

Survival of mineral-bound peptides into the Miocene

Beatrice Demarchi¹, Meaghan Mackie^{2,3}, Ziheng Li⁴, Tao Deng⁴, Matthew J. Collins^{2,5} & Julia Clarke⁶

¹ Department of Life Sciences and Systems Biology, University of Turin, Turin, Via Accademia Albertina 13, 10123 Turin, Italy

² The Globe Institute, Faculty of Health and Medical Sciences, University of Copenhagen, Øster Farimagsgade 5, 1353 Copenhagen, Denmark

³ The Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3b, 2200 Copenhagen, Denmark

⁴ Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, 100044 Beijing, China

⁵ McDonald Institute for Archaeological Research, University of Cambridge, Cambridge CB2 3ER, United Kingdom

⁶ Department of Geological Sciences, The University of Texas at Austin, Austin, TX 78712

Abstract

The oldest authenticated peptide sequences to date were reported in 2016 from 3.8 Ma old ostrich eggshell (OES) from the site of Laetoli, Tanzania (Demarchi et al., 2016). Here we demonstrate survival of the same sequences in 6.5-9 Ma OES recovered from a palaeosteppe setting in northwestern China. The eggshell is thicker than those observed in extant species and consistent with the Liushu *Struthio* sp. ootaxon. These findings push the preservation of ancient proteins back to the Miocene and highlight their potential for paleontology, paleoecology and evolutionary biology.

Keywords

Miocene, ostrich eggshell (OES), peptide, deep time, palaeoproteomics

Introduction

The oldest authenticated peptide sequences to date were reported in 2016 from 3.8 Ma old ostrich eggshell (OES) from the site of Laetoli, Tanzania (Demarchi et al., 2016), with a thermal age of the peptides estimated to be equivalent to ~16 Ma at a constant 10°C. This finding had great scientific impact since it integrated computational chemistry (molecular dynamics simulations) as well as experimental data to propose a mechanism of preservation, concluding that mineral binding ensures survival of protein sequences. Importantly, this study demonstrated that peptide-bound amino acids could survive into deep time even in hot environments. This has fuelled the analysis of ancient proteins from other mineral matrices, namely tooth enamel (Cappellini et al., 2019; Welker et al., 2020, 2019) in order to reconstruct phylogenies.

The recovery of peptides has so far been limited to biomineral samples of Plio-Pleistocene age: our attempts at retrieving intact peptides from Cretaceous eggshell were unsuccessful, despite the fact that a genuine intracrystalline fraction of amino acids was preserved in the same sample (Saitta et al., 2020). Here we show that the same peptide sequences that exhibit strong binding to the calcite surface - and that were recovered from the Laetoli OES - also persist into the late Miocene, in a > 6.5-Ma-old eggshell sample from the Linxia Basin, northeastern Tibetan Plateau, China, Liushu Formation.

Results and discussion

The ostrich eggshell (IVPP V26107) was recovered from an area between Xinji and Songming towns, Hezheng County, close to the border with Guanghe County, from mudstone facies of the Liushu Formation (Figure 1).

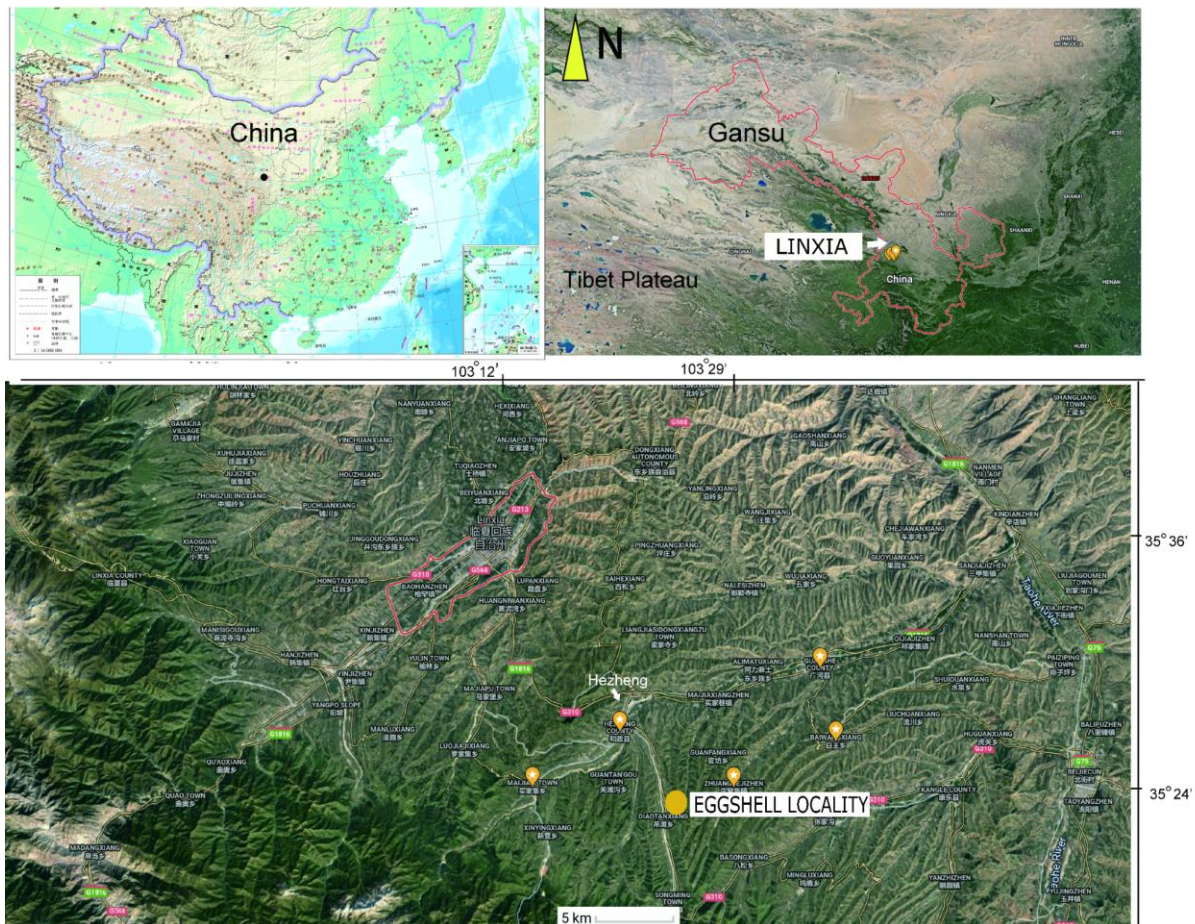


Figure 1: Map showing the location of the fossil eggshell site.

