Developments in accurate and traceable chemical measurements

Richard J. C. Brown* and Martin J. T. Milton

Received 20th March 2006
First published as an Advance Article on the web 9th February 2007
DOI: 10.1039/b507452p

In recent years there has been considerable interest in the application of the principles of measurement science to chemistry. This has led to the recognition of 'metrology in chemistry' as an area of relevance to analytical chemistry research. This *tutorial review* describes the benefits to chemistry of the implementation of the principles of measurement science and explains how they are able to improve the reliability and accuracy of chemical measurements.

1 Introduction

Metrology is the science of measurement. It is the underpinning study of how to make measurements and how to express and use the results. Although much of the infrastructure developed to support metrology was intended to underpin physical measurements, such as mass, length and time, it is now recognised that it can usefully be extended to encompass accurate chemical measurements.[1] These are needed to quality assure better and higher throughput automated analytical measurements, to underpin trade and industry, and to support regulation which requires increasingly low and more rigorously-enforced limit values.

Chemical measurements are made on an enormous number of species in a large range of matrices; this makes the development of fully validated measurement methodologies and suitable certified reference materials to support every possible measurement a nearly intractable task. Additionally it is often a feature of chemical measurements that the measurand (the quantity which is actually being measured) is difficult to define, for example, the 'amount' of DNA.[2] The study of how these measurements are made and their application is known as 'metrology in chemistry'.[3]

In this review, we highlight three generic properties of a measurement result that make it of value to a user. These are that it should be: stable if repeated at a later time, comparable with results from other laboratories and coherent with the results of measurements of the same quantity made with different methods. This review explores what can be done to ensure that chemical measurements have these three generic properties.

The review includes a discussion of the mole and the rationale for its inclusion amongst the base units of the International System (SI). We then discuss the benefits of carrying out measurements that are traceable to the mole and more widely to the SI in terms of ensuring that their results are comparable, stable and coherent; and how these objectives can be achieved.

Analytical Science Group, National Physical Laboratory, Teddington, Middlesex, UK TW11 0LW. E-mail: richard.brown@npl.co.uk; Fax: +44 (0)20 8614 0423; Tel: +44 (0)20 8943 6409

Richard J. C. Brown

Richard J. C. Brown was born in Epsom, UK, in 1975. He was educated at Epsom College, and Imperial College London, from where, in 1997, he received his BSc in Chemistry, together with the Governors' Prize for Chemistry. He remained at Imperial and was awarded his PhD, in Electrochemistry, in 2000. He then joined the Analytical Science Group at the National Physical Laboratory (NPL), in Teddington, UK, where his early work on ultra-low reflectance surfaces attracted widespread scientific and media attention as 'NPL Super Black'. His current research focuses on: measurements of the chemical composition of ambient particulate matter; the use of Surface Enhanced Raman Spectroscopy as a quantitative trace analysis technique, and primary pH measurements made using the Harned cell.

Martin J. T. Milton

Martin J. T. Milton was born in Eastbourne, UK. He received the BA degree in physics from Oxford University in 1981 and the PhD in 1991 as an external student of Southampton University. Since 1981 he has been with the National Physical Laboratory. Initially he worked on the development of laser sources for use in atmospheric remote sensing. Subsequently, his work has concerned the provision and comparisons of high accuracy standard gases and new methods for high accuracy analytical measurements including mass spectrometry. In 2003, he won the award for best paper from the Worshipful Company of Scientific Instrument Makers. He is a fellow of the International Union of Pure and Applied Chemistry (IUPAC) and the UK Institute of Physics.

2 Amount of substance and the mole

Metrology in chemistry is largely concerned with the measurement of 'amount of substance', which is the quantity used to characterise a number of entities.[4] Its use occurs largely, but not exclusively, in chemical studies. The name is sometimes shortened to 'amount' and, when the substance being measured is known, it can be made specific; for example 'amount of lead nitrate'.

Amount of substance is an extensive quantity (because its magnitude is proportional to the size of the system it describes). It is usually related to a measurement of a mass, which is also an extensive quantity. For example, it is related to the mass of a pure substance by the relative molecular mass.[5]

The SI unit of amount of substance (symbol n) is the mole, abbreviated to mol. It is one of the seven base units of the International System.[6] It is defined as the number of entities equal to the number of atoms in 0.012 kg of carbon-12 (which are understood to be unbound atoms, at rest and in their ground state), and therefore corresponds to the same number of entities as the Avogadro constant (N_A). The definition also states that when the mole is used, the elementary entities must be specified and may be "atoms, molecules, ions, electrons, other particles, or specified groups of such particles".[7]

It is not possible to realise the definition of the mole directly to any useful level of accuracy at present. That is to say, the mole cannot be realised as a physical artefact, like the prototype kilogram held at the BIPM in Paris, neither can it be defined accurately by a reproducible experimental method, for instance like the metre.[6] New approaches to define and realise the mole are currently being explored.[8,9] The absence of any direct method for realising the mole can cause some conceptual difficulties when making measurements of amount of substance. We will return to this in Section 4.

