# The Demodifier: a tool for screening modification-induced alternate peptide taxonomy in palaeoproteomics

**Miranda Evans**[1,2,3*]

*** Corresponding Author.**
**ORCID:** 0000-0002-9284-9268
**Email address: mae52@cam.ac.uk**
**1:** Department of Archaeology, University of Cambridge, Cambridge, UK
**2:** Department of Zoology, University of Cambridge, Cambridge, UK
**3:** BioArCh, Department of Archaeology, University of York, York, UK

**Key words:** Palaeoproteomics, Taxonomy, Authentication, Bioinformatics

# Abstract

In palaeoproteomic research, the accuracy of taxonomic matches is crucial, as research questions frequently hinge on which species were utilised by ancient people. However, protein modifications including deamidation of glutamine and asparagine, and conversion of glutamine or glutamic acid to pyroglutamic acid, can change the sequence of peptides, leading to spurious taxonomic detections and potentially inaccurate archaeological interpretations. While a handful of examples of this phenomenon have been reported in the literature, the issue is potentially much wider reaching than currently realised. In reality, any time a peptide containing a deamidated glutamine or asparagine residue, an unmodified glutamic acid or aspartic acid residue, or a pyroglutamic acid modification is detected by proteomic search software, the sequence, and therefore potentially its taxonomy, may be incorrect, which could potentially lead to unsound archaeological interpretations. `The Demodifier` is a fast, open source tool which solves this issue by screening for modification-induced alternate peptide taxonomy, enabling archaeologists to make informed interpretations of the taxonomies of peptides detected in ancient samples. To assess its utility, `the Demodifier` is tested against an archaeological dataset containing all unique peptides reported in palaeoproteomic studies of dental calculus and vessels. The results reveal that modification-induced alternate peptide taxonomies are severely under-reported, occurring 15% more frequently than previously understood. Modifications were found to

produce three different types of inaccurate taxonomic matches: those which yielded completely different taxonomic lowest common ancestors to the input peptide, those which were more taxonomically specific than the input peptide, and those which were less taxonomically specific than the input peptide. `The Demodifier` therefore enables the rapid detection of potentially inaccurate peptide taxonomies, avoiding spurious archaeological interpretations in future studies.

# Introduction and Background

In palaeoproteomics, the reported taxonomy of ancient peptides is assessed by analysing the lowest common ancestor (LCA) of all the taxa to which a given peptide matches. It is crucial that these taxonomic assignments are secure, because they frequently form the basis of archaeological interpretations, e.g. the detection of newly introduced plant taxa (Scott et al., 2021), multitaxa dairying (Evans et al., 2023), or mixed ingredient processing (Hendy et al., 2018a). Yet post-translational modifications (PTMs) including deamidation of glutamine (Q) and asparagine (N) and N-terminal pyroglutamic acid formation can alter peptide sequences, potentially leading to inaccurate taxonomic matches.

Deamidation is a PTM in which asparagine (N) is converted to aspartic acid (D), and glutamine (Q) is converted to glutamic acid (E) (Figure 1). This reaction involves the conversion of an amide group ($-CONH_2$) into a carboxylic acid (-COOH), resulting in a mass shift of approximately +0.984 Da, a change commonly observed in both peptide mass fingerprinting and tandem mass spectrometry analyses. Deamidation of both asparagine and glutamine can proceed via two pathways (Figure 1). At acidic pH, the direct hydrolysis pathway predominates, while under neutral to basic conditions, intramolecular cyclisation to a cyclic intermediate occurs (succinimide for asparagine, glutarimide for glutamine), although this pathway is less favourable for glutamine than for asparagine (Capasso et al. 1991; Catak et al. 2009; Li et al. 2010). The cyclic intermediate pathway can generate isoform residues, isoaspartic acid and isoglutamic acid, which have identical masses to their non-isoform counterparts (Tyler-Cross and Schirch 1991; Li et al. 2010). Asparagine (N) has been found to deamidate at a higher rate than glutamine (Q), which deamidates considerably more slowly (Robinson et al., 2004). Given that deamidation occurs at a relatively regular rate under constant conditions, it has been proposed as a molecular clock (Robinson et al., 2004), a marker of preservational quality (Schroeter and Cleland, 2016; van Doorn et al., 2012), or even an indicator of processing (Nair et al., 2023). Deamidation rates have been used as a tool for establishing the authenticity of palaeontological, archaeological

or heritage proteins, initially using bulk deamidation rates (Hill et al., 2015; Orlando et al., 2013; van Doorn et al., 2012; Welker et al., 2017, 2015), and later using site-specific measures (Ramsøe et al., 2020). Deamidation rates are also impacted by protein structure (Kossiakoff, 1988; Robinson et al., 2004; Xie and Schowen, 1999), temperature (Stratton et al., 2001), and pH (Hao et al., 2011; Wilson et al., 2012).

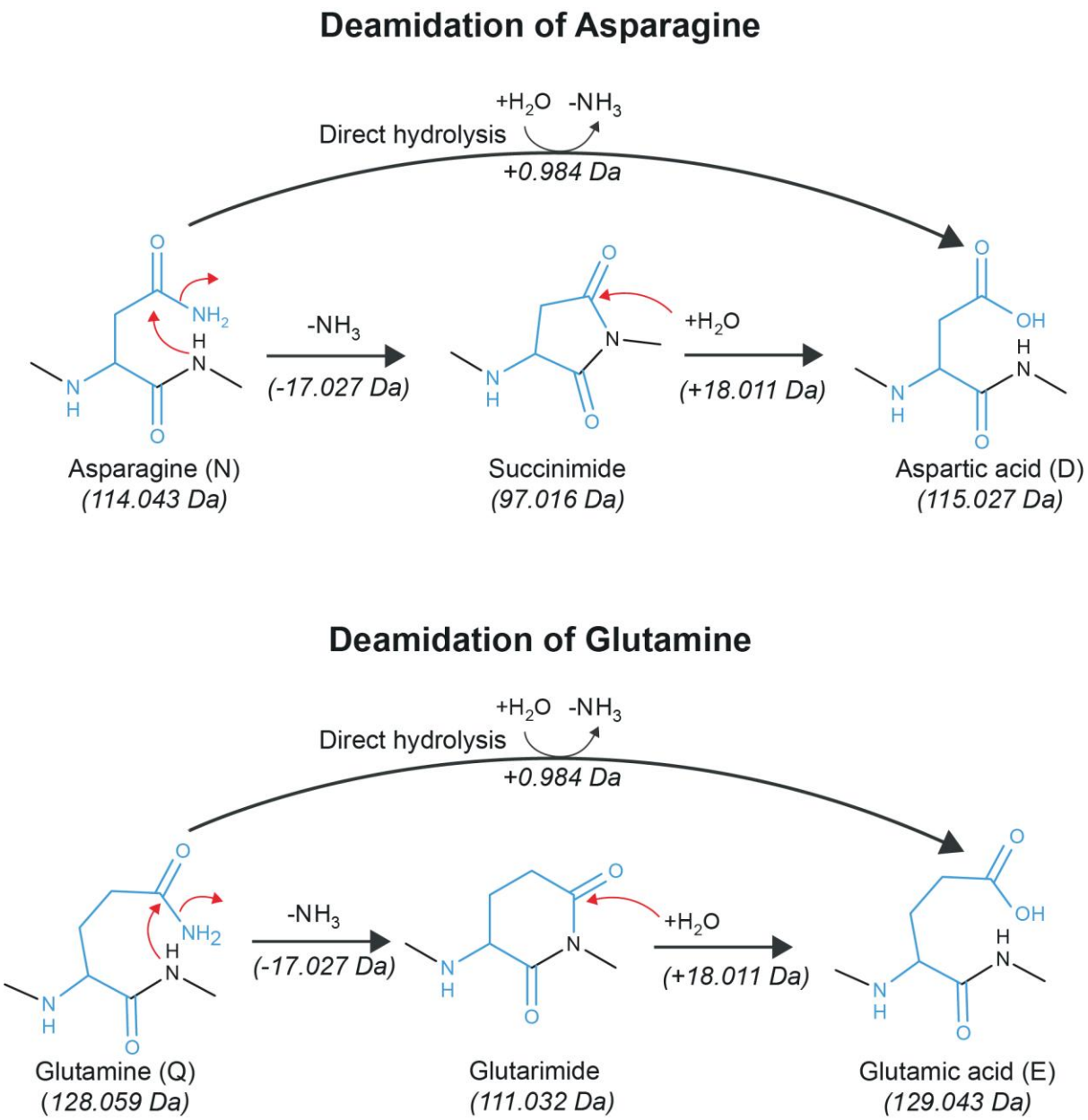


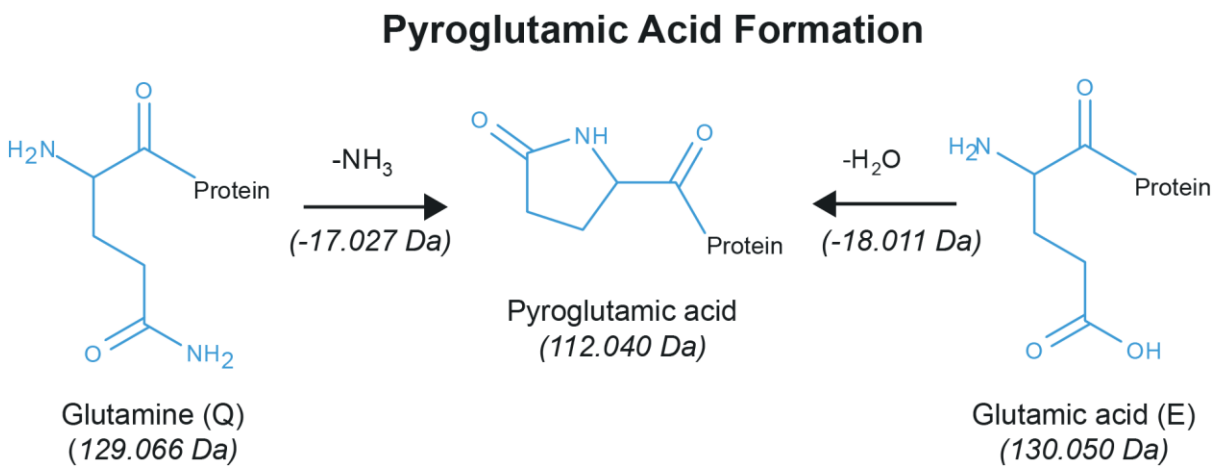

**Figure 1: diagram showing deamidation of asparagine (N) to aspartic acid (D) (top) and glutamine (Q) to glutamic acid (E) (bottom), illustrating both direct hydrolysis and cyclic intermediate pathways. Note: the formation of isoaspartic acid and isoglutamic acid is not depicted.**