Both eggshell and skeletal remains are previously known from the Late Miocene Liushu Formation, Linxia Basin, of Gansu Province (Hou et al., 2005; Li et al., 2021; Wang, 2008). Age control on the highly fossiliferous units of the Liushu Formation is good (Deng et al., 2019, 2013; Zhang et al., 2019), with detailed correlations across China as well as with Neogene deposits globally. These units have been assessed via new magnetostratigraphic data; most of the *Hipparion* fauna is estimated to be 6.4-6.5 Ma, with a lower Liushu transition to a *Platybelodon* fauna yielding age estimates of 11.1-12.5 Ma (Zhang et al., 2019). These dates agree with the prior estimates from biostratigraphic and magnetostratigraphic data (Deng et al., 2019, 2013). The recovery site is closer to the southern part of the basin near the Heilinding section of Zhang et al. (2019), which had an estimated minimum age of 6.4-6.5 Ma (Chron 3An). In contrast, the candidate stratotype section, Guoniguo, that exposes lower parts of the Liushu, is north of the recovery area (Deng et al., 2019; Zhang et al., 2019).

Ostrich (*Struthio*) remains have a long history of recovery from the late Miocene of northwest China (Buffetaut and Angst, 2021; Hou et al., 2005; Li et al., 2021; Lowe, 1931; Mikhailov and Zelenkov, 2020). The first reported ostrich fossils from China were from units of the “*Hipparion* Clay” (Red Clay) dated to between 6.54-7.18 Ma (Lowe, 1931; Zhu et al., 2008). Mikhailov and Zelenkov (2020) concluded that the Wang (2008) Liushu taxon is referable to *Struthio* and not presently supported as a

distinct but related genus; it is proposed to be from a taxon larger than extant *Struthio* species with a thickness of ~2.4mm consistent with the eggshell sampled here (Figure 2C).

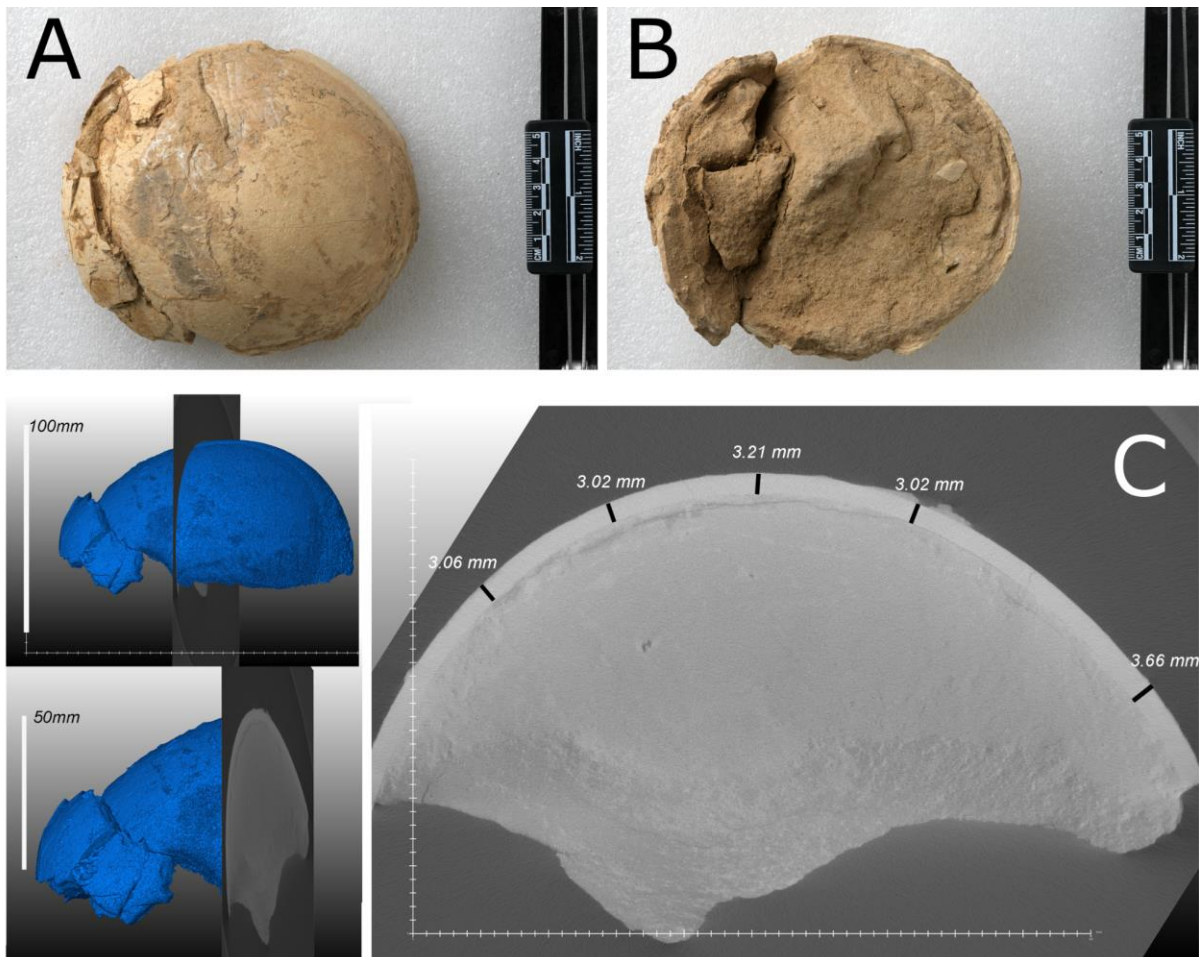


Figure 2: Eggshell specimen (IVPP V26107): photographs (A, B) and CT scan (C)

Thermal age calculations for the OES sample yielded a crude estimate of 6.8 Ma at a constant 10°C, thus thermal age and numerical age are broadly equivalent. Given the extreme antiquity of the samples, protein extraction was performed in an ultra-clean facility at the University of Copenhagen in order to minimise any chance of contamination, following Hendy et al. (2018). Furthermore, OES powders were bleached extensively to isolate the intracrystalline fraction only (as described by Demarchi et al. (2016)), samples were analysed by LC-MS/MS using a new LC column, and six blanks (analytical and procedural) were included in the run, flanking the sample. Data analysis was equally cautious: raw tandem mass spectrometry data were used to reconstruct potential peptide sequences starting from raw product ion spectra: 25 sequences were found with an ALC (Average Local Confidence) score $\geq 80\%$. This is the most stringent threshold that can be applied, and it signifies that all potential peptide sequences with lower confidence are discarded. Of these sequences, 14 contained the typical Asp-rich motif “DDDD” (Figure 3), which had been shown in our previous paper to be the mineral-binding peptide belonging to the sequence of struthiocalcin-1 (SCA-1).

