

Referee #2

I would like to thank the referee for the helpful comments. The referee expressed some concern about the comparability of original XANES studies. This concern relates to measurement variations, which can be caused by particle size effects, self-adsorption, detector response rates and different reference materials. The referee further expressed concern that potential errors may propagate in subsequent comparisons. However, these concerns are not justified in view of the following four facts:

- (i) All results gave no indication for significant error propagation; and the referee's comments do not provide any indication that measurement variations adversely affected results. Results of the systematic comparison show striking similarities among spectra belonging to the same species and statistically significant differences between spectra of different species. Moreover, these statistically significant differences were related to the specific thermodynamic stabilities of the different Ca-P species. Assuming that there was exceptionally high measurement variability and subsequent error propagation, these striking similarities and statistically significant differences would certainly not occur.*
- (ii) Results and conclusions, obtained by the use of the normalization method, are in line with results and conclusions of the original studies.*
- (iii) To assess the degree of uncertainty in original studies is the objective of a systematic comparison. Thus, the above-stated measurement variations have been the reasons for this systematic comparison. The inclusion of measurement variations was consequently not a disadvantage, but on the contrary the prerequisite and objective of this study. To completely present measurement variations in original studies as adversely affecting a systematic comparison is therefore incomprehensible.*
- (iv) Virtually all reference materials used by the original studies were certified reference materials or independently confirmed reference materials. XANES studies are characterized by detailed pre-analysis of standards due to the always-restricted beam time at synchrotron facilities.*

Specific comments. The author's response is given in italics.

1) It would be difficult for readers of this manuscript to apply normalization procedures from the brief description given. For example, use of terms such as "unit edge jump" and "equal species" add confusion because they are not clearly defined. Also, when it is said that main (white line) peaks were adjusted to 2150 eV and tertiary peaks to ca. 2167 eV, it is unclear what exactly was done (Page 18730 Section 2.2). It sounds like the positions of both peaks were moved. Certainly the overall position of spectra may need to be moved due to differences in energy calibration between different studies. However, adjusting the position of multiple peaks in the same spectra to specific positions would greatly alter the original data. Subtracting multiple baseline segments

from data altered to shift multiple peaks is troubling. Perhaps this is not what was done, but again the writing of this section is not detailed enough. Other examples of vague description of the method - What does “ca. 2167” specifically mean? What does “normalized to a step of one” mean?

Response 1: The term ‘unit edge jump’ is a common term in XANES. Several web search engines return numerous links to XANES publications and figures using this term and this kind of normalization. Further, the manuscript introduction includes a brief introduction to the theory, which also explains the jump in the edge. Moreover, the manuscript introduction begins by citing a reference (p. 18726 L27-28: ‘for an introduction to the fundamental theory see e.g. Newville, 2004’), which includes a simple step-by-step description of this normalization procedure. Thus, care was taken to provide the unfamiliar reader with high quality references and practical analysis guides, which include simple step-by-step descriptions for commonly applied normalization procedures. The same facts apply to comment #9.

The phrase “normalized to a step of one” (p.18730 L18) was substituted with “normalized to an edge step of one”. This phrase is commonly used in XANES. Newville (2004) is now cited at the end of this sentence (p.18730 L18) to direct the reader, once again, to a high quality introduction to XANES for detailed and practical step-by-step descriptions of this specific normalization procedure. The phrase “spectra of equal species” (p 18730 L16) was replaced by “spectra belonging to the same P species”. The phrase “among equal species” (p 18730 L20,26) was replaced by “for each species”.

Furthermore, the referee states “However, adjusting the position of multiple peaks in the same spectra to specific positions would greatly alter the original data”. The adjustment was conducted by normalization. Normalization procedures are conducted by virtually all XANES studies. For example, the applications in the software packages, which are mentioned by the referee in comment #2 and #5 and which are commonly used, greatly alter original data using a series of sequentially applied procedures. Normalization procedures are conducted to alter raw data in order to facilitate data comparison.