Recognising that the mass of aspartic acid (D) is indistinguishable from deamidated asparagine (N), Warinner et al. (2014) observed that such modification induced sequence changes could result in potentially incorrect LCA matches. In particular, they found that deamidation of an asparagine (N) residue in the *Ovis* whey protein β-lactoglobulin (βLG) peptide TPEVD**N**EALEK causes it to be indistinguishable from the Bovinae βLG peptide TPEVD**D**EALEK, meaning that the detection of TPEVD**D**EALEK could equally be derived from Bovinae or *Ovis* dairy in ancient samples. Here I propose the term Modification Induced Sequence Permutation (MISP) to describe such peptide variants. Crucially in this case, the detection of inaccurate dairy species could lead to spurious interpretations of dairy exploitation and species introduction. Given that the peptide in question is more frequently detected than any other in ancient dental calculus and ceramic residues (Hendy et al., 2018a, 2018b; Jeong et al., 2018; Wilkin et al., 2020), this discovery has gained widespread application (e.g. Charlton et al., 2019; Evans et al., 2023; Hendy et al., 2018a, 2018c).

Several similar MISPs have been noted in the interim, especially for other βLG peptides. Jersie-Christensen et al. (2018) reported that the βLG peptide PTPEG**D/N**LEILLQK which they detected in medieval dental calculus samples could match to either Bovinae or *Ovis* respectively. Analysing βLG on neonate dog bones from 430–960 CE Japan, Tsutaya et al. (2019) reported that the peptide TME**D/N**LDLQK could belong to either Canidae, the most parsimonious interpretation, or alternatively to *Callorhinus ursinus* (northern fur seal), *Odobenus rosmarus divergens* (Pacific walrus), and *Leptonychotes weddellii* (Weddell seal). Lastly, in their analysis of a well-preserved Early Bronze Age wooden container from Switzerland, Colonese et al. (2017) reported a MISP for the peptide TLFGS**D/N**PPIPTPVLTK, which could render its initial LCA match to *Hordeum vulgare* insecure. To date, no glutamine (Q) deamidation MISPs have been reported in the ancient protein literature, which is perhaps unsurprisingly given the lower rate of glutamine (Q) deamidation compared to asparagine (N), however, deamidation of glutamine could also potentially result in incorrect taxonomic detection. It is also possible that further examples have been recognised but not reported in the literature, i.e. if researchers who do screen for MISPs simply provide additional LCAs in their results, without explaining that these were caused by modification-induced sequence changes.

Another PTM which causes amino acid mass shifts that could lead to MISPs and thus potentially inaccurate LCAs is the conversion of N-terminus glutamine or glutamic acid to pyroglutamic acid (Gln →pyroGlu or Glu →pyroGlu). Pyroglutamic acid forms via two different pathways (Figure 2). The loss of an amide group (i.e. deamidation) converts glutamine to pyroglutamic acid, or alternatively, the loss of a water molecule (i.e.

116 dehydration) converts glutamic acid to pyroglutamic acid. Pyroglutamic acid is formed
117 through the intramolecular cyclisation of N-terminal glutamine (Q) or glutamic acid (E)
118 residues via both non-enzymatic (Haitinger, 1882) and enzymatic processes (Connell et al.,
119 1956; Woodward and Reinhart, 1942). This transformation can occur under physiological
120 conditions or after an organism's death, and is influenced by factors such as temperature,
121 pH, enzymatic activity, and pressure (e.g., Chelius et al., 2006; Dick et al., 2007; for reviews
122 see Gazme et al., 2019; Kumar & Bachhawat, 2012). Due to the identical masses of
123 glutamine (Q) minus $NH_3$, pyroglutamic acid, and glutamic acid (E) minus $H_2O$, any time a
124 peptide with the Gln→pyroGlu (N-term Q) or Glu → pyroGlu (N-term E) modification is
125 detected by search software, the peptide's N-terminus amino acid could equally be glutamic
126 acid (E) or glutamine (Q). So far, the only instances of this PTM causing MISPs reported in
127 the archaeological literature are in Evans (2024). However, its implications could potentially
128 be far reaching.
129




130
131 **Figure 2: Conversion of N-terminal Glutamine and Glutamic Acid to Pyroglutamic**
132 **Acid.**
133
134 Given the lack of software tools for MISP detection, researchers seeking to screen for MISPs
135 are left to manually calculate peptide permutations and their taxonomies, which is a time-
136 consuming task, and is yet to be standard practice. It is thus perhaps unsurprising that none
137 of the reported examples involve multiple modification sites, despite the equal mass of multi-
138 modification site MISPs, those with single modifications, and the original sequence. Peptides
139 with numerous potential deamidation sites have drastically increased numbers of MISPs and
140 therefore potentially taxonomies, making their detection time-consuming without
141 programmatic tools. Despite the handful of reported examples, the issue of modification-
142 induced changes to taxonomic detection is likely much more widespread than previously
143 realised. In reality, any time a peptide containing a glutamic acid (E) or aspartic acid (D), a

deamidated glutamine (Q) or asparagine (N), or a Gln/Glu→ pyroGlu modification is detected by the search software, the sequence and potentially its taxonomy may be incorrect.

To enable researchers to combat this issue, I present `The Demodifier` (1.4.0). This tool screens for all possible MISPs of a given list of peptides and their modifications, and assigns taxonomy using the Unipept API (Vande Moortele et al. 2025; Mesuere et al., 2016). The aim of `The Demodifier` is to reveal peptides with MISPs which yield additional alternate taxonomies, enabling the researcher to more accurately interpret the taxonomies of peptides represented in their samples. I then use `The Demodifier` (1.4.0) to screen published ancient dietary peptides for alternative, equally possible taxonomic detections, with the aim of assessing its utility on a sample of real ancient peptides. The results reveal many MISPs resulting in equifinal taxonomic detections, which have yet to be recognised in previous palaeoproteomic studies. As more ancient metaproteomic analyses are published, further modification-induced sequence permutations will likely be identified.

# Materials and Methods

## The Demodifier

**Overview**

`The Demodifier` (1.4.0) screens for potential modification-induced sequence changes, which yield alternate taxonomies to the input peptide. It consists of a Python3 script which can be run either as an executable on Windows or Linux, or on the command line: https://doi.org/10.5281/zenodo.16961714. A tutorial for `The Demodifier` is available at https://github.com/miranda-e/Demodifier. `The Demodifier` requires the external libraries: requests (2.28.1) and urllib3 (1.26.7) and was written in Python 3.9.7.
Figure 3 provides a schematic overview of `The Demodifier's` process. Briefly, `The Demodifier` accepts as input a CSV with one column called “Sequence” or “pep_seq” containing peptide sequences, and one column called “Modifications” or “pep_var_mod” containing their corresponding variable modifications (if detected) in either Mascot or Maxquant format. `The Demodifier` first parses the input data, extracting the number of modifications for each peptide (Figure 3 step A). It then calls the Unipept API (Vande Moortele et al. 2025; Mesuere et al., 2016) using the pept2lca function to gather the LCAs for each input peptide (Figure 3 step B), before generating all possible deamidation-induced

177 permutations of that input sequence for glutamine (Q), asparagine (N), glutamic acid (E) and
178 aspartic acid (D) residues, and all possible pyroGlu induced permutations for N-terminal
179 glutamine (Q) and glutamic acid (E) (Figure 3 step C, see also further details below). `The`
180 `Demodifier` then removes duplicate MISPs (Figure 3 step D) before querying the Unipept
181 API to gather the LCAs of each MISP (Figure 3 step A). `The Demodifier` summarises the
182 total number of MISPs, total number of MISPs which yield LCAs and whether the input
183 peptides resultant MISPs yield identical LCAs (Figure 3 step F) and decides which input
184 peptides require manual checking to assess which LCAs are supported by peptide ion
185 coverage (Figure 3 step G). Finally, the results are output in two CSV files and a JSON
186 described below (Figure 3 step H). The researcher is then able to quickly detect input
187 peptides with additional LCAs by consulting the results CSV, before returning to standard
188 palaeoproteomic analysis process (BLAST search peptides, determine overall protein LCA
189 by interpreting the LCAs of all peptides in the sample, make archaeological interpretations).
190 The script incorporates retry logic for transient failures and manages API sessions through
191 coordinated handling. It is multi-threaded, executing tasks in concurrently for optimised
192 performance, and allows the user to specify the number of processors.
193
194 **Output**
195
196 `The Demodifier` produces a results summary CSV, a permutations CSV and a results
197 JSON. The results summary CSV contains all input peptides and modifications, annotated
198 with their input LCA, their total number of MISPs, the number of MISPs that yielded LCAs,
199 the specific MISPs that yielded LCA results, their resulting LCAs, whether all resulting LCAs
200 are identical and whether inspection of peptide fragment coverage is required to determine
201 which permutations are supported by a particular peptide spectral match (PSM). The
202 permutations CSV contains the input peptide, modifications, each MISP generated, and
203 each MISP's LCA where they were available. The JSON contains all the above information
204 in a nested format.
205

