



Quantifying collagen quality in archaeological bone: Improving data accuracy with benchtop and handheld Raman spectrometers

Odile Madden*, Dora Man Wai Chan, Morgan Dundon, Christine A.M. France

Smithsonian Museum Conservation Institute, 4210 Silver Hill Road, Suitland, MD 20746, USA



ARTICLE INFO

Keywords:

Raman spectroscopy
Handheld
Bone
Collagen
Protein
Hydroxyapatite
Diagenesis

ABSTRACT

Chemical analysis of collagen and other endogenous protein in excavated bones is ever more common in paleontology and archaeology to determine dietary ecology, migration patterns, age, and diagenetic pathways. Decisions to sacrifice valuable samples to destructive stable isotope, radiogenic isotope, and proteomic analyses are easier to make if one knows ahead of time whether sufficient undegraded organic material is preserved in the bone for chemical analysis. Recent advances with near-infrared Raman spectroscopy to non-destructively pre-screen bones for the presence of undegraded endogenous protein can indicate whether original isotopic or proteomic signatures have remained intact in long-buried bones. However, it is crucial to examine the different emerging approaches with benchtop and handheld Raman spectrometers and to establish uniform data reporting standards that will allow laboratories to accurately compare results. Three recently published studies are compared to understand the factors that affect protein screening success and to establish best practices for the technique in the laboratory and in the field. A set of 37 archaeological human bone samples analyzed previously by the authors using FT-Raman spectroscopy are re-analyzed with a handheld 1064 nm Raman spectrometer such that the laser power was similar but the spectral resolution and signal-to-noise were lower. Two methods to identify peak heights were evaluated: 1) peak height fixed at a specific wavenumber and baseline anchored at fixed points bracketing the band and 2) averaging peak intensities over a short wavenumber range while also setting the baseline of the band by an average of several anchor points. This second method appeared to better classify whether the bones contained well-preserved protein signatures. In particular, peak height ratios of 960 cm^{-1} to 1636 cm^{-1} or 1450 cm^{-1} can indicate quality and abundance of bone protein, but have benefits and trade-offs that depend on the instrument used. The 960 cm^{-1} : 1636 cm^{-1} ratio is more descriptive of collagen quality as defined by an extracted carbon:nitrogen ratio of 2.8–3.6, but the 1636 cm^{-1} peak can be difficult to resolve well with an inherently less sensitive handheld Raman spectrometer. The 1450 cm^{-1} peak is more prominent in the spectrum and therefore more easily resolved with a handheld spectrometer, but it describes a common C–H stretch found in collagen and other organic molecules.

1. Introduction

Chemical analysis of excavated bones is increasingly popular in paleontology and archaeology. Bones frequently are analyzed using stable isotopes, radiogenic isotopes and proteomic techniques to determine dietary ecology, migration patterns, age, and diagenetic pathways. Decisions to sacrifice valuable samples to these destructive techniques are easier to make if one knows ahead of time whether sufficient undegraded, endogenous protein is preserved in the bone for chemical analysis. Recent advances have been made with near-infrared Raman spectroscopy to non-destructively pre-screen bones for good collagen preservation (France et al., 2014a; Halcrow et al., 2014; Pestle et al., 2014, 2015), which is a strong indicator that original isotopic or

proteomic signatures have remained intact in long-buried bones. However, these researchers have focused on different criteria for collagen quality and analytical parameters. As this technique continues to develop it is important to recognize these different approaches, examine their attributes, and establish uniform data reporting standards that will allow laboratories to compare results. This is a crucial step toward becoming a mature science with generally agreed upon standards that yield accurate and precise protein data (Killick, 2015). Furthermore, the advent of relatively inexpensive, easy to operate, handheld Raman spectrometers sets up new challenges for assessing data quality. The spectral sensitivity and resolution of these handheld instruments is impressive but nonetheless lower than research-grade benchtop spectrometers, and these instruments are likely to be

* Corresponding author.

E-mail address: MClweb@si.edu (O. Madden).

deployed in the field by archaeologists with varying expertise in spectroscopy. Here, three recently published studies are examined, and the set of bone samples evaluated by France et al. in 2014 (2014a) with a high-resolution, benchtop FT-Raman spectrometer is re-analyzed with a handheld 1064 nm Raman spectrometer for comparison.

1.1. Raman spectroscopy of bone

Raman spectroscopy describes the energy distribution of light scattered inelastically by a bone. The organic and inorganic components of bone are made up of atomic bonds that vibrate in characteristic movements. The bonds might stretch or bend, for example, and each movement requires a small, very specific amount of energy. In Raman spectroscopy a laser is directed at the bone. Most of the light reflects off the bone with the same energy (i.e., color) as the laser light, and no energy is exchanged. However, a small amount of the laser light does interact with the bonds in hydroxyapatite, other inorganic salts, type I collagen, and other proteins in the bone. These bonds siphon off small packets of energy to make their characteristic movements, and the photons scattered back to the instrument's detector have slightly shifted wavelengths. A Raman spectrum is a plot of those shifted energies. Each peak describes a type of movement, called a *vibrational mode*, of a specific atomic bond or group of bonds. Together the peaks in a Raman spectrum are like a fingerprint of the material. The presence, position, relative size, or absence of certain peaks can indicate whether a molecule such as collagen is present, intact, or altered such that it has become unsuitable for further chemical analysis. A typical bone spectrum includes many peaks, the most prominent of which are indicated with their associated bonds in France et al. (2014a) and are reproduced here (Fig. 1).

A key factor in Raman spectroscopy of bone is the choice of laser. A laser is monochromatic, meaning it emits a single color of light, typically described by its wavelength in nanometers (nm). Visible light lasers (i.e., 532 nm green and 785 nm red) can trigger fluorescence signals that distort spectral baselines and obscure the relatively weak Raman peaks. Instead near infrared lasers are chosen for bone analysis to avoid interference from fluorescence. For this reason, each of the recently published studies uses a 1064 nm laser. Other important instrument variables are detector sensitivity, spectral resolution, and instrument portability, and these are interrelated. The most sensitive spectrometers with the best ability to discern different vibrations (i.e., adjustable resolution up to 1 cm^{-1}) currently are designed for laboratory research and are not easily portable. This type of instrument is described in this article as high-resolution, research-grade, or benchtop. In addition, the specific instruments used in France et al. (2014a) and Halcrow et al. (2014) are Fourier-transform Raman (FT-Raman) spectrometers; the name describes the mechanism by which the instruments create a spectrum. Increasingly popular portable and handheld

instruments are lightweight and less expensive, but they have less sensitive detection, and the resolution is lower (i.e., $8\text{--}15\text{ cm}^{-1}$) and usually not adjustable. This affects the ability to collect Raman spectra without damaging a bone and the quality of those spectra. While these smaller instruments can be used for some research purposes, their ability to discern subtle differences between materials is more limited.

1.2. Bone collagen

Bone is a composite material composed of inorganic mineral grains in *collagen*, which is the main structural protein in bone and also the traditional catch all term for other organic compounds that are also present in lesser proportion. As the science of proteins develops ever more quickly, the ability to distinguish the minor organic components is improving and outpacing the language traditionally used to describe the bulk organic matrix. Thus the term *collagen* is still the convention, but it is not entirely accurate. In this article the authors have chosen to use the term *collagen* in the conventional way, but type I collagen is specified where appropriate and other endogenous proteins and organic compounds are as well.

In chemical studies of excavated bone, the presence of undegraded *collagen* has long been taken to indicate the overall preservation of original isotopic, radiogenic, and other chemical signatures. If the proteinaceous component of the bone is present, intact, and uncontaminated, it can be used to establish dates for skeletal remains or as evidence of an individual's environment or diet during life.