To energy-normalize the spectra, most distant features were used, that is, the white-line peak and the oxygen oscillation. The adjustment of white-line peaks to 2150 mV was straightforward and very precise due to relatively thin and intense peaks (Fig. 1a and 3; therefore ‘ca.’ was omitted). For a few oxygen oscillations the adjustment was less precise due to specific spectral shapes. For example, the selected sample spectrum (see HAP-4 in Figure 1a) indicates a flattened peak of the oxygen oscillation. In case of less precisely definable center positions of peaks, the normalization was optimized by also matching the secondary peak of such a spectrum to the range of secondary peaks of spectra belonging to the same species (see secondary peak of HAP-4 in Figure 1a). However, the normalization of a spectrum to the given intensities was always conducted by means of a suitable normalization factor, which was applied to the investigated spectral range. In summary, the adjustment of oxygen oscillations was still precise (see Figure 1a) but the adjustment of white line peaks was more precise (Figure 1a). The text was modified accordingly.

2) It needs to be clearly demonstrated that such procedures are vastly superior to currently available spectral processing and identification routines. ATHENA and SIXPACK are powerful and freely downloadable spectral processing and comparison software for XANES spectra. In contrast, a potential user of the methods presented in this paper would be faced with writing their own code to process the data based on unclear and rather vague instructions. It may be possible to directly compare standards from a database to unknowns using ATHENA and SIXPACK to achieve nearly identical results. A direct comparison between approaches in deconvoluting mixtures of known composition (see comment 5 below) would help convince the reader that the methods presented in this paper are worth adopting.

Response 2: Comments #2 and #5 are similar and suggest a comparison of the proposed method with ATHENA and SIXPack. Hence, the following reply addresses both comments:

Both comment #2 and #5 suggest that the proposed method of using peak height ratios for the identification of calcium phosphate species has to be ‘vastly superior’ to spectral processing and identification routines of ATHENA and SIXPack. First of all it needs to be stated that these are software packages encompassing the entire range of XAS analysis and not methods (the proposed method could also be implemented). Thus, it remains unclear what the referee is referring to. Which of the several routines (or combinations of these routines) are to be compared with the proposed method?

However, to be able to compare P k-edge XANES spectra of different studies the proposed normalization method was specifically developed because, to the best of my knowledge and in agreement with comment #1, the mentioned processing routines are not yet developed for this purpose. From this it follows that a reliable comparison of methods/routines is not yet possible without implementing additional processing routines in the software packages. Moreover, the development of suitable processing routines would likely follow an approach very similar to that proposed in the current manuscript. Apart from this, even if other tools for data handling could be used, this would neither affect the new concept described in the manuscript (the normalization procedure for cross-study comparisons; the method of using peak height ratios) nor the results and the conclusions of this study.

Moreover, there were two aims of this study and these were clearly defined: (i) To establish a spectral library that encompasses a compiled suite of P K-edge XANES spectra of various Ca-P species published in the literature; and (ii) to conduct a systematic comparison of spectra included in this library in order to identify diagnostic spectral features (p.18728 L 24-27). Hence, it was not intended to develop a ‘rival method’ or to criticize other methods. Yet, the opposite is the case. For example, some authors determined OCP contents of complex samples by linear combination fitting (which is implemented in the above-mentioned software packages) while other authors speculated that pure OCP spectra may not be distinguishable from pure apatite spectra. Hence, the specificity of XANES detected in this study indicates that linear combination fitting approaches are likely more reliable than it might be assumed and therefore supports these approaches (and related software packages).

In summary, the proposed method follows a different objective than other currently used methods. Hence, these methods should be used in a complementary and not in a competitive way. For example, the findings of this study may well be helpful to optimize fitting approaches. As suggested in comment #5, it would be therefore interesting to figure out how different XAS methods can be used to complement each other. However, I would like to point out that the study aims were defined and achieved; and that the manuscript describes a self-contained study, which is of great interest (see the introduction of the first referee). A detailed comparison of methods (or the evaluation or refinement of software packages) was not intended, would go beyond the scope of this work and, thus, belongs to a separate study.

