Comments and responses - Round 1

Recommender:

I have now received three reviews for your preprint entitled “PARCHMENT GLUTAMINE INDEX (PQI): A NOVEL METHOD TO ESTIMATE GLUTAMINE DEAMIDATION LEVELS IN PARCHMENT COLLAGEN OBTAINED FROM LOW-QUALITY MALDI-TOF DATA”

C[1]: I am pleased to say that the reviews are mostly positive and recognise the importance of the work, and I agree with them. The reviewers raise valid points on some aspects of the manuscript, mainly concerning the choice of not screening out poorest-quality data. I share such concerns: after all, poor quality data is poor quality data, and no amount of post-acquisition work can really change that (and I agree with the reviewers that most of the data in Fig 2 looks very poor). My suggestion is that the authors provide an explanation of the reasons why looking at poor-quality data (rather than re-analysing samples in order to acquire better spectra and/or discarding the dataset) is deemed so important as to warrant a whole new application for assessing Q → E.

R[2]: The purpose of the developed package MALDIpqi is to maximise the information from the MS1 spectra. Poor peak shapes still carry information and by weighting the poor quality data appropriately we can gain more information than by simply ignoring it. A striking feature of our approach is the very consistent errors that we can obtain on the whole system considering PQI even when noisy data is included because this data is weighted appropriately.

C: The other main question I have is on the reproducibility of the analytical settings across samples: the manuscript does not provide information on how the dataset was obtained, i.e. sample extraction, preparation, instrument, mass range, suppression, matrix type/concentration, operator(s), number of analytical/biological replicates, laser intensity, number of shots, etc…Maybe I could not find the information, but in any case it should be in the main text.

R: We thank the referee for this comment and a detailed description of sample preparation and data acquisition has been added to the manuscript (line 164-182).

---

[1] C refers to Comments
[2] R refers to Response

C: Another issue I would want to see addressed is that a few years ago Simpson et al published a paper demonstrating that MALDI-TOF is not ideal for Q → E because it underestimates E in peptide mixtures containing the Q and E forms (https://doi.org/10.1002/rcm.8441). While this might not be crucial if comparing relative difference between samples, I would mention this bias and justify why you still want to use MALDI, i.e you want to compare data that's already been acquired/costs etc.

R: The reviewer makes a good point and we have added a reference to Simpson et al's work in our revised text. We justify the use of MALDI for the obvious reason that it is the most widely used proteomics-based method in archaeology (notably as it is the basis of ZooMS), and MALDIpqi adds a further level of information to these datasets conducive to delving into information hidden in archaeological samples.

C: Finally, can the authors explain how they are separating the signal of parchment quality from that of time/preservation histories? The samples come from the same site, but they will have different post-production "biographies", which may explain some of the variability observed. Can the authors comment on this? I appreciate that the focus of the manuscript is on the PQI index, but section "Applications of PQI" is a bit generic - the ms would benefit from a more nuanced interpretation of the data.

R: At present, there is no evidence that there is any link between PQI and time/preservation histories in the case of parchment. This is probably because we are comparing an aggressive chemical treatment (liming) which is conducted specifically to deamidate collagen against a much slower process (age/time-dependent hydrolysis) which is further slowed by the controlled storage conditions of parchment documents (line 395-401).

C: A general suggestion: refer to J. Gross's manual for doubts on mass spectrometry terminology https://link.springer.com/book/10.1007/978-3-319-54398-7

R: Changes have been made to the manuscript according to general terminology in mass spectrometry.

I would ask you to revise the manuscript within one month, according to the comments of the three reviewers and the general points I made above, and to submit the revised preprint, along with a detailed point-by-point response. I shall be happy to recommend it, pending suitable revision.

Looking forward to receiving your revised manuscript.