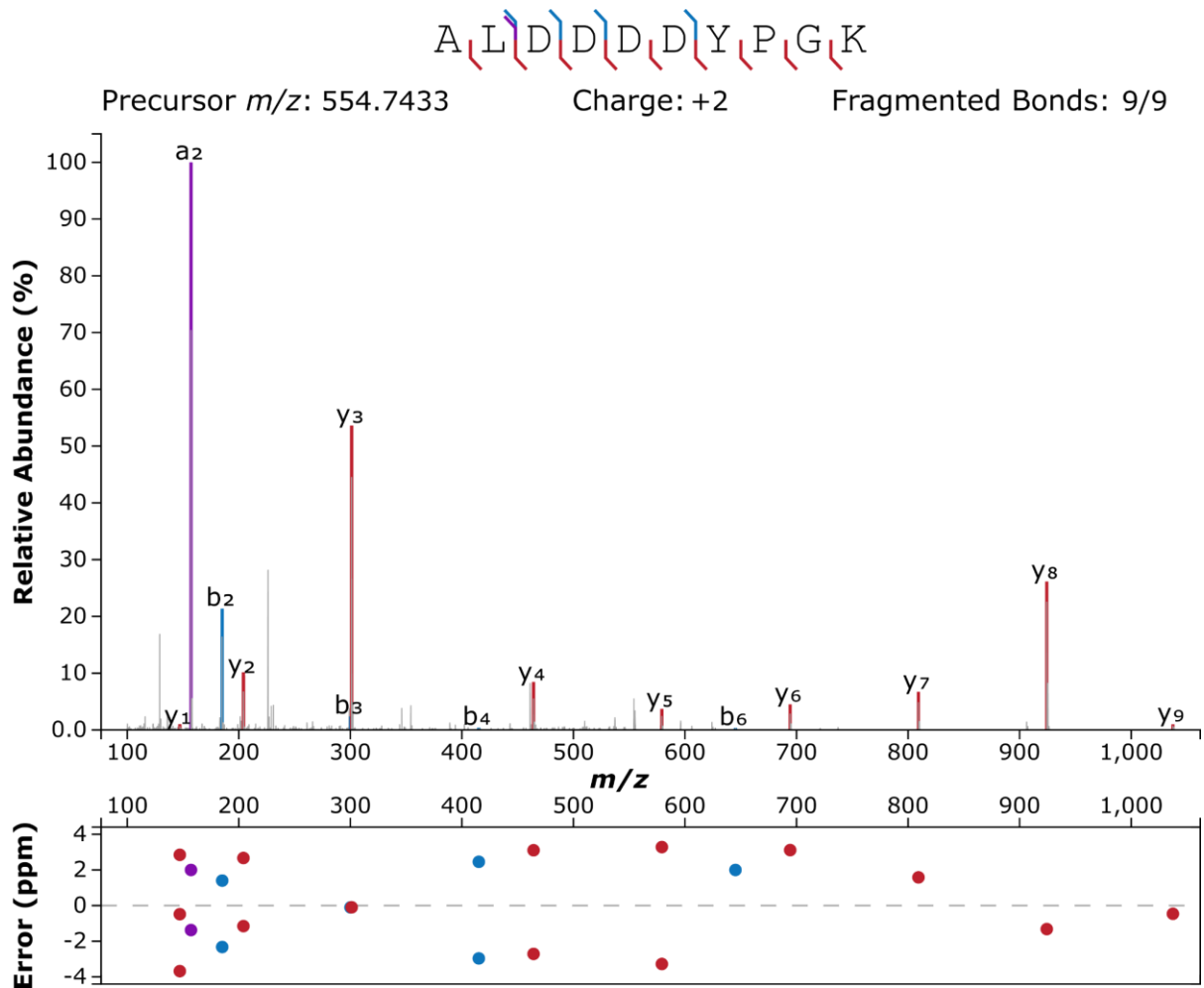


Figure 3: Annotated product ion spectrum of peptide ALDDDDYPGK. Figure created using www.interactivepeptidespectralannotator (Brademan et al., 2019).

The *de novo* peptides were searched against the Uniprot/Swissprot database (containing 565,254 manually annotated and reviewed protein sequences). A database of common laboratory contaminants (cRAP) was included in the search. The results were filtered according to the parameters detailed in the Methods section. The only protein identified was struthiocalcin-1. Figure 4 reports the coverage of SCA-1, including number of peptide-spectrum matches (PSM) and *de novo* tags (individual annotated tandem mass spectra are shown as Supplementary Figures).

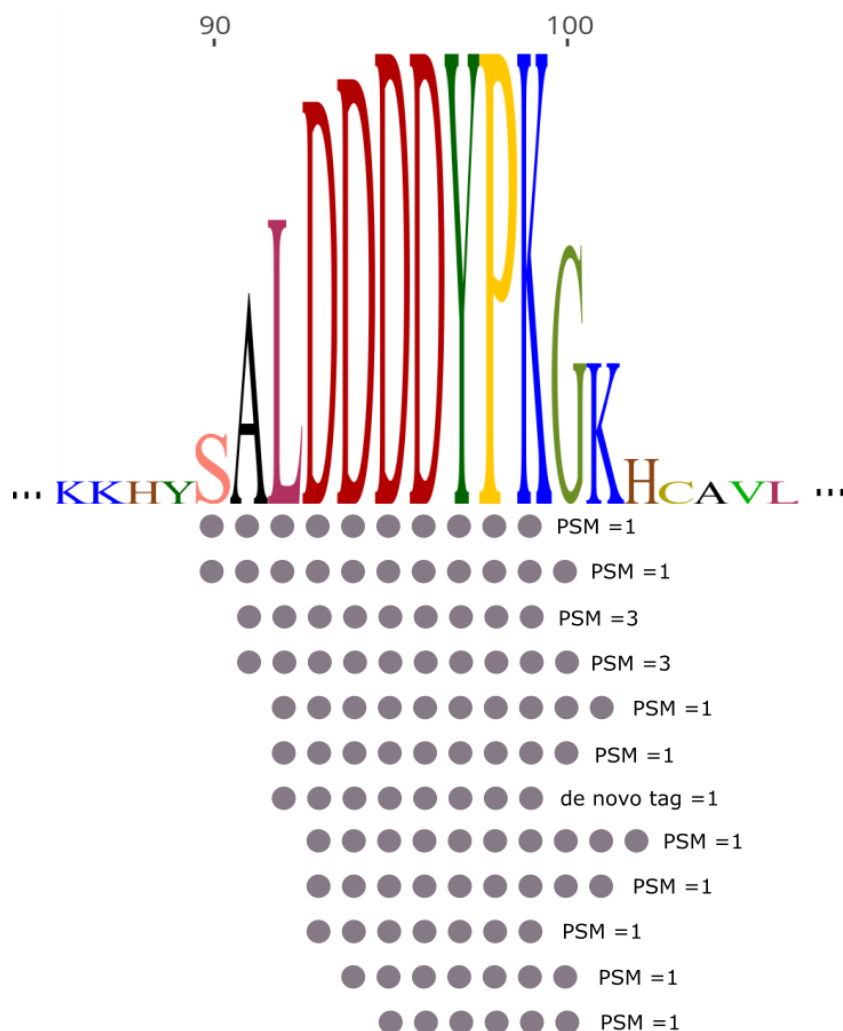


Figure 4: Coverage of struthiocalcin-1 (SCA-1) in Miocene OES specimen IVPP V26107.