3) Using a data set constructed from various publications is potentially problematic as it may contain artifacts leading to erroneous conclusions. For example, white line intensities relative to tertiary peak intensities could be affected by detector response rates. Slight overload of the detector at the white line could strongly affect ratios of different peaks at different energies. Furthermore, self-adsorption can strongly affect the intensity of certain spectral regions. This concern is addressed in the manuscript by saying most studies claim to reduce self adsorption (Page 18732 Section 3.1). Even small differences in particle size used between different studies are likely to lead to spectral differences especially at higher energies in samples of the same phosphorus phase. The effect of self-adsorption is probably non-linear over the energy range of a spectrum. Finally, identification of different phosphate phases, especially different apatite minerals, is difficult. Unless the composition of the different forms of apatite used in the data compilations were independently verified, it reduces confidence in the conclusions drawn from this study. Similarity of spectra, as is invoked in the manuscript, does not prove that the standards used were of the same composition. Natural specimens of calcium phosphate minerals are often misidentified by mineral dealers and in mineral collections. Such errors can propagate through the various studies used in this compilation.

Response 3: The referee expresses some concern about the comparability of original studies due to measurement variability in the original studies. However, to interpret measurement inaccuracies in original studies as adversely affecting systematic comparisons, which by definition allow for heterogeneity in original studies, is incomprehensible. Moreover, in view of the referee's suggestion, it would be interesting to know how the detected systematic relationships between peak height ratios and thermodynamic stabilities (e.g. Figure 4), the statistical evidence for these relationships and the congruence with findings of other studies can be explained by error propagation and poor reproducibility. However, the comment does not refer in any way to any result of the current study. Hence, instead of providing any indication that measurement variations adversely affected results the comment independently confirms that the prerequisites for systematic comparisons were met. The compilation of libraries was assumed to be representative of common measurement variations (p.18729 L5-8). Hence, each point in the list of reasons for measurement variability, which I do not call into question, is indeed a proof of the necessity and correctness of the approach. The systematic comparison indicated distinctive spectral features despite measurement variations mentioned in the comment.

It needs to be also emphasized that some studies developed standard libraries without measuring any complex sample, such as the comprehensive compilation of P k-edge XANES spectra of mineral standards by Ingall et al. (2010). The goal of these studies is to detect distinctive spectral features of certain P species in order to facilitate their identification in subsequent studies. Assuming XANES measurements would not be reproducible during subsequent studies, as implied by the comment, those libraries would be unsuitable. However, those libraries are widely recognized and the current systematic comparison highlights their importance.

Further, comment #3 explains in detail known issues of self-absorption and that these issues may adversely affect spectral comparisons. Indeed, most studies describe measures taken to reduce self-absorption during FY measurements. Moreover, TEY measurements prevent self-absorption. Spectra of FY and TEY measurements were compared in Figure 4 and results of these measurements were comparable. Hence, this study showed that using diluted or finely grounded and thinly mounted samples during FY measurements efficiently mitigates self-absorption effects (p. 18739 L17-20).

The referee also expresses concern about incorrect identification of natural specimens in the original studies. However, 69% of the standards in the compiled library were synthesized standards (including numerous certified reference materials of chemical supply companies; Table 1); and 31% were natural specimens (see Table 1). As a large number of spectra of both synthetic and natural specimens were available, the compiled library enabled a comparison of these two types of standards. For example, differences between natural and certified, synthetic apatites were discussed in Sect. 3.4 (18737 L14-21). Further, the referee states: “Natural specimens of calcium phosphate minerals are often misidentified by mineral dealers and in mineral collections”. No further information is given on the general procedure and therefore this statement implies that these specimens were obtained and then analyzed at synchrotron facilities without further verification. As a general rule, however, natural specimens were independently verified by other methods prior to XANES analysis due to restricted beam time. For example, 67% of the spectra of natural standards were drawn from a large spectral library of natural specimens (Ingall et al., 2010; see Table 1) and all of these specimens were independently confirmed. Hence, the suggestion that XANES studies could depend on the identification of standards by mineral dealers is certainly incorrect.