In practical usage, amount of substance is usually combined with other extensive quantities such as volume or mass to derive quantities that measure the composition of mixtures. These quantities are intensive because their magnitude is independent of the size of the system they describe.[10] Examples include amount concentration (amount of substance per unit volume) and amount content (amount of substance per unit mass). It is also combined with other amounts to form an amount ratio or fraction. These are defined as the ratio of the amount of a particular analyte in a sample to that of another in the sample (amount ratio), and the amount of a particular analyte in a sample to that of all constituents in the sample (amount fraction).

Table 1 lists nine quantities that are used to characterise composition. Although it appears that these quantities may be related by simple factors, such as density and relative molecular mass, there are important differences in how completely they define a sample. For example, all of the quantities described in Table 1 provide the complete information required to describe the composition of a sample when a single component is in a mixture of two or more components. The ratio quantities, which are the ratio of a property of a particular component in a mixture to the same property of another component in the mixture, do not provide complete information to describe the composition of a component in a mixture, when the mixture contains three or more components.

Table 1 Quantities used to characterise composition. The meaning of the symbols in the table is as follows: m_i, V_i, and n_i are the mass, volume (prior to mixing) and the amount of component i; Σm_j, ΣV_j and Σn_i are the sums of the masses, volumes (prior to mixing) and the amounts of all components; m_{solv} is the mass of the solvent only; m and V are the total mass and total volume (after mixing) of the mixture. As explained in the text, fractions describe how much of the total property of a sample is contributed by one of its constituent substances; concentrations describe the ratio of one extensive quantity of a single substance to the total volume of the mixture; and contents describe the ratio of one extensive quantity of a substance to the total mass of the mixture. Molality describes the amount of solute entities divided by the mass of the solvent. Adapted from[10]

Name	Symbol	Definition	SI unit
Mass fraction	w	$w_i = m_i/\Sigma m_j$	kg kg^{-1}
Volume fraction	φ	$\varphi_i = V_i/\Sigma V_j$	m^3 m^{-3}
Amount fraction	x	$x_i = n_i/\Sigma n_j$	mol mol^{-1}
Mass concentration	γ	$\gamma_i = m_i/V$	kg m^{-3}
Volume concentration	σ	$\sigma_i = V_i/V$	m^3 m^{-3}
Amount concentration	c	$c_i = n_i/V$	mol m^{-3}
Molality	b	$b_i = n_i/m_{solv}$	mol kg^{-1}
Volume content	κ	$\kappa_i = V_i/m$	m^3 kg^{-1}
Amount content	k	$k_i = n_i/m$	mol kg^{-1}

The quantity mass fraction has the property that it can be used to describe the composition of a component in a mixture when only knowledge of the mass of the mixture and the mass of the component are available. Amount content and molality require the same information together with the molar mass. The use of any of these quantities does not require detailed knowledge of the composition of other components within the mixture, nor are they sensitive to temperature or pressure. Therefore, these quantities are the most suitable for routine use in the accurate description of the composition of mixtures. In contrast quantities involving amount as a denominator require an exact knowledge of all components within the mixture.

Quantities involving volumes are more limited in their application because their magnitude is dependent on temperature and pressure (for gaseous mixtures). It should be noted that when gravimetric preparation is involved, the expression of composition with these quantities also requires knowledge of the density of the mixture to perform an accurate buoyancy correction.[11] Additionally, the use of volume fraction and volume concentration is discouraged for accurate work without reiteration of the full description of the quantity, since definitions vary as to whether the volume of the whole is measured before or after the mixing of individual components.[12]

The requirement to use amount of substance to form intensive quantities illustrates one of the principal reasons that the mole was adopted as a base unit of the SI. Although amount of substance is fundamentally a number, which could be used without any unit, it is necessary to include the mole as a part of the SI in order to highlight that it refers to an extensive quantity, which can be combined into a ratio with another extensive quantity to form an intensive quantity.[13] Since the units used for some of these quantities such as amount fraction and mass fraction are mol mol^{-1}, and kg kg^{-1}, respectively, they are often referred to as being 'dimensionless'.[14] More properly we should refer to these quantities as having the dimension 'one' since the expression for their units simplifies to unity (mol/mol = 1) rather than to zero. Indeed, it has been proposed that the numeral 1 should itself be regarded as an SI unit,[15] referred to by the term 'uno'.[16] These proposals have thus far not been widely adopted.[17]