Conclusions

We present the first evidence for peptide survival into the Miocene, confirming our previous (Pliocene) data and providing further support to the mechanism of preservation based on the binding of Asx-rich peptides to calcite surfaces (Demarchi et al., 2016). The sequence recovered from the Chinese specimen is identical to the peptides found in the 3.8 Ma Laetoli OES, thus also supporting the attribution of the Liushu ootaxon to genus *Struthio*. While the four Asp residues are conserved across several avian taxa, in some species, including other ratites, Asp can be substituted by Glu and some of the flanking residues are also variable (Demarchi et al., 2022, fig. S1); therefore, variability within this sequence could be informative for evolutionary relationships between extinct and extant taxa.

Materials and Methods

Sample preparation and analysis

A subsample of OES specimen IVPP V26107 was prepared in the ultra-clean facility at the University of Copenhagen, following the protocol of Demarchi et al. (2022, 2016) and omitting the digestion step. In brief, the fragment was powdered, bleached for 72 hours (NaOCl , 15% w/v) and demineralised in just enough cold weak hydrochloric acid (0.6 M HCl). The extracts were exchanged in ammonium

bicarbonate buffer (pH = 7.8) using 3kDa MWCO ultrafilters and the peptides purified and concentrated using C18 Stage Tips (Cappellini et al., 2019; Rappsilber et al., 2007).

Prepared StageTips of the procedural blank and sample were eluted with 30 μ L of 40% acetonitrile (ACN) 0.1% formic acid into a 96-well plate prior to LC-MS/MS analysis. To remove the ACN, the plate was vacuum centrifuged until approximately 5 μ L remained. Samples were then resuspended with 6 μ L of 0.1% trifluoroacetic acid (TFA) 5% ACN. Based on protein concentration results at 205 nm (NanoDrop, Thermo Scientific), 5 μ L of each sample and procedural blank was then separated over a 77 min gradient by an EASY-nLC 1200 (Proxeon, Odense, Denmark) attached to a Q-Exactive HF-X mass spectrometer (Thermo Scientific, Germany) using a 15 cm column. The column (75 μ m inner diameter) was made in-house, laser pulled, and packed with 1.9 μ m C18 beads (Dr. Maisch, Germany). Parameters were the same as those already published for historical samples (Mackie et al., 2018). In short, MS1: 120k resolution, maximum injection time (IT) 25 ms, scan target 3E6. MS2: 60k resolution, top 10 mode, maximum IT 118 ms, minimum scan target 3E3, normalized collision energy of 28, dynamic exclusion 20 s, and isolation window of 1.2 m/z . Wash-blanks consisting of 0.1% TFA 5% ACN were also run in order to hinder cross-contamination. The LC-MS/MS run included, in this order: two wash blanks, one procedural blank, one wash blank, the OES sample, and two wash blanks. Data are available via ProteomeXchange with identifier PXD035872.

Data analysis

Bioinformatic analysis was carried out using PEAKS Studio 8.5 (Bioinformatics Solutions Inc (Zhang et al., 2012)). The Uniprot_swissprot database (downloaded 11/08/2021) was used for carrying out the searches and common contaminants were included (cRAP; common Repository of Adventitious Proteins: <http://www.thegpm.org/crap/>). No enzyme was specified for the digestion and the tolerance was set to 10 ppm on the precursor and 0.5 Da on the fragments. The thresholds for peptide and protein identification were set as follows: peptide score $-10\lg P \geq 30$, protein score $-10\lg P \geq 20$, *de novo* sequences scores (ALC%) ≥ 80 .

Thermal age estimates

We assumed an age for the eggshell to be Late Miocene, 6.4–6.5 Ma, in other words at the very youngest of the estimated age range of the Lishu Formation. The formation lies in the Linxia Basin of Gansu Province, where the Mean Annual Temperature (MAT) is $\sim 11^\circ\text{C}$ (Liu et al., 2016). A Late Miocene age would place the birds in a period in which, following uplift of the Tibetan Plateau during a phase of climate cooling (Holbourn et al., 2018), there was an increase in the intensification of the East Asian monsoon associated with a northwestern expansion of the humid belt (Musser et al., 2019). Local range of lithological palaeobotanical, fauna and isotopic data discussed in Deng (2006) suggest dryer conditions and a shift from broad-leaved woodland to a more seasonal open grassland environment.

A compilation of proxies suggests a drop in MAT $\sim 6^\circ\text{C}$ over High-Mountain Asia during the Last Glacial Maximum (Yan et al., 2018), while PMIP3 predicts a decrease of $\sim 5^\circ\text{C}$ at the LGM, similar to the similar to drop in Global surface air temperature used by Holden et al. (2019) in their simulation of Pliocene–Pleistocene climate (giving an estimated MAT from Late Miocene of 14.8°C). Assuming a present day MAT of $\sim 11^\circ\text{C}$ and a temperature at LGM of 6°C , and using the temperature data from Holden et al. (2019), extending the end Pliocene to 6.4 Ma as the date of the eggshell, we can provide a very crude estimate of thermal age (i.e. age at a constant 10°C) of 6.8 Ma. This effectively means that the additional warming in the Miocene is partly offset by Pleistocene cooling. Thus whilst this sample is almost twice as old as the sample from Africa (Demarchi et al., 2016), this estimate suggests it is little over half the thermal age of eggshell from Laetoli, Tanzania. For example, fossils on present

day Ellesmere Island, Nunavut (Wang et al., 2017) using the same model would have a thermal age 50 times younger, suggesting the possibility of much greater antiquity of proteins in high latitude sites.

Acknowledgements

B.D.'s work was supported by a "Young Researchers - Rita Levi Montalcini" grant from the Italian Ministry of University and Research. M.M. and M.J.C. were supported by the Danish National Research Foundation Award PROTEIOS (DNRF128). J.C. acknowledges the Alexander von Humboldt Foundation and the Jackson School of Geosciences and Z.L. and D.T. the Chinese National Science Foundation. We thank Prof. Jesper Velgaard Olsen at the Novo Nordisk Center for Protein Research for providing access and resources, which were also funded in part by a donation from the Novo Nordisk Foundation (Grant No. NNF14CC0001).