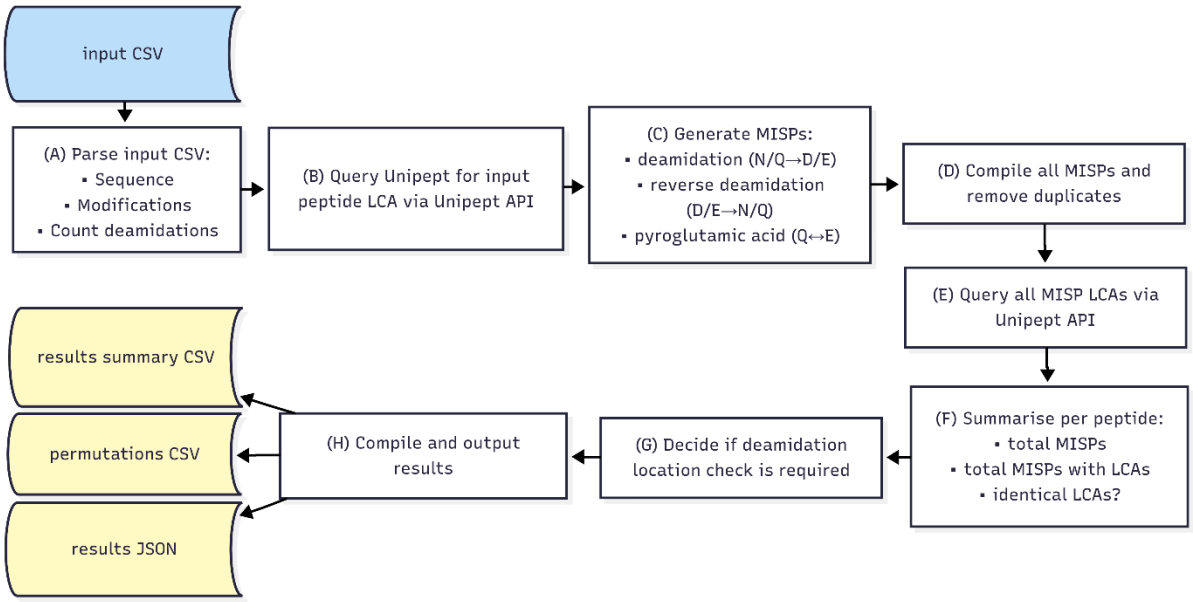



**Figure 3: Schematic overview demonstrating The Demodifier's pipeline. Figure created in Mermaid.live**

## Permutation generation

`The Demodifier` simulates six potential substitutions (Figure 4). First, `The Demodifier` simulates all permutations of deamidation of glutamine (Q) and asparagine (N) (Figure 4, step B) by replacing these residues with glutamic acid (E) and aspartic acid (D), respectively, up to the number of deamidations detected for a given input peptide by search software, and ignoring N-terminal glutamine (Q) if the pyroGlu modification is present. Only permutations with an identical mass to the input peptide are created: i.e. those with the same number or fewer deamidated residues than the input peptide, given that only deamidated glutamine (Q) and asparagine (N) residues have identical masses to glutamic acid (E) and aspartic acid (D), respectively. For each resultant deamidation permutation, `The Demodifier` then simulates all possible *reverse* deamidation permutations, excluding positions which were previously altered in the first stage, and N-terminal glutamic acid (E) if the pyroGlu modification was detected (Figure 4, step C). That is, it *reverses* deamidation of glutamic acid (E) and aspartic acid (D), *returning* them to glutamine (Q) and asparagine (N) respectively with no maximum number of substitutions (given that glutamic acid and aspartic acid residues have identical masses to deamidated glutamine and asparagine residues respectively). Finally, for each resultant permutation, `The Demodifier` then simulates glutamine (Q) to glutamic acid (E) substitutions for peptides bearing N-terminal Q with the Gln → pyroGlu modification detection, and E to Q substitutions for peptides bearing N-terminal E with Glu → pyroGlu modification detection (Figure 4, step C). Any duplicate peptide permutations for a given input peptide are removed prior to downstream analysis.

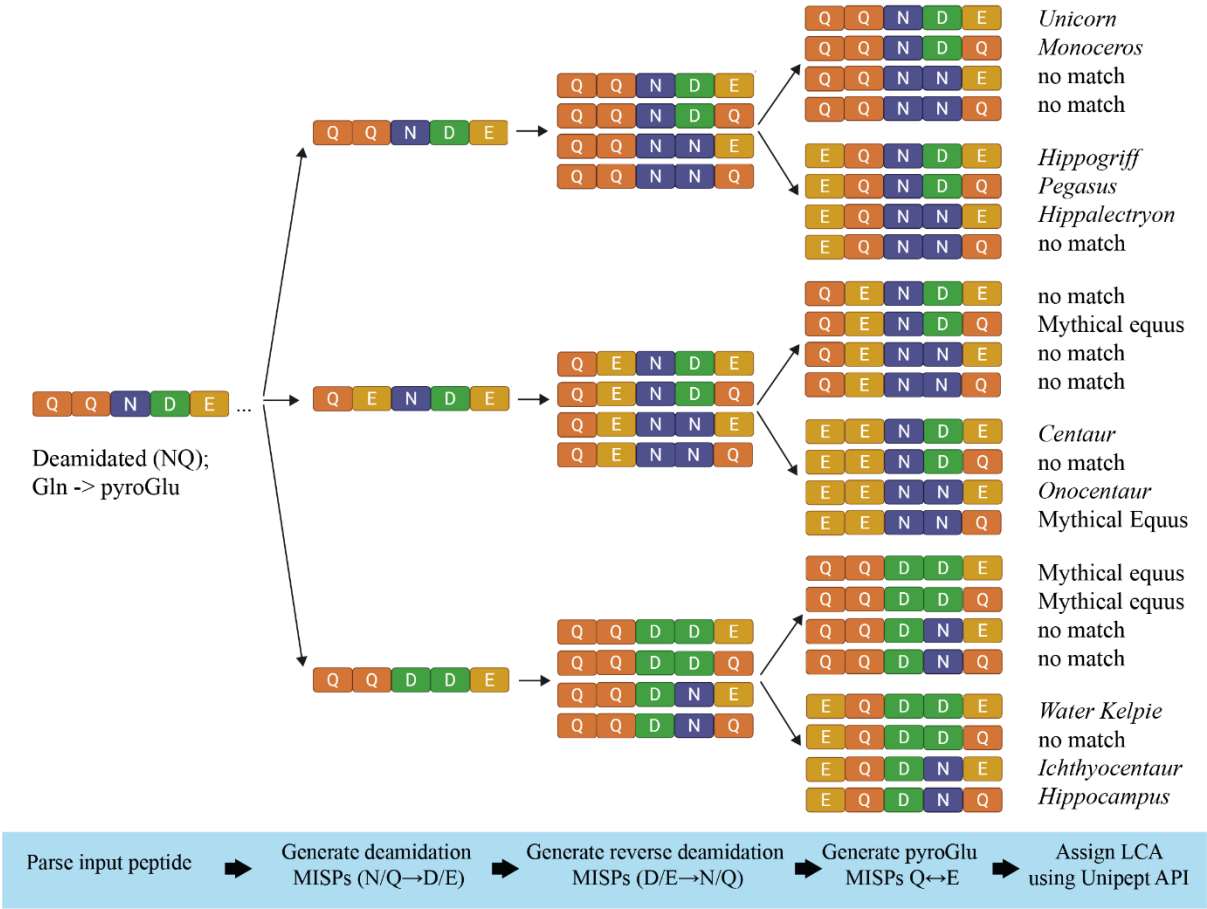



**Figure 4: The Demodifier's process of making sequence permutations for an imaginary peptide. First, all possible deamidations of N/Q to D/E are simulated up to the number detected by search software and ignoring N-terminal Q if the pyroGlu modification was detected, next reverse deamidation permutations of D/E to N/Q are simulated, ignoring previously altered positions and N-terminal E if the pyroGlu modification was detected, and then if N-term pyroGlu formation was detected, Q to E or E to Q substitutions are simulated. Peptide permutation duplicates are removed (if present) before LCA is assigned for each MISP using Unipept pept2LCA.**
**In the case of this imaginary peptide, the LCA of the input peptide is Unicorn, however the MISPs yield LCAs matching to several other Mythical Equus. Figure created in Biorender.com**

## Flagging potentially unsupportable LCAs

Certain MISPS would only be possible if deamidation was detected at a particular position in the sequence during the proteomic search. `The Demodifier` only considers the peptide sequence when generating MISPs, as opposed to the ion coverage of a particular PSM, and therefore does not know the location of deamidations. As a result, where deamidation was

detected by search software at a particular position `The Demodifier` may generate MISPs at other positions which are not actually possible for a particular PSM. This possibility can only occur when the number of deamidations detected by the search software is less than the number of potentially deamidatable residues (NQ), meaning the deamidations could be in several different locations, some of which may not be supported by a PSMs ion coverage. Only peptides which satisfy this criterion *and* generate multiple MISPs which match real protein sequences could be impacted. To flag these cases, the Demodifer outputs a column in the results CSV flagging the input peptides which require manual verification of possible sites of deamidation (i.e. through scrutiny of MS2 spectra, or consulting deamidation locations and probabilities if these are provided by the search software). Peptides are flagged as requiring further investigation if the number of deamidations in the “Modifications” column is less than the total number of asparagine (N) and glutamine (Q) residues in a given input peptide (excluding N-term Q if a pyroGlu modification was detected), and the number of deamidations detected is not zero, and the resulting MISPS generate multiple different LCAs which are not identical.

