wrongly placed “,”.

Line 588: I think it is difficult to make too many assumptions on the HS and FA results, because there is large variability in the literature with respect to the available dFe binding sites. Again, what is the approximate “true” value we can compare these model ligand titrations to? Did you determine your own dFe binding capacity measurements for the batches of FA and HS you used?

Answer:

We agree, there is no true value here. Still Laglera et al 2009 showed that CLE-AdCSV and straight Fe titration of the HS until the end of a linear response gave close results. Luis Laglera did determine the binding capacity for FA (Sukekava et al., 2018), we used the same batch. For HA, the binding capacity was measured in 2009 and re analyzed recently in a different batch (unpublished results, Laglera) with a consistent result slightly over 30 nmol Fe (mg SRHA)-1.

Line 645: You can still have mercury adsorption on the drop even with smaller drop sizes. The drop on the Metrohm itself is smaller, therefore having the same sized drops in the bottom of the cell, relative to the active drop, will still give you the same issue.

Answer:

Correct we changed the sentence in (line 678): “Possible contributions due to decreasing oxygen were excluded , and due to adsorption on mercury on the cell bottom were minimized”

Table 1: Some of this table is cutoff, so it is difficult to understand what is being displayed.

Answer:

We are very sorry that this happened. We made a new table.

Randie Bundy

Answer References

Abualhaja, M. M., Van den Berg, C. M. G. Chemical speciation of iron in seawater using catalytic cathodic stripping voltammetry with ligand competition against salicylaldoxime. *Mar. Chem.* 164: 60-74, 2014.

Ardiningsih, I., Zu, K., Lodeiro, P., Gledhill, M., Reichart, G-J., Achterberg, E.P., Middag, R., Gerringa, L.J.A. Iron speciation in Fram Strait and over the northeast Greenland shelf: An inter-comparison study of voltammetric methods. *Front. Mar. Sci.* 7:609379. doi: 10.3389/fmars.2020.609379., 2021.

Bayens, W. Gao, Y., Davison, W., Galceran, J., Leermakers, M., Puy, J., Superville, P.-J., Beguery, L. In situ measurements of micronutrient dynamics in open seawater show that complex dissociation rates may limit diatom growth. *Scientific reports* 8:16125 | DOI:10.1038/s41598-018-34465-w, 2018

Buck, K.N., Lohan, M.C., Berger, C.J., Bruland, K.W. Dissolved iron speciation in two distinct river plumes and an estuary: Implications for riverine iron supply. *Limnol. Oceanogr.* 52: 843-855, 2007.

Gerringa, L.J.A., Rijkenberg, M.J.A., Schoemann, V., Laan, P., de Baar, H.J.W. Organic complexation of iron in the West Atlantic Ocean. *Mar Chem.* 177:434-446 .doi.org/10.1016/j.marchem.2015.04.007, 2015.

Hiemstra, T., van Riemsdijk, W.H., 2006. Biogeochemical speciation of Fe in ocean water. *Mar. Chem.* 102, 181–197. <https://doi.org/10.1016/j.marchem.2006.03.008>

Lodeiro, P., Rey-Castro, C., David, C., Achterberg, E.P., Puy, J., Gledhill, M. Acid-base properties of dissolved organic matter extracted from the marine environment. *Science of The Total Environment*, 729: 138437, 2020.

Rapp, I, Schlosser, C., Rusiecka, D., Gledhill, M., Achterberg, E.P. Automated preconcentration of Fe, Zn, Cu, Ni, Cd, Pb, Co, and Mn in seawater with analysis using high-resolution sector field inductively-coupled plasma mass spectrometry. *Analytica Chimica Acta*, 976:1-13, 2017. <http://dx.doi.org/10.1016/j.aca.2017.05.008>.

Schlosser, C., Croot, P.L., 2008. Application of cross-flow filtration for determining the solubility of iron species in open ocean seawater. *Limnology and Oceanography: Methods* 6, 630-642.