Reviewer 1

C: The manuscript describes a method to calculate the level of glutamine deamidation for several peptides by modelling the convoluted isotope distributions and combine the results to give an overall estimate of the deamidation for a parchment sample. The method is described as a simple three-step workflow, where the modelling occurs in the second and the third steps. The modelling of individual peptides uses weighted least squares regression which seems sensible for the limited data although the weights are calculated according to the noise level estimated from at most 3 observations. The details of the regression in this section somehow suggest more data than is available, possibly due to the matrix X (line 208). It could be worth linking back to the fact that n is at most 3 (sometimes just 1), and (according to Figure 2) m is at most 6 (should this be 5?). The notation is confusing as xij suggests the element in the ith row of the jth column in the matrix X, but in fact this is used to denote the model for the ith isotopic peak in the jth replicate. Perhaps the latter could be denoted yij. Also γ is of length k rather than k+1.

R: The weighted linear regression is used to calculate individual *q* values for each peptide and replicate. In the mathematical formulation of the method, we assume *n* replicates and describe a way to aggregate the replicates using weights w_{ij}. Although we mention this, we consider at this stage separate replicates (i.e. *n=1*) that are combined in the third step of the linear mixed effect model. One could use these aggregate replicates, but we proceed differently. We have made this clearer in the text. We thank the reviewer for the suggestion for a change on the notation of x_{ij}, and we have changed this to y_{ij}. Also, γ is actually of length k+1 as we start the indices from 0; we have corrected the notation in the manuscript to avoid confusion.

C: The third step of the process uses a mixed effects model to combine the individual deamidation estimates and the mathematical details are provided. My worry is that, for all the statistical theory, the results may not be meaningful if the data are poor. The authors perform residual analysis and apart from a few badly fitted values, seem to have obtained good results. However, these results check the fit of the model to the peaks extracted from (as the title advertises) low quality data. This depends entirely on the first step in the workflow which involves pre-processing the MALDI spectra and the only way to assess the performance here is from Figure 2. Of the 5 randomly chosen spectra, 4 appear to show only noise in all but the first peptide in Figure 2A. Maybe the much greater intensities for sample 67_19_2 make the other spectra look worse than they are. In any case, apart from this sample, Figure 2B shows just how few peaks were identified (assuming that the solid circles

indicate identified peaks) and it is difficult to see why the peaks in other samples were identified. In the case of 59_I_4_2, the intensities are just too low to see anything yet peaks were identified for several peptides. Maybe the weighted (normalised) spectra should be shown instead? However, even for sample 67_19_2, the pre-processing for the last peptide does not look convincing. Are these typical spectra?

R: We show the spectra without normalization, simply because the preprocessing that precedes the estimation of *q* does not normalize the spectra, as this is only affected by the internal proportions of the isotopic peaks. However, it is true that we can show the normalized spectra after preprocessing so it is easier to visualize the peptides from different samples together (see figure below). We have replaced Figure 2 with the normalized spectra.

We want to point out that we picked for this plot some of the spectra with higher residual values when *q* was estimated. While these are not typical, the higher mass peaks are always the more problematic. However, it is important to notice that the same parameters for preprocessing should be used for all samples, as otherwise they can create artificially different PQI values afterwards.

Reviewer 2

This paper presents a novel method for quantifying the deamidation of collagen peptides in parchment derived from animal skins. They present a well-written paper and a thorough, if dense, overview of the methods and results. The R-package and model developed here provide interesting new avenues to examine the treatment of animal skins in the past. I do not have the background to speak to the suitability of the statistical models developed here and will focus on the peptides.

C: Peptide naming should be checked against reviewed collagen sequences to ensure correct and consistent naming throughout the paper. As this paper follows Brown et al. (2020), the peptide positions start counting from the beginning of the collagen-specific three-letter pattern (G-X-Y). Thus COL1a2 756-789 and COL1a2 535-567 in this paper are both COL1a2 756-789 and COL1a1 9-42 should be 10-42. Peptide COL1A1 375 is also referred to at certain points as 376. The reference for Brown et al. (2020) is also missing from the reference list.

R: We thank the reviewer for spotting this. The numbers have been corrected and the reference added to the reference list.