# Test dataset

A test dataset consisting of peptides and modifications previously detected in archaeological samples was assembled to explore `The Demodifier's` ability to detect modification-induced sequence variations and their resulting taxonomies. The test dataset consists of all unique peptides reported in ancient protein studies of dental calculus and vessels which are readily accessible (see Supplementary Information 1), and which include corresponding peptide modifications on a per-peptide basis. The dataset was filtered to exclude identical peptide and modification combinations (with PTMs other than deamidation and pyroglutamic acid formation excluded), and sequence and modification terminology was converted to Mascot format (Supplementary Information 2). The resulting dataset consisted of 1076 unique peptide and modification combinations drawn from 19 published studies (Bleasdale et al., 2021; Charlton et al., 2019; Chowdhury et al., 2021; Evans et al., 2023; Geber et al., 2019; Hendy et al., 2018a, 2018b; Jeong et al., 2018; Lyu et al., 2024; Scott et al., 2021; Siano et al., 2022; Solazzo et al., 2008; Tan et al., 2024; Tanasi et al., 2021; Ventresca Miller et al., 2023; Warinner et al., 2014; Wilkin et al., 2024, 2023, 2021) and my PhD thesis (Evans 2024). The PRIDE accession numbers for these datasets can be found in Supplementary Information 3, where they were available. Of these 20 datasets, 19 included deamidation (NQ) as a variable modification in their search parameters, while nine included pyroglutamic acid formation from glutamine (Q), and three included pyroglutamic acid formation from glutamic acid (E), (Supplementary Information 3).

The focus of the case study was to assess the utility of The Demodifier at detecting all deamidation and pyroglutamic acid-based MISPs and their LCAs from a large dataset of archaeologically plausible peptides, rather than to re-assess the taxonomic interpretations of particular PSMs from specific studies. Therefore, the original peptide ion coverages were not consulted, and no MISPs were excluded from this discussion as all would be potentially possible, given different PSMs of the same peptide.

# Results and Discussion

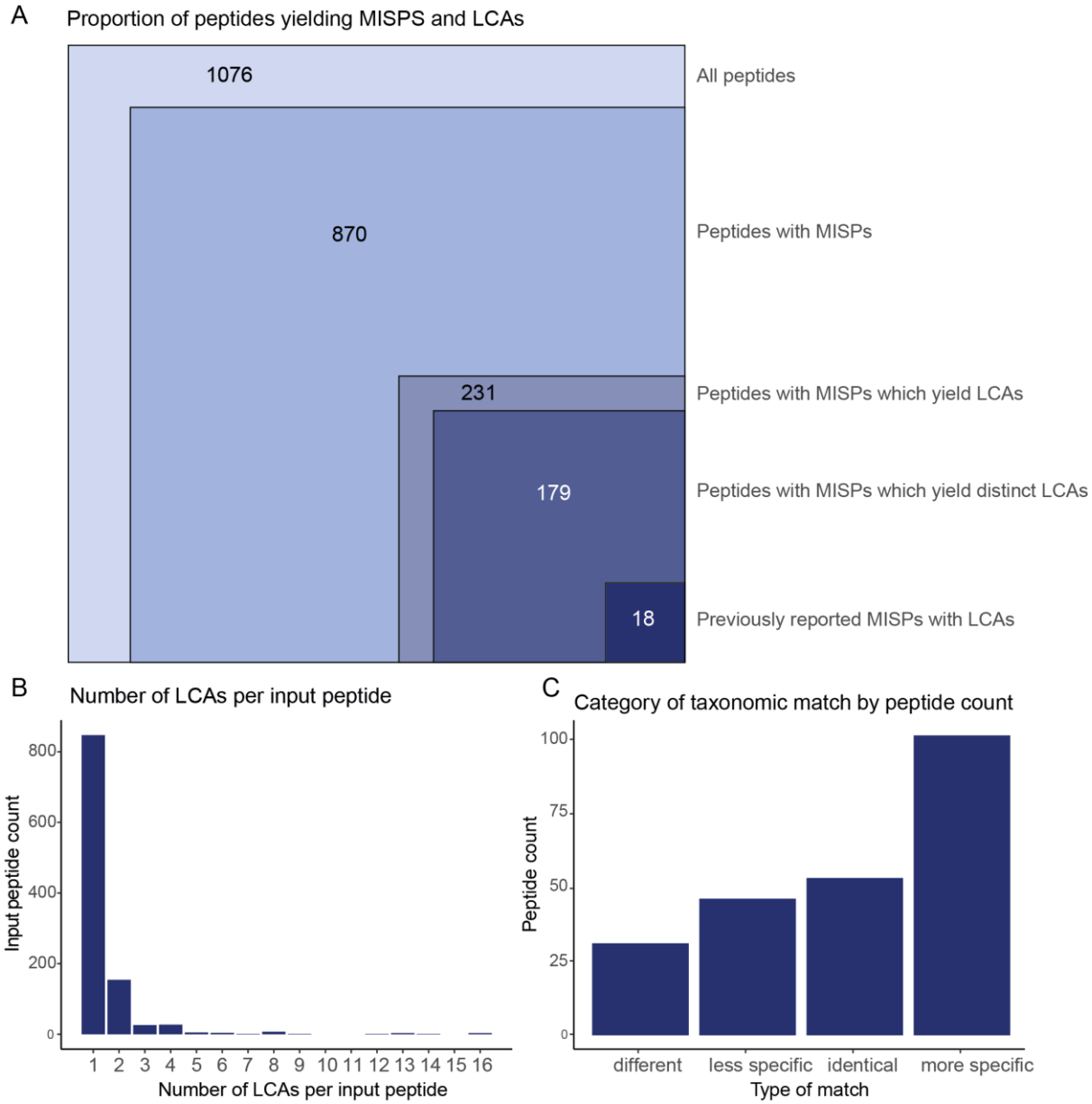
## How common were modification induced sequence permutations?

Modification-induced sequence permutations were identified for 80.9% of input peptides, with 870 of the 1076 input peptides generating at least one MISP, meaning the remaining 206 input peptides contained no D/E, deamidated N/Q or pyroGlu residues (Figure 3). A total of 14629 MISPs were generated from the 1076 input peptides (Supplementary Information 4 and 5). The median number of MISPs generated for an input peptide was 4, with a mode of 2 and mean of 13.5, although the mean was substantially increased by one input β-casein peptide, FQSEEQQQTEDELQDKIHPFAQTQ + 3 deamidated (NQ), which yielded 4096 MISPs (only one of which matched to a real protein sequence).

## How frequently did modification induced sequence permutations result in multiple taxonomic matches?

The results revealed that 16.6% of peptides (179 of 1076) generated at least one MISP which resulted in a taxonomic identification different from that of the input peptide (Figure 4A). Only 18 of these have been reported previously, meaning that modification induced alternate peptide taxonomy is 15% more common than currently reported. While some peptides had multiple MISPs yielding distinct taxonomies in addition to their input peptide taxonomy, most only had one additional taxonomic match (Figure 4B). The vast majority of MISPs yielding LCAs resulted from deamidation (220 of 231), while a small minority was caused by a pyroglutamic acid modification (11 of 231). This is probably largely due to the fact that deamidation can occur multiple times in a peptide, while pyroglutamic acid is only formed at the N-terminus. However, it could also be because fewer studies in the test dataset included the formation of N-terminal pyroglutamic acid as a variable modification in

316 their search parameters (9 of 20), compared to deamidation (19 of 20) (Supplementary
317 Information 3).
318




319
320 **Figure 4: A: Proportional figure showing the total number of peptides in the test**
321 **dataset, number of peptides with MISPs, number of peptides with MISPs that match to**
322 **real protein sequences, those which additionally match to multiple distinct**
323 **sequences, and those which have previously been reported, B: Number of LCAs per**
324 **input peptide by input peptide count C: MISP LCA categories by input peptide count**
325 **including: LCAs which were different to the input peptide's LCA, identical to the input**
326 **peptide's LCA, less taxonomically specific than the input peptide's LCA, or more**
327 **taxonomically specific than the input peptide's LCA.**

## What was `The Demodifier's` run time?

`The Demodifier's` mean run time on the test dataset using 8 processors was 7.003 seconds (range= 6.91 to 7.15 seconds, number of replicates=10). A minimal test was undertaken to assess whether run time would increase in a linear or exponential fashion with increased input size (Supplementary Information 6). This revealed that run time increased approximately linearly with increased peptide number input, meaning that `The Demodifier` would be applicable to larger datasets than that tested here. I note, however, that the Unipept response time varies which can impact total run time.

