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

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. 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 method-specific perspective on the continuum of iron ligands in seawater. 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. 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). 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.

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). 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. 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 $Fe(SA)_2$ 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 $Fe(SA)_2$ is formed under the conditions of the conditioning, and this complex has different adsorption properties to Teflon than the Fe(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 over shadowed by these potential conditioning issues, rather than a lack of equilibrium of Fe with SA.

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 a 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).

<u>Comparison of the TAC and SA methods with model ligands</u>: 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 $\log K^{cond}$ 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 $\log K^{cond}$ 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). 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.

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.

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.

Line 104: Add a space between "knowledge" and "In"

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.

Line 129: Add a space between "[L]" and "and"

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

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

Line 215: Was all kinetic conditioning done in UVSW?

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

Line 227: Do you have a reference for this?

Line 260: Bundy et al. (2018) determine the conditional stability constant of ferrioxamine E in seawater (as well as ferrioxamine B). The log$K^{cond}_{FeL,Fe'}$ for ferrioxamine E was 14.05 and the $\alpha_{Fe'}$ used was 10. These measurements were performed using the SA method at 5 µM.

Line 308: Were blanks also in absence of Fe?

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).

Line 321: I think you mean “prior to 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?

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.

Line 410: Change to “ligands like siderophores point to biases”

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 ligan titrations to? Did you determine your own dFe binding capacity measurements for the batches of FA and HS you used?

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.

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

Randie Bundy

References

Abualhaija, 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.