Nearly all the protein in a bone is type I collagen, which makes up approximately 20% of bone weight and is composed of strong and complex triple helical molecules organized into arrays of fibrils that protect the hydroxyapatite mineral grains from exposure to ground water and other diagenetic influences under most burial conditions (Fratzl et al., 2004; Nelson et al., 1986; Person et al., 1996; Shoulders and Raines, 2009; Tütken et al., 2008; Veis, 2003; Weiner and Wagner, 1998). The remainder of the bone is inorganic minerals including hydroxyapatite (approx. 65%), non-collagenous proteins (< 5%), and water (approx. 10%) (Olszta et al., 2007; Veis, 2003; Dorozhkin, 2011). The quality of type I collagen is typically assessed by measuring its elemental composition and the total amount of material extracted from a bone through a chemical process. Its unique amino acid profile contains ~11–16% nitrogen, ~30–45% carbon, and a carbon to nitrogen (C:N) ratio of 2.8–3.6 (Ambrose, 1990; DeNiro, 1985; DeNiro and Weiner, 1988; McNulty et al., 2002; van Klinken, 1999). Partially degraded type I collagen, other bone proteins, and post-burial humic, fulvic, and bacterial products have different compositions with C:N ratios outside this range (Balzer et al., 1997; DeNiro and Weiner, 1988; Harbeck and Grupe, 2009; Hare, 1980; van Klinken and Hedges, 1995). The C:N ratio is the typical criterion of choice for determining the presence of type I collagen with a well-preserved structure and whether

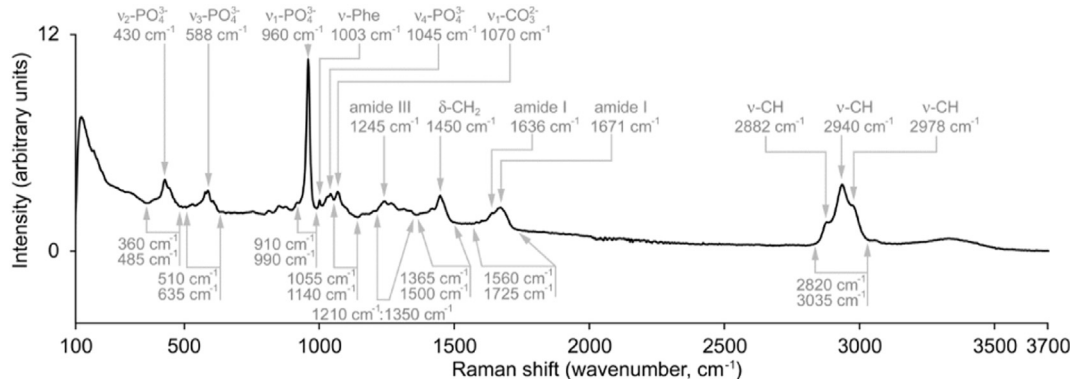


Fig. 1. Characteristic Raman spectrum collected from a well-preserved historic bone sample using a Thermo NXR FT-Raman spectrometer and 1064 nm laser excitation. Corresponding vibrational and stretching modes are indicated with wavenumber. (Reproduced from France et al., 2014a with permission from Elsevier.)

isotope values are original to the bone or the result of burial contamination. It is widely considered a more accurate measure of quality than the independent proportions of carbon and nitrogen in a sample. Furthermore, obtaining the elemental yields and C:N ratio requires only a small aliquot of extracted organic material, and the errors tend to be well-constrained.

The amount of organic material obtained from a bone by chemical extraction is obviously a quantitative measure but is also a predictor of type I collagen quality. It has been shown that intact type I collagen is likely to be present when the chemical extraction process yields 1–22% of the sample weight, though opinions on the lower end cut-off vary (Ambrose, 1990; McNulty et al., 2002; van Klinken, 1999; France et al., 2014a). The wide range of what is considered a sufficient overall yield, the fact that current extraction methods typically are only accurate to $\pm 1\%$ at best, and the fact that these organic residues are not identified as type I collagen analytically means that the proportion of extracted collagen, the so-called *collagen yield*, is not an entirely reliable measure of type I collagen quality. However, knowledge of overall collagen yield can be useful for proteomic techniques or selection of samples in diagenetic studies where a range of preservation quality is acceptable. Choosing to focus on C:N versus overall collagen yield depends on the research question at hand and the desired accuracy of predicted collagen quality. The C:N ratio is the more reliable quality indicator when high accuracy is preferred.

2. Comparison of recent studies

In the last two decades, several studies have used Raman spectroscopy to qualitatively assess the presence or absence of collagen in ancient bones (Morris and Mandair, 2011; France et al., 2014a; Pestle et al., 2014), but only a few have used Raman to quantitatively assess collagen preservation quality (France et al., 2014a; Halcrow et al., 2014; Pestle et al., 2015). These studies vary in experimental set up and findings. France et al. (2014a) used a high-resolution, research-grade FT-Raman spectrometer to analyze bone outer surfaces and unpolished, fresh fractured cross sections. They measured collagen quality by correlating Raman peak height ratios to C:N ratios obtained by chemical extraction. Halcrow et al. (2014) also used a high-resolution FT-Raman spectrometer to analyze outer bone surfaces, but they compared Raman peak area ratios (as opposed to peak heights) to the proportions of collagen, nitrogen, and carbon extracted from a bone sample (i.e., % collagen yield, % nitrogen yield, and % carbon yield), but not the C:N ratio. Pestle et al. (2015) used a handheld spectrometer to compare Raman peak height ratios to the collagen yield in powdered bone samples and did not consider C:N ratio. Each of the three studies offers criteria by which Raman spectra can determine the presence and quality of protein in the bone. All chose to evaluate the relative proportions of protein and hydroxyapatite in the bone as a ratio of two Raman peaks. All chose the peak at 960 cm^{-1} to represent the mineral, but they chose three different parts of the spectrum to represent the organic matrix. The studies also differ in the instrumentation used and the sample preparation. As we move forward in developing Raman spectroscopy to evaluate bone protein, it bears questioning if the criteria from one approach hold true for all data sets. A more detailed examination of the three studies highlights their relative strengths and the difficulties of comparing results in the absence of uniform reporting standards.

2.1. France et al. (2014a)

The most complete sample and data set available to the authors of this current paper is their own as originally published in France et al. (2014a). Therefore, it is used as the primary basis for comparison here. In that work, organic material was chemically extracted from 44 modern and 19th-century archaeological bone samples from humans and other mammals. The C:N ratio of each extracted residue was

calculated. Bones from which the extracted organic material had a C:N ratio of 2.8–3.6 were considered to contain well-preserved type I collagen, and C:N values outside that window were construed to mean the extracted organic material was not type I collagen or was poorly preserved. FT-Raman spectra were also collected of the bones, and the spectra were subjected to a rigorous statistical analysis to determine a best Raman indicator of a C:N ratio of 2.8–3.6. Seventy-eight possible peak height ratios, 28 peak area ratios, and multivariate statistical methods (i.e., partial least squares discriminant analysis (PLSDA)) were considered. It was determined that a bivariate 960 cm^{-1} : 1636 cm^{-1} peak height ratio of ≤ 19.4 in freshly exposed cross-sections provided the highest successful prediction rate (95%) of well-preserved samples with a C:N ratio of 2.8–3.6. For outer surfaces of the bone they divided the results into three groups. A peak height ratio ≥ 26.8 was not expected to yield intact type I collagen. A ratio < 26.8 indicated that 61% of samples in that Raman range contained well-preserved type I collagen, and for ratios < 19.4 the proportion of well-preserved samples jumped to 67%.