## Archaeological implications of modification-induced alternative peptide taxonomy

When scrutinised, it was apparent that some, but not all, of the 231 input peptides with MISPs which yielded LCAs could lead to inaccurate archaeological interpretations. Excluding the 52 MISPs which rendered identical LCAs to the input sequence (Supplementary Information 7), the remaining 179 peptides with MISPs could be classified into three categories (Figure 4C, Supplementary Information 7):

1) Peptides that could equally be derived from two or more totally different taxonomies
2) Peptides that could be more taxonomically specific than originally thought
3) Peptides that could be less taxonomically specific than originally thought

Of the input peptides with multiple LCA matches, if they had at least one “less specific” match, they were categorised into the third category, even if they also had “more specific” matches, given the more serious implications of less specific taxonomy on an archaeological interpretation. A multiclass confusion matrix comparing input peptide LCA to MISP peptide LCAs can be seen in Supplementary Information 8.

### Peptides yielding different LCAs to input peptide

The first category included peptides which could belong to two completely different taxonomic lineages, as is the case for the well-known peptide TPEVD**D/N**EALEK, which could equally be derived from *Ovis* or Bovinae βLG. This class consisted of 31 peptides derived from 14 proteins (Supplementary Information 7). Sometimes, such peptides would be easily detectable because they initially appear to belong to an archaeologically unlikely

taxonomy, such as the haemoglobin peptide VDQVGAEALGR (deamidated) detected by Evans (2024), which initially matched to *Triturus cristatus* (Northern crested newt), but has a MISP which matches to Laurasiatheria (a superorder of placental mammals which includes all true insectivores, bats, carnivorans, pangolins, even-toed ungulates and odd-toed ungulates). Conversely, sometimes this category of MISPs produce additional LCAs which would be unlikely in most archaeological contexts. For example, the peptide βLG TKIPAVFKID, which initially matches to Bovidae, but also has a MISP matching specifically to *Physeter catodon* (Sperm whale). However, other examples are archaeologically plausible and could easily be missed because their input peptide LCAs seem perfectly reasonable. For example, the βLG peptide IPAVFKIDALNEN + 2 Deamidated (NQ) (Also detected by Evans (2024)), initially appears to match Pecora (an infra-order of even-toed hoofed ruminants), but also has a MISP which matches to *Sus scrofa*. This class of peptides can result in potentially spurious taxonomic detections, which could impact archaeological interpretation. Ancient samples often yield only a few peptides of a particular protein, of which fewer still are likely to be taxonomically specific. In such cases, unscreened MISPs could alter archaeological interpretations by leading the researcher to believe that a potentially spurious taxon was utilised on their site, causing them to miss evidence for a taxon which was actually present.

## Peptides yielding more specific LCAs than input peptide

The second class of MISPs includes peptides that could potentially be more taxonomically specific than indicated by the input peptide LCA (Supplementary Information 7). This class comprised 101 peptides derived from 60 proteins. For instance, the α-lactalbumin peptide NNGKTEYGLFQINNK + 3 Deamidated (NQ) initially appears taxonomically conserved, matching to Laurasiatheria, however `The Demodifier` revealed a MISP belonging specifically to *Equus caballus* (Horse), a species for which the archaeological sample bore other peptide matches (Ventresca Miller et al., 2023). It's worth noting that given the higher number of detected deamidations than deamidatable residues in this peptide, scrutiny of deamidation locations would be required to confirm whether this MISP LCA was supported in each specific PSM. Another example is the βLG region VYVEELKPTPE (identified in seven of the test dataset studies), which initially has an LCA of Pecora, but in fact generates several MISPs matching to more specific taxonomies (*bos taurus, Moschus moschiferus* and *Odocoileus virginianus texanus).* Peptides from plant proteins were also identified in this category, including the 12S seed storage globulin 1 peptide ALPVDVLANAYR (unmodified) detected by Hendy et al. (2018b), which initially matches to Magnoliopsida (a large class of flowering plants) but has a MISP matching specifically to *Avena sativa* (Oat) (although at the time of their study, both MISPs would have matched to *Avena* due to subsequent database

396 additions). Another example is the Gamma-hordein-1 peptide VMQQQCCLQLAQIPEQYK +
397 2 Deamidated (NQ), detected by Hendy et al. (2018a), which initially matches to Triticeae (a
398 tribe of grasses which includes many domesticated grains including wheat, barley and rye)
399 but has a MISP matching specifically to *Hordeum vulgare subsp. vulgare* (domesticated
400 barley); a taxon for which there was additional proteomic evidence in this sample. This is
401 another example of a MISP LCA which would require scrutiny of deamidation locations to
402 confirm which LCA or LCAs were supported by a specific PSM.
403
404 This class of MISPs may not change archaeological interpretation as the researcher can't
405 exclude the possibility that the peptide only bears evidence for the less specific taxonomy.
406 However such findings may offer additional insight when viewed in the context of other
407 peptides of a given protein in an ancient sample. For example, the alpha S1 casein peptide
408 region IGVNQELAYFYP**E**LFR (unmodified), detected by Tanasi et al. (2021) and Siano et al.
409 (2022) in Maltese and Italian contexts respectively, initially appears to belong to Bovinae but
410 has a MISP (IGVNQELAYFYP**Q**LFR) which belongs specifically to *Bubalus Bubalis* (water
411 buffalo). Similarly, the frequently detected βLG peptide VLVLDTDYKK (unmodified), which
412 was reported in 13 of the test-dataset studies, initially appears to belong to Pecora, yet also
413 has a MISP belonging to *Bubalus Bubalis*. Water buffalo are a major dairy animal, meaning
414 this finding would be important in archaeological contexts where they are present, especially
415 considering that water buffalo βLG only has one species-specific tryptic peptide, and it is not
416 in a sequence region that has frequently been detected.

## 417 Peptides yielding less specific LCAs than input peptide

418 Conversely, the third class of MISP taxonomies are those that are less taxonomically
419 specific than the input peptide taxonomy. In the test dataset, this class consisted of 46
420 peptides derived from 30 different proteins (Supplementary Information 7). These examples
421 have the potential to change archaeological interpretations by reducing the taxonomic
422 specificity of detected proteins, meaning that archaeological evidence for more specific (but
423 potentially inaccurate) taxa may be removed. However, I note that in the test dataset, the
424 samples often had additional taxonomically-specific peptides, meaning the protein taxonomy
425 would not have changed. For example, while it initially appears that the α-amylase inhibitor
426 peptide SGPWSWCNPATGYK + Deamidated (NQ) detected by Tanasi et al. (2021)
427 matches *Triticum aestivum* (Common wheat), `The Demodifier` revealed a MISP with an
428 LCA of Triticeae (a tribe of grasses which includes many domesticated grains such as
429 wheat, barley and rye). Similarly, the β casein peptide GVAGEPGRNGLPGGPGLR +
430 Deamidated (NQ) detected by Siano et al. (2022) initially appears to match *Bos taurus*
431 (cattle) but in fact has a MISP with the LCA of Artiodactyla (an order of even-toed ungulates

which contains most domestic and several wild species). Another example is the 11S globulin isoform 4 peptide QEEEPYYGR (unmodified) detected by Scott et al. (2021), which initially matches *Sesamum indicum* (Sesame) but has a MISP belonging to Eukaryota. While, as mentioned above, in each of these studies the presence of additional species-specific peptides in the sample meant that the protein taxonomic assessment would not have been altered, if such peptides had been the lynchpin in a protein's taxonomic assessment, this could have impacted the archaeological interpretation.

As with the first category of matches, occasionally matches of this class may be readily detectable due to archaeologically unlikely taxonomic matches yielded by the input peptides. For instance, Evans (2024) detected the Haemoglobin subunit beta peptide QEVGGEALGR + Gln→pyroGlu in a Romano-British vessel that initially appears to belong to *Hippopotamus amphibius* (Hippopotamus), which is highly improbable in this context. Upon further inspection this peptide generated a MISP matching to Sarcopterygii (a clade of vertebrates which contains all tetrapods). As can be seen, this class of MISPs could have serious implications on archaeological interpretation of proteomic results, as they may reveal the initial peptide taxonomy to be inaccurately specific.

This analysis revealed that `The Demodifier` was effective at screening for MISPS in the test dataset, which were shown to be 15% more common than previously reported. The results revealed that modification-induced alternate peptide taxonomies could potentially be pivotal in archaeological analyses, as they can completely change taxonomic detections or provide more (or less) specific taxonomic detections than initially thought. I reiterate that the risk of incorrect taxonomic matches due to MISPs would have been alleviated in many cases in the test dataset due to the presence of other taxonomically specific peptides contributing to overall protein LCA in the ancient samples in question. However, this will not always be the case, and since taxonomic specificity is one of the significant advantages of metaproteomics over other analytic methods, recognising these alternate peptide taxonomies in archaeological interpretations is crucial in avoiding spurious or overly specific taxonomic detections.

# Challenges and Future Directions

## LCA attribution

The biggest challenge to `The Demodifier` is that automated LCA assignment using Unipept pept2LCA can only sometimes capture the nuances of manual BLAST searching.