The terms 'parts per million (ppm)', 'parts per billion (ppb)', and 'per cent (%)' are often used in place of units to express mass, volume or amount fractions. These terms are not units and simply represent the multipliers 10^{-6}, 10^{-9} and 10^{-2} respectively. Although mass, volume and amount fractions are quantities with the dimension one, they should always be accompanied by units, since the use of the correct unit for a measurement result conveys useful information about the quantity being expressed. The abbreviations 'ppm' and 'ppb' do not convey this information, and are not part of the SI system of units.[18] For example, it is correct to state that an amount fraction $w = 2.3 \times 10^{-6}$ mol mol^{-1} or 2.3 μmol mol^{-1}. Reference to an "amount fraction being 2.3 ppm" does not convey any specific information to confirm the units of the measurement result.

Moreover, the terms 'ppm' and 'ppb' are often incorrectly, and confusingly, used as units to describe concentrations. By its definition, a concentration does not have the dimension one (except for the rather obscure quantity volume concentration), therefore, the use of 'ppm' and 'ppb' in this context should also be avoided. Equally, the addition of extra labels to ppm and similar descriptors, such as in the case of 'ppmv' (parts per million by volume) should be strictly avoided. The term 'ppm' and other abbreviations should never be used in combination with other units. Since all concentrations are expressed per unit volume, 'ppmv' does not unambiguously distinguish between a mass concentration, amount concentration, number concentration or volume fraction, multiplied by 10^{-6}.

3 Traceability

3.1 Hierarchies of measurement

The principle that is most fundamental to the metrological approach described here is that of traceability†. It is defined[19] as "the property of a measurement result or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons ('the traceability chain'), all having stated uncertainties". The concept is often expressed using the adjective 'traceable'. Traceability is concerned primarily with establishing the relationship between a measurement result and a stated reference. It is an expression of the fact that a measurement result is not simply the statement of a value that exists in isolation. In principle, the stated reference may be any type of standard or a method that is specified precisely, but in order to bring additional value to a measurement result, the stated reference should be widely recognised and practically accessible. Hence traceability can only exist as a property of a measurement result with respect to a framework that establishes a specific 'stated reference'. Therefore traceability can, to some extent, be considered to be an *a priori* property of a measurement process because it depends on a pre-existing infrastructure. In practice it is difficult to establish such an infrastructure after the measurement has been made.

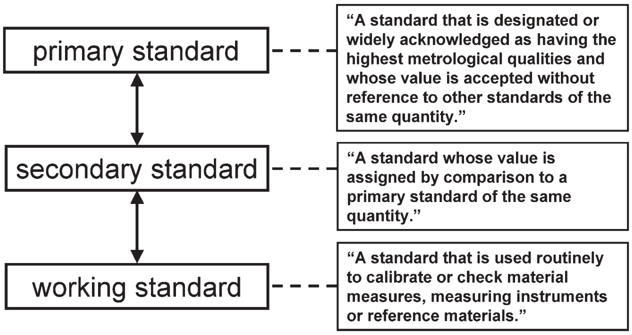

Fig. 1 A possible set of relationships between primary, secondary and working standards (with their definitions[19]) exemplifying a traceability chain.

The concept of traceability has often been exemplified by reference to the relationships within a rigid hierarchical structure. For example, a hierarchy linking primary, secondary and working standards is shown schematically in Fig. 1. A review of the definitions of the classifications of standards given in Fig. 1, shows that this approach to rationalising traceability is qualitative and explanatory, but does not provide sufficient information for it to be applied to a variety of hierarchies or for the relationships involved to be quantified.

In order to develop a more practical understanding of how traceability might work in chemical measurement it is useful to consider the mechanisms by which measurement results acquire the properties of stability, comparability and coherence, referred to in the introduction. Fig. 2a illustrates a straightforward hierarchy where two measurement results, A and B, have been produced by reference to the same standard. The measurement results are therefore each traceable to the same standard and consequently, they are comparable with each other. However, they have no wider comparability unless the standard is made more widely available. When repeated at a different time it is clear that this arrangement will lead to results that have similar stability properties to those of the standard.

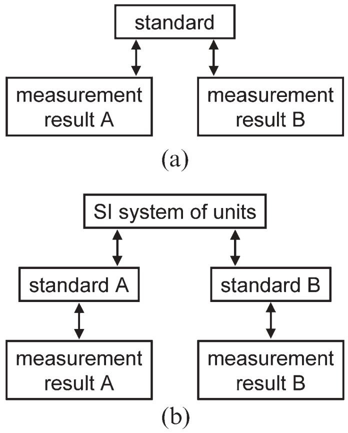

Fig. 2 (2a) The production of two measurement results using the same standard. (2b) The production of two measurement results using different standards, both traceable to the SI.

