



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/570/>

Article:

Nielsen-Marsh, C M, Richards, M P, Hauschka, P V et al. (2005) Osteocalcin protein sequences of Neanderthals and modern primates. *Proceedings of the National Academy of Sciences of the United States of America*. pp. 4409-4413. ISSN: 1091-6490

<https://doi.org/10.1073/pnas.0500450102>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Osteocalcin protein sequences of Neanderthals and modern primates

Christina M. Nielsen-Marsh^{a,b}, Michael P. Richards^{a,c}, Peter V. Hauschka^d, Jane E. Thomas-Oates^e, Erik Trinkaus^f, Paul B. Pettitt^g, Ivor Karavanich^h, Hendrik Poinarⁱ, and Matthew J. Collins^j

^aDepartment of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany; ^bDepartment of Archaeological Sciences, University of Bradford, Bradford, West Yorkshire BD7 1DP, United Kingdom; ^cDepartments of Orthopaedic Surgery and Oral and Developmental Biology, Harvard Schools of Medicine and Dental Medicine, Boston, MA 02115; ^dDepartment of Chemistry, University of York, York YO10 5DD, United Kingdom; ^eDepartment of Anthropology, Washington University, Campus Box 1114, St. Louis, MO 63130; ^fDepartment of Archaeology, University of Sheffield, Northgate House, West Street, Sheffield S1 4ET, United Kingdom; ^gDepartment of Archaeology, Faculty of Philosophy, University of Zagreb, Ivana Lucica 3, 10000 Zagreb, Croatia; ^hDepartment of Anthropology, Pathology, and Molecular Medicine, McMaster University 1280 Main Street West, Hamilton, ON, Canada L8S 4L9; and ⁱBioArch, Departments of Biology and Archaeology, University of York, York YO10 5YW, United Kingdom

Contributed by Erik Trinkaus, January 20, 2005

We report here protein sequences of fossil hominids, from two Neanderthals dating to $\approx 75,000$ years old from Shanidar Cave in Iraq. These sequences, the oldest reported fossil primate protein sequences, are of bone osteocalcin, which was extracted and sequenced by using MALDI-TOF/TOF mass spectrometry. Through a combination of direct sequencing and peptide mass mapping, we determined that Neanderthals have an osteocalcin amino acid sequence that is identical to that of modern humans. We also report complete osteocalcin sequences for chimpanzee (*Pan troglodytes*) and gorilla (*Gorilla gorilla gorilla*) and a partial sequence for orangutan (*Pongo pygmaeus*), all of which are previously unreported. We found that the osteocalcin sequences of Neanderthals, modern human, chimpanzee, and orangutan are unusual among mammals in that the ninth amino acid is proline (Pro-9), whereas most species have hydroxyproline (Hyp-9). Posttranslational hydroxylation of Pro-9 in osteocalcin by prolyl-4-hydroxylase requires adequate concentrations of vitamin C (L-ascorbic acid), molecular O₂, Fe²⁺, and 2-oxoglutarate, and also depends on enzyme recognition of the target proline substrate consensus sequence Leu-Gly-Ala-Pro-9-Ala-Pro-Tyr occurring in most mammals. In five species with Pro-9-Val-10, hydroxylation is blocked, whereas in gorilla there is a mixture of Pro-9 and Hyp-9. We suggest that the absence of hydroxylation of Pro-9 in *Pan*, *Pongo*, and *Homo* may reflect response to a selective pressure related to a decline in vitamin C in the diet during omnivorous dietary adaptation, either independently or through the common ancestor of these species.

MALDI-TOF/TOF | dietary adaptation | biomolecular preservation | vitamin C | evolution

Biomolecules rarely survive intact in fossil bone and teeth, yet they potentially provide key information on phylogenetic relationships and adaptations. DNA survival in fossil bones and teeth is especially rare in older samples, or in those from hotter climates (1). However, proteins, and therefore phylogenetically informative protein sequences, do persist in much older samples (2, 3). Of the various bone proteins, osteocalcin has been shown to survive particularly well in fossil bone, as demonstrated recently when a complete sequence of fossil *Bison priscus* osteocalcin was determined with protein extracted from exceptionally well preserved permafrost bones (4).