As such, I stress that `The Demodifier` was designed to screen for MISPs that yielded alternate taxonomies but not to replace manual BLAST-based LCA analysis. When a peptide matches to select (but not closely related) taxonomies, the LCA can appear needlessly broad, particularly where peptide matches to archaeologically impossible or unlikely species are present. Therefore, once peptides bearing MISPs have been detected by `The Demodifier`, users should further assess taxonomic matches using tools that enable scrutiny of each sequence match (e.g. NCBI protein-protein BLAST GUI). By scrutinising matched protein sequences manually, the user is able to investigate the quality of sequences and their relevance to the context. For example, several bacterial sequences in NCBI NR appear to have become contaminated with common food proteins such as grain protein sequences (Scott et al., 2021); or those used in cell culture including milk proteins (Wilkin et al., 2021) and ovalbumin (Geber et al., 2019). This means that automatic LCAs may be assigned at a much higher taxonomy than is actually accurate. Similarly, sometimes manually scrutinising the BLAST matched sequences allows the user to flag potentially low quality matches. For example, Wilkin et al. (2023) note that certain Bovinae and Caprinae blood peptides additionally match to white tailed deer and elk, but only in predicted protein sequences. Lastly, when addressing certain archaeological questions, the taxonomy to which a sequence does *not* match can sometimes be just as meaningful as the taxonomy to which a sequence does match. For example, in their analysis of dietary proteins present in human dental calculus from British Neolithic individuals, Charlton et al. (2019) discovered that one sample harboured evidence for βLG derived from *Capra sp.*, as well as evidence for Bovidae excluding *Capra*, meaning that the individual had consumed dairy from multiple distinct taxa. As a result of these challenges, while automating LCA in `The Demodifier` was necessary to detect permutations with real sequence matches, manually BLAST searching and interrogating the taxonomy trees in the web version of NCBI protein-protein BLAST can yield valuable information, and despite its time-consuming nature should be seen as the gold standard of LCA attribution in ancient protein studies.

The decision to use Unipept rather than a BLAST-based tool was made due both to recent advances in Unipept, and difficulties and limitations of other tools. Previously Unipept was only capable of generating LCA for tryptic peptides (or those with missed cleavages), meaning that it could not assign LCA for many of the degraded peptides identified in palaeoproteomic studies. However, a recent update allows semi-tryptic and non-tryptic peptides to be matched (Vande Moortele et al., 2025). While many popular BLAST-based LCA-assignment tools were investigated for use in `The Demodifier`, none fulfilled the criteria of generating LCA for peptides (rather than whole proteins), on the command line

502 (rather than through a GUI), and which do not require the end user to download the entire
503 NCBI/ TrEMBL database or use paid cloud servers (neither of which are accessible to many
504 users).This meant that it was not possible to integrate these BLAST-based tools.
505 Nevertheless, throughout this study it became apparent that the peptide LCA generated
506 using Unipept does not always match the LCA generated by NCBI BLAST searching. This is
507 partly a function of the different databases the two tools use (NCBI protein-protein BLAST
508 uses the NR database by default, while Unipept uses UniprotKB (Swissprot + TrEMBL) by
509 default. As such, cross checking is advisable, and a larger conversation about taxonomic
510 assignment in palaeoproteomics is warranted. I also note that the taxonomic results of the
511 case study are not intended to be compared directly with author-attributed LCAs from the
512 original studies. This would be an inaccurate comparison as, with the addition of new protein
513 sequences to the databases over time and the different tools used for taxonomic
514 assessment, LCAs may have changed since studies were published. Even between the
515 initial submission of this manuscript and its revision, minor changes were observed in the
516 LCAs yielded by the test dataset for a few peptides.

## 517 Additional modifications

518

519 `The Demodifier` currently screens for deamidation and pyroglutamic acid-formation
520 induced sequence permutations, however it is probable that other PTMs could cause similar
521 issues. For example, Engels et al. (2025) very recently created a tool for collagen species
522 classification, which screens for isobaric alanine-to-serine substitutions in addition to
523 deamidation. Future research could expand `The Demodifier` to detect this and other
524 additional PTMs.

## 525 Probability of modification

526

527 The current version of `The Demodifier` treats all modification sites as equally probable,
528 despite the fact that certain modifications and modification sites may be more likely than
529 others. For instance, deamidation of glutamine (Q) is less common given that it occurs over
530 a much slower timescale than deamidation of asparagine (N) (Robinson et al., 2004), and
531 residues in flexible regions have been found to be more likely sites of asparagine
532 deamidation (Kossiakoff, 1988). In the case of pyroglutamic acid formation, deamidation of
533 N-terminal glutamine (Q) occurs at a higher rate than dehydration of N-terminal glutamic acid
534 (E)), however factors such as temperature, pH, and enzymatic activity also play a role (for
535 reviews see Gazme et al., 2019; Kumar & Bachhawat, 2012). Therefore, future iterations

could introduce some form of probabilistic scoring. However, while some modification scenarios may be more likely than others, given the unknown taphonomic and cultural history of many archaeological samples, I decided to employ a conservative approach, allowing `The Demodifier` to output all MISPs regardless of probability, with the rationale that if there is any possibility that a modification could occur, it should be screened for.

## Unsupported MISPs

`The Demodifier` calculates all possible MISPs for a given peptide, regardless of deamidation locations. Therefore, as noted in the methods section, certain MISPs generated by `The Demodifier` may not be possible for a particular PSM detection and are flagged as requiring manual scrutiny of deamidation locations. In this case study, this issue only impacted 26 of the 1076 unique input peptides (2.4%). One possible solution may be to adapt `The Demodifier` to accept input peptides with deamidation locations embedded within the sequence and restrict `The Demodifier` from generating permutations for any other locations, however this is not without its problems. Firstly, depending on which search software is used, in some cases multiple deamidation positions remain possible due to incomplete ion coverage, yet only the most probable is annotated in the peptide sequence in search software outputs. In such cases, excluding all non-annotated positions may exclude MISPs that are actually possible. Secondly, not all software outputs a peptide sequence that is annotated with the modification positions, meaning that this solution may reduce accessibility of the Demodifier to some users. Another approach may have been for `The Demodifier` to accept fragment mass information and isolate the possible deamidation positions from this, however this would be drastically more complicated and resource-intensive than the current approach. Moreover, it would complicate user experience as, rather than simply inputting a (potentially filtered or curated) list of peptides of interest and their modifications, the user would have to input all fragment mass data, which may be vast and contain many contaminants. For these reasons and considering the low number of cases requiring scrutiny in the test dataset, I chose not to implement this approach.

## Archaeological usage

An important disclaimer is that `The Demodifier` is only designed to aid in detection of MISPs and any additional LCAs which they yield. It is not designed to replace any other part of the palaeoproteomic or analysis of interpretation pipeline, including BLAST searching or overall protein LCA interpretation. In particular, while the focus of this analysis was

scrutinising MISPs which yielded alternate taxonomies on a per-peptide basis, it is critical to note that in a usual archaeological use-case researchers must then interpret the resulting LCAs of all peptides of a detected protein together, to understand which protein LCAs are supported in their sample.

This study provided an opportunity to explore the utility of `The Demodifier` across a large dataset of archaeologically detected peptides, which will hopefully garner awareness of the field-wide risk of modification induced alternate peptide LCAs, and equip researchers with a rapid open source screening tool to combat this issue.

# Conclusion

In this paper I present `The Demodifier 1.4.0`: an open source tool to screen for PTM-induced alternate peptide taxonomies. The results of this study reveal that modification-induced alternate peptide taxonomies are much more common than previously recognised in ancient protein studies. Often these multimorphic peptide sequences are only discovered when unexpected taxonomic hits are unearthed, while “normal” seeming taxonomic results do not always garner scrutiny. Screening of all sequences which could have MISPs is therefore necessary, regardless of initial input peptide taxonomy. The results revealed that modification-induced alternate peptide taxonomy is 15% more common than reported in ancient protein studies, and can result in peptide permutations with completely different, more specified or less specific taxonomic LCAs compared to the input peptide. It is likely that future proteomic analyses will reveal additional peptides with alternate taxonomies, meaning screening for MISPs should be standard practice in protein studies. `The Demodifier` provides a simple and efficient open source tool for detecting such sequences, thus avoiding spurious or overly specific taxonomic detections which could yield inaccurate archaeological interpretations.

# Data Availability

A static version of `The Demodifier` (1.4.0), used in the case study analysis is available at https://doi.org/10.5281/zenodo.16961714. A tutorial for `The Demodifier` is available at https://github.com/miranda-e/Demodifier. All other data are available within the article or its supplementary information.

# Acknowledgements

I would like to thank Matthew Collins for providing valuable feedback on a version of this manuscript, and Oliver Boyd for his assistance in converting the script to an executable. I would also like to thank Meaghan Mackie, Shevan Wilkin and two anonymous reviewers for their extremely helpful reviews.