---

† It has been proposed that the term 'metrological traceability' be used instead of 'traceability' in the applications discussed here.

The 'metrological approach' to establishing comparability and stability depends on calibration with respect to a fully traceable calibration standard. Fig. 2b shows a hierarchy that is fully traceable to the SI. Two measurement results, A and B, have been produced by reference to two different standards, which are traceable to the SI system of units. (More complex diagrammatic representations of traceability chains are available.[20]) In this way the measurement results are traceable to their individual standards and then to the SI system and are therefore comparable with each other. In this case, the measurement results may be compared with other measurements taken elsewhere, or at different times, providing they are also made with standards traceable to the SI. By virtue of this link to the SI, these measurement results exhibit full comparability and will remain stable over time, provided they are always made with standards that are traceable to the SI. Additionally, there is no requirement for the same calibration standard to be used. For either measurement, a calibration standard of different content could have been used providing that it was fully traceable to the SI. This is ensured by the principle that the link between the calibration standards and the SI unit of the mole is invariant over time. In Section 6 we will discuss the technical and administrative infrastructure provided nationally and internationally to ensure this.

The principle of coherence is exemplified by the hierarchy shown in Fig. 3, which shows the analysis of the same sample by three different analytical methods. The method listed as 'gravimetry' can be thought of as referring indirectly to methods that analyse unknown samples by comparison with known, gravimetrically prepared standards or reference materials. Different methods have different traceability routes. For instance: 'gravimetric' methods have a traceability route which is entirely mass based (standards prepared for ICP-MS, ion chromatography, *etc.*); whilst coulometry relies on traceability to standards of both electric current and mass; and freezing point depression (FPD) has a temperature and mass based traceability reliance. Therefore despite these techniques being highly dissimilar in chemical terms, they all rely on a stable relationship between the amp, kelvin, and kilogram, to ensure their mutual accuracy and comparability—this is an example of coherence. They depend on this stable relationship between units which is governed by the agreed SI System of Units.[6]

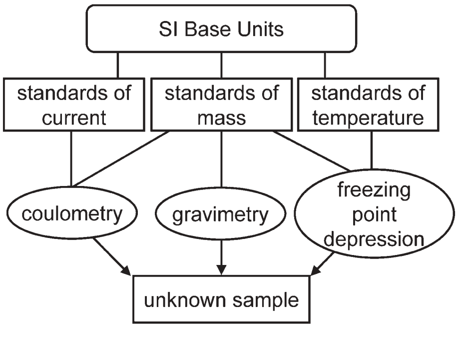

Fig. 3 The analysis of an unknown sample using different measurement methods.

3.2 Primary methods of measurement

In the case of measurements of amount of substance, where traceability to the mole is sought, there is a conceptual difficulty caused by the fact that the definition of the mole does not lead to a practical realisation.[21] Therefore it cannot be used physically in a comparison as required by the definition of traceability. The resolution of this difficultly lies in the use of primary methods of measurement. These are defined as methods having "the highest metrological qualities, whose operation can be completely described and understood, for which a complete uncertainty statement can be written down in terms of SI units, and whose results are, therefore, accepted without reference to a standard of the quantity being measured".[21] There is no formal process by which methods are reviewed and declared to be primary. This is because the definition of a primary measurement method must be interpreted in the context of a specific measurement requirement. This context provides the appropriate interpretation of 'highest metrological quality'. Primary methods, therefore, require no external calibration with standards or reference materials. They have measurement equations of the form

$$I = g(K,n) \tag{1}$$

where I is the measured output quantity from the method (for example mass or charge) in response to the amount of substance n. K is a parameter that does not depend on n, or on any other measurement of amount of substance.

Primary methods play an essential role in the practical realisation of the base units of the SI and in establishing traceability to the SI throughout metrology. They are the indispensable first step in the traceability chain since they do not rely on any other measurements of the same kind-of-quantity. Since primary methods allow quantities to be expressed in terms of a particular SI unit without reference to a standard or measurement already expressed in the same unit, they are therefore independent of any measurement of the same kind-of-quantity (as that being measured) and can only be dependent on measurements expressed in other SI units. By definition, the operation of a primary method is described completely by a measurement equation which leads to results that are unbiased and allows a rigorous calculation of the accompanying measurement uncertainty. In practice this means that a primary method allows SI units to be 'realised'; that is to say, it provides a means to transform the abstract definition of an SI unit into practical measurements made in terms of that unit.