The protein itself is a small (M_r of 5.9 kDa), negatively charged (pI of 4.0) molecule that is the second most abundant protein in bone. The role of osteocalcin in bone metabolism is uncertain, but it has α -helical domains forming a tightly packed charged molecule that coordinates Ca²⁺ at the surface of the hydroxyapatite-like lattice of bone mineral crystals (5, 6). In all vertebrate species so far investigated, the osteocalcin protein sequence is highly conserved within the central portion of the

molecule, containing the three γ -carboxyglutamic acid (Gla) residues and Ca²⁺ binding sites (Fig. 1) (6–8), whereas the N terminus displays considerable variation and is the most genetically informative region of the molecule.

Using a modified version of the methods described in ref. 4, we have attempted to extract osteocalcin from three Neanderthals (specimens 2, 4, and 6) from Shanidar Cave, Iraq, dating to $\approx 75,000$ years old (12), and one Neanderthal [Vi-76/228 (Vi13.1), layer G3] from Vindija Cave, Croatia, dating to $\approx 40,000$ years old (13). These data allowed comparison with both published modern human osteocalcin sequences and those we produced from archaeological, modern humans, and other primates by using both MALDI-TOF/TOF and Edman sequencing.

Methods

Using the method modified from ref. 4, osteocalcin was extracted and purified from bone powder by using Amprep C18 MiniColumns after demineralization (0.5 M sodium EDTA, pH 8.0, 4 h, 25°C). An Applied Biosystems 4700 Proteomics Analyzer was used for MALDI-TOF/TOF analyses. Because the fossil samples are rare and precious specimens, we used minimal sample sizes (≈ 55 mg of bone); therefore, the results reported here for the Neanderthals are for one extraction per sample. Chimpanzee (*Pan troglodytes*) and gorilla (*Gorilla gorilla gorilla*) samples were obtained by H.P. from the Swedish Natural History Museum. *De novo* sequencing of the samples from the modern species was performed via a combination of peptide mass mapping and collision-induced dissociation (CID) product ion MS of selected peptides after tryptic digestion. For Edman sequencing, chimpanzee (*Pan troglodytes*) and orangutan (*Pongo pygmaeus*) specimens were obtained by permit from Yerkes Research Center (Emory University, Atlanta). Osteocalcin was extracted and purified from bone powder as described in ref. 7. Tryptic digestion, HPLC purification of peptides, amino acid analysis, and Edman sequencing followed ref. 7, except that 50- to 500-pmol peptide samples were analyzed on an ABI-477A pulsed-liquid phase protein sequencer (Applied Biosystems).

Results

Unlike MS data from the permafrost samples (4), none of the mass spectra of the fossil extracts contained the protonated

Abbreviations: CID, collision-induced dissociation; HIF-1 α , hypoxia-inducible factor 1 α ; P4H, prolyl-4-hydroxylase;

Data deposition: The sequences reported in this paper have been deposited in the Swiss-Prot and TrEMBL database [accession nos. P84348 (*Pan troglodytes*), P84349 (*Gorilla gorilla gorilla*), P84350 (*Pongo pygmaeus*), and P84351 (Neanderthal)].

^bTo whom correspondence should be addressed. E-mail: nielsen@eva.mpg.de.