4) It seems that spectra that did fit the models were excluded from the data compilations for a particular species. “A few spectra of the reference library were excluded from the averaging in Fig. 2a due to a limited energy range (Beauchemin et al., 2003) or inconsistency with common species specific spectra.” Inconsistent spectra could be misidentified calcium phosphates in the original papers or the few samples that were dissimilar to the others could be correct ones.

Response 4: The citation in the comment omitted the cross-reference “(see next section)” from the original sentence. The reasons for the exclusion of a few spectra were explained in detail in the next section (p. 18734 L7-10; p. 18734 L17 – p. 18735 L6). The sentence was revised: ‘A few spectra of the reference library were excluded from the averaging in Fig. 2a due to a limited

energy range (Beauchimin et al., 2003) or inconsistency with common species specific spectra (see explanations in Sect. 3.3).'

To address the referee's concern: Firstly, the spectra of Beauchimin et al. (2003) were not excluded due to dissimilarities, but due to the limited energy range (described in detail: p.18729 L10-14). Secondly, XANES studies generally obtain DCPD from chemical supply companies and, hence, these are certified reference materials. It was also explained why the exclusion of the two atypical DCPD spectra was possible without difficulty (Sect. 3.3; p. 18734 L7-10). Thirdly, two ACP spectra were excluded because of a likely OCP contamination (see very detailed discussion about this likely OCP contamination; Sect. 3.3; p. 18734 L17 – p. 18735 L6). The detection of this OCP contamination was acknowledged as an important finding of this study (see comments of the first referee). If the assumption in the comment above would be true (i) the two apparently OCP-contaminated ACP samples would be correctly synthesized calcium phosphates, which is very unlikely (cf. plots for OCP and ACP in Fig. 3; see also Eveborn et al., 2009); (ii) the other reliable ACP spectra of reliable ACP standards (certified ACP reference material of a chemical supply company; ATR-FTIR/XRPD-confirmed ACP standard verified to be not contaminated by OCP; Eveborn et al., 2009) would be misidentified calcium phosphates, which is almost impossible. That these two very unlikely events occur at the same time is even more unlikely. The latter example is already the one with the highest uncertainty because the DCPD samples cannot be misidentified calcium phosphates (see above).

5) In order to really convince a reader of the utility of the spectral normalization techniques presented in this paper, characterization of phosphorus in samples of precisely known composition is necessary. While the techniques are applied to samples of calcined bones, the wording of the section (3.4) indicates the composition of these samples was not precisely known. It would be more useful to show that the presented techniques are either superior to ATHENA and SIXPACK software in apportioning calcium phosphorus minerals in mixtures of known composition or show how these software packages can be used in tandem with the described normalization procedure. I believe one of the papers by Ajiboye cited in this manuscript contains spectra of known mixtures that would make an excellent test case.

Response 5: See response to comment 3

6) The results and conclusions could also be strengthened if a theoretical basis for the difference in peak height ratios was offered.

Response 6: I agree with this suggestion. Several authors observed that post-edge shoulders were more well-defined for species with higher thermodynamic stability and the current study substantiates this observation. The detected differences in peak height ratios, which were acknowledged in this comment, parallel those earlier observations: The relative secondary peak height was larger for species with higher thermodynamic stability. Averaged relative secondary peak heights were larger for FAP, CFAP/CHAP, HAP (synthetic, natural and poorly crystalline),

TCP and OCP than for DCPD, DCPA/MCP and ACP (Table 2). This is an important finding and the manuscript will be revised to include this finding.

7) Errors or ranges of values should be presented for secondary peak height % in table 2. This would give the reader a better idea of the strength of the findings.