3.2.1 Gravimetry. Gravimetry is probably the most common method used for providing measurements of amount of substance that are traceable to the SI. It relates amount of substance directly to mass by use of the relative molecular or atomic mass of the substance in question. The measurement is made with a balance that is calibrated with mass standards that are ultimately traceable to the international prototype of the kilogram. Using gravimetry it is routinely possible to produce mixtures with compositions with uncertainties of approximately 1 part in 10^4 when expressed in units of mass fraction (kg kg^{-1}). In order to convert this mass fraction into an amount fraction it is necessary to specify the chemical and isotopic purities of the material. The factor that limits the accuracy of gravimetric preparations is the requirement for pure materials. The uncertainties in the mass of the substance being weighed, and its purity, (for the solute and the solvent) both contribute to the overall uncertainty of gravimetric preparation. The uncertainty in the molar mass also contributes to the overall uncertainty, but is usually insignificant compared to the other contributions. Gravimetric preparation is used routinely for high accuracy analytical work, usually in the preparation of gaseous or liquid mixtures for use as primary standards. In the case of solution preparation, gravimetry can be up to an order of magnitude more accurate than preparation by volumetric methods. The requirement of gravimetry for pure materials depends on the availability of materials with their purity measured using a primary method. In the following sections we describe how freezing point depression and coulometry are used to achieve this.

3.2.2 Freezing point depression. Freezing point depression (FPD) belongs to a group of effects known as the 'colligative properties' (denoting 'depending on collection'). Colligative properties stem from the reduction of the chemical potential of a liquid solvent as a result of the presence of a solute. These phenomena can be used to calculate the effect of a solute on the freezing point of mixtures. The greater the amount of solute present, the greater the change in the observed freezing point of the liquid. The lowering of a liquid's chemical potential has a greater effect on the freezing point than on the boiling point, therefore freezing point depression measurements (cryoscopy) are more sensitive than boiling point elevation measurements (ebullioscopy). Notwithstanding this, freezing point depression is only used rarely as a primary method for amount of substance measurement. The technique is now mostly employed for applied analytical or comparative analysis where there is not such a need for high accuracy. Freezing point depression measurements require values for the latent heat of fusion of the material being measured (determined by differential scanning calorimetry, often with an uncertainty of less than a few parts in 10^3), and an accurate absolute determination of the depression of the freezing point. This requires traceability to standards of temperature and the overall uncertainty of the technique for amount of substance measurement is several parts in 10^3.

3.2.3 Coulometry. Coulometry (or coulometric titrimetry) involves the measurement of the number of electrons required to react with an unknown amount of an analyte.[22] It is essential to have some means of identifying the stoichiometric point, where equal amounts of analyte and charge have reacted with each other. This is known as the 'end point' of the titration. Specifically in coulometric titrimetry, the titrant is generated electrochemically and the amount of titrant is determined from measurements of the current and time with traceability to their appropriate SI units. Either constant potential or, more usually, constant current implements can be used. Coulometry has the advantage over titrimetry that it generates reactant *in situ* and is therefore not subject to the dilution effect found with volumetric titration.[23] The technique is also suitable for use with unstable analytes such as bromine and Ag^{2+}. The uncertainty of the technique is approximately 2 to 3 parts in 10^4. The technique is routinely used in National Measurement Institutes for purity analysis and for the certification of pure materials.

3.3 Primary ratio methods

Primary direct methods measure extensive quantities. However analytical chemistry is mainly concerned with the measurement of composition, which is represented by intensive quantities. Intensive quantities may be developed and maintained ratiometrically (*i.e.* without reference to any external standard) using primary ratio methods. These are methods that measure ratios of amounts of substance whose operation can be "completely described and understood". They can generally be considered to be methods that operate in two parts each of which is a primary 'direct' method.[24] If the first part can be described by:

$$I_1 = g(K,n_1) \qquad (2)$$

where I is the response of the measuring system, often a current, potential or intensity to an amount n_1, and the second can be described by an equation of the same form:

$$I_2 = g(K,n_2) \qquad (3)$$

The two parts of the method can be combined to give:

$$I_2/I_1 = G(K,n_2/n_1) \qquad (4)$$

Since this type of method also plays an important part in chemical metrology, a definition that encompasses the ratio as well as the direct implementation of the primary method of measurement has been agreed:

A primary measurement method for amount of substance is universally accepted as[21] "a primary method of measurement is a method having the highest metrological qualities, whose operation can be completely described and understood, for which a complete uncertainty statement can be written down in terms of SI units.

A primary direct method: measures the value of an unknown without reference to a standard of the same quantity.

A primary ratio method: measures the value of a ratio of an unknown to a standard of the same quantity; its operation must be completely described by a measurement equation."