In addition, France et al. (2014a) asserted that Raman spectra should also be evaluated visually for the presence of bone spectral features before evaluating the spectra mathematically. They found PLSDA also predicted type I collagen quality successfully but requires significant time and effort to establish an internal laboratory dataset for comparison. Finally, they showed that samples predicted by Raman to contain poorly preserved collagen also had low collagen yields ($< 5\%$).

2.2. Halcrow et al. (2014)

In contrast, Halcrow et al. (2014) correlated a ratio of Raman peak areas at $3060\text{--}2800\text{ cm}^{-1}$ and $983\text{--}930\text{ cm}^{-1}$ to the extracted collagen yield ($r = 0.716$), % N yield ($r = 0.706$), and % C yield ($r = 0.630$). They provided the C:N values for their samples but did not compare these to their Raman data. By applying the method of France et al. (2014a) to calculate a *classification rate*, described below, one can determine the percentage of Halcrow's samples that would be classified correctly as well-preserved or poorly preserved based on C:N ratios and their proposed $3060\text{--}2800\text{ cm}^{-1}$: $983\text{--}930\text{ cm}^{-1}$ peak area ratio. Using a cut-off for this peak area ratio of 0.205 allows successful prediction of collagen quality $\sim 77\%$ of the time, where ratios > 0.205 should indicate well-preserved samples. This is a higher success rate for outer bone surfaces than the method proposed by France et al. (2014a).

Halcrow et al. (2014) also visually examined Raman spectra and stated that all bones with a collagen yield $> 1\%$ showed the presence of a C–H stretching band in the $3060\text{--}2800\text{ cm}^{-1}$ region. However, this band is not necessarily well-preserved type I collagen. In the France et al. (2014a) data set, 88% of samples that were poorly preserved (i.e. C:N outside the range of 2.8–3.6) showed a small Raman peak group in the same region. This peak group sometimes had a different shape from spectra of well-preserved samples, which suggests a different arrangement of C–H bonds in the sample. Bonds between carbon and hydrogen are hallmarks of most organic matter. Many substances, be they type I collagen, degradation products of type I collagen, other bone proteins, or something else, have Raman peaks in this region.

2.3. Pestle et al. (2015)

Pestle et al. (2015) provided the first study of a handheld Raman spectrometer to quantitatively determine collagen yields. The study used the Raman peak at 1450 cm^{-1} as the collagen quantity indicator. The samples were mostly powders plus a small number of outer bone surfaces. The study determined that samples with a peak height ratio of 1450 cm^{-1} : $960\text{ cm}^{-1} > 0.1$ could be expected (with 95% confidence) to contain $> 1\%$ collagen.

The 1450 cm^{-1} : 960 cm^{-1} peak height ratios from France et al. (2014a) show a fairly close agreement to this cut-off. In terms of C:N ratios, approximately 91% of well-preserved samples (i.e., C:N

= 2.8–3.6) show $1450\text{ cm}^{-1}:960\text{ cm}^{-1} > 0.1$, while 88% of poorly preserved samples show $1450\text{ cm}^{-1}:960\text{ cm}^{-1} < 0.1$. Approximately 10% of the France et al. (2014a) samples would be misclassified using Pestle's criteria which are based solely on quantitative collagen yield. The France et al. (2014a) data show a moderate correlation between the collagen yield and $1450\text{ cm}^{-1}:960\text{ cm}^{-1}$ ratio ($R^2 = 0.604$). In fact, the $1636\text{ cm}^{-1}:960\text{ cm}^{-1}$ ratio provides a slightly better correlation ($R^2 = 0.649$) for the France et al. (2014a) data.

Pestle et al. (2015) also offer an equation to predict the potential collagen yield of a sample based on $1450\text{ cm}^{-1}:960\text{ cm}^{-1}$ ratio. However, application of this formula to the France et al. data set fails to accurately predict collagen yield within two standard deviations for almost 30% of their samples. Pestle et al. (2015) do not provide C:N ratios, so a direct comparison to the France et al. $960\text{ cm}^{-1}:1636\text{ cm}^{-1}$ cut-off of ≤ 19.4 for well-preserved samples is not possible.

3. Re-evaluation of France et al. (2014a) data with a handheld Raman spectrometer

The advent of miniaturized Raman spectrometers with near infrared excitation is a boon for bone analysis. Smaller instruments like that used by Pestle et al. can be deployed at archaeological sites or in the laboratory at a fraction of the cost of a research-grade benchtop spectrometer. Some of these handheld instruments are ruggedized for use in the field. However, there are important differences between the spectra obtained with research-grade and handheld instruments that must be taken into account when evaluating archaeological bone. This is a potential obstacle to comparing the data set collected by Pestle et al. to those of Halcrow et al. and France et al. To bridge this divide, the set of bones analyzed by France et al. was re-examined with a Rigaku Progeny handheld 1064 nm Raman spectrometer. The data is summarized here and compared to results obtained by Pestle et al. (2015). All spectra and calculations are available as Electronic Supplemental Material.

3.1. Materials and methods

Samples: Thirty-seven human bone specimens analyzed by France et al. (2014a) were re-analyzed with the Progeny (Table 1). Briefly, all bones are archaeological human remains from 19th-century North American burial sites (France et al., 2014b). The remains were stored in a cool dry location, typically a storage drawer, in keeping with typical museum and university collections storage environments. Two bones tested in France et al. (2014a) were unavailable as they have since been repurposed for a different project. To the extent possible, samples were analyzed on an outer surface ($n = 10$), on an unpolished cross section that was either naturally exposed during excavation or prepared with a chisel or bone saw ($n = 29$), and as a powder ($n = 35$). For outer surfaces, soil-like accretions were removed with a scalpel, but the bone was otherwise unmodified. Unlike the benchtop FT-Raman spectrometer used in the previous study, the handheld Progeny does not have an automated microstage or live camera to help with sample positioning. As a result, many of the smaller bone samples did not have sufficient surface area for the Progeny to analyze on both an outer surface and in cross section.

Raman spectroscopy: Spectra were collected with a Rigaku Progeny 1064 nm handheld Raman spectrometer. The 3.5 lb. instrument features a continuous wave, unpolarized Nd:YAG laser, thermoelectrically cooled InGaAs 512 pixel detector, and an optional docking station for static measurements (when hand holding is not desired). Spectra span $200\text{--}2500\text{ cm}^{-1}$ at fixed $8\text{--}11\text{ cm}^{-1}$ spectral resolution. Most spectra were a co-addition of twelve three-second scans and used the instrument's proprietary "turbo" mode. Laser power was determined empirically starting from the power density of FT-Raman measurements from the previous study. The equivalent laser power was calculated for the Progeny, which has a smaller analytical spot ($25\text{ }\mu\text{m}$ diameter) (see Electronic Supplemental Material: Laser power density conversion). For

nearly all samples, the Progeny instrument's 125 mW setting ($2.5 \times 10^4\text{ W/cm}^2$) achieved sufficient Raman scattering without damaging the samples. Three spectra were collected for each solid bone orientation (outer surface and cross section) and two spectra for each powder.

3.2. Data processing and analysis

Whereas France et al. (2014a) compared bivariate peak ratios and multivariate approaches, this current study only considers the bivariate methods recommended in recent literature. Spectra were evaluated visually followed by calculation of the ratio of the tallest hydroxyapatite peak (960 cm^{-1}) to the heights of two prominent organic peaks: 1450 cm^{-1} (advocated by Pestle et al.) and 1636 cm^{-1} (advocated by France et al.). The peak group advocated by Halcrow et al., $3060\text{--}2800\text{ cm}^{-1}$, lies beyond the spectral range of today's handheld Raman spectrometers, so it could not be evaluated.