© 2005 by The National Academy of Sciences of the USA

Osteocalcin Primary Structure

	1	10	20	30	40	49
		*	γ	γ		
Human	YLYQWLGAPVYPDPLEPRREVCELNPDCEDELADHIGFQEAYRRFYGPV					
Neanderthal	YLYQWLGAPVYPDPLEPRREVCELNPDCEDELADHIGFQEAYR...					
Chimpanzee	YLYQWLGAPVYPDPTLEPRREVCELNPDCEDELADHIGFQEAYRRFYGPV					
Orangutan	YLYQWLGAPVYPDPLEPKREVCELNPDCEDELA...					
Gorilla	YLYQWLGAVOVPYPDPLEPRREVCELNPDCEDELADHIGFQEAYRRFYGPV					
Monkey	YLYQWLGAVOVPYPDPLEPKREVCELNPDCEDELADHIGFQEAYRRFYGPV					
Cow	YLDHWLGAVOVPYPDPLEPKREVCELNPDCEDELADHIGFQEAYRRFYGPV					

* Residue-9 O=Hypro, P=Pro, Gorilla position 9 is O/P

Chimpanzee Edman Sequences:

Intact Osteocalcin	YLYQWLGAPVYPDPTLEPRREVCELN...
OC989	
Tryptic Peptides	
16A-T9	YLYQWLGAPVYPDPTLEPRR...
16A-T6	REVCELNPDCEDELADHIGFQEAYRR
16A-T1	
16A-T2	
16A-T7	REVCELNPDCEDELADHIGFQ...
Chimpanzee	YLYQWLGAPVYPDPTLEPRREVCELNPDCEDELADHIGFQEAYRRFYGPV

Orangutan Edman Sequence:

Intact Osteocalcin	YLYQWLGAPVYPDPLEPKREVCELNPDCEDELA...
--------------------	--------------------------------------

Fig. 1. Osteocalcin amino acid sequences for modern human (9, 10) and Neanderthal (Shanidar 2 Neanderthal). Chimpanzee (*Pan troglodytes*, Tai Forest, Ivory Coast), gorilla (*G. gorilla gorilla*, East Africa), and orangutan (*Pongo pygmaeus*) sequences are also presented, as are selected data for monkey (*Macaca fascicularis*) (7) and cow (*Bos taurus*) (11). γ -Carboxyglutamate residues at positions 17, 21, and 24 are indicated by " γ "; hydroxyproline at position 9 is indicated by "O." Neanderthal residues shaded in gray were unable to be confirmed on CID MS; however, results of peptide mass mapping of tryptic digests are consistent with the first 19 residues being identical to those in the human protein. The new protein sequence data from the primates and Neanderthal were deposited in the Swiss-Prot and TrEMBL database under accession numbers P84348 (*Pan troglodytes*), P84349 (*G. gorilla gorilla*), P84350 (*Pongo pygmaeus*), and P84351 (Neanderthal).

molecule for osteocalcin. However, after tryptic digestion, masses corresponding to expected tryptic peptides were observed (Fig. 2). Of the Neanderthal samples analyzed, only Shanidar 2 and 6 yielded measurable osteocalcin (Fig. 2).

Because of the exceptional limits of detection afforded by the MALDI-TOF/TOF mass spectrometric instrumentation, ion intensities from a tryptic peptide from the Shanidar 2 protein extract were sufficiently high for sequencing after high-energy

