Quaternary Geochronology 16 (2013) 183–197

Contents lists available at SciVerse ScienceDirect

Quaternary Geochronology

journal homepage: www.elsevier.com/locate/quageo

Research paper

Results from an amino acid racemization inter-laboratory proficiency study; design and performance evaluation

Joanne Powell [a,*], Matthew J. Collins [a], James Cussens [c], Norman MacLeod [d], Kirsty E.H. Penkman [b]

[a] BioArCh, Department of Archaeology, University of York, Wentworth Way, Heslington, York YO10 5DD, UK
[b] BioArCh, Department of Chemistry, University of York, Heslington, York YO10 5DD, UK
[c] YCCSA, Department of Computer Science, University of York, Heslington, York YO10 5GH, UK
[d] Department of Palaeontology, Natural History Museum, Cromwell Road, London SW7 5BD, UK

ARTICLE INFO

Article history:
Received 14 November 2011
Received in revised form
31 October 2012
Accepted 1 November 2012
Available online 27 November 2012

Keywords:
Amino acid racemization
Geochronology
Inter-laboratory comparison
Proficiency test
Accuracy
Precision
Bias
Measurement uncertainty

ABSTRACT

It is nearly thirty years since the last inter-laboratory study was carried out for amino acid racemization (AAR) analysis using powdered fossil material (Wehmiller 1984). Since then there have been major changes in sample preparation and instrumentation, and it was considered timely to coordinate a new inter-laboratory study in support of current methodologies. In 2010, two such studies were undertaken. The first of these, coordinated by Wehmiller (this edition), used homogeneous hydrolysates of Pleistocene mollusc and eggshell materials and focused on the agreement of analytical measurements between laboratories, without interference from differing sample preparation procedures. The second (this study) was designed specifically as a proficiency test. Unlike previous inter-laboratory comparisons that have focussed on precision estimates, the purpose of this study was to carry out an evaluation of measurement bias by comparing the measurement results of laboratories carrying out their routine methods, including extraction, against the consensus values. Participants were sent one dried sample of a mixed amino acid standards solution and five homogeneous powders: two Pleistocene mollusc test materials prepared from material (ILC-A) supplied and used by Wehmiller in previous inter-laboratory studies (1984; and this edition), one Pleistocene opercula test material from the terrestrial gastropod, *Bithynia tentaculata*, and two heat-treated modern ostrich eggshell test materials. Results from this study demonstrate that whilst individual laboratory precision may be excellent, suggesting good control of random error influences (less than 1% for replicate measurements by some individual laboratories), agreement between methods, or even between laboratories carrying out the same method, may be very different. Trueness evaluation (determined as the relative percentage bias) reveals the extent of the disagreement reflected by the inter-laboratory variability. Individual laboratory *D*/*L* value biases of 10–30% or more when compared to the consensus values are not uncommon. We demonstrate why bias contributions should also be included in AAR uncertainty estimation and recommend that the preparation of defined reference materials are seen as a priority in order to control and correct for systematic error influences in the analytical system.

1. Introduction

The last 30 years have seen significant changes in amino acid racemization (AAR) analysis. Early research based on ion-exchange liquid chromatography (IEx) was able to separate L-isoleucine from its diastereomer D-alloisoleucine, yielding a D-Aile/L-Ile value, or often termed A/I value. As methods developed, it became possible to detect and measure increasing numbers of chiral pairs of amino acids, from six or seven using gas chromatography (GC) to ten or more routinely determined today using reverse-phase HPLC (RP). These developments have continued to advance its application in routine analysis. AAR now requires mg sample sizes, is relatively fast and with inexpensive preparation and analytical costs, is a useful dating method with the potential to provide age estimates that cover the entire Quaternary (Wehmiller and Miller, 2000).

However, the last 30 years have also seen significant changes in the determination of measurement uncertainty and the introduction of widely accepted international guidelines for its evaluation (ISO/IEC 98, 1993, JCGM 100, 2008; ISO 21748, 2010) in an attempt to harmonise procedures and avoid the use of ad hoc methodologies (Lira, 2002). These principles recognise that both precision and

* Corresponding author. Tel.: +44 (0) 1904 328806.
E-mail addresses: jp588@york.ac.uk, jozone@btinternet.com (J. Powell).

1871-1014/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.quageo.2012.11.001

trueness (expressed as the analytical bias) are essential components of measurement accuracy, and considerations of both are necessary for the proper reporting of measurement uncertainty. Previous inter-laboratory studies have tended to focus only on the comparison of precision data (Bada et al., 1979; Kvenvolden, 1980; McCartan et al., 1982; Wehmiller, 1984). However, inter-laboratory studies also provide a unique opportunity to compare results and evaluate bias. To date, there has been no formalised inter-laboratory evaluation of trueness, but with the changes in AAR analytical methodology that have taken place, it is appropriate for the extent of agreement between the different methods/laboratories to be evaluated.

1.1. Accuracy or precision?

The accuracy of any measurement result is influenced by both random and systematic error effects, evaluated as measurement (im)precision and measurement bias respectively (Thompson, 2000). Both precision and bias are essential elements of measurement uncertainty, and usually determined during method validation before the method is brought into routine use (Barwick and Ellison, 2000; Thompson et al., 2002).