# Bibliography


Bleasdale, M., Richter, K.K., Janzen, A., Brown, S., Scott, A., Zech, J., Wilkin, S., Wang, K., Schiffels, S., Desideri, J., Besse, M., Reinold, J., Saad, M., Babiker, H., Power, R.C., Ndiema, E., Ogola, C., Manthi, F.K., Zahir, M., Petraglia, M., Trachsel, C., Nanni, P., Grossmann, J., Hendy, J., Crowther, A., Roberts, P., Goldstein, S.T., Boivin, N., 2021. Ancient proteins provide evidence of dairy consumption in eastern Africa. Nat. Commun. 12, 1–11. https://doi.org/10.1038/s41467-020-20682-3

Capasso, S., Mazzarella, L., Sica, F., Zagari, A., 1991. First evidence of spontaneous deamidation of glutamine residue via cyclic imide to α- and γ-glutamic residue under physiological conditions. J. Chem. Soc. Chem. Commun. 1667–1668. https://doi.org/10.1039/c39910001667

Catak, S., Monard, G., Aviyente, V., Ruiz-López, M.F., 2009. Deamidation of asparagine residues: direct hydrolysis versus succinimide-mediated deamidation mechanisms. J. Phys. Chem. A 113, 1111–1120. https://doi.org/10.1021/jp808597v

Charlton, S., Ramsøe, A., Collins, M., Craig, O.E., Fischer, R., Alexander, M., Speller, C.F., 2019. New insights into Neolithic milk consumption through proteomic analysis of dental calculus. Archaeol. Anthropol. Sci. 11, 6183–6196. https://doi.org/10.1007/s12520-019-00911-7

Chowdhury, M.P., Campbell, S., Buckley, M., 2021. Proteomic analysis of archaeological ceramics from Tell Khaiber, southern Iraq. J. Archaeol. Sci. 132, 105414. https://doi.org/10.1016/j.jas.2021.105414

Colonese, A.C., Hendy, J., Lucquin, A., Speller, C.F., Collins, M.J., Carrer, F., Gubler, R., Kühn, M., Fischer, R., Craig, O.E., 2017. New criteria for the molecular identification of cereal grains associated with archaeological artefacts. Sci. Rep. 7, 6633. https://doi.org/10.1038/s41598-017-06390-x

Connell, G.E., Hanes, C.S., 1956. Enzymic formation of pyrrolidone carboxylic acid from gamma-glutamyl peptides. Nature 177, 377–378. https://doi.org/10.1038/177377a0

Dick, L.W., Jr, Kim, C., Qiu, D., Cheng, K.-C., 2007. Determination of the origin of the N-terminal pyro-glutamate variation in monoclonal antibodies using model peptides. Biotechnol. Bioeng. 97, 544–553. https://doi.org/10.1002/bit.21260

Engels, I., Burnett, A., Robert, P., Pironneau, C., Abrams, G., Bouwmeester, R., Van der Plaetsen, P., Di Modica, K., Otte, M., Straus, L.G., Fischer, V., Bray, F., Mesuere, B., De Groote, I., Deforce, D., Daled, S., Dhaenens, M., 2025. Classification of collagens via peptide ambiguation, in a paleoproteomic LC-MS/MS-based taxonomic pipeline. J. Proteome Res. 24, 1907–1925. https://doi.org/10.1021/acs.jproteome.4c00962

641 Evans, M., Lundy, J., Lucquin, A., Hagan, R., Kowalski, Ł., Wilczyńki, J., Bickle, P.,
642 Adamczak, K., Craig, O.E., Robson, H.K., Hendy, J., 2023. Detection of dairy products
643 from multiple taxa in Late Neolithic pottery from Poland: an integrated biomolecular
644 approach. R Soc Open Sci 10, 230124. https://doi.org/10.1098/rsos.230124
645 Evans, M (2024, June 28). The proteomics of pottery and its residues: applications to
646 ancient foodways (PhD). University of Cambridge. https://doi.org/10.17863/CAM.115779
647 Gazme, B., Boachie, R.T., Tsopmo, A., Udenigwe, C.C., 2019. Occurrence, properties and
648 biological significance of pyroglutamyl peptides derived from different food sources.
649 Food Sci. Hum. Wellness 8, 268–274. https://doi.org/10.1016/j.fshw.2019.05.002
650 Geber, J., Tromp, M., Scott, A., Bouwman, A., Nanni, P., Grossmann, J., Hendy, J.,
651 Warinner, C., 2019. Relief food subsistence revealed by microparticle and proteomic
652 analyses of dental calculus from victims of the Great Irish Famine. Proc. Natl. Acad. Sci.
653 U. S. A. 116, 19380–19385. https://doi.org/10.1073/pnas.1908839116
654 Haitinger, L., 1882. Vorläufige mittheilung über glutaminsäure und pyrrol. Monatsh. Chem. 3,
655 228–229. https://doi.org/10.1007/bf01516801
656 Hao, P., Ren, Y., Alpert, A.J., Sze, S.K., 2011. Detection, evaluation and minimization of
657 nonenzymatic deamidation in proteomic sample preparation. Mol. Cell. Proteomics 10,
658 O111.009381. https://doi.org/10.1074/mcp.O111.009381
659 Hendy, J., Colonese, A.C., Franz, I., Fernandes, R., Fischer, R., Orton, D., Lucquin, A.,
660 Spindler, L., Anvari, J., Stroud, E., Biehl, P.F., Speller, C., Boivin, N., Mackie, M., Jersie-
661 Christensen, R.R., Olsen, J.V., Collins, M.J., Craig, O.E., Rosenstock, E., 2018a.
662 Ancient proteins from ceramic vessels at Çatalhöyük West reveal the hidden cuisine of
663 early farmers. Nat. Commun. 9, 1–10. https://doi.org/10.1038/s41467-018-06335-6
664 Hendy, J., Warinner, C., Bouwman, A., Collins, M.J., Fiddyment, S., Fischer, R., Hagan, R.,
665 Hofman, C.A., Holst, M., Chaves, E., Klaus, L., Larson, G., Mackie, M., McGrath, K.,
666 Mundorff, A.Z., Radini, A., Rao, H., Trachsel, C., Velsko, I.M., Speller, C.F., 2018b.
667 Proteomic evidence of dietary sources in ancient dental calculus. Proc. Biol. Sci. 285.
668 https://doi.org/10.1098/rspb.2018.0977
669 Hendy, J., Welker, F., Demarchi, B., Speller, C., Warinner, C., Collins, M.J., 2018c. A guide
670 to ancient protein studies. Nat Ecol Evol 2, 791–799. https://doi.org/10.1038/s41559-
671 018-0510-x
672 Hill, R.C., Wither, M.J., Nemkov, T., Barrett, A., D'Alessandro, A., Dzieciatkowska, M.,
673 Hansen, K.C., 2015. Preserved proteins from extinct bison latifrons identified by tandem
674 mass spectrometry; Hydroxylysine glycosides are a common feature of ancient
675 collagen. Mol. Cell. Proteomics 14, 1946–1958.
676 https://doi.org/10.1074/mcp.m114.047787
677 Jeong, C., Wilkin, S., Amgalantugs, T., Bouwman, A.S., Taylor, W.T.T., Hagan, R.W.,

Bromage, S., Tsolmon, S., Trachsel, C., Grossmann, J., Littleton, J., Makarewicz, C.A., Krigbaum, J., Burri, M., Scott, A., Davaasambuu, G., Wright, J., Irmer, F., Myagmar, E., Boivin, N., Robbeets, M., Rühli, F.J., Krause, J., Frohlich, B., Hendy, J., Warinner, C., 2018. Bronze Age population dynamics and the rise of dairy pastoralism on the eastern Eurasian steppe. Proceedings of the National Academy of Sciences 115, E11248–E11255. https://doi.org/10.1073/pnas.1813608115

Jersie-Christensen, R.R., Lanigan, L.T., Lyon, D., Mackie, M., Belstrøm, D., Kelstrup, C.D., Fotakis, A.K., Willerslev, E., Lynnerup, N., Jensen, L.J., Cappellini, E., Olsen, J.V., 2018. Quantitative metaproteomics of medieval dental calculus reveals individual oral health status. Nat. Commun. 9, 4744. https://doi.org/10.1038/s41467-018-07148-3

Kossiakoff, A.A., 1988. Tertiary structure is a principal determinant to protein deamidation. Science 240, 191–194. https://doi.org/10.1126/science.3353715

Kumar, A., Bachhawat, A.K., 2012. Pyroglutamic acid: throwing light on a lightly studied metabolite. Curr. Sci. 102, 288–297.

Li, X., Lin, C., O'Connor, P.B., 2010. Glutamine deamidation: differentiation of glutamic acid and gamma-glutamic acid in peptides by electron capture dissociation. Anal. Chem. 82, 3606–3615. https://doi.org/10.1021/ac9028467

Lyu, N., Yan, L., Wang, T., Lin, L., Rao, H., Yang, Y., 2024. Characterization of pottery foodcrusts through lipid and proteomic analyses: A case study from the Xiawan site in Yixing city, East China. J. Archaeol. Sci. 161, 105902. https://doi.org/10.1016/j.jas.2023.105902

Mackie, M., Rüther, P., Samodova, D., Di Gianvincenzo, F., Granzotto, C., Lyon, D., Peggie, D.A., Howard, H., Harrison, L., Jensen, L.J., Olsen, J.V., Cappellini, E., 2018. Palaeoproteomic Profiling of Conservation Layers on a 14th Century Italian Wall Painting. Angew. Chem. Int. Ed Engl. 57, 7369–7374. https://doi.org/10.1002/anie.201713020

Mesuere, B., Willems, T., Van der Jeugt, F., Devreese, B., Vandamme, P., Dawyndt, P., 2016. Unipept web services for metaproteomics analysis. Bioinformatics 32, 1746–1748. https://doi.org/10.1093/bioinformatics/btw039

Nair, B., Palomo, I.R., Markussen, B., Wiuf, C., Fiddyment, S., Collins, M.J., 2023. Parchment Glutamine Index (PQI): A novel method to estimate glutamine deamidation levels in parchment collagen obtained from low-quality MALDI-TOF data. Peer Community J. 3. https://doi.org/10.24072/pcjournal.230

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., Johnson, P.L.F., Fumagalli, M., Vilstrup, J.T., Raghavan, M., Korneliussen, T., Malaspinas, A.-S., Vogt, J., Szklarczyk, D., Kelstrup, C.D., Vinther, J., Dolocan, A., Stenderup, J., Velazquez, A.M.V., Cahill, J., Rasmussen,