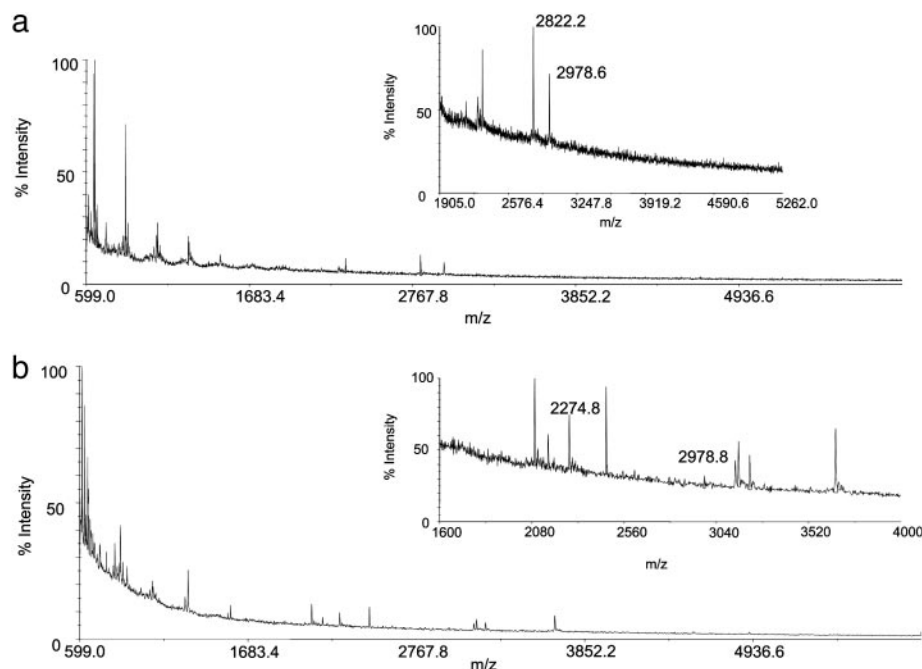


Fig. 2. MALDI mass spectra of tryptic digests of bone protein extracts from Neanderthals Shanidar 2 (a) and Shanidar 6 (b). Insets show the expanded regions of the spectra. Only the ion with m/z 2,822.2 from the Shanidar 2 sample was intense enough to perform CID MS, but peaks observed in a MALDI spectrum of a modern human bone protein digest corresponding to osteocalcin tryptic fragments were observed in both Shanidar 2 (m/z 2,822.2, 20–43; m/z 2,978.6, 20–44) and Shanidar 6 (m/z 2,274.8, 1–19; m/z 2,978.8, 20–44) sample spectra.

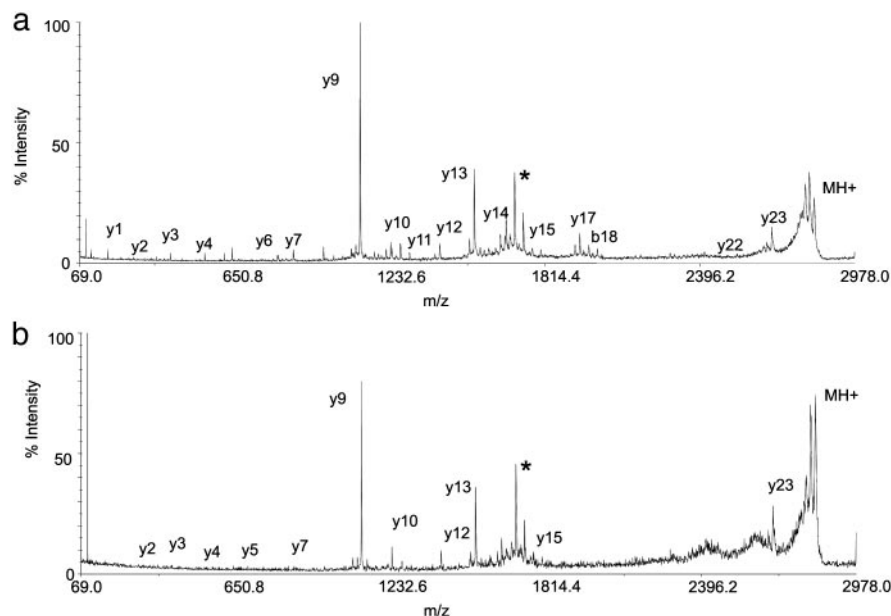


Fig. 3. CID product ion spectra of peptide Arg-20–Arg-43 from modern human (a) and Shanidar 2 (b) Neanderthal. Precursor ion, MH⁺, is represented by the peak at the highest *m/z* (2,822). Peptide sequence ions produced on cleavage of peptide bonds are labeled by using Roepstorff nomenclature (4). The spectrum from the Neanderthal sample contains a less complete *y* ion series than that of the modern human, probably as a result of a much smaller quantity of protein surviving in the fossil. The *y* ion series given by both samples are consistent with the sequence Arg-Glu-Val-Cys-Glu-Leu-Asn-Pro-Asp-Cys-Asp-Glu-Leu-Ala-Asp-His-Ile-Gly-Phe-Gln-Glu-Ala-Tyr-Arg. The peak marked with * is unassigned.