3.3.1 Isotope dilution-mass spectrometry. Isotope dilution-mass spectrometry (IDMS) is recognised as a primary ratio method of measurement. It is a complex method that has been applied to the measurement of a wide range of organic and elemental analytes in various matrices. It works by combining isotope ratio mass spectrometry measurements of a sample with unknown amount content before and after the addition of a known amount of an isotopically enriched 'spike' material. The amount of substance in the original sample is then calculated solely from the measured isotope ratios and the amount of spike added. IDMS fulfils the definition of a primary ratio method because it is completely described by a measurement equation, but it cannot establish a link to the definition of the mole without the use of some other primary method. A primary direct method can be combined with a primary ratio method to produce measurements that retain their primary qualities (for example IDMS with a gravimetric assay of the pure spike). In this way, a measurement traceable to the SI can be made using a primary ratio method in combination with a reference of the same quantity that is itself traceable to the SI. The uncertainty of the IDMS method depends on the concentration of the unknown and the purity of available isotopic spike materials, and can be minimised using 'exact matching' techniques.[24] It can be as low as 3 parts in 10^4.

Much is written about the benefits of traceability, but its importance relates to generating measurement results that have the three key properties referred to earlier: comparability, stability and coherence.[25] This means that when chemical measurements adhere to the practice of implementing traceability their results should be stable and comparable with each other over time (for example, when measurements are repeated at one laboratory, possibly by different analysts) and at different locations (the same measurement made at a different laboratory) and with different measurement methods (ultimately traceable to different SI units), because they have all derived their measurement results through a chain of comparisons that link them to the SI.

3.4 Routine measurement methods

The vast majority of routine measurements of amount of substance are not made using primary measurement methods; they are made with methods that have measurement equations that are not completely described from first principles. Often, they have terms that have been determined empirically. Such measurement equations are sufficiently accurate for many applications in the laboratory, but their operation is dependent upon calibration with standards, or reference materials, of the quantity being measured. The calibration standards thereby determine the empirical terms in the measurement equation, such as the sensitivity of the detector—effectively determined from the slope of the calibration relationship. By analogy with the discussion of primary ratio methods above, these routine measurement methods are described by an empirical equation of the form:

$$I_1 = h(K,n_1) \tag{5}$$

where K represents environmental influences that may be significant, and the function h may not be characterised accurately. When combined with a second step, the equation for the ratio method is of the form

$$I_2/I_1 = H(K,n_2/n_1) \tag{6}$$

but, in this case, the behaviour of the function H or some of the terms within may not be entirely understood.

Such techniques may be thought of as 'secondary methods' since they are only able to provide full traceability when calibrated with standards that have 'primary' status (those produced by a primary method). Whilst the majority of chemical measurements will be made using these secondary methods, it is the primary measurement methods that ultimately provide traceability for the standards used to calibrate them.

An example of such a method is gas chromatography (GC) where the sensitivity of the GC instrument is not described in terms of a measurement equation, but it is known to be stable and linear over a specified range. Therefore, it cannot be a candidate primary ratio method, but it can be used to provide traceability in the manner described here. Other examples would include colourimetry, many mass spectrometry methods, ion chromatography and stripping voltammetry. As with other non-primary methods, GC is an instrumental technique whose response to an unknown sample is indicative of the amount, or amount content of an analyte within the sample. The sensitivity of the measuring instrument is calibrated with respect to a known standard. Therefore, such a measurement method relies on standards underpinned by primary gravimetric preparation. The accuracy of the GC method will depend, amongst other parameters, on the accuracy of the standards used for calibration, and on the repeatability of the measuring instrument.

One possible set of relationships between the primary direct and ratio methods described in this section is shown schematically in Fig. 4. It shows how the link between the SI and calibration standards is made by primary direct methods. Primary ratio methods are then used to make measurements of matrix reference materials that are traceable to these calibration standards.

3.5 A quantitative framework of traceability

At each stage in the hierarchy illustrated in Fig. 1, the relationship between a standard and the standard 'above and below' it is established by a comparison process that must operate according to a measurement equation that is "completely described and understood". Each comparison introduces some variability (or 'uncertainty') into the traceability chain. This necessarily means that as successively more comparisons are required to link any two standards, the traceability from one to the other becomes weaker, and more uncertain.