Shaked, Y., Buck, K.N., Mellett, T., Maldonado, M.T. Insights into the bioavailability of oceanic dissolved Fe from phytoplankton uptake kinetics. *ISME Journal* 14:1182–1193 <https://doi.org/10.1038/s41396-020-0597-3>, 2020.

Shaked, Y., Twining, B.S., Tagliabue, A. Maldonado, M.T. Probing the bioavailability of dissolved iron to marine eukaryotic phytoplankton using *in situ* single cell iron quotas. *Biogeochemical cycles.* <https://doi.org/10.1029/2021GB006979>, 2021.

Slagter, H.A., Laglera, L.M., Sukekava, C., Gerringa, L.J.A. Fe-binding organic ligands in the humic-rich TransPolar Drift in the surface Arctic Ocean using multiple voltammetric methods. *Journal of Geophysical Research: Oceans.* DOI: 10.1029/2018JC014576, 2019.

Stockdale, A., Tipping, E., Lofts, S., 2015. Dissolved trace metal speciation in estuarine and coastal waters: Comparison of WHAM/Model VII predictions with analytical results. *Environ. Toxicol. Chem.* 34, 53–63. <https://doi.org/10.1002/etc.2789>

Sukekava, C. Downes, J., Slagter, H.A., Gerringa, L.J.A., Laglera, L.M. Determination of the contribution of humic substances to iron complexation in seawater by catalytic cathodic stripping voltammetry. *Talanta* 189, 359-364. <https://doi.org/10.1016/j.talanta.2018.07.021>, 2018

Town, R.M., Chakraborty, P., van Leeuwen, H.P. Dynamic DGT speciation analysis and applicability to natural heterogeneous complexes. *Environmental Chemistry* 6(2) 170-177 <https://doi.org/10.1071/EN08091>, 2009

Wuttig, K., Heller, M. I., and Croot, P. L. Reactivity of inorganic Mn and Mn desferrioxamine B with O₂, O₂–, and H₂O₂ in seawater. *Environ. Sci. Technol.* 47, 10257–10265. doi: 10.1021/es4016603, 2013.

Ye, Y., Völker, C., & Gledhill, M. Exploring the iron-binding potential of the ocean using a combined pH and DOC parameterization. *Global Biogeochemical Cycles*, 34, e2019GB006425. <https://doi.org/10.1029/2019GB006425> , 2020.

Zhang, H., Davison, W. Use of diffusive gradients in thin-films for studies of chemical speciation and bioavailability *Environ. Chem.* 2015, 12, 85–101 <http://dx.doi.org/10.1071/EN14105>, 2015

References

Abualhija, M.M. and van den Berg, C.M., 2014. Chemical speciation of iron in seawater using catalytic cathodic stripping voltammetry with ligand competition against salicylaldoxime. *Marine Chemistry*, 164, pp.60-74.

- Bruland, K.W., Rue, E.L., Donat, J.R., Skrabal, S.A. and Moffett, J.W., 2000. Intercomparison of voltammetric techniques to determine the chemical speciation of dissolved copper in a coastal seawater sample. *Analytica Chimica Acta*, 405(1-2), pp.99-113.
- Buck, K.N., Lohan, M.C., Berger, C.J. and Bruland, K.W., 2007. Dissolved iron speciation in two distinct river plumes and an estuary: Implications for riverine iron supply. *Limnology and Oceanography*, 52(2), pp.843-855.
- Bundy, R.M., Boiteau, R.M., McLean, C., Turk-Kubo, K.A., McIlvin, M.R., Saito, M.A., Van Mooy, B.A. and Repeta, D.J., 2018. Distinct siderophores contribute to iron cycling in the mesopelagic at station ALOHA. *Frontiers in Marine Science*, 5, p.61.
- Tagliabue, A., Bowie, A.R., Boyd, P.W., Buck, K.N., Johnson, K.S. and Saito, M.A., 2017. The integral role of iron in ocean biogeochemistry. *Nature*, 543(7643), pp.51-59.