collision of the ion at *m/z* 2,822 (Fig. 3). A characteristic pattern of structurally diagnostic fragment ions was observed in product ion spectra from Arg-20–Arg-43 from human (Fig. 3), chimpanzee, and gorilla extracts (data not shown). Submission of the fragment ion data from the Shanidar 2 peptide to a MASCOT MS/MS ion search of the NCBI database yielded a match to peptide Arg-20–Arg-43 of human (and other) osteocalcin with an “ions score” of 83 and a probability of $<1e^{-0.5}$ that this is a random match. For human, chimpanzee, and gorilla samples, CID spectra were obtained from Tyr-1–Arg-19 (human data not shown) and Arg-20–Arg-43 (chimpanzee and gorilla data not shown) peptides to identify the sequences (Figs. 1 and 3–6).

Discussion

The absence of the protonated molecule, and the success of MS after tryptic digestion, suggests that ancient osteocalcin may be partially preserved by condensation into a higher-molecular-weight matrix. The trypsin cleavage product (Arg-20–Arg-43), which spans most of the mineral binding domain (5, 6), was able to be sequenced after CID MS (Figs. 1 and 3). The persistence of the hydrophobic core and calcium binding surface provides direct evidence that adsorption to mineral enhances long-term preservation of ancient proteins. In addition to providing phylogenetic information, MALDI-MS also provides clues as to how the osteocalcin molecule degrades. Osteocalcin peptides extracted from samples from both temperate and permafrost burial environments demonstrate that in fossils with little or no collagen remaining, the first 14 residues of osteocalcin are sometimes missing (C.M.N.-M., unpublished data). These first 14 residues are proline-rich; this finding supports the hypothesis that this region could play a role in binding osteocalcin to collagen (14). Previous attempts to extract collagen from both Shanidar and Vindija Neanderthals proved unsuccessful (M.P.R., unpublished data), and although a peptide with an *m/z* corresponding to the protonated Tyr-1–Arg-19 fragment observed in spectra of the human sample was detected in the Shanidar 6 sample (but not in Shanidar 2), this was too weak for successful CID MS (Fig. 2).

This 1–19 peptide is where most variations between mammal species are observed (8) (Fig. 1), and whereas modern humans and (based on the peptide mass determinations) Neanderthals have identical sequences, gorilla was found to differ by only a single posttranslationally modified amino acid, with a mixture of Hyp and Pro at the ninth position (Figs. 1 and 4).

Chimpanzee shares Pro-9 with orangutan and *Homo* but has Thr at position 15 (Figs. 1 and 5). Substitution of Thr-15 for Pro-15 is readily explained by a single-base change of C → A at the first position on the 5′ side of the codon for Pro-15 in the chimpanzee osteocalcin gene. This mutation must have occurred in the chimpanzee branch after divergence from the ancestral hominid lineage, because the Pro-15 has been conserved in *Homo* as well as in primates with earlier divergence: gorilla, orangutan, and Old World monkey, *M. fascicularis* (Fig. 6). At position 19, Lys-19 occurs in Old World monkey and orangutan but has mutated to Arg-19 in chimpanzee, gorilla, and *Homo*. The simplest explanation is a single A → G base change in the second position of the Lys codon, yielding the Arg codon, which must have occurred after divergence of orangutan. At position 10, Ala-10 occurs in Old World monkey but changed to Val-10 in orangutan, gorilla, chimpanzee, and *Homo*. Again, a single-base change in the osteocalcin gene of C → T in the second position of the Ala-10 codon would yield the Val codon (Fig. 6).