Response 7: The peak height ratios are now presented together with the standard deviation for all Ca-P species in Table 2 to address the referee's suggestion (see revised Table 2). Moreover, the statistical differences between peak height ratios for all Ca-P species are now given in Table 2. Table 2 was revised to provide a probability matrix that reports different probability levels for all possible combinations of the different Ca-P species.

8) The figures presenting spectra are extremely hard to read. In general there are so many overlapping spectra presented with extremely fine lines that it is impossible to readily distinguish different features or phases.

Response 8: Spectra of different phases are generally shown in separate plots (Fig. 3; Fig 4). It thus follows that most spectra of different phases do not overlap. Further, it is possible to distinguish the different spectral features mentioned in the body of the text. The different secondary peak heights for several P phases, the distinctive secondary peak positions for DCPDs and the absent post-edge shoulders for DCPAs and MCPs are immediately visible in Figure 3, to mention just a few examples. Hence, it is generally not necessary, but optional, to zoom in for more detail. The different width of the lines in Figure 4 was adjusted to account for the relevance of the plotted spectra. Hence, the grey lines (which indicate the raw data for the model spectra; blue lines) are thin. However, Figures 2 and 3 were revised (increased width of some lines and additional colour discrimination) to further improve these Figures (see revised Figures).

9) Much of the standard data out there in the literature is presented as figures that are not useful for spectral comparisons. The author should be commended for scanning and digitizing many of these figures. If such raw spectral data for all standards analyzed were made available as a supplement this paper would be potentially much more useful even with the problems listed above. Furthermore, if a detailed step-by-step description of their normalization procedures with a sequence of clear, simple figures would be helpful to readers as a supplement. Also inclusion of the computer code used for normalization with detailed comment lines inserted in the code may be helpful. The techniques presented in this paper may be step forward in the interpretation of XANES spectra but the issues above need to be thoughtfully addressed to truly convince a reader that it is worthwhile to apply them. In terms of the analysis of natural samples, this method may be limited to only samples dominated by calcium phosphate phases.

Response 9: The inclusion of all raw spectral data would likely not be justified because the inclusion of published data is an unnecessary repetition. Further, used information of studies that did not provide these data is permanently stored, all references are provided and the methodology of digitizing plots is described. Further, it may be generally considered to be improper to present raw spectral data of original studies in a subsequent systematic comparison,

whether the data were published or not. Further, the comment suggests including even more simple step-by-step descriptions of the normalization methods. As this comment renews a suggestion in comment #1, I would also like to refer to the corresponding reply. The introduction cites a reference (p. 18726 L27-28: Newville, 2004), which includes simple step-by-step descriptions of normalization procedures used in XANES. The revised version of the manuscript will include another citation of this reference in the method section to direct the unfamiliar reader, once again, to those step-by-step descriptions (see reply #1). Moreover, Fig. 1 includes an example of the employed normalization procedure, which shows a clear and simple step-by-step description of the normalization procedure. In reply to the next suggestion of this comment, the manuscript will be revised to include the information that the digitized data were imported into a spreadsheet application and subsequently normalized by appropriate normalization factors using this spreadsheet application. A computer code was not required.

The referee also suggests that the proposed method may be limited to samples dominated by calcium phosphates. However, the manuscript includes an example for the identification of calcium phosphates in sediment (spectra of Brandes et al., 2007). This sample was surface marine sediment from an oxic basin in the Effingham Inlet (Brandes et al., 2007). Such samples generally contain relatively high contents of aluminium- and iron-bound P, and less Ca-P. Brandes et al. (2007) concluded that XANES will be capable to identify and quantify minor and exotic P phases in complex samples, provided that a comprehensive library is available. The current study established such a library and identified important diagnostic spectral features in this library.