M., Wang, X., Min, J., Zazula, G.D., Seguin-Orlando, A., Mortensen, C., Magnussen, K., Thompson, J.F., Weinstock, J., Gregersen, K., Røed, K.H., Eisenmann, V., Rubin, C.J., Miller, D.C., Antczak, D.F., Bertelsen, M.F., Brunak, S., Al-Rasheid, K.A.S., Ryder, O., Andersson, L., Mundy, J., Krogh, A., Gilbert, M.T.P., Kjær, K., Sicheritz-Ponten, T., Jensen, L.J., Olsen, J.V., Hofreiter, M., Nielsen, R., Shapiro, B., Wang, J., Willerslev, E., 2013. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. Nature 499, 74–78. https://doi.org/10.1038/nature12323

Ramsøe, A., van Heekeren, V., Ponce, P., Fischer, R., Barnes, I., Speller, C., Collins, M.J., 2020. DeamiDATE 1.0: Site-specific deamidation as a tool to assess authenticity of members of ancient proteomes. J. Archaeol. Sci. 115, 105080. https://doi.org/10.1016/j.jas.2020.105080

Robinson, N.E., Robinson, Z.W., Robinson, B.R., Robinson, A.L., Robinson, J.A., Robinson, M.L., Robinson, A.B., 2004. Structure-dependent nonenzymatic deamidation of glutaminyl and asparaginyl pentapeptides. J. Pept. Res. 63, 426–436. https://doi.org/10.1111/j.1399-3011.2004.00151.x

Schroeter, E.R., Cleland, T.P., 2016. Glutamine deamidation: an indicator of antiquity, or preservational quality? Rapid Commun. Mass Spectrom. 30, 251–255. https://doi.org/10.1002/rcm.7445

Scott, A., Power, R.C., Altmann-Wendling, V., Artzy, M., Martin, M.A.S., Eisenmann, S., Hagan, R., Salazar-García, D.C., Salmon, Y., Yegorov, D., Milevski, I., Finkelstein, I., Stockhammer, P.W., Warinner, C., 2021. Exotic foods reveal contact between South Asia and the Near East during the second millennium BCE. Proc. Natl. Acad. Sci. U. S. A. 118. https://doi.org/10.1073/pnas.2014956117

Siano, F., Picariello, G., Caruso, T., Esposito, S., Rescigno, C., Addeo, F., Vasca, E., 2022. Proteomics and Integrated Techniques to Characterize Organic Residues in Funerary Findings from Italic Populations of the First Millennium BC. J. Proteome Res. 21, 1330–1339. https://doi.org/10.1021/acs.jproteome.2c00093

Solazzo, C., Fitzhugh, W.W., Rolando, C., Tokarski, C., 2008. Identification of Protein Remains in Archaeological Potsherds by Proteomics. Anal. Chem. 80, 4590–4597. https://doi.org/10.1021/ac800515v

Stratton, L.P., Kelly, R.M., Rowe, J., Shively, J.E., Smith, D.D., Carpenter, J.F., Manning, M.C., 2001. Controlling deamidation rates in a model peptide: effects of temperature, peptide concentration, and additives. J. Pharm. Sci. 90, 2141–2148. https://doi.org/10.1002/jps.1165

Tanasi, D., Cucina, A., Cunsolo, V., Saletti, R., Di Francesco, A., Greco, E., Foti, S., 2021. Paleoproteomic profiling of organic residues on prehistoric pottery from Malta. Amino Acids 53, 295–312. https://doi.org/10.1007/s00726-021-02946-4

Tan, L., Fan, X., Rao, H., Zhang, H., Yang, Y., 2024. The scientific analysis of the bronze mous excavated from Wushan, Chongqing, China: new perspectives from alloy composition analysis and proteomic method. Heritage Science 12, 88. https://doi.org/10.1186/s40494-024-01203-7

Tsutaya, T., Mackie, M., Koenig, C., Sato, T., Weber, A.W., Kato, H., Olsen, J.V., Cappellini, E., 2019. Palaeoproteomic identification of breast milk protein residues from the archaeological skeletal remains of a neonatal dog. Sci. Rep. 9, 12841. https://doi.org/10.1038/s41598-019-49183-0

Tyler-Cross, R., Schirch, V., 1991. Effects of amino acid sequence, buffers, and ionic strength on the rate and mechanism of deamidation of asparagine residues in small peptides. J. Biol. Chem. 266, 22549–22556. https://doi.org/10.1016/s0021-9258(18)54607-x

van Doorn, N.L., Wilson, J., Hollund, H., Soressi, M., Collins, M.J., 2012. Site-specific deamidation of glutamine: a new marker of bone collagen deterioration: Site-specific deamidation of glutamine. Rapid Commun. Mass Spectrom. 26, 2319–2327. https://doi.org/10.1002/rcm.6351

Vande Moortele, T., Devlaminck, B., Van de Vyver, S., Van Den Bossche, T., Martens, L., Dawyndt, P., Mesuere, B., Verschaffelt, P., 2025. Unipept in 2024: Expanding metaproteomics analysis with support for missed cleavages and semitryptic and nontryptic peptides. J. Proteome Res. 24, 949–954. https://doi.org/10.1021/acs.jproteome.4c00848

Ventresca Miller, A.R., Wilkin, S., Bayarsaikhan, J., Ramsøe, A., Clark, J., Byambadorj, B., Vanderwarf, S., Vanwezer, N., Haruda, A., Fernandes, R., Miller, B., Boivin, N., 2023. Permafrost preservation reveals proteomic evidence for yak milk consumption in the 13th century. Commun Biol 6, 351. https://doi.org/10.1038/s42003-023-04723-3

Warinner, C., Hendy, J., Speller, C., Cappellini, E., Fischer, R., Trachsel, C., Arneborg, J., Lynnerup, N., Craig, O.E., Swallow, D.M., Fotakis, A., Christensen, R.J., Olsen, J.V., Liebert, A., Montalva, N., Fiddyment, S., Charlton, S., Mackie, M., Canci, A., Bouwman, A., Rühli, F., Gilbert, M.T.P., Collins, M.J., 2014. Direct evidence of milk consumption from ancient human dental calculus. Sci. Rep. 4, 1–6. https://doi.org/10.1038/srep07104

Welker, F., Collins, M.J., Thomas, J.A., Wadsley, M., Brace, S., Cappellini, E., Turvey, S.T., Reguero, M., Gelfo, J.N., Kramarz, A., Burger, J., Thomas-Oates, J., Ashford, D.A., Ashton, P.D., Rowsell, K., Porter, D.M., Kessler, B., Fischer, R., Baessmann, C., Kaspar, S., Olsen, J.V., Kiley, P., Elliott, J.A., Kelstrup, C.D., Mullin, V., Hofreiter, M., Willerslev, E., Hublin, J.-J., Orlando, L., Barnes, I., MacPhee, R.D.E., 2015. Ancient proteins resolve the evolutionary history of Darwin's South American ungulates. Nature 522, 81–84. https://doi.org/10.1038/nature14249

Welker, F., Soressi, M.A., Roussel, M., van Riemsdijk, I., Hublin, J.-J., Collins, M.J., 2017. Variations in glutamine deamidation for a Châtelperronian bone assemblage as measured by peptide mass fingerprinting of collagen. Sci. Technol. Archaeol. Res. 3, 15–27. https://doi.org/10.1080/20548923.2016.1258825

Wilkin, S., Bayarsaikhan, J., Ganbold, A., Batsuuri, A., Ishtseren, L., Nakamura, D., Eregzen, G., Ventresca-Miller, A., Miller, B.K., 2024. Cauldrons of Bronze Age nomads reveals 2700 year old yak milk and the deep antiquity of food preparation techniques. Sci. Rep. 14, 11625. https://doi.org/10.1038/s41598-024-60607-4

Wilkin, S., Hommel, P., Ventresca Miller, A., Boivin, N., Pedergnana, A., Shishlina, N., Trifonov, V., 2023. Curated cauldrons: Preserved proteins from early copper-alloy vessels illuminate feasting practices in the Caucasian steppe. iScience 26, 107482. https://doi.org/10.1016/j.isci.2023.107482

Wilkin, S., Ventresca Miller, A., Fernandes, R., Spengler, R., Taylor, W.T.-T., Brown, D.R., Reich, D., Kennett, D.J., Culleton, B.J., Kunz, L., Fortes, C., Kitova, A., Kuznetsov, P., Epimakhov, A., Zaibert, V.F., Outram, A.K., Kitov, E., Khokhlov, A., Anthony, D., Boivin, N., 2021. Dairying enabled Early Bronze Age Yamnaya steppe expansions. Nature 598, 629–633. https://doi.org/10.1038/s41586-021-03798-4

Wilkin, S., Ventresca Miller, A., Taylor, W.T.T., Miller, B.K., Hagan, R.W., Bleasdale, M., Scott, A., Gankhuyg, S., Ramsøe, A., Uliziibayar, S., Trachsel, C., Nanni, P., Grossmann, J., Orlando, L., Horton, M., Stockhammer, P.W., Myagmar, E., Boivin, N., Warinner, C., Hendy, J., 2020. Dairy pastoralism sustained eastern Eurasian steppe populations for 5,000 years. Nat Ecol Evol 4, 346–355. https://doi.org/10.1038/s41559-020-1120-y

Wilson, J., van Doorn, N.L., Collins, M.J., 2012. Assessing the extent of bone degradation using glutamine deamidation in collagen. Anal. Chem. 84, 9041–9048. https://doi.org/10.1021/ac301333t

Xie, M., Schowen, R.L., 1999. Secondary structure and protein deamidation. J. Pharm. Sci. 88, 8–13. https://doi.org/10.1021/js9802493