The lack of hydroxylation of Pro-9 in chimpanzee, orangutan, and human/*Homo* osteocalcin is intriguing (*i*) because it is a posttranslational variation distinct from the evolution of the osteocalcin gene sequence, and (*ii*) because of the distinctly different properties of the two classes of prolyl-4-hydroxylases (P4Hs) that may target osteocalcin. P4Hs (EC 1.14.11.2) are discriminated by their protein substrate preferences, *K_m* and kinetic properties, and subcellular localization. P4Hs all require molecular O₂, L-ascorbic acid, 2-oxoglutarate, and Fe²⁺ to catalyze the hydroxylation of Pro in peptides (15, 16). Collagen P4Hs (Col-P4Hs) act in the endoplasmic reticulum to modify nascent procollagen chains. The tetrameric enzyme comprises two α subunits (P4HA1, P4HA2, or P4HA3) and two β subunits

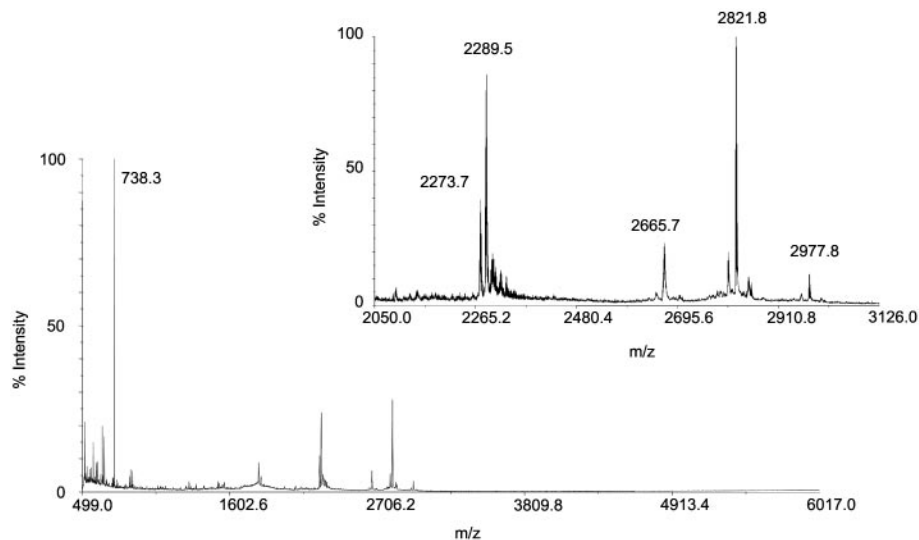


Fig. 4. MALDI mass spectrum of tryptic digest from gorilla (*G. gorilla gorilla*) bone protein extract. The *Inset* shows the expanded region of the spectrum. Tryptic fragments were observed suggesting a mixture of Hyp (m/z 2,289.5) and Pro (m/z 2,273.7) at the ninth position in the 1–19 peptide. Other tryptic peptides observed support the assignment of the proposed sequence (Fig. 1) (m/z 738.3, 44–49; m/z 2,665.7, 21–43; m/z ; 2,821.8, 20–43; m/z 2,977.8, 20–44).

(P4HB/protein disulfide isomerase). Hydroxyproline accounts for $\approx 10\%$ of the total amino acid composition of type I collagen and is essential for stability of the collagen triple helix. HIF-1 α (hypoxia-inducible factor 1 α) P4Hs (HIF-P4Hs) are cytoplasmic and nuclear, targeting Pro-524 and Pro-402 in HIF-1 α (17–19). Hydroxyproline at one or both of these sites facilitates binding to the von Hippel–Lindau (VHL) protein that then initiates the ubiquitination and proteasome-mediated degradation of HIF-1 α (20). HIF-P4Hs serve as critical sensors for the concentration of dissolved atmospheric molecular O_2 because of their poor affinity for O_2 ($K_m = 230\text{--}250 \mu\text{M}$); in contrast, Col-P4H1 binds O_2 strongly ($K_m = 40 \mu\text{M}$) (16). Under hypoxic conditions, HIF-P4Hs are inactive and thus the nonhydroxylated form of HIF-1 α accumulates in cells and activates the transcription of genes essential for cell and tissue responses to hypoxia.