References

- Brademan DR, Riley NM, Kwiecien NW, Coon JJ. 2019. Interactive Peptide Spectral Annotator: A Versatile Web-based Tool for Proteomic Applications. *Mol Cell Proteomics* **18**:S193–S201.
- Buffetaut E, Angst D. 2021. A Giant Ostrich from the Lower Pleistocene Nihewan Formation of North China, with a Review of the Fossil Ostriches of China. *Diversity* **13**:47.
- Cappellini E, Welker F, Pandolfi L, Ramos-Madrigal J, Samodova D, Rütther PL, Fotakis AK, Lyon D, Moreno-Mayar JV, Bukhsianidze M, Rakownikow Jersie-Christensen R, Mackie M, Ginolhac A, Ferring R, Tappen M, Palkopoulou E, Dickinson MR, Stafford TW Jr, Chan YL, Götherström A, Nathan SKSS, Heintzman PD, Kapp JD, Kirillova I, Moodley Y, Agusti J, Kahlke R-D, Kiladze G, Martínez-Navarro B, Liu S, Sandoval Velasco M, Sinding M-HS, Kelstrup CD, Allentoft ME, Orlando L, Penkman K, Shapiro B, Rook L, Dalén L, Gilbert MTP, Olsen JV, Lordkipanidze D, Willerslev E. 2019. Early Pleistocene enamel proteome from Dmanisi resolves Stephanorhinus phylogeny. *Nature* **574**:103–107.
- Demarchi B, Hall S, Roncal-Herrero T, Freeman CL, Woolley J, Crisp MK, Wilson J, Fotakis A, Fischer R, Kessler BM, Rakownikow Jersie-Christensen R, Olsen JV, Haile J, Thomas J, Marean CW, Parkington J, Presslee S, Lee-Thorp J, Ditchfield P, Hamilton JF, Ward MW, Wang CM, Shaw MD, Harrison T, Domínguez-Rodrigo M, MacPhee RDE, Kwekason A, Ecker M, Kolska Horwitz L, Chazan M, Kröger R, Thomas-Oates J, Harding JH, Cappellini E, Penkman K, Collins MJ. 2016. Protein sequences bound to mineral surfaces persist into deep time. *Elife* **5**. doi:10.7554/eLife.17092
- Demarchi B, Stiller J, Greal A, Mackie M, Deng Y, Gilbert T, Clarke J, Legendre LJ, Boano R, Slicheritz-Pontén T, Magee J, Zhang G, Bunce M, Collins MJ, Miller G. 2022. Ancient proteins resolve controversy over the identity of Genyornis eggshell. *Proc Natl Acad Sci U S A* e2109326119.
- Deng T. 2006. Chinese Neogene mammal biochronology. *Vertebrata Palasiatica* **44**:143–163.
- Deng T, Hou S, Wang S. 2019. Neogene integrative stratigraphy and timescale of China. *Science China Earth Sciences* **62**:310–323.
- Deng T, Qiu Z-X, Wang B-Y, Wang X, Hou S-K. 2013. Late Cenozoic Biostratigraphy of the Linxia Basin, Northwestern China Fossil Mammals of Asia. Columbia University Press. pp. 243–273.
- Hendy J, Welker F, Demarchi B, Speller C, Warinner C, Collins MJ. 2018. A guide to ancient protein studies. *Nat Ecol Evol* **2**:791–799.
- Holbourn AE, Kuhnt W, Clemens SC, Kochhann KGD, Jöhnck J, Lübbers J, Andersen N. 2018. Late Miocene climate cooling and intensification of southeast Asian winter monsoon. *Nat Commun* **9**:1584.
- Holden, Edwards, Rangel. 2019. PALEO-PGEM v1. 0: a statistical emulator of Pliocene–Pleistocene

- climate. *Geosci Model Dev*.
- Hou L, Zhou Z, Zhang F, Wang Z. 2005. A Miocene ostrich fossil from Gansu Province, northwest China. *Chinese Science Bulletin* **50**:1808–1810.
- Liu, Li, Song, Yu, Peng, Hui. 2016. Palynological evidence for late Miocene stepwise aridification on the northeastern Tibetan Plateau. *Clim Past*.
- Li Z, Bailleul AM, Stidham TA, Wang M, Deng T. 2021. Exceptional preservation of an extinct ostrich from the Late Miocene Linxia Basin of China. *Vertebrata Palasiatica* **59**:229–244.
- Lowe PR. 1931. Struthious Remains from Northern China and Mongolia: With Descriptions of *Struthio Wimani*, *Struthio Anderssoni* and *Struthio Mongolicus*, Spp. Nov. Geological Survey of China.
- Mackie M, R  ther P, Samodova D, Di Gianvincenzo F, Granzotto C, Lyon D, Peggie DA, Howard H, Harrison L, Jensen LJ, Olsen JV, Cappellini E. 2018. Palaeoproteomic Profiling of Conservation Layers on a 14th Century Italian Wall Painting. *Angewandte Chemie International Edition* **57**:7369–7374.
- Mikhailov KE, Zelenkov N. 2020. The late Cenozoic history of the ostriches (Aves: Struthionidae), as revealed by fossil eggshell and bone remains. *Earth-Sci Rev* **208**:103270.
- Musser G, Li Z, Clarke JA. 2019. A new species of Eogruidae (Aves: Gruiformes) from the Miocene of the Linxia Basin, Gansu, China: Evolutionary and climatic implications. *Auk* **137**:ukz067.
- Rappsilber J, Mann M, Ishihama Y. 2007. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2**:1896–1906.
- Saitta ET, Vinther J, Crisp MK, Abbott GD, Kaye TG, Pittman M, Bull I, Fletcher I, Chen X, Collins MJ, Sakalauskaite J, Mackie M, Bello FD, Dickinson MR, Stevenson MA, Donohoe P, Heck PR, Demarchi B, Penkman KEH. 2020. Non-Avian Dinosaur Eggshell Calcite Contains Ancient, Endogenous Amino Acids. *bioRxiv*. doi:10.1101/2020.06.02.129999
- Wang S. 2008. Reexamination of taxonomic assignment of *Struthio linxiaensis* Hou et al., 2005. *Acta Palaeontol Sin* **47**:362–368.
- Wang X, Rybczynski N, Harington CR, White SC, Tedford RH. 2017. A basal ursine bear (*Protarctos abstrusus*) from the Pliocene High Arctic reveals Eurasian affinities and a diet rich in fermentable sugars. *Sci Rep* **7**:17722.
- Welker F, Ramos-Madrigal J, Gutenbrunner P, Mackie M, Tiwary S, Rakownikow Jersie-Christensen R, Chiva C, Dickinson MR, Kuhlwil M, de Manuel M, Gelabert P, Martin  n-Torres M, Margvelashvili A, Arsuaga JL, Carbonell E, Marques-Bonet T, Penkman K, Sabid   E, Cox J, Olsen JV, Lordkipanidze D, Racimo F, Lalueza-Fox C, Berm  dez de Castro JM, Willerslev E, Cappellini E. 2020. The dental proteome of *Homo* antecessor. *Nature* **580**:235–238.
- Welker F, Ramos-Madrigal J, Kuhlwil M, Liao W, Gutenbrunner P, de Manuel M, Samodova D, Mackie M, Allentoft ME, Bacon A-M, Collins MJ, Cox J, Lalueza-Fox C, Olsen JV, Demeter F, Wang W, Marques-Bonet T, Cappellini E. 2019. Enamel proteome shows that *Gigantopithecus* was an early diverging pongine. *Nature* **576**:262–265.
- Yan Q, Owen LA, Wang H, Zhang Z. 2018. Climate constraints on glaciation over high-mountain Asia during the last glacial maximum. *Geophys Res Lett* **45**:9024–9033.
- Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, Zhang W, Zhang Z, Lajoie GA, Ma B. 2012. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol Cell Proteomics* **11**:M111.010587.
- Zhang W, Appel E, Jiuyi W, Xiaomin F, Jinbo Z, Yibo Y, Yunfa M, Xiaoli Y. 2019. New paleomagnetic constraints for *Platybelodon* and *Hipparion* faunas in the Linxia Basin and their ecological environmental implications. *Glob Planet Change* **176**:71–83.
- Zhu Y, Zhou L, Mo D, Kaakinen A, Zhang Z, Fortelius M. 2008. A new magnetostratigraphic framework for late Neogene *Hipparion* Red Clay in the eastern Loess Plateau of China. *Palaeogeogr Palaeoclimatol Palaeoecol* **268**:47–57.