We now introduce a framework within which traceability can be quantified. In order to do this, we introduce the concept

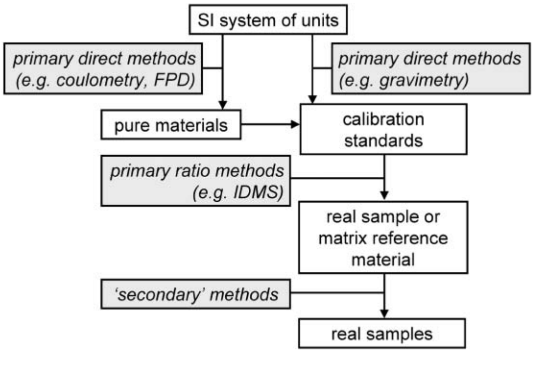

Fig. 4 The practical links between the SI system, primary methods, pure materials and routine calibration.

of the uncertainty of a measurement, which is a parameter defined as characterising "the dispersion of values that could reasonably be attributed to the measurand".[19]

Suppose, the comparison of an amount n_A with an amount n_B is described by the measurement equation:

$$n_A = n_A(n_B, K_B) \tag{7}$$

where K_B represents the influence of environmental parameters (for example temperature or pressure).

The sensitivity of the value of n_A to changes in the values of the quantities involved in the measurement equation are given by the partial derivatives $\frac{\partial n_A}{\partial n_B}$ and $\frac{\partial n_A}{\partial K_B}$. These partial derivatives are also the sensitivity coefficients used in the Guide to the Expression of Uncertainty in Measurement (GUM) uncertainty framework[26] to calculate the standard uncertainty of n_A arising from the measurement:

$$u(n_A)^2 = \left(\frac{\partial n_A}{\partial n_B}\right)^2 u(n_B)^2 + \left(\frac{\partial n_A}{\partial K_B}\right)^2 u(K_B)^2 \tag{8}$$

Where $u(n_A)$, $u(n_B)$ and $u(K_B)$ are the uncertainties in n_A, n_B and K_B respectively. We can also use these partial derivatives to illustrate how the concept of traceability can be interpreted quantitatively. For example, if $\frac{\partial n_A}{\partial n_B} = 0$, then the value n_A is insensitive to n_B and we can say that it is not traceable to it. A consequence of this is that if $\frac{\partial n_A}{\partial K_B} \neq 0$, then the value n_A is sensitive to the environmental influence parameter K_B and we must recognise that, at least to some extent, it is traceable to it.

This reasoning can be extended further by stating that if $\frac{\partial n_A}{\partial n_B} u(n_B)$ is very much larger than $\frac{\partial n_A}{\partial K_B} u(K_B)$ (and both are non-zero), then the influence of n_A on n_B is sufficiently strong as to make the influence of K_B irrelevant. In this case, n_A is only trivially traceable to K_B.

This is an example of a well-designed experiment in which the result is traceable to n_B. The opposite case, where $\frac{\partial n_A}{\partial n_B} u(n_B)$ is very much smaller than $\frac{\partial n_A}{\partial K_B} u(K_B)$ is an example of a less well designed experiment in which the measurement result is dominated by the influence of environmental parameters. In this case, traceability to n_B is not an important issue.

Now we consider an unbroken chain formed of two of these comparisons. Suppose that, in turn, the mass, n_B has been compared with n_C according to the measurement equation:

$$n_B = n_B(n_C, K_C) \tag{9}$$

hence, n_A can be related to n_C by

$$n_A = n_A(n_B(n_C, K_C), K_B) \tag{10}$$

And using the GUM uncertainty framework again, leads to:

$$\begin{aligned} u(n_A)^2 = {} & \left(\frac{\partial n_A}{\partial n_B}\right)^2 \left(\frac{\partial n_B}{\partial n_C}\right)^2 u(n_C)^2 + \\ & \left(\frac{\partial n_A}{\partial K_B}\right)^2 u(K_B)^2 + \left(\frac{\partial n_A}{\partial n_B}\right)^2 \left(\frac{\partial n_B}{\partial K_C}\right)^2 u(K_C)^2 \end{aligned} \tag{11}$$

We can now see that n_A is traceable to n_C. Although $u(n_B)$ does not appear in this equation, n_A is still traceable to n_B because $\frac{\partial n_A}{\partial n_B} \neq 0$. If n_A was not traceable to n_B, then $\frac{\partial n_A}{\partial n_B}$ would be zero. Hence, $\frac{\partial n_A}{\partial n_C}$ would also be zero, and $n_A$would not be traceable to n_C (as expected). Therefore measurement equations must be interpreted with some care when seeking to evaluate the origins of traceability, since some of the partial derivatives may have been substituted or cancelled.

Since it is difficult to confirm whether any specific measurement result is traceable to a specific standard (since it might not even appear in the uncertainty calculation), it is more convenient to use the convention of referring to measurement results as being 'traceable to the SI'. This reflects the fact that they have been measured within the framework of the SI, in the sense that all appropriate influence parameters are correctly traceable to the SI.