Spectra were downloaded and converted to SPA (proprietary format of Thermo Fisher Scientific) and CSV formats for interpretation using TQ Analyst chemometric software (version 8.4.257, Thermo Fisher Scientific, Inc.) and for publication here as Electronic Supplemental Material, respectively.

Raman spectra of archaeological bone can be affected by sample heterogeneity (France et al., 2014a). For each sample, the best of three Raman spectra for each outer surface and cross section, and the best of two spectra for each powder was chosen with a quick visual assessment by the instrument operator. Visual selection is an intuitive process, but the relevant criteria of a *best* spectrum are most likely low noise, low fluorescence, well-defined peaks, and the absence of extraneous peaks. Next, a classification rate was calculated. The classification rate measures the success of each peak height ratio to predict the quality of type I collagen (based on C:N ratio) for a group of Raman spectra. Spectra were grouped into six spectral datasets and compared: (1) all cross section spectra, (2) all outer surface spectra, (3) all powder spectra, and (4) best cross section spectra, (5) best outer surface spectra, and (6) best powder spectra (France et al., 2014a). The *all* versus *best* spectral datasets were compared to evaluate whether selecting spectra visually influenced data reliability.

Classification rate = (# correctly assigned spectra \div total number of spectra) \times 100.

3.2.1. Baseline correction

Peak heights and baselines were calculated two ways in this study using automated methods developed in TQ Analyst. The first method is that used in France et al. (2014a), whereby each peak height is measured at a single specified position, and each baseline is defined as a line drawn across the base of the peak and anchored on either side by a single point along the spectrum (Method 1) (Fig. 2A) (Table 2). This reduces the influence of sample fluorescence, which results in spectra that slope upwards.

However, this approach does not consider the error associated with spectral noise, a term that describes fluctuations along the spectrum that are not related to Raman or fluorescence signals and which are usually unwanted. In a so-called noisy spectrum the magnitude of these fluctuations is large relative to the height of the Raman peaks. Noise can distort a baseline that is defined by only two anchor points and also the height of a peak that is measured at a single discrete point (Fig. 2B). The peak height is a sum of the Raman signal (a), the baseline noise error (b), and the peak noise error (c). This effect is more pronounced for smaller Raman peaks, where the percentage noise error can be quite significant and even obscure a peak or create the illusion of a peak that is not there. Noise tends to be more pronounced with portable and handheld Raman spectrometers than research-grade instruments (Fig. 3).

The influence of noise on peak height ratios was addressed with a second, modified method of calculating peak heights and baselines

Table 1
Sample list and elemental abundance data organized by whether collagen was well- or poorly preserved and by C:N ratio.^a

Sample	Museum Designation ^b	Burial Location	Bone	% Collagen	Wt%N	Wt%C	C:N	Collagen Preservation ^c	Orientation ^c
394	6CT58-5-AMM03	Walton Family cemetery plot, CT	Femur	21.2	15.0	41.7	3.2	Well	cp
396	6CT58-5-AMM05	Walton Family cemetery plot, CT	Femur	17.6	14.6	40.3	3.2	Well	ocp
400	6CT58-5-AMM11	Walton Family cemetery plot, CT	Femur	19.0	14.5	40.2	3.2	Well	cp
281	29LA1091-BOR-42	Ft. Craig Cemetery, NM	Metatarsal	4.9	14.2	40.4	3.3	Well	op
343	51RICHARDS-CC-07	Congressional Cemetery, DC	Metacarpal	12.1	14.8	41.8	3.3	Well	op
345	GLO-099-2A	Glorieta Pass Battlefield, NM	Femur	13.3	14.2	39.7	3.3	Well	cp
350	GLO-099-2C	Glorieta Pass Battlefield, NM	Femur	6.3	14.2	40.0	3.3	Well	c
417	51KEYWORTH-CC-04	Congressional Cemetery, DC	Radius	14.9	14.7	41.0	3.3	Well	cp
420	51KEYWORTH-CC-07	Congressional Cemetery, DC	Metatarsal	21.5	14.3	40.4	3.3	Well	cp
426	51WHITE-CC-02	Congressional Cemetery, DC	Fibula	16.4	14.5	40.8	3.3	Well	ocp
442	51CAUSTEN-CC-11	Congressional Cemetery, DC	Metacarpal	23.5	14.2	40.4	3.3	Well	cp
443	51CAUSTEN-CC-13	Congressional Cemetery, DC	Metacarpal	17.5	14.9	41.8	3.3	Well	cp
454	18PR224-06-300	Family tomb, MD	Femur	9.7	14.3	40.3	3.3	Well	cp
455	18PR224-99-400	Family tomb, MD	Tibia	7.3	12.4	35.3	3.3	Well	cp
466	51KEYWORTH-CC-08	Congressional Cemetery, DC	Tibia	16.8	12.3	34.9	3.3	Well	cp
508	ELMINA-A-L49B	Elmina Settlement, Ghana, Africa	Tibia	5.0	12.1	34.6	3.3	Well	op
525	ELMINA-A-Y51L	Elmina Settlement, Ghana, Africa	Femur	8.0	10.2	28.8	3.3	Well	c
558	7NCE98A-WOODVILLE-10	Woodville Cemetery, DE	Metatarsal	16.1	13.6	38.5	3.3	Well	op
560	7NCE98A-WOODVILLE-SLOPEB	Woodville Cemetery, DE	Ulna	29.7	13.8	39.1	3.3	Well	cp
566	FABC-08-107a	First African Baptist Church Cemetery, PA	Metacarpal	15.3	14.5	40.8	3.3	Well	cp
471	31FOSCUE-ECU-1	Foscue Plantation family plot, NC	Ulna	7.2	14.1	40.7	3.4	Well	cp
556	7NCE98A-WOODVILLE-04	Woodville Cemetery, DE	Fibula	5.1	12.3	36.5	3.5	Well	op
598	44PWKINCHELOE-SI9114-C	Kincheloe Plantation family plot, NC	Temporal	3.6	12.9	38.8	3.5	Well	cp
401	6CT58-5-AMM13	Walton Family cemetery plot, CT	Femur	0.5	12.9	23.3	2.1	Poor	cp
330	7BAYVISTA-DHCA-118A	Ground burial, DE	Femur	0.5	7.0	22.0	3.7	Poor	cp
501	ELMINA-A-E51B	Elmina Settlement, Ghana, Africa	Femur	4.4	10.0	32.5	3.8	Poor	cp
255	GETTYS-NPS-965	Gettysburg Battlefield, PA	Temporal	0.6	8.2	27.9	4.0	Poor	cp
305	TRINITY-EAST-22	Trinity Church Cemetery, DC	Temporal	0.8	9.7	34.8	4.2	Poor	cp
307	TRINITY-EAST-25	Trinity Church Cemetery, DC	Mandible	0.5	7.1	25.9	4.3	Poor	op
297	TRINITY-EAST 06	Trinity Church Cemetery, DC	Temporal	0.4	1.9	8.3	5.0	Poor	cp
555	7NCE98A-WOODVILLE-02	Woodville Cemetery, DE	Metatarsal	1.6	6.4	27.8	5.0	Poor	op
303	TRINITY-EAST-14	Trinity Church Cemetery, DC	Temporal	< 0.2 ^d	7.7	35.2	5.4	Poor	cp
557	7NCE98A-WOODVILLE-06	Woodville Cemetery, DE	Femur	0.8	6.5	32.7	5.9	Poor	cp
266	29LA1091-BOR-23B	Ft. Craig Cemetery, NM	Talus	1.0	6.5	34.0	6.1	Poor	cp
267	29LA1091-BOR-23C	Ft. Craig Cemetery, NM	Talus	0.5	4.3	25.9	7.0	Poor	cp
268	29LA1091-BOR-23D	Ft. Craig Cemetery, NM	Talus	0.2	5.3	33.9	7.5	Poor	cp
309	TRINITY-EAST-04C	Trinity Church Cemetery, DC	Metacarpal	< 0.2	5.2	47.2	10.5	Poor	op

^a Elemental abundance data from these bone specimens has been published previously in France and Owsley (2015) or France et al. (2014a and 2014b) and reappears here.

^b Historic specimens are accessioned objects in the collection of the Smithsonian National Museum of Natural History.

^c Samples with “well” preserved collagen had a C:N ratio of 2.8–3.6, and “poor” preservation fell outside that range.

^d % Collagen yields ~0% neared the limit of resolution of our analytical scale and are noted as < 0.2% to encompass the inherent error.

^e Orientation(s) of sample from which Raman spectra were collected (o = outer surface, c = cross section, p = powder).

(Method 2) (Fig. 2C) (Table 2). Peak heights were defined as an average height across a narrow range bracketing the peak of interest. For example, the 960 cm⁻¹ peak was defined as the average peak height across 961.5–958.5 cm⁻¹. Each baseline anchor was also defined as an average height across a range. Post-processing smoothing algorithms were not applied to the spectra as they did not reliably improve, and often worsened, classification predictions during the France et al. (2014a) study of these same samples.

3.3. Results

Classification rates for the six Raman spectral datasets and two baseline modeling methods are presented in Table 3. The best success was achieved with Raman spectra of bone cross sections and a 960 cm⁻¹:1450 cm⁻¹ averaged peak height ratio; 86% of samples were classified correctly. In all cases, selectivity was improved by collecting more than one spectrum and selecting the best-looking one according to the criteria given in the Materials and Methods section. Furthermore, calculating peak heights and the anchor points as average values across a small range of wavenumbers almost always improved selectivity. Selectivity always improved when doing both. All spectra from which these scores are calculated are included as Electronic Supplemental Material.

Selectivity for Raman spectra of outer surfaces and powdered samples with the 960 cm⁻¹:1450 cm⁻¹ ratio was less successful than

for cross sections. In all Raman datasets, there were threshold values below and above which all samples contained well- or poorly preserved collagen, and results between these thresholds were a mix of well-preserved samples and poorly preserved samples (see Electronic Supplemental Information: Spectral Interpretation). The Raman peak height ratios also seem to have some predictive value in this zone of mixed results (Table 4). For example, considering the sample set Best Outer Surface Spectra with ratios calculated by baseline modeling Method 2, ratios ≤ 8.1 were all well-preserved. For ratios in the range of 8.3–20.2, 60% of samples contained well-preserved collagen (i.e., C:N 2.8–3.6). All ratios ≥ 25.8 indicated poorly preserved samples. For the Best Powdered Sample Spectra calculated by baseline Method 2, ratios ≤ 6.2 were well-preserved. For ratios in the range 6.3–8.9, 92% of samples contained well-preserved collagen. For ratios in the range 9.9–17.7, that proportion dropped to 57%. All ratios > 20.2 were poorly preserved. A similar trend was observed with the 960 cm⁻¹:1636 cm⁻¹ peak height ratio data (Table 4).

Several spectra had negative ratio values, and these would have been classified incorrectly by France et al. (2014a), where values below < 19.4 are considered well-preserved. All samples with which this occurred were poorly preserved. Negative ratio values occurred in two situations. In the first situation, seen with the cross section spectrum of sample 557, the collagen indicator peak was so short as to have a negative height relative to the calculated baseline (i.e., the noise was greater than the peak height). The other scenario was due to

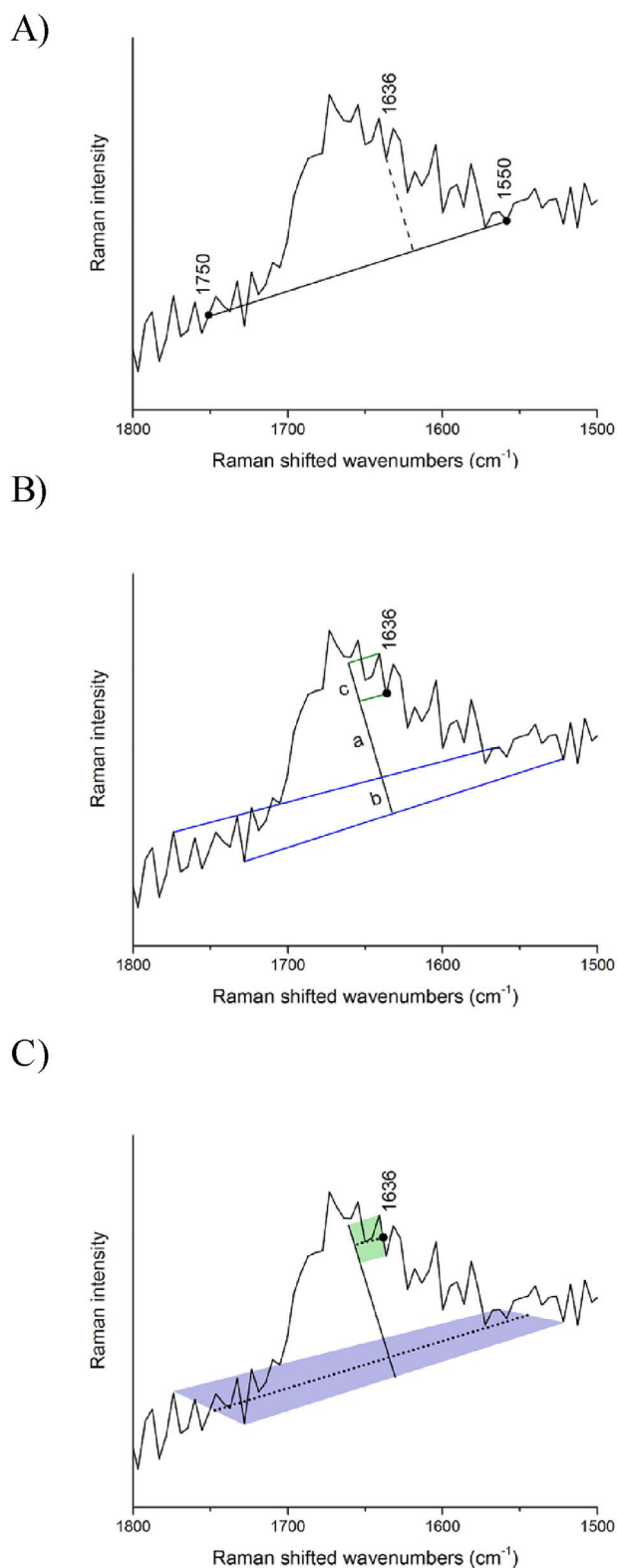


Fig. 2. Raman spectrum from specimen 560, a historic human ulna and an example of a bone with well-preserved collagen. A) The 1636 cm^{-1} peak with baseline defined as two discrete points. B) The measured peak height varies significantly depending on how the baseline is drawn (two possibilities shown in violet) and on the magnitude of noise at the top of the peak (two possibilities shown in green). C) Averaging the baseline and peak apex y-values across a range of wavenumbers (shown as dotted lines) reduces the effect of noise, resulting in more reliable peak measurements.

contamination; the spectrum had unidentified peaks that skewed the baselines upwards. This underscores the importance of visually examining spectra before calculating collagen quality mathematically or with an automated process.