C: While I believe the methods touch on this topic, I feel this paper requires a clearer explanation of how missing peaks are dealt with at the replicate and sample levels. If these are being estimated, as the methods seem to suggest, what are the implications of this to the final PQI value and why is this preferred over removing replicates where peaks are missing? It seems that many of the issues discussed in the paper, such as the high proportion of PQI values above 1 and the deviation of tails on the qq plots could be alleviated by screening out poorer quality spectra or increasing the signal to noise ratio. Given these issues, it is unclear whether background noise or missing values may be conflated with the deamidation models in the paper.

R: The missing peaks are not estimated. When a given peak from a peptide is missing, the *q* is just calculated with the rest of the present peaks. We do however check for coherence of

the isotopic envelope. For example, we exclude the complete peptide (i.e. it will be missing *q*) if a peak other than the first or the last or just the first is missing, or the first peak is not identified. At the last stage, the estimation of the PQI uses the available *q* values for the peptides and replicates.

As discussed earlier, PQI includes all possible data from each spectra but gives each of the peptides a weight based upon the quality of the spectra. We do not think that PQI values of > 1 are entirely caused by poor spectra - indeed we are surprised that so many are. Further research is needed to fully explain this, for example, a peak may conflate more than one peptide, resulting in incorrectly assigned deamidation values. The PQI code is flexible and we can select any specific peptide masses. A new analysis of PQI using different masses and /or characterisation of the peaks selected by Fiddyment using MALDI-TOF MS/MS may help to solve these issues. Although we subjected few different samples to MS/MS, no sequences were identified due to low intensity peaks in MS1 spectra for the peptides under study.

C: One of the two masses of peptide 9-42 is suggested to be incorrectly identified due to different rates of deamidation (lines 318-322). From Table 3, it is unclear which value is being referred to as the relative rates are very close in number. In contrast COL1a2 756-789 and its variant, referred to as COL1a2 535-567 have different relative rates of deamidation. As these are variants of the same peptide, is this difference important and if so, how might it impact interpretations about the use of goat skins? The finding that goat hides were more heavily limed is intriguing. However, given that goats have a different variant of peptide COL1a2 756-789, I would like to see evidence that goat parchment is more heavily deamidated across multiple peptides to rule out any effect of the different rates of deamidation across the two variants of COL1a2 756-789. \

R: We refer the reviewer to the main text, where it is explained that COL1a1 508-519 is used as a reference as it is the slowest peptide to estimate relative rates of deamidation (line 342). For peptide COL1a1 756-789, we do find the estimates from table 3 to be similar indeed, compared to the other peptides. We assume that the goat version of the peptide COL1a2 756-789 deamidates at a very similar rate, given that the two amino acids flanking Q are the same in the two versions, and it is known that these are the main factors influencing deamidation (see (Robinson *et al.*, 2004)). Table 3 also shows that they both have similar *q* values and relative values to COL1a1 508-519.

Assuming the statistical analysis is sound, this study will make a valuable contribution to parchment production studies and the development of novel applications of MALDI-TOF in cultural heritage contexts.

Reviewer 3

This is an interesting paper that looks at deamidation as a marker of parchment quality, and allows the comparison of deamidation levels with other variables such as species and thickness of parchment to gain insight into the parchment making process. It is interesting that this baseline correction appears to improve the S/N of some peaks, but not others. I think the paper could be improved by acknowledging the limitations of the background correction technique and if possible, to draw some conclusions as to why this works for some, but not other peaks, is it to do with the S/N for the peak in the raw data? I am unable to comment on the mathematical/statistical aspects as this is outside my area of expertise. Some minor corrections are recommended.

C: Lines 116-119: I don't think this is true? You have a charge in MS2 as well, you can only see charged product ions and the product ion spectrum would be used for confirming the peptide sequence. The reason you have overlapping distributions in MS1 is due to the resolution of the instrument not the fact is the parent ion. In the future, since you have identified clear biomarkers it would be interesting to make a targeted MS method to carry out absolute quantitation using SRM on your 8 peptides.