References

- Beauchemin, S., Hesterberg, D., Chou, J., Beauchemin, M., Simard, R. R., and Sayers, D. E.: Speciation of phosphorus in phosphorus-enriched agricultural soils using X-ray absorption near-edge structure spectroscopy and chemical fractionation, *J. Environ. Qual.*, 32, 1809-1819, 2003.
- Brandes, J. A., Ingall, E., and Paterson, D.: Characterization of minerals and organic phosphorus species in marine sediments using soft X-ray fluorescence spectromicroscopy, *Mar. Chem.*, 103, 250-265, 2007.
- Eveborn, D., Gustafsson, J. P., Hesterberg, D., and Hillier, S.: XANES speciation of P in environmental samples: An assessment of filter media for on-site wastewater treatment, *Envir. Sci. Tech.*, 43, 6515-6521, 2009.
- Ingall, E. D., Brandes, J. A., Diaz, J. M., de Jonge, M. D., Paterson, D., McNulty, I. et al.: Phosphorus K-edge XANES spectroscopy of mineral standards, *J. Synchrotron Rad.*, 18, 189-197, 2011.
- Newville, M.: *Fundamentals of XAFS revision 1.7*, Consortium for Advanced Radiation Sources, University of Chicago, Chicago, 2004.

Revised Table 2. Summary of diagnostic P K-edge XANES features for various Ca-P species.

Ca-P species	Other distinctive features			Secondary peak height (%)													
	Post edge shoulder	Post edge shoulder height	Oxygen osc.	Secondary peak height ¹ % (n)	SD	CFAP CHAP	HAP Σ	HAP synth.	HAP nat.	TCP	OCP	DCPD	DCPA MCP	ACP	P on CaCO ₃	HAP p. cryst. +am.	
FAP	broad	very high	narrow	8.7 (8)	8.5	1.1	N	**	***	N	*	****	****	****	****	***	***
CFAP CHAP	broad	very high	narrow	8.1 (5)	8.0	1.3		N	**	N	N	****	****	***	**	**	**
HAP Σ	broad	high	narrow	7.1 (17)	7.0	1.4			N	N	N	****	****	****	****	***	**
HAP synth.	broad	high	narrow	6.6 (7)	6.3	0.7			**	N	**	****	****	****	**	**	***
HAP nat.	broad	high	narrow	8.6 (5)	8.5	1.3				N	****	***	***	**	**	*	
TCP	broad	medium	medium	6.3 (4)	5.9	1.6					**	*	*	*	N	N	
OCP	broad	medium	narrow	4.2 (8)	4.0	0.5						N \surd	N \surd	****	*	N \surd	
DCPD	mixed	low	medium	3.8 (7) ²	3.7	1.0							N \surd ?	*	N \surd ?	N \surd ?	
DCPA MCP	missing	n/a	broad	2.6 (16) ³	2.5	2.8								N \surd	N	N	
ACP	narrow	medium	broad	1.6 (2)	1.6	0.1									N	**	
P on CaCO ₃	mixed	low	broad	2.4 (2)	2.4	1.5										N	
HAP p. cryst.	narrow	low	broad	4.2 (1)	n/a	n/a											
HAP am.	broad	low	narrow	3.8 (1) ²	n/a	n/a											

¹ Averaged relative secondary peak height (% of white-line peak height). Left column: averaged secondary peak height (%) of individual spectra and number of averaged spectra in parentheses. Right column: secondary peak height (%) of model spectra (averages for multiple curves; Figure 4)

² Secondary peak shifted towards main peak.

³ Includes four spectra, which lacked a secondary peak (see high standard deviation).

N: Difference of secondary peak heights (%; population means) is non-significant (Two Sample Independent t-Test)