Supplementary Images

Figure S1. Annotated product ion spectrum, $m/z = 554.7487$.

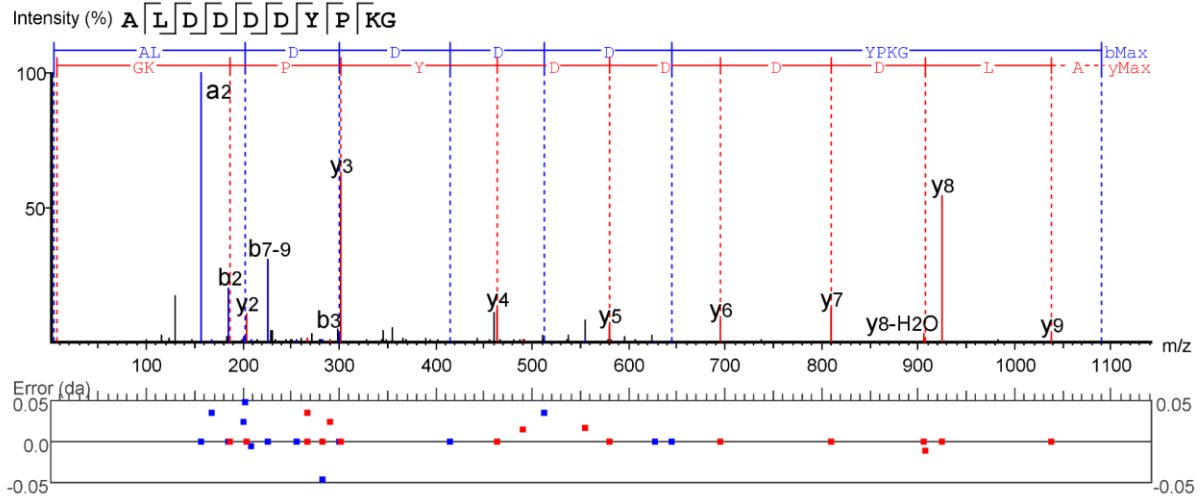


Figure S2. Annotated product ion spectrum, $m/z = 526.2355$

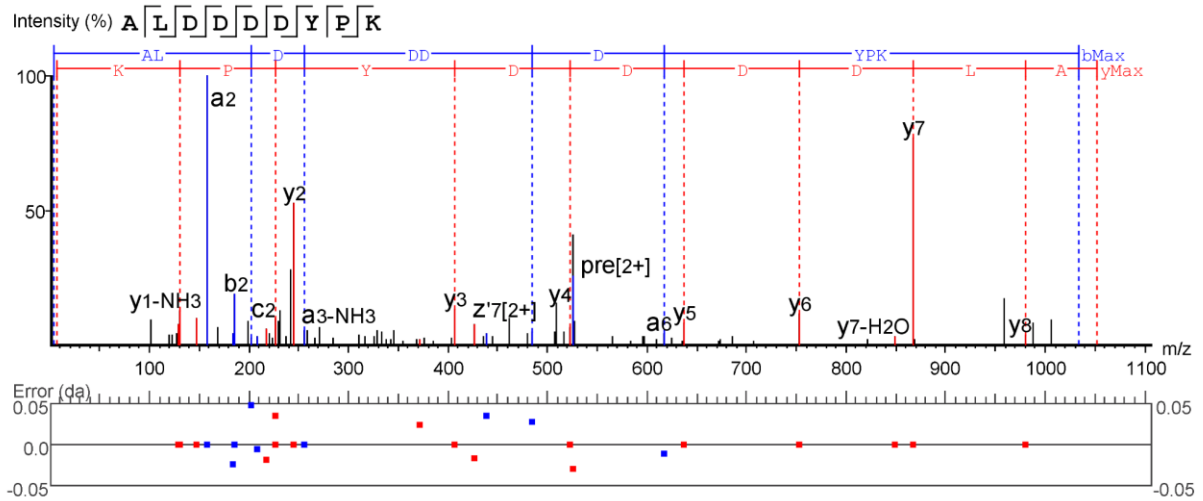


Figure S3. Annotated product ion spectrum, $m/z = 598.2663$

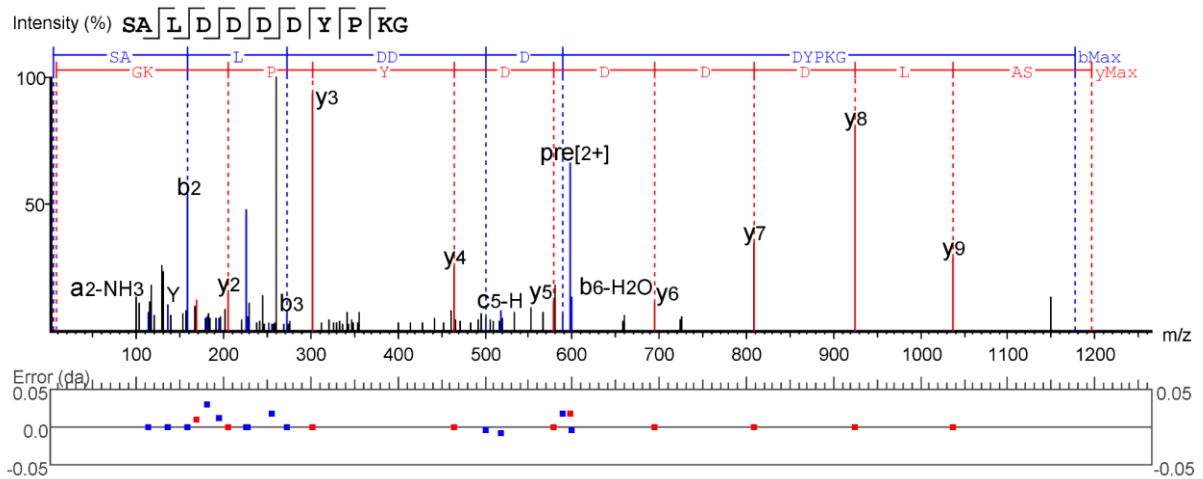


Figure S4. Annotated product ion spectrum, $m/z = 397.1782$