Precision characteristics for target matrices and concentration/value ranges are defined under repeatability and reproducibility conditions (Barwick and Ellison, 2000; ISO 21748, 2010). Repeated measurements on suitable materials might include commercially available reference materials (i.e. certified reference materials; CRMs) or in-house reference materials of sufficient quantity, homogeneity and stability (Thompson et al., 2002). Whilst knowledge of precision estimates for in-house standard solutions is an important aspect of internal quality control, analysis of simple solutions free from matrix effects are not necessarily representative of the precision of solid matrix bound analytes. Consequently, their use to derive uncertainty values for routine samples risks underestimation.

Bias determination however, involves the repeated analysis of a matrix-matched CRM or other suitably defined reference material (Thompson et al., 2002). Unless the method is empirical and by definition makes no correction (JCGM 100, 2008; EURACHEM/CITAC, 2000), significant systematic error influences should always be corrected for, or included with the precision estimate to reflect the overall doubt or the uncertainty associated with a measurement result (Barwick and Ellison, 2000; EURACHEM/CITAC, 2000; JCGM 100, 2008). Currently, the absence of defined AAR reference materials makes bias evaluation challenging. For this reason, AAR uncertainty estimation currently only focuses on the precision of analytical results.

For AAR geochronology and aminostratigraphic studies, it is the relative differences between the *D*/*L* values, analysed within a single laboratory, which are most important. Therefore ensuring internal consistency within an individual laboratory is often all that is required and measurement precision becomes the principle concern. However, even precision estimates will vary depending on sample type and analytical conditions, and require further qualification to enable direct comparability (see supplementary information).

Nonetheless, the inability to evaluate laboratory and method bias routinely has important implications for analytical accuracy and the proper reporting of the measurement uncertainty. In the absence of comparable materials, the only alternative means of evaluating bias may be through cooperation between laboratories and comparability against other analytical data.

1.2. Previous AAR inter-laboratory studies

Several authors have previously observed important inter-laboratory and method related differences in *D*/*L* values from previous comparability studies (Bada et al., 1979; Kvenvolden, 1980; McCartan et al., 1982; Wehmiller, 1984; Hollin and Hearty, 1990; Bakeman, 2006; Wehmiller (this edition)). Early inter-laboratory comparisons focused on gas chromatography (GC) method variations (Bada et al., 1979) with ion exchange liquid chromatography (IEx) also being used for isoleucine determination (Kvenvolden, 1980; Wehmiller, 1984). In contrast, reverse-phase HPLC (RP) is the method more commonly used today. In Wehmiller's original study, eleven laboratories (using three GC methods and one IEx method) were each given six different materials to analyse: three marine mollusc shell powders (inter-laboratory comparison materials or ILC A, B and C) and their respective desalted hydrolysates. Performance evaluation was carried out by a qualitative comparison of *D*/*L* CV values (coefficient of variation or relative standard deviation expressed as a percentage) achieved by each laboratory. For example, for alanine, aspartic acid and glutamic acid, precision estimates ranged between 3–8%, for leucine and phenylalanine, 5–10% and for isoleucine, proline and valine, between 10–18%. Wehmiller (1984) reports that whilst CVs for powders did not indicate significant differences from that of the hydrolysates, the median CV from all the results of 9.6% for powdered samples and 6.5% for liquid samples, were higher than the 2–5% typically reported by an individual laboratory. Further, it was observed that significant differences between laboratories' results could lead to 25% differences in estimated age. As a result, Wehmiller called for the need for reference standards in routine analysis to ensure comparability more than twenty-five years ago. More recent intra-laboratory studies (Bakeman, 2006; Bakeman and Wehmiller, 2006) reported a 0.4% bias between GC and RP *D*/*L* values for aspartic acid (with RP giving the higher readings); while for D-alloisoleucine/L-isoleucine a 6.8% higher systematic offset was observed for GC compared to RP, and 1.9% compared to IEx. A further 4.6% difference was observed between GC and RP for glutamic acid *D*/*L*, with as large as 25% for valine *D*/*L*, with RP giving the higher readings in both cases.

Clearly there are noticeable discrepancies between the closeness of the intra-laboratory precision estimates achievable, and the comparability of data between different methods and/or laboratories. This strongly suggests the presence of additional uncertainty contributions, due to unaccounted-for bias arising from analytical differences between methods and/or laboratories. For this reason, AAR dating is predominantly currently carried out by laboratories independently from each other, precluding direct comparison of *D*/*L* data.

1.3. Trueness (bias) determination

CRMs or other defined standard reference materials are frequently used for calibration as they eliminate laboratory, method and even run bias (Thompson, 2000), thus correcting analytical results for systematic error. However, traceability back to standard materials with reference values and known uncertainty is currently impossible in AAR geochronology. Wehmiller's original ILC powders are used routinely by some laboratories for internal quality control. However, issues regarding method and laboratory bias have made defining reference values and their use for calibration, thus far problematic.

In the absence of reference values, comparability against other analytical data may be the only remaining option. This may be an intra-laboratory comparison against data determined using a published or reference method, an inter-laboratory comparison such as a collaborative trial, or results from proficiency tests (Thompson et al., 2002). A method specific inter-laboratory collaborative trial eliminates method bias and incorporates the individual laboratory bias components into the between-laboratory precision estimate. It is thus designed to evaluate both repeatability and overall precision, expressed as the reproducibility of a method (Horwitz, 1995; ISO