4. Discussion

The three studies published in 2014 and 2015 come to different conclusions about the best indicator of collagen quality and quantity in a Raman spectrum. Re-evaluation of the [France et al. \(2014a\)](#) bone samples with a handheld spectrometer supports the finding of [Pestle et al. \(2015\)](#) that a peak height ratio of $960\text{ cm}^{-1}:1450\text{ cm}^{-1}$ can indicate the presence of intact collagen better than $960\text{ cm}^{-1}:1636\text{ cm}^{-1}$ when using a handheld 1064 nm Raman spectrometer. However, earlier findings by France et al. with a high-resolution FT-Raman spectrometer are likely to have measured type I collagen quality more accurately. Pestle et al. posit that the difference between these studies' results "may be a consequence of, among other possibilities, differences in the thermal histories of the samples used in each study and consequent differences in the degree of glutamine deamination within those samples" ([Pestle et al., 2015](#) p. 116). Instead, the data presented here indicate that the difference is probably a consequence of instrument resolution and the Raman signal-to-noise ratio rather than sample composition.

In a Raman spectrum of a modern bone unaltered by diagenesis, the signature of hydroxyapatite is stronger than that of the organic matrix ([Fig. 1](#)). The 960 cm^{-1} peak, which represents a symmetrical stretching vibration of phosphate ($\nu_1[\text{PO}_4]^{3-}$), is the strongest in the spectrum. In the wavelength range of today's handheld Raman spectrometers, which cannot see the C–H stretches at $3060\text{--}2800\text{ cm}^{-1}$, the strongest organic peak is at 1450 cm^{-1} and represents a scissoring motion of C–H bonds. The peak at 1636 cm^{-1} is the third strongest peak and part of a peak group that represents amide I, a composite feature that is peculiar to proteins, is mostly due to carbonyl stretching, and informs about crosslinking of the collagen strands and protein conformation ([Barth and Zscherp, 2002](#); [Carden et al., 2003](#)). Here the difference in sensitivity between a research-grade spectrometer and a handheld instrument becomes important. The baselines of bone spectra, and particularly bones that have been buried, are noisy, and the noise buries a significant portion of these relatively weak Raman peaks. The 1450 cm^{-1} peak is taller and can be measured more reliably than the 1636 cm^{-1} peak, which has a higher percentage of noise error ([Fig. 3](#)). This error increases as the peak height shrinks, so bones with little remaining collagen will be affected the most. It follows that handheld instruments, with their noisier spectra, are less able to distinguish small peaks than more sensitive, less noisy research-grade instruments.

While the 1450 cm^{-1} peak may be a more successful indicator with handheld instruments because it is taller, there are qualitative concerns about suggesting it to identify type I collagen and other endogenous bone protein. As with the $3060\text{--}2800\text{ cm}^{-1}$ peak group described by [Halcrow et al.](#), the 1450 cm^{-1} peak describes vibrations of C–H bonds, which are a hallmark of organic material. This sets up the possibility that the source of a 1450 cm^{-1} peak could be something other than protein. For example, the most common formulations that have been used by conservators to consolidate or adhere archaeological bone, including Acryloid B-72 acrylic resin, polyvinyl acetate, cellulose nitrate, and soluble nylon, have peaks at $3060\text{--}2800\text{ cm}^{-1}$ and 1450 cm^{-1} . The 1636 cm^{-1} peak is specific to proteins and therefore a better indicator of collagen with a C:N ratio of 2.8–3.6, if the spectrometer can resolve that peak well. The potential for organic consolidants, adhesives, and burial contamination underscores the importance of qualitatively evaluating the entire Raman spectrum by eye.

Negative ratio values are another potential source of error in the methods published thus far. These can occur when the magnitude of spectral noise exceeds the height of a relevant Raman peak, or when a contamination peak occurs at one of the baseline anchors. If there is a

Table 2

List of Raman peaks used in peak height ratios and the intervals over which corrected baselines were modeled.

METHOD 1: SPECIFIC POINTS			
Bond vibration probed	Peak position (Raman shifted cm^{-1})	Modeled baseline region (Raman shifted cm^{-1})	
$\nu_1\text{-[PO}_4\text{]}^{3-}$	960	990–910	
$\delta\text{-CH}_2$	1450	1530–1150	
Amide I	1636	1750–1550	
METHOD 2: AVERAGE HEIGHT			
Bond vibration probed	Peak region (Raman shifted cm^{-1})	Baseline: left anchor region (Raman shifted cm^{-1})	Baseline: right anchor region (Raman shifted cm^{-1})
$\nu_1\text{-[PO}_4\text{]}^{3-}$	961.5–958.5	1175–1135	920–890
$\delta\text{-CH}_2$	1451.5–1448.5	1550–1510	1170–1130
Amide I	1637.5–1634.5	1770–1730	1570–1530

sharp peak at 960 cm^{-1} , and there are no additional peaks at the baseline anchor points, then any negative ratio value must indicate a poorly preserved sample. In any case, these Raman spectra should be examined individually and, if needed, collected again before the collagen quality is evaluated mathematically.

For this sample set choosing the best-looking spectrum out of two or three usually improved the collagen quality prediction rate over collecting a single spectrum. Differences between spectra are likely due to variation in the bone from life processes or diagenesis (Weiner and Wagner, 1998) or changes in the position of the instrument during the measurement. The value of this practice should be investigated further as selectivity improved as much as 300% when calculating the ratio using the smaller 1636 cm^{-1} peak (Table 3). Furthermore, calculating the peak heights and baseline anchors as average heights over a narrow range usually improved selectivity by reducing the influence of noise. Selectivity was always improved when doing both.

These practices improved our ability to choose bones that contain well-preserved collagen, but the answers are not black-and-white. In most cases, the Raman ratios identify some bone samples that obviously contain good collagen and others that obviously do not. Many bone samples fall somewhere in between. If one only considers the results as presented in Table 3, Raman pre-screening does not look very helpful. However, considering this data in terms of the likelihood of finding well-preserved collagen is much more useful. Table 4 shows ranges of Raman ratios in which 92%, 87%, or 20% of samples were well-preserved, for example. That is valuable information for an archaeologist deciding

whether or not to sacrifice a bone to destructive analysis. The task now is to build and publish many more bone samples and build Raman spectroscopy into a statistically relevant predictive tool.

Given the statistical nature of many collagen studies, for example with stable isotope analysis, a future goal could be to automate these calculations with a software interface that controls the spectrometer and automatically evaluates the collagen quality with a numerical value, probability that a sample is well preserved, or a pass/fail system. The task now is to rigorously test the method on more bones in order to build a statistically relevant sample set and to determine the best parameters for spectral collection and evaluation. To date only around 200 bone samples have been published that correlate Raman spectra to C:N ratios or collagen yield. To develop this technique into a reliable predictive tool, more samples must be published. Gaps in the comparisons of France, Halcrow, and Pestle above demonstrate the importance of uniformity in data reporting. For each sample, a full set of elemental and overall yield information (i.e. collagen yield, % N yield, % C yield, and C:N) is required as well as the Raman spectra (in a universal format such as JCAMP or CSV), and the calculated peak height and peak area ratios. Given the ubiquitous option to publish supplementary data tables with any reputable journal, there is little reason to withhold this information. By working together and testing methods against many data sets, a consensus of best practices is likely to emerge.