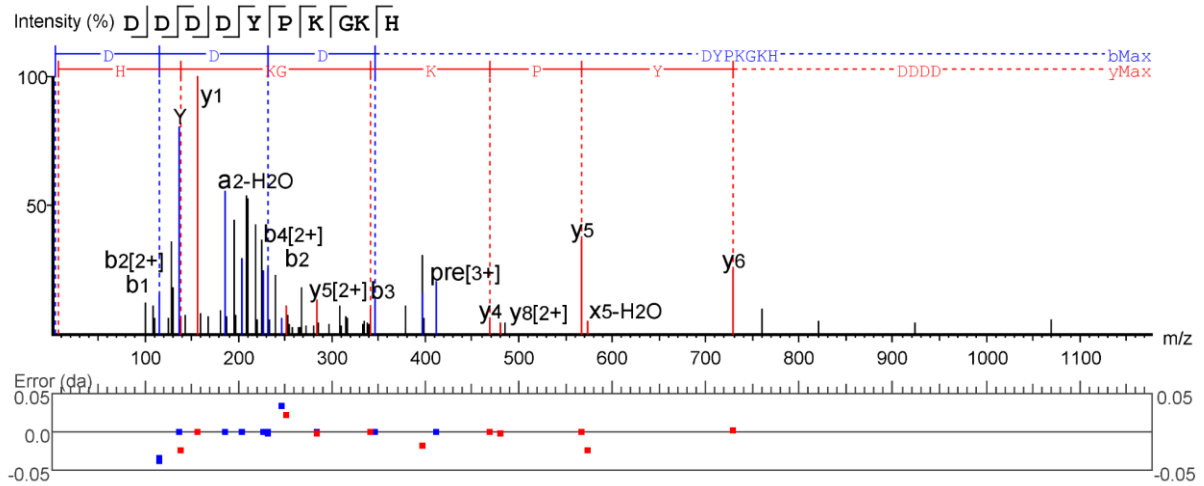


Figure S5. Annotated product ion spectrum, $m/z = 583.2767$

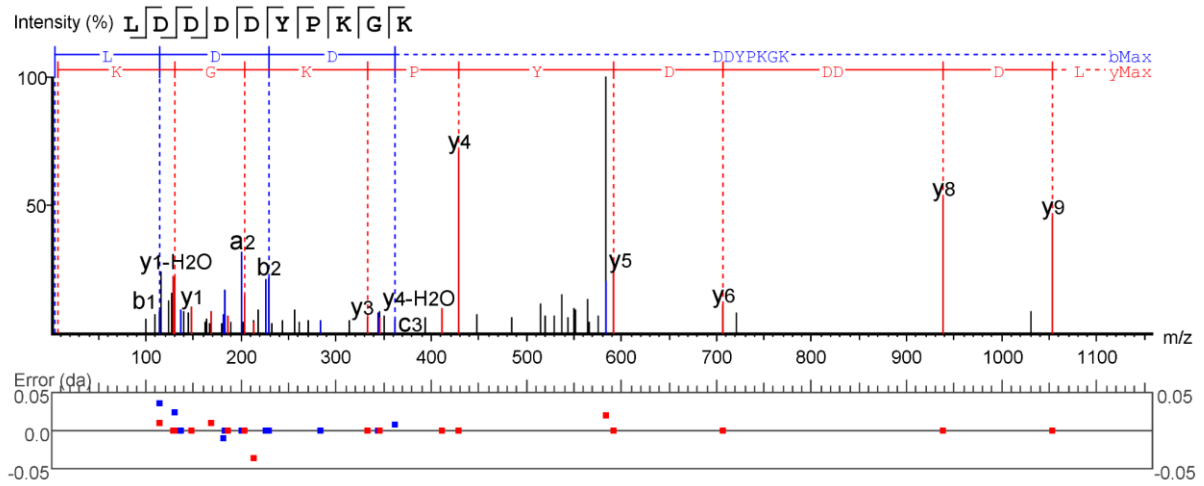


Figure S6. Annotated product ion spectrum, $m/z = 569.7542$

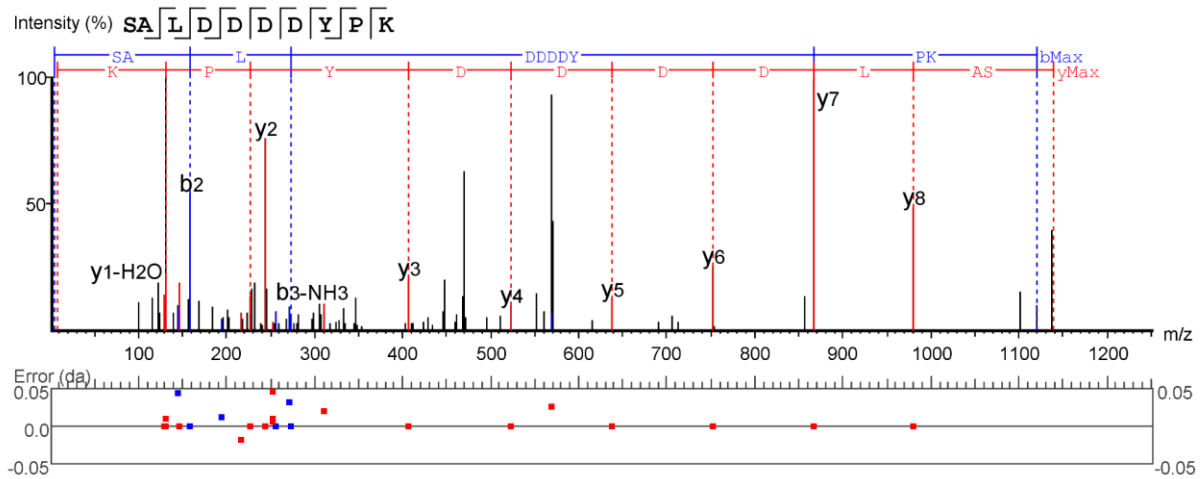


Figure S7. Annotated product ion spectrum, $m/z = 526.7343$

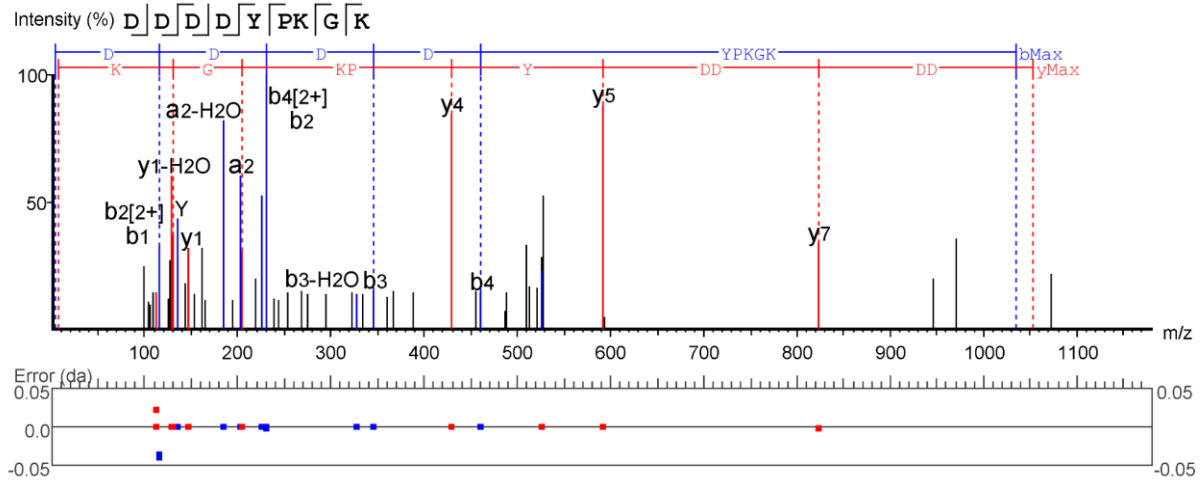