R: We thank the reviewer for this correction and have updated the manuscript (line 120-123). Yes, it would be interesting to perform absolute quantitation of these peptides using selective reaction monitoring (SRM).

C: Line 183: A S/N ratio of 1.5 is very low. Usually this is set at 3 for identification and higher than this for quantitative analysis (around 10) and what you are trying to do here is semi-quantitative.

R: S/N is indeed very low, but we are trying to capture as much information as possible, e.g. using a higher S/N ratio, we would remove the signal from the last isotopic case. The S/N ratio is also dependent on the method used to estimate the noise.

C: Line 318 and 322: You could de-novo this to confirm?

R: We have not analysed any of the data, we are using published studies. However, we agree it would be useful to run the samples using MALDI-TOF MS/MS.

C: Line 365: I'm not sure if authors mean that the base line correction did not work for this peptide in the sense that even after baseline correction, the S/N was too low?

R: We want to point out that baseline correction does generally work for poor quality spectra given the parameters we specify to perform the process. However, there are still peptides within some samples for which the baseline correction and peak detection is still problematic. However, in general, the same parameters for preprocessing should be used for all samples, as otherwise they can create artificially different PQI values afterwards.

C: Line 308: Why would a high level of deamidation in a peptide result in higher background if the background is caused by chemical interference from the MALDI matrix?

R: This would seem to be inherently unlikely as the high background would interfere equally with all the masses in the isotopic envelope which is used to estimate individual levels of peptide deamidation.

C: Line 384: MADI-TOF and ZooMS aren't interchangeable. One is describing the ionisation technique/mass analyser and one is an application of MS to a sample type. Maybe better to say something like in the field of bioarcheology the application of MALDI-TOF to ancient samples is commonly referred to as ZooMS.

R: The referee is correct and we have amended the manuscript to reflect this.

C: Line 391: I wonder if this variation due to the poor quality spectra or biological variation?

R: We thank the referee for pointing this out. The variation arises due to differences in species.

C: Line 400: It would be very difficult, if not impossible to validate this to EMA or FDA regulations.

R: We do not see this as a regulatory tool, rather we see it assisting producers in assessing the quality of collagen-based products as part of manufacture. At present deamidation levels are assessed using wet chemical methods (e.g. (Langmaier *et al.*, 2002)).

C: Figure 2: It's interesting that for some peptides you can see an improvement in signal to noise when you compare them with the spectra in A vs B, but some still look poor quality after the background correction. Do you see any patterns in peptides where the correction works well for and which is doesn't? I suspect that rather than the ionisation type, it's the amount of collagen you are extracting in terms of S/N. If you remove deamidation levels from poor S/N spectra do your error bars change?

R: We thank the reviewer for this valid observation. The peptides for which preprocessing works worse don't necessarily reflect that in the fitted vs. residual or quantile-quantile plots.

The idea was to develop a method that can process a large number of samples and calculate a deamidation index, without having to remove samples *a posteriori*. Although we agree that in the future, effort should be put into developing a pre-screening step that allows for removing poor samples from the model. We expect that the amount of collagen extracted as well as the particularities of the crystalization play an important role in S/N values.

C: Figure 6: I'm not familiar with what level of spread of deamidation values you would expect for these samples but the error bars here seem quite high.

R: We request the reviewer to refer to our response above. The error bars are low given the nature of the raw data used for analysis.

C: General comments: m/z should be italicised throughout.

R: We have corrected this throughout the manuscript.

References

Langmaier, F. *et al.* (2002) 'Isolation of elastin and collagen polypeptides from long cattle tendons as raw material for the cosmetic industry', *International journal of cosmetic science*, 24(5), pp. 273–279.

Robinson, N.E. *et al.* (2004) 'Structure-dependent nonenzymatic deamidation of glutaminyl and asparaginyl pentapeptides', *The journal of peptide research: official journal of the American Peptide Society*, 63(5), pp. 426–436.