The peptide substrate specificities for the HIF-P4Hs (Leu-X-X-Leu-Ala-Pro*-Ala/Tyr) (16, 19) suggest that they may hydroxylate Pro-9 in osteocalcin (Leu-Gly-Ala-Pro*-Ala/Val) (Fig. 1) more efficiently than the Col-P4Hs, which target the Gly-X-Pro*-Gly sequence (15). Both P4H classes are expressed in osteoblasts where osteocalcin and collagen type I are synthesized, but the subcellular localization of osteocalcin hydroxylation has not been established. Interestingly, 14 of 18 vertebrate osteocalcins with a conserved Pro-9 are hydroxylated to Hyp-9, and 13 of these 14 have the sequence Hyp-9–Ala-10 (refs. 4, 7, and 8; P.V.H., unpublished data), whereas gorilla is the lone exception with Hyp-9–Val-10. In all four species where Pro-9 is not hydroxylated, the Pro-9–Val-10 sequence occurs. Thus, hydroxylation of Pro-9 is strongly favored by Ala-10 and resisted by Val-10, pointing to properties of the responsible P4H enzyme.

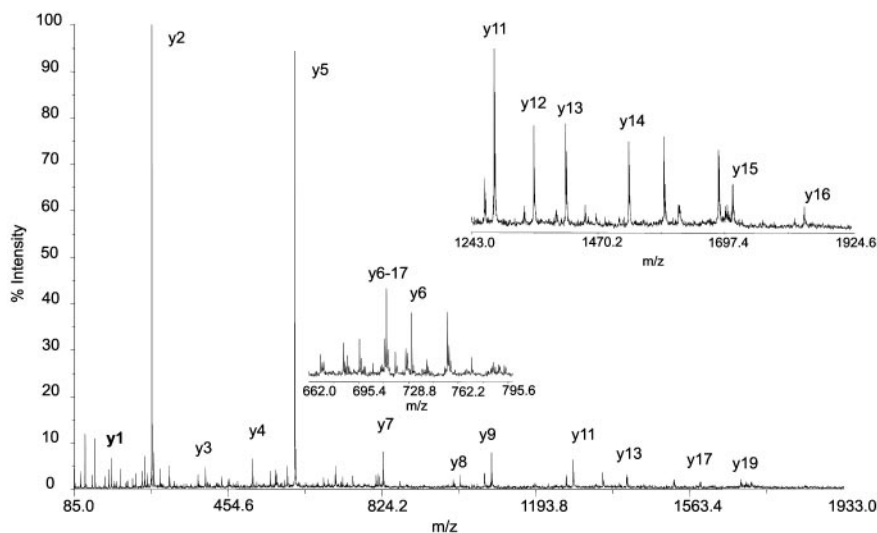


Fig. 5. CID product ion spectrum of peptide Tyr-1–Arg-19 from chimpanzee (*Pan troglodytes*). Precursor ion, MH^+ , is represented by the peak at the highest m/z (2,278.0). Peptide sequence ions produced on cleavage of peptide bonds are labeled by using Roepstorff nomenclature (4); only y ions are labeled. *Insets* show the expanded region of the spectrum. The sequence given by the y ions series is Tyr-Leu-Tyr-Gln-Trp-Leu-Gly-Ala-Pro-Val-Pro-Tyr-Pro-Asp-Thr-Leu-Glu-Pro-Arg.