It may be that a combination of tests might assess collagen quality more reliably than any single test. For example, if two distinct peak height ratios indicate that both C:N ratios and collagen yield “pass” pre-

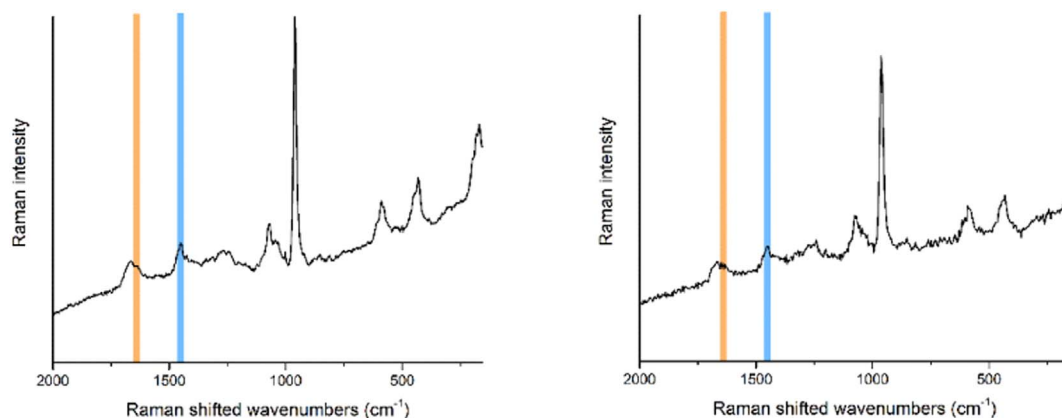


Fig. 3. Raman spectra collected from Sample 525, a historic human femur and an example of bone with well-preserved collagen by C:N and a collagen yield toward the low end of well-preserved samples (8.0%), with (A) Thermo NXR research-grade FT-Raman and (B) Rigaku Progeny handheld 1064 nm Raman spectrometers. The amplitude of noise is visibly higher in the spectrum collected with the handheld instrument, and the effect on peaks of interest is highlighted around 1636 cm^{-1} and 1450 cm^{-1} .

Table 3

Summary of classification rates for the samples listed in Table 1. The success of the 960 cm^{-1} : 1450 cm^{-1} and 960 cm^{-1} : 1636 cm^{-1} Raman peak height ratios to predict the presence of well-preserved collagen in 37 bone samples is shown as percentages of correctly classified samples. The data were assessed by whether all Raman spectra collected from a sample or only the best-of-three spectra (for cross sections and outer surfaces) or best-of-two spectra (for powders) were considered. The two baseline modeling methods described in Table 2 are also compared.

Raman ratio	All cs spectra (n=87)		Best cs spectra (n=29)	
	Method 1	Method 2	Method 1	Method 2
960 cm^{-1} : 1450 cm^{-1}	75%	79%	83%	86%
960 cm^{-1} : 1636 cm^{-1}	28%	36%	45%	48%
Raman ratio	All o spectra (n=30)		Best o spectra (n=10)	
	Method 1	Method 2	Method 1	Method 2
960 cm^{-1} : 1450 cm^{-1}	9%	9%	40%	50%
960 cm^{-1} : 1636 cm^{-1}	13%	10%	50%	60%
Raman ratio	All p spectra (n=70)		Best p spectra (n=35)	
	Method 1	Method 2	Method 1	Method 2
960 cm^{-1} : 1450 cm^{-1}	21%	21%	37%	46%
960 cm^{-1} : 1636 cm^{-1}	20%	29%	20%	37%

Table 4

Summary of Raman ratio cut-offs and the incidence of well-preserved collagen for the bone samples described in Table 1.

Raman peak height ratio	Raman ratio range ^b	Samples in range ^c	Well-preserved samples in range ^d	% classified as well-preserved ^e		
960 cm^{-1} : 1450 cm^{-1}	<i>Best cross section spectra^a</i>	≤ 9.26	15	15	100	
		9.33-18.3	4	3	75	
		≥ 21.5	10	0	0	
	<i>Best outer surface spectra</i>	≤ 8.1	4	4	100	
		8.3-20.2	5	3	60	
		≥ 25.8	1	0	0	
	<i>Best powder spectra</i>	≤ 6.2	6	6	100	
		6.3-9.0	12	11	92	
		9.9-17.7	7	4	57	
		≥ 20.2	10	0	0	
	960 cm^{-1} : 1636 cm^{-1}	<i>Best cross section spectra</i>	≤ 10.6	5	5	100
			10.7-35.0	15	13	87
≥ 35.8			9	0	0	
<i>Best outer surface spectra</i>		≤ 13.4	5	5	100	
		14.1-23.3	4	2	50	
		≥ 47.1	1	0	0	
<i>Best powder spectra</i>		≤ 12.9	9	9	100	
		13.0-17.0	12	10	83	
		17.1-38.0	10	2	20	
		≥ 39.7	4	0	0	

^a Dataset is the best of three Raman spectra collected for each sample, based on a visual assessment.

^b Range of values for the given Raman peak height ratio.

^c Number of samples with Raman peak height ratios in range.

^d Number of samples with peak height ratios in range that had well-preserved type I collagen. Well-preserved type I collagen is defined as extracted organic material with C:N ratio of 2.8-3.6.

^e Percentage of samples in range that were classified as containing well-preserved type I collagen.

determined cut-offs, then the sample has a very high probability of yielding well-preserved collagen. Similarly, if all criteria “fail”, then the sample is unlikely to produce well-preserved collagen. This approach would give more robust information to decide if precious samples should be sacrificed to destructive analysis. It is also possible that no consensus will emerge. All studies agree that fluorescence and

heterogeneity within samples are issues that increase the error in data. Different studies employ different Raman spectrometers, each with a specific signal-to-noise ratio, spot size, and optimum resolution. Individual instruments may require custom criteria to determine collagen quality. Furthermore, the current methods of collagen extraction can produce significant differences and error in collagen yield data

(Jorkov et al., 2007). All studies discussed above use a common acid-base-acid extraction ultimately modeled after Longin (1971) which has the disadvantage of minor sample loss during the process. It is unavoidable and results in collagen yield data that can be slightly inaccurate. Differences between the extraction practices in each lab may result in Raman ratio thresholds that are unique to that lab. Therefore it is recommended that C:N ratios also be reported and correlated to peak height or area ratios because C:N is less susceptible to lab-specific procedural differences.

5. Conclusions

Raman spectroscopy is the most promising quantitative method for non-destructive determination of collagen quality in recent decades. Portable and handheld 1064 nm Raman spectrometers are certainly a boon to archaeological fieldwork and analysis of organic materials in general. However, improvements in portability and price are still offset by lower sensitivity and signal-to-noise than research-grade instruments. That tradeoff must be considered when evaluating collagen in archaeological bone. Three recent quantitative studies predict collagen quality or quantity by comparing the proportion of mineral and organic bonds indicated in a Raman spectrum. The studies share many similarities but differ in the peaks chosen to represent the organic component. The authors of this article found that the 1636 cm^{-1} peak, which represents amide I and is a characteristic of proteins, was the best collagen quality indicator using a highly sensitive research-grade FT-Raman spectrometer. However, the authors also found that this peak is difficult to resolve with a handheld spectrometer, especially when there is little collagen present in the sample (i.e., the peak is small). Conversely, the peak at 1450 cm^{-1} is more likely to be resolved well by a handheld spectrometer, but it is not as specific to collagen and could describe other organic matter. While similarities are emerging in the Raman indicators of yield and quality, differences in equipment and chemical processing remain to be considered.

Raman evaluation of collagen will improve as a technique as more sample sets are published and compared. This will require uniform and complete data reporting to be effective. The present study and that of France et al. (2014a) examined the same set of archaeological bones with research-grade and handheld Raman spectrometers. Based on those findings, the following procedure is recommended for assessing collagen quality:

1. Collect multiple spectra of each bone in the desired orientation (cross section, outer surface, or powder). Cross sections that are freshly exposed intentionally with tools or unintentionally during excavation have consistently been most successful for evaluating the presence of well-preserved bone collagen.
2. Examine the spectra visually for fluorescence, noise, and any extraneous peaks.
3. Select the best-looking spectrum (a complex intuitive choice based on low fluorescence, low noise, well-defined peaks, and no extraneous peaks, among other possible considerations).
4. Calculate $960\text{ cm}^{-1}:1450\text{ cm}^{-1}$ and $960\text{ cm}^{-1}:1636\text{ cm}^{-1}$ peak height ratios using baseline modeling Method 2 (Table 3). Peaks and baseline anchors should be modeled as averages over narrow wavenumber ranges.
5. Assess the likelihood of finding well-preserved collagen by comparison to past results (i.e., Table 4).