*: Difference of secondary peak heights (%) is significant at the 0.05 level

**: Difference of secondary peak heights (%) is significant at the 0.01 level

***: Difference of secondary peak heights (%) is significant at the 0.001 level

****: Difference of secondary peak heights (%) is significant at the 0.0001 level

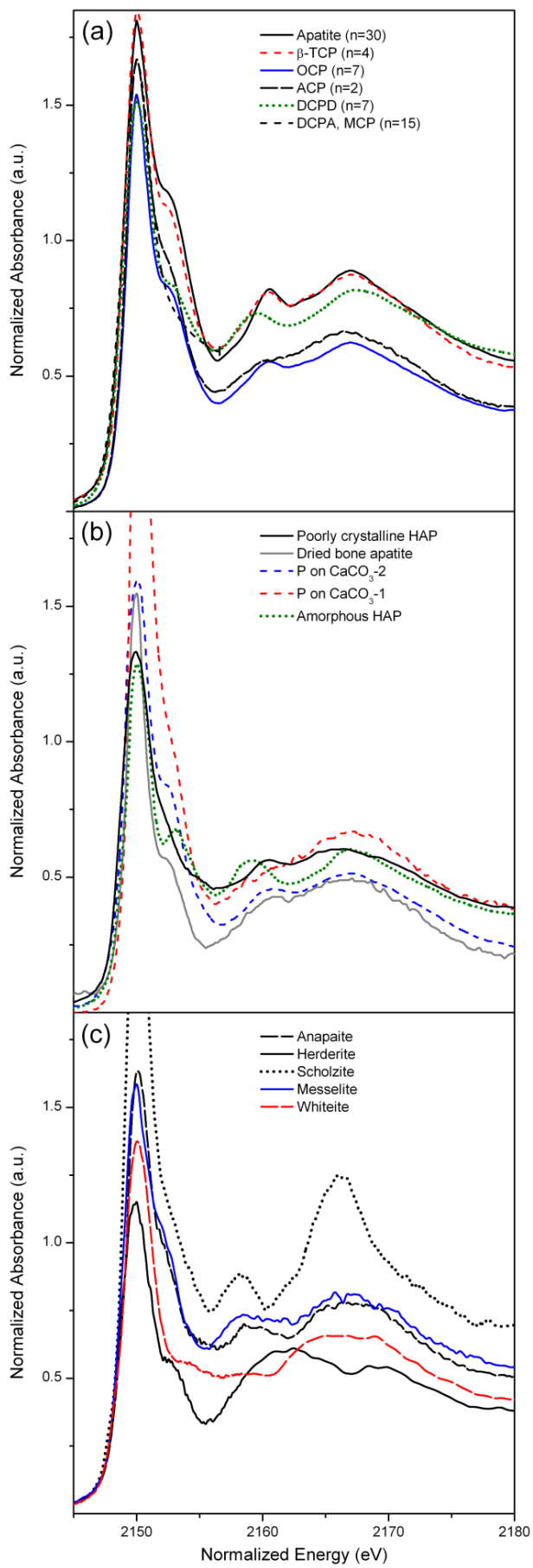
N \surd : Denotes other distinctive features for two species, which cannot be distinguished by their secondary peak heights (note that other distinctive features are not denoted for pairs with significantly different secondary peak heights). OCP distinguishable from DCPD (secondary peak shift or missing post-edge shoulder), DCPA (missing post-edge shoulder), MCP (missing post-edge shoulder), poorly crystalline HAP (missing post-edge shoulder, broad oxygen oscillation) and amorphous HAP (secondary peak shift). ACP distinguishable from DCPD/MCP (missing post-edge shoulder).

N \surd ?: Denotes other potentially distinctive features for two species, which cannot be distinguished by their secondary peak heights. DCPD likely (note the two atypical DCPDs in Figure 3) distinguishable from DCPA (absent secondary peak shift), P on CaCO₃ (absent secondary peak shift, broad oxygen oscillation) and poorly crystalline HAP (absent secondary peak shift)

Grey: Note that P on CaCO₃ was probably contaminated by crystalline Ca-P phases (see text) and that poorly crystalline HAP and amorphous HAP require further verification by additional spectra.

Abbreviations: n/a (not applicable); SD (standard deviation); oxygen osc. (oxygen oscillation); HAP synth. (synthetic HAP); HAP nat. (natural HAP); HAP p. cryst. (poorly crystalline HAP); HAP am. (amorphous HAP)

Revised Figure 2:



Revised Figure 3:

