Fresenius J Anal Chem (2001) 370 : 194–199

SPECIAL ISSUE PAPER

B. King · S. Westwood

GC–FID as a primary method for establishing the purity of organic CRMs used for drugs in sport analysis

Received: 12 October 2000 / Revised: 27 December 2000 / Accepted: 3 January 2001

Abstract The National Analytical Reference Laboratory has synthesized and characterized 67 anabolic steroid marker metabolites, both unlabelled and deuterated, and 37 key glucuronide and sulfate steroid conjugate pure substance reference materials. Work is also in process to establish their full traceability so that they can be issued as certified and primary reference materials. Both identity and purity have been rigorously characterized using a number of techniques and a primary method for purity assessment developed, based gas chromatography combined with flame ionization detection for the parent steroids and HPLC with evaporative light scattering detection for non-volatile steroid conjugates. Strategies for establishing traceability and for estimating measurement uncertainty are reported. The strategies described are considered applicable to a wide range of organic pure substance reference materials.

Introduction

"The spirit of relentless competition that reigns in elite sport, the glory of victory, and its attendant financial and social advantages, can push athletes to employ all possible means to improve their performance". The above quote from an International Olympic Committee (IOC) study report [1] succinctly describes the driver for the growing problem of drug abuse in sport. To ensure equal chances for all competitors, and for medical and ethical reasons, the IOC have banned a number of substances, methods, and manipulations. This in turn requires an extensive monitoring program.

B. King (✉) · S. Westwood
National Analytical Reference Laboratory,
PO Box 385, Pymble, NSW 2073, Australia
e-mail: bernard.king@agal.gov.au
e-mail: steven.westwood@agal.gov.au

During the Sydney Olympics in September 2000, the Australian Sports Drug Testing Laboratory (ASDTL) was in the eye of the storm, where it was responsible for monitoring competing athletes for banned performance-enhancing drugs. Sophisticated regimes are employed to cheat and equally sophisticated chemical measurement is required to ensure that cheats are detected, while the innocent are not falsely accused.

Prohibited substances include stimulants, narcotics, anabolic agents, diuretics and peptide hormones. ASDTL and other IOC laboratories have developed methods for detecting the drugs and their metabolites so that drug abuse can be detected many weeks after the drugs have been taken. One of the main problem areas is the analysis of anabolic steroids. The GC–MS methods [2] rely on the availability of over one hundred pure substance reference materials which have being synthesized and are being certified by the National Analytical Reference Laboratory (NARL). These include 67 parent steroids and their metabolites and deuterates as well as 37 glucuronide and sulfate conjugates of these materials. The latter comprise the major form in which anabolic steroids are found in urine samples [3]. The types of reference material of interest are illustrated by the metabolic pathway for testosterone shown in Fig. 1. Figure 1 also illustrates the high level of structural detail that it is necessary to address in this area. Steroids and their marker metabolite RMs are required for method validation, for calibration of measurements designed to detect low levels of substances associated with synthetic anabolic steroid abuse, for operational quality control, and as positive controls in confirmation analyses. In addition, deuterated steroid conjugate RMs are required as internal standards for GC–MS analysis of endogenous steroid profiles, which are used to detect drug abuse. For example, the presence of a testosterone to epitestosterone ratio greater than 6 to 1 is strongly indicative of drug abuse and constitutes an offence. The potential damage to the Olympic ideal and to diplomatic and trade relations that could result from any failure in this program is incalculable. Hence the effort devoted to ensuring that the measurements

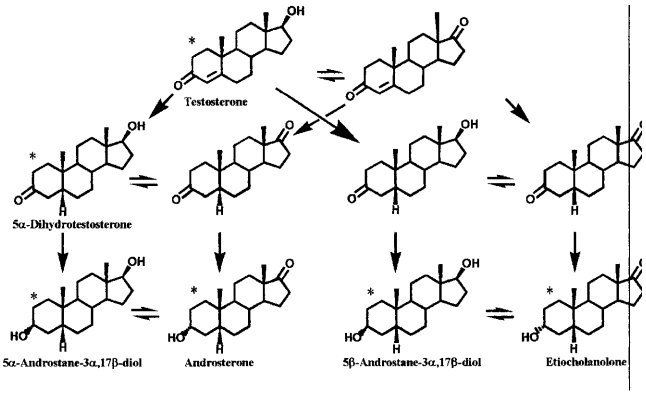

Fig. 1 Metabolic pathway for testosterone. Compounds marked * have been prepared as deuterated CRMs

and the reference materials on which they depend are valid.

Developments in metrology are leading to robust strategies for establishing the traceability of chemical measurements to SI, with full uncertainty budgets. Pure substance and matrix-certified reference materials (CRM) are important tools in the process. They help analysts establish their traceability and validity. CRMs must by definition [4] be traceable to the units in which the property value is expressed. Although pure substance CRMs are normally characterized in terms of their purity, a dimensionless property, the underlying unit of measurement is the mole. The mole is defined in terms of "specified entities" or identity and relative "number of entities" or amount. Thus the requirement is to establish in a traceable manner the identity of the material and to assign the amount of substance present in the form of a purity value. To do this it is necessary to establish the traceability of the measurements used for the CRM characterization.

The normal metrological way of establishing traceability to SI is to use a primary method. Sadly however, few of the primary methods identified by CCQM [5] are readily applicable to the characterization of pure substance organic compounds. On the other hand chromatography is extensively used and it is, therefore, important to evaluate the degree to which it can be considered a primary method. A primary method [4] is one capable of operation at the highest metrological level, which can be completely described and for which a complete uncertainty budget can be prepared in SI units. Gas chromatography with flame-ionization detection (GC–FID) has been extensively used within NARL for the characterization of organic CRMs over many years and more recently its metrological pedigree has been established. It will be argued that at an appropriate level of uncertainty GC–FID can form the basis of a primary method.

This paper reports the strategy being used by NARL to certify both the identity and the purity of a range of steroid materials. The basic strategy is also used for other types of CRM, such as forensic drugs and agrochemicals, and is considered to be applicable to a wide range of organic materials.

Impurities in a CRM can interfere when they are used in subsequent analysis and thus need to be kept to a minimum. As a working target a purity of at least 99% was sought and generally achievement of this level of purity was not a problem.

The uncertainty associated with purity will be transferred to any analysis where the CRM is used. The target was based on an assessment that the uncertainty associated with the CRM should not contribute more than one third of the overall uncertainty of a subsequent measurement. As a guide for NARL work, the following target relative measurement uncertainty values [U (k2)] were sought: CRMs for use as a standard for assay work U = 0.03–0.3%; CRMs for use as a standard for primary measurements of the material in a matrix U = 0.2–1.0%; CRMs for use as a standard for routine measurements of the material in a matrix U = 1.0–5.0%. Generally uncertainties close to the bottom end of the range were readily achieved. However, a constraint was the small amount of material synthesized, typically 0.5–1.0 g. This limited the range of techniques that could be used and the size of sample that could be employed for characterization. This in turn impacted on the uncertainty of the purity values obtained. In practice, the best possible material with the smallest feasible uncertainty was prepared, without resource to inappropriately long or costly procedures

Characterization of identity

As indicated above, establishing the identity of steroids and their metabolites is a key aspect of the work. There is a need for well characterized CRMs to help ensure that the correct identity is established in routine analysis of these analytes. The aim was to establish the identity of the CRMs with an uncertainty that is for practical purposes zero, by employing three types of information, namely: information about the synthetic route, structural analysis from first principles, and comparison with reference data.

The synthetic route used to prepare the candidate material was reviewed by an expert synthetic chemist to confirm that the desired product was expected to result from the stated starting materials and reaction sequence. A strong literature precedent for the correct outcome of all reactions crucial to the establishment of the correct stereochemical framework of the material was considered important and reference to this information is included in the final certification report for the compound.

The identity of all materials was independently confirmed by analysis. The characterization was made from data obtained by use of a combination of the key techniques:

– MS, IR, and $^{1}H/^{13}C$ NMR;
– GC and/or HPLC;
– TLC;
– measurement of melting point or boiling point; and
– C, H, N microanalysis.

When the physical, spectroscopic, and chromatographic characterization of the material had been published in detail, in peer-reviewed literature, and independently synthesized samples were available for direct comparison studies, the structure could be readily verified. Materials whose physical and spectroscopic properties had not been fully reported or for which no direct comparison materials were available required a greater amount of spectroscopic evidence and skilled interpretation of data to establish a comparable degree of confidence in the structural assignment.

Identification was accepted as established if the GC–MS and IR spectra of the material matched published and reviewed reference spectra. The minimum criteria for establishing a match using a given identification technique were in line with criteria used by other workers [6]. If reference spectra were not available for both of these techniques then the ^{1}H NMR spectrum was also obtained and all three spectra were required to be consistent for unambiguous identification. The spectra for other NMR-active nuclei (e.g. ^{13}C, ^{2}D, ^{19}F, ^{31}P) were obtained if required to resolve ambiguities. At least one other corroborative test (e.g. melting point, boiling point, elemental analysis, TLC retention factor) was also undertaken and agreement with literature data established. Where a well-characterized sample of the compound was available from another source, the two materials were compared for the congruence of their GC–MS, HPLC, TLC, and melting point properties, as appropriate.

For some candidate materials comparison samples were not available or there were limited published reference data. In these cases structural identification was based on review of the synthetic sequence used to prepare the material together with evidence derived from first principles from the spectroscopic and physical properties of the compound.

When a limited amount of candidate material was available the minimum amount of information required for acceptable qualitative identification was the MS, IR, and ^{1}H NMR spectra, one chromatographic identification (GC, HPLC or TLC), and a consistent microelemental analysis.

Occasionally, when ambiguity still existed, further information was obtained by use of two-dimensional NMR experiments or X-ray crystallography. Where necessary characterization data were corrected for known impurities. The essential requirement was that *all* characterization data had to be consistent with the expected structure.

Purity assessment

The purity of a nominally pure substance can be established by either or a combination of direct assay of the material or measurement of *all* the impurities and subtracting these from 100%. The techniques that are, in varying ways, used within NARL, together with an indication of their type of use, are listed in Table 1. The assay approach has the advantage that it measures directly the property of interest. It is, however, limited to an uncertainty of ca 0.1%. The measurement of impurities approach can yield much smaller uncertainties, but runs the risk of serious error if significant components are missed. A combination of both approaches offers the best solution and is used by NARL. The assigned purity value was based on the impurity data, but cross-checked by assay.

Table 1 Techniques used for purity analysis by NARL

Technique	Assay	Impurities
Chromatography: GC, HPLC, TLC	✓	✓
NMR	✓	✓
Titrimetry	✓	×
DSC	×	✓
MS	✓	✓
TGA	×	✓
Karl Fischer	×	✓
IR	×	✓

GC–FID using capillary columns is a well tried and trusted technique as it offers good resolution of the impurities, good limits of detection and good dynamic range, so that the major component and the impurities can be determined in a single run. Although the technique can miss both very volatile and non-volatile impurities, it will be argued that GC–FID in combination with HPLC, NMR, DSC, and TGA to correct for impurities not detected by GC–FID, constitutes a primary method. The algorithms used to calculate the purity and the associated uncertainty are:

$$\text{Purity} = 100 - I_T\%$$

where $I_T = I_{GC} + I_{NR} + I_{ND} + I_{OT}$

$$U_{IT} = (U_{IGC}^2 + U_{INR}^2 + U_{IND}^2 + U_{IOT}^2)^{0.5}$$

and $U_P = U_{IT}$

I_T is the total amount of impurities and I_{GC} is the amount of impurities determined by GC–FID. The remaining three corrections that need to be applied (I_{NR}, I_{ND}, I_{OT}) take account of undetected impurities, namely:

– One or more impurities could be co-eluted with the main peak. A low level of impurity would lead to a small error in the area of the major peak, but the impurity peaks could be significantly underestimated. However, given the high resolution of capillary columns, the error can be expected to be small. Also any undetected error of this type would be revealed by analysis using other techniques. A correction (I_{NR}) was obtained by consideration of NMR, HPLC, MS, and other analytical data.
– If several impurities were to be present at less than the limit of detection for the GC–FID method, which was approximately 0.02%, they would not be detected but their combined effect could be significant. Correction (I_{ND}) was based on experience of the number of impurities typically detected and the expectation that a significant level of impurity of this type would be revealed by other techniques.
– GC–FID will not detect volatile impurities (solvents or water) or non-volatile materials. Such impurities (I_{OT}) were determined using TGA.

The uncertainties associated with the various issues are designated by U_{IT}, etc. U_P is the uncertainty associated with the purity estimate.

In detail, 1–2 µL of a solution of the candidate CRM was injected onto a suitable GC and the chromatogram obtained. The peak areas of the main peak and the impurity peaks were measured. The GC–FID impurities (% mole/mole and % weight/weight) were calculated as follows:

For a reference material 'a' containing 'n' impurities (b, c...n) the actual amount of impurities measurable by GC (I_{GC}) is defined as:

$$I_{GC} = \frac{\sum_{i=b}^{i=n} A_i F_i}{A_a \times F_a + \sum_{i=b}^{i=n} A_i F_i} \times F_H \times 100\% \quad (1)$$

where A_i is the integrated area of the GC–FID response for the individual impurity I expressed as area percent of total peak area, F_i is the GC–FID response factor for individual impurity i, A_a is the peak area for reference material 'a' expressed as percentage of total peak area, F_a is the GC–FID response factor for reference material 'a', and F_H is the correction factor for homogeneity of material, taken to be 1.0 but which will have an associated uncertainty.

The following assumptions are considered to be justified and enable the simplification of Eq. (1) to give Eq. (2)

1. The impurities are structurally very similar to the RM. That is why they could not be completely separated during the purification stage. Typically one to three impurities were detected at a total level of 0.1–0.5% w/w.
2. Because the level of impurities was low (≤ 0.5%), any errors in estimating their concentration and the associated uncertainty will have relatively little impact on the purity data.
3. It can be expected that the molecular weight of the impurities and the RM were all very similar. GC–MS analysis indicates that impurities were isomeric or differed from the RM by only a few hydrogen atoms or at most an oxygen (or nitrogen) atom. On the basis of three impurities, present in equal amounts, having differences in molecular weight from the RM of ±3, ±3 and ±18, it can be simply shown that the variation between the % purity calculated on a % mole/mole or % weight/weight basis is ±0.01%. If appropriate, an allowance for this effect can be included in the uncertainty budget, assuming a rectangular distribution for the variation in molecular weight (u_m = 0.00006). For practical purposes the % mole/mole and % weight/weight purity values are equal.
4. The concentration of a particular impurity (C %w/w) can be defined as: C = A .F; where A = peak area and F is the response factor covering all the GC and FID detector effects. Given that the impurities were similar the differences in behavior can be expected to be small and the response factor for all the impurities can be expressed as a single factor, F_I. The summation of the individual impurity peaks can be represented by A_I.

The above can be justified because the following apply:

- The FID response is proportional to the number of carbon atoms in the analyte and is relatively insensitive to small structural differences, particularly for a given class of compounds. Thus the response for the impurities in the RM can be expected to be very similar and this is supported by experience.
- The behavior of the RM and impurities in the GC process (injector and column) can be expected to be very similar for the reasons explained above.

Thus Eq. (1) can be simplified to:

$$I_{GC} = \frac{A_I \times F_I}{A_a \times F_a + A_I \times F_I} \times F_H \times 100\% \quad (2)$$

Given that $F_I \approx F_a$, because of the similarity of structures, by definition $A_a + A_I = 100\%$ and A_I is small relative to A_a, Eq. (2) can be further simplified by substituting F_a for F_I in the denominator as follows:

$$I_{GC} \cong \frac{A_I \times F_I}{(A_a + A_I) \times F_a} \times F_H \times 100\%$$

$$= \frac{A_I \times F_I}{100 \times F_a} \times F_H \times 100\%$$

$$= A_I \times \frac{F_I}{F_a} \times F_H \%$$

The above establishes the credentials of this approach as a primary method, because it can be completely described as required by the definition. However, small effects such as those discussed above and, for example, changes in the response factors with concentration, could contribute to the measurement process and in practice an empirical approach was adopted. The empirical equation was generated from calibration data. A calibration curve was prepared by plotting the impurity levels calculated assuming an F_I/F_a ratio of 1.0, versus the actual gravimetrically determined impurity levels. This data were obtained from analysis of solutions spiked with known concentrations of different impurities over a range of concentrations and using different major components. One hundred and one data points obtained over a three-year period were plotted and all fell on the same straight line. The data were obtained by different operators using different column types and operating conditions and thus includes components for reproducibility, repeatability, and ruggedness. The same GC instrument was used for all analyses. These combined data give a measure of the precision and bias inherent in the raw GC–FID data. It can be seen from the example given below that the difference between I'_{GC} and I_{GC} was less than 7% relative, justifying Eq. (2), the empirical approach, and the primary method claim. Further studies showed that the system response was linear over a concentration range of 1/2000 but with some curvature over the range of 1/5000, enabling cross checking of the purity by direct assay based on the measurement of the major GC–FID peak.

Linear regression analysis of this combined data gave an equation relating the observed response for the impurity A_I to the actual impurity level I'_{GC}. The empirically derived equation has the form:

$A_I = I'_{GC} \times 0.905 + 0.007$ which rearranges to: $I'_{GC} = (A_I – 0.007) \times 1.105$. The uncertainty $u_{I'GC}$ was calculated for a y on x regression as detailed in the Eurachem Guide [7]. Allowance was made within the regression analysis uncertainty calculations for the case where A_I is the mean value of repeat analyses.

The equations for uncertainty calculations, where the homogeneity factor F_H is taken as 1.0, are: $I_{GC} = I'_{GC}$ and

$$u_{GC} = I_{GC} \times \sqrt{\left(\frac{u_{I'GC}}{I'_{GC}}\right)^2 \pm \left(\frac{u_{FH}}{1.0}\right)^2}$$

The uncertainty in the homogeneity factor is the uncertainty associated with the value determined for a single sample, because of the heterogeneity of the material, and is assigned as the standard deviation of the repeat analyses used to obtain A_I.

Example: d_3-epitestosterone

The data for a real example of d_3-epitestosterone are:

1. Calculation of I'_{GC} and U_{GC}
   - $A_I = 0.260\%$ (mean of seven replicates)
   - $I'_{GC} = (A_I – 0.007) \times 1.105 = 0.279\%$
   - $u_{I'} = 0.011\%$ (from regression analysis uncertainty calculation)
   - $u_{FH} = 0.073\%$ (std. deviation of individual determinations)

$$u_{GC} = 0.279 \times \sqrt{\left(\frac{0.011}{0.279}\right)^2 + \left(\frac{0.073}{1.0}\right)^2} = 0.023$$

2. Calculation of I_{NR} and its associated uncertainty u_{NR}

There was no evidence of any I_{NR} from either GC using a different column or from HPLC, but it is prudent to make some allowance for these effects and an impurity is therefore assumed to be present at a value between zero and the HPLC limit of detection of 0.1%.

The uncertainty u_{NR} based on a triangular distribution of ±0.05% is $\frac{0.05}{\sqrt{6}} = 0.020\%$ and $I_{NR} = 0.05\%$.

3. Calculation of I_{ND} and its associated uncertainty, u_{ND}
   - A number of impurities present at less than the limit of detection could collectively be significant and it was assumed that there were three impurities present at less than the LOD of 0.02% and the data treated as triangular distributions. HPLC and NMR analysis indicated that impurity levels greater than the assumed level were unlikely. Thus $I_{ND} = 3 \times 0.01 = 0.03\%$ and $u_{ND} = 0.0071\%$

4. Calculation of I_{OT} and its associated uncertainty u_{OT}
   - The best estimate of impurities not detectable by GC–FID was obtained by thermogravimetric analysis (TGA). Invariably there was no evidence of any I_{OT} but it was considered prudent to make some allowance for these effects. The limit of detection for volatile impurities and non-volatile residues by TGA when only small sample sizes were available (5–10 mg of material per assay) was estimated as 0.1% w/w for both classes of impurity. Assuming a triangular distribution for each impurity gave $I_{OT} = 0.1\%$ and $u_{OT} = 0.029\%$

5. Calculation of I_T and its associated uncertainty u_T

$$I_T = I_{GC} + I_{NR} + I_{ND} + I_{OT} = (028 + 0.05 + 0.03 + 0.1)\% = 0.46\%$$

$$u_{I_r} = \sqrt{u_{GC}^2 + u_{NR}^2 + u_{ND}^2 + u_{OT}^2} = \sqrt{(0.023)^2 + (0.020)^2 + (0.007)^2 + (0.029)^2} = 0.0427$$

With a coverage factor of 2, which gives a level of confidence of approximately 95%,

$U_{IT} = 2 \times 0.0427 = 0.0854$

Rounding the data to one decimal place, the purity value for d_3-epitestosterone was obtained by subtracting the determined I_T from 100% and Purity=99.5 ±0.1% at a level of confidence of approximately 95%.

Stability

Accelerated stability studies indicate that the parent steroids are fully stable for at least two years at 4 °C when stored under an inert gas in sealed ampoules. As a precaution, however, materials are stored at –5 °C. If stability proves to be a problem in future NARL work, a strategy has been devised to allow for decreasing purity and increasing uncertainty, along the lines described by Pauwels [8].

Certification of CRMs

Identity was established using the procedure described above and then the data submitted to an expert external review panel. A material is only certified for identity when the review panel is satisfied with the material and the characterization data.

Purity certification for the parent steroids has to-date been based on GC–FID data, with cross-checks by estimating I_T by HPLC and ^{1}H NMR and assay by GC–MS. As explained above, the GC–FID method of estimating impurities has been shown to be the most accurate under the conditions employed and to yield the smallest uncertainties. This is not to say that the other methods could not be further developed. To-date, all analytical data have

been consistent within the levels of uncertainty, as illustrated by the data shown below:

– purity by GC–FID = 99.5%; U (k2) = 0.1
– purity by HPLC = 99.7%; U (k2) = 0.3
– purity by NMR = 99%; U (k2) = 1

In the event of data not agreeing, within the estimated levels of uncertainty, the strategy is to first try to identify the causes of the differences and to correct the results accordingly. Where unresolved differences exist, a hidden bias would be assumed to be the cause and a weighted mean of the data would be used as the basis of purity certification and estimation of the uncertainty, along the lines described in the VAM Guidelines [9].

As with identity, the purity data are submitted to expert external review and a material only certified when the panel is satisfied with the data.

The pure substance RM work of NARL has been audited by an international team of experts, to ISO 34 and ILAC G12:2000 requirements and the work has been accredited by the Australian accreditation body, NATA.

Conclusions

The process described has been devised to ensure that NARL CRMs can be absolutely relied upon by users. It is based on a balance between technical excellence and cost-effectiveness. The aim is to provide the best estimates, but where doubt exists, to underestimate, rather than over-estimate the characterization claims.

The process has also been developed as a demonstration of good metrological practice and it is claimed that GC–FID analysis, when combined with appropriate corrections, can be considered a primary method for the certification of a number of organic reference materials. Certainly we are not aware of suitable methods of higher metrological order.

The materials characterized by the reported process can be considered to be fully compliant with the definition of a CRM and to be primary reference materials.

References

1. European Commission Report EUR 19076
2. Olympic Movement Anti-Doping Code 1999.
3. Westwood S, Noble B, Moule C (2000) Proceedings of the 17th Cologne Workshop on Dope Analysis, Sport und Buch Strauß Köln:181–192
4. International Vocabulary of Basic and General terms in Metrology, ISO, Geneva, 2nd edn. (1993)
5. Kaarls R, Quinn JJ (1997) Metrologia 34:7–11
6. De Ruig W, Stephany R, Dijkstra G (1989) J Chromatogr 489:89
7. Eurachem/CITAC Guide: Quantifying Uncertainty in Analytical Measurement, 2nd edn. (2000)
8. Pauwels J (1998) Accred. Qual. Assur. 3:51–55
9. VAM Guidelines for the Production of Certified Reference Materials, part 4 – Certification, LGC, Teddington, UK, (1997)