Figure S8. Annotated product ion spectrum, $m/z = 519.2303$

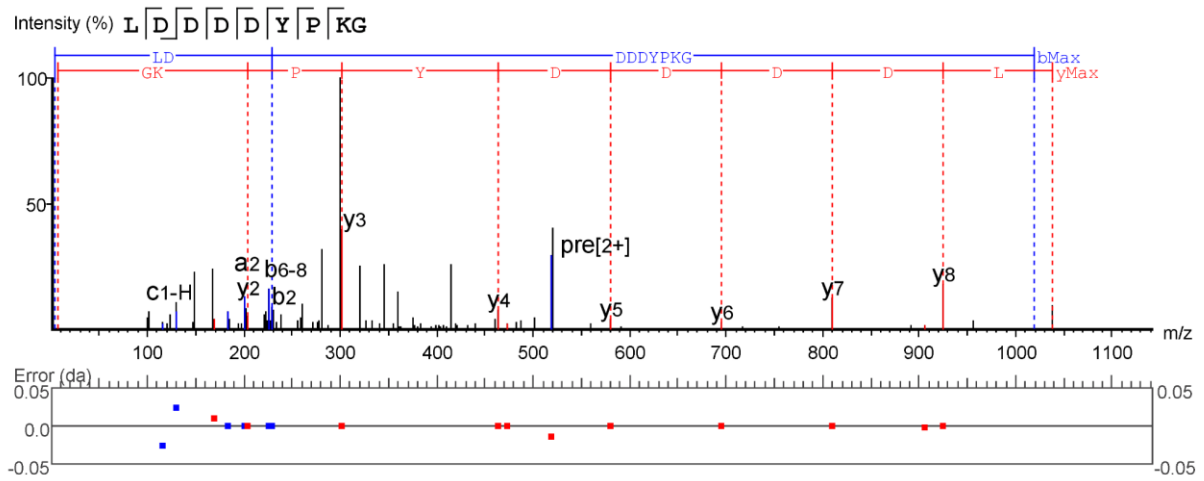


Figure S9. Annotated product ion spectrum, $m/z = 469.2212$

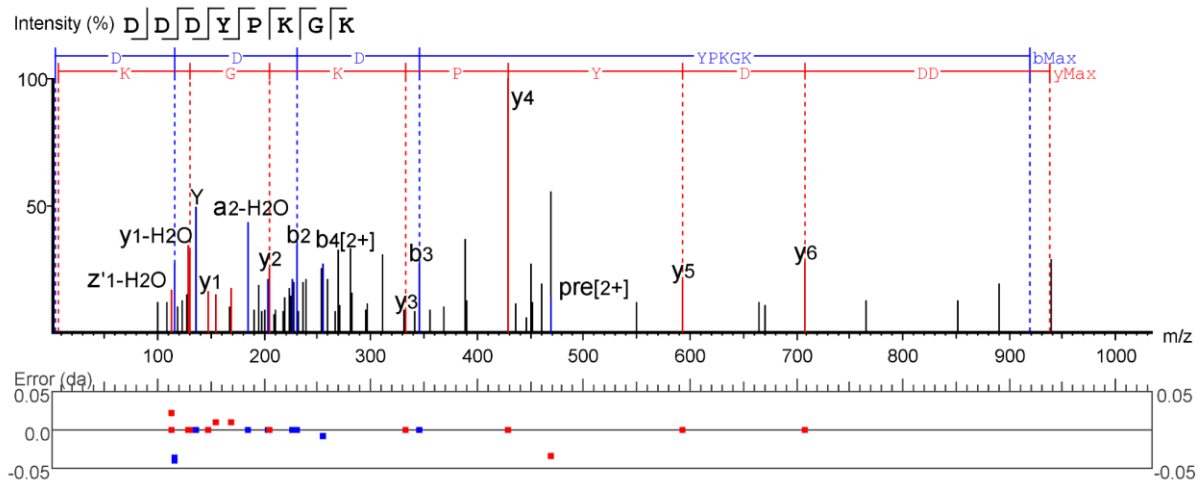


Figure S10. Annotated product ion spectrum, $m/z = 434.1741$

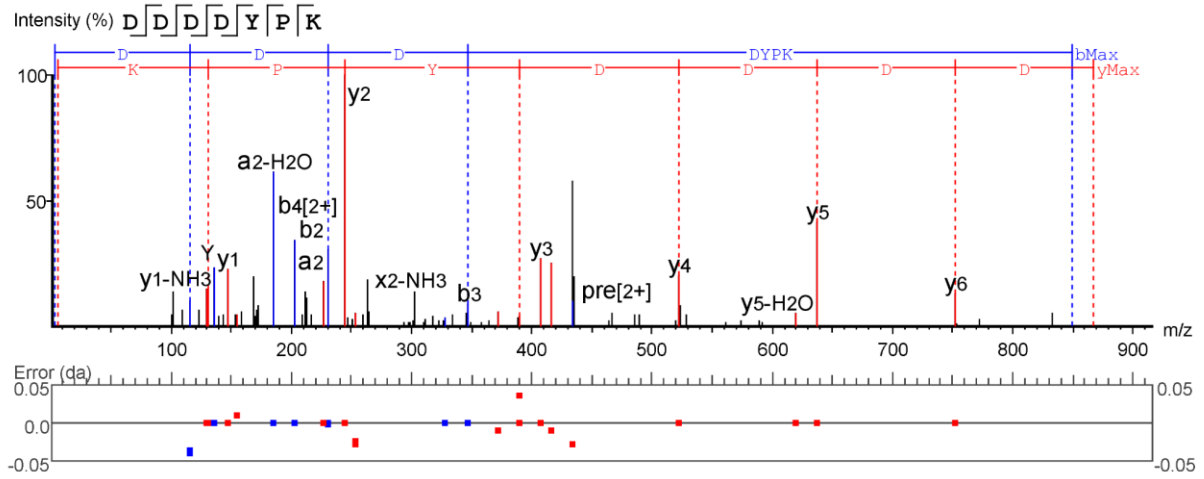


Figure S11. Annotated product ion spectrum, $m/z = 411.7059$