This general method is applicable to any 1064 nm Raman spectrometer, with the exception of the probabilities in Step 5 that are specific to human bones analyzed with a handheld instrument.

Acknowledgments

The authors acknowledge D. Owsley for access to the historic bone samples, and D. B. Thomas and C. R. Doney for their previous work and insights. National Science Foundation Research Experience for Undergraduates Award #SMA-1156360 provided financial support for D. M. W. Chan and M. Dundon and did not otherwise contribute to the research. All analyses were performed at the Smithsonian Modern Materials Laboratory at MCI and the Smithsonian MCI Stable Isotope Mass Spectrometry Laboratory in Suitland, Maryland.

References

- Ambrose, S.H., 1990. Preparation and characterization of bone and tooth collagen for isotopic analysis. *J. Archaeol. Sci.* 17, 431–451.
- Balzer, A., Gleixner, G., Grupe, G., Schmidt, H.-L., Schramm, S., Turban-Just, S., 1997. In vitro decomposition of bone collagen by soil bacteria: the implications for stable isotope analysis in archaeometry. *Archaeometry* 39, 415–429.
- Barth, A., Zscherp, C., 2002. What vibrations tell us about proteins. *Q. Rev. Biophys.* 35, 369–430.
- Carden, A., Rajachar, R.M., Morris, M.D., Kohn, D.H., 2003. Ultrastructural changes accompanying the mechanical deformation of bone tissue: a Raman imaging study. *Calcif. Tissue Int.* 72, 166–175.
- DeNiro, M.J., 1985. Postmortem preservation and alteration of in vivo bone collagen isotope ratios in relation to palaeodietary reconstruction. *Nature* 317, 806–809.
- DeNiro, M.J., Weiner, S., 1988. Chemical, enzymatic and spectroscopic characterization of “collagen” and other organic fractions from prehistoric bone. *Geochim. Cosmochim. Acta* 52, 2197–2206.
- Dorozhkin, S.V., 2011. Calcium orthophosphates: occurrence, properties, biomineralization, pathological calcification and biomimetic applications. *Biomater* 1, 121–164.
- France, C.A.M., Owsley, D.W., 2015. Stable carbon and oxygen isotope spacing between bone and tooth collagen and hydroxyapatite in human archaeological remains. *Int. J. Osteoarchaeol.* 25, 299–312.
- France, C.A.M., Thomas, D.B., Doney, C.R., Madden, O., 2014a. FT-Raman spectroscopy as a method for screening collagen diagenesis in bone. *J. Archaeol. Sci.* 42, 346–355.
- France, C.A.M., Owsley, D.W., Hayek, L.C., 2014b. Stable isotope indicators of provenance and demographics in 18th and 19th century North Americans. *J. Archaeol. Sci.* 42, 356–366.
- Fratzl, P., Gupta, H.S., Paschalis, E.P., Roschger, P., 2004. Structure and mechanical quality of the collagen-mineral nano-composite in bone. *J. Mater. Chem.* 14, 2115–2123.
- Halcrow, S.E., Rooney, J., Beavan, N., Gordon, K.C., Tayles, N., Gray, A., 2014. Assessing Raman spectroscopy as a prescreening tool for the selection of archaeological bone for stable isotope analysis. *PLoS One* 9, 1–9.
- Harbeck, M., Grupe, G., 2009. Experimental chemical degradation compared to natural diagenetic alteration of collagen: implications for collagen quality indicators for stable isotope analysis. *Archaeol. Anthropol. Sci.* 1, 43–57.
- Hare, P.E., 1980. Organic geochemistry of bone and its relation to the survival of bone in the natural environment. In: Behrensmeyer, A.K., Hill, A.P. (Eds.), *Fossils in the Making: Vertebrate Taphonomy and Paleoecology*. The University of Chicago Press, Chicago, pp. 208–219.
- Jorkov, M.L.S., Heinemeier, J., Lynnerup, N., 2007. Evaluating bone collagen extraction methods for stable isotope analysis in dietary studies. *J. Archaeol. Sci.* 34, 1824–1829.
- Killick, D., 2015. The awkward adolescence of archaeological science. *J. Archaeol. Sci.* 56, 242–247.
- van Klinken, G.J., 1999. Bone collagen quality indicators for palaeodietary and radiocarbon measurements. *J. Archaeol. Sci.* 26, 687–695.
- Longin, R., 1971. New method of collagen extraction for radiocarbon dating. *Nature* 230, 241–242.
- McNulty, T., Calkins, A., Ostrom, P., Ganghi, H., Gottfried, M., Martin, L., Gage, D., 2002. Stable isotope values of bone organic matter: artificial diagenesis experiments and paleoecology of Natural Trap Cave, Wyoming. *PALAIOS* 17, 36–49.
- Morris, M.D., Mandair, G.S., 2011. Raman assessment of bone quality. *Clin. Orthop. Relat. Res.* 469, 2160–2169.
- Nelson, B.K., DeNiro, M.J., Schoeninger, M.J., DePaolo, D.J., Hare, P.E., 1986. Effects of diagenesis on strontium, carbon, nitrogen, and oxygen concentration and isotopic composition of bone. *Geochim. Cosmochim. Acta* 50, 1941–1949.
- Olszta, M.J., Cheng, X., Jee, S.S., Kumar, R., Kim, Y.-Y., Kaufman, M.J., Douglas, E.P., Gower, L.B., 2007. Bone structure and formation: a new perspective. *Mater. Sci. Eng. R Rep.* 58, 77–116.
- Person, A., Bocherens, H., Mariotti, A., Renard, M., 1996. Diagenetic evolution and experimental heating of bone phosphate. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 126, 135–149.
- Pestle, W.J., Ahmad, F., Vesper, B.J., Cordell, G.A., Colvard, M.D., 2014. Ancient bone collagen assessment by hand-held vibrational spectroscopy. *J. Archaeol. Sci.* 42, 381–389.
- Pestle, W.J., Brennan, V., Sierra, R.L., Smith, E.K., Vesper, B.J., Cordell, G.A., Colvard,

- M.D., 2015. Hand-held Raman spectroscopy as a pre-screening tool for archaeological bone. *J. Archaeol. Sci.* 58, 113–120.
- Shoulders, M.D., Raines, R.T., 2009. Collagen structure and stability. *Annu. Rev. Biochem.* 78, 929–958.
- Tütken, T., Vennemann, T.W., Pfretzschner, H.-U., 2008. Early diagenesis of bone and tooth apatite in fluvial and marine settings: constraints from combined oxygen isotope, nitrogen and REE analysis. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 266, 254–268.
- van Klinken, G.J., Hedges, R.E.M., 1995. Experiments on collagen-humic interactions: speed of humic uptake, and effects of diverse chemical treatments. *J. Archaeol. Sci.* 22, 263–270.
- Veis, A., 2003. Mineralization in organic matrix frameworks. In: Dove, P.M., DeYoreo, J.J., Weiner, S. (Eds.), *Reviews in Mineralogy and Geochemistry. Biomineralization* 54. The Mineralogical Society of America, Washington, pp. 249–289.
- Weiner, S., Wagner, H.D., 1998. The material bone: structure-mechanical function relations. *Annu. Rev. Mater. Sci.* 28, 271–298.