The increase in uncertainty along the 'chain' of comparisons required for traceability is an unavoidable by-product of the comparison processes needed for the dissemination of standards. It is the consequence of the process that makes the standard more widely available, and leads to an obvious feature of measurement standards that, as they are disseminated more widely, they become less accurate, or less strongly traceable to the stated reference. Consequently, when very small uncertainties and very high accuracies are required, for instance in the preparation or comparison of national standards, methods of the highest metrological quality will be used. Where it is not so important to have such low uncertainties, for example when performing routine field measurements, the standards used will have larger uncertainties, which are acceptable for the purpose of the measurement.

4 Methods of measurement

4.1 Measurement methods in chemistry

Prior to the definition of the primary method of measurement by the CCQM (the Consultative Committee for Amount of Substance—Metrology in Chemistry; whose role is explained in Section 6), measurements of amount of substance were often classified according to a hierarchy using principles and terminology similar to those applied in mass metrology. This approach is the same as that illustrated in Fig. 1 and denotes standards as being primary, secondary or working. Consequently, the use of the adjective 'primary' has become associated with methods at the top of such a hierarchy rather than being a reference to their operation in accordance with a measurement equation as specified in the definition in the previous section. Many other international organisations have established definitions of different types of measurement methods including the International Union of Pure and Applied Chemistry (IUPAC), the International Organization for Standardization's Committee on Reference Materials (ISO-REMCO) and the International Federation of Clinical Chemistry (IFCC). Each of these have developed definitions for use in specific applications and applied them for different purposes. For example, the IUPAC definitions of absolute, definitive, and reference measurements emphasize how the measurement is calibrated.

A 'definitive method' is defined as a method of "exceptional scientific status that is sufficiently accurate to stand alone in the determination of a given property for the certification of a reference material. Such a method must have a firm theoretical foundation so that systematic error is negligible relative to the intended use".[27,28] These methods are expected to require highly skilled personnel and be time consuming as well as expensive. A method of determination can be considered to be 'absolute' when the measurement equation that describes its operation contains only fundamental physical constants and universal quantities.[28] For such methods, any instrumental constants should be evaluated using theoretical principles. Absolute methods are considered to be of particular practical relevance when they offer an analysis result with superior accuracy and precision than methods based on calibration. A 'reference method', by definition,[28,29] has a small, estimated inaccuracy relative to the end use requirement. Reference methods of analysis may have an empirical basis but should have good accuracy, which should be demonstrated through direct comparison with a 'definitive method' or be based on the measurement of (certified or standard) reference materials.

An 'absolute method' (which uses only values of fundamental constants in the measurement equation) and a 'definitive method' (which may also incorporate well-known empirical constants) are similar to the 'primary method' described here, but neither makes reference to the use of SI units or to the traceability of the result. Similarly, schemes intended to establish a hierarchy of methods according to their uncertainty do not correspond exactly with the traceability approach advocated here. For example, the preference for the use of a definitive method over a reference method for labelling reference materials has no direct correspondence with whether the method fulfils the definition of a primary method of measurement.

The objective of the CCQM in defining the primary method of measurement, was to establish how measurements of amount of substance can be made that are traceable to the SI system. This objective is unique to the CCQM (see Section 6), so the definition of a primary method of measurement that it has developed does not necessarily correspond with definitions developed for other purposes. Consequently, there is no exact relationship between these different definitions.

4.2 Empirically defined measurands

Amount of substance measurements are often complicated by the fact that many of the measurands of greatest importance in chemistry are difficult to define precisely. For example, one technique may respond to the total concentration of an analyte whilst another technique may only be sensitive to a specific chemical state of the analyte or whether it is present in labile or strongly-bound complexes. In these cases the measurand cannot be fully defined without reference to the method by which the measurement is being made. Collectively, these measurands are known as 'empirically-defined measurands', or '(reference) method-defined measurands'. The methods used for determination of these measurands, often known as 'empirical methods of analysis', differ from 'reference methods', as defined above. They cannot be fully compared with a definitive method, or with a primary reference material, because, for empirical methods of analysis, it is the measurement method itself that defines the quantity being measured. For these measurements the uncertainty depends, not only on the repeatability and reproducibility of the measurement results, but also on how well the standard measurement method has been implemented. An example of an empirically defined measurand would be the polycyclic aromatic hydrocarbon (PAH) content of ambient air particulates collected on a filter measured following an extraction procedure. These PAHs are extracted by use of an organic solvent mixture at elevated temperatures and pressures, and then the extracts are analysed, perhaps by GC-MS. In this situation the measurement result would depend on the composition of the organic solvent mixture used, and the temperature, pressure and time of the extraction procedure used in the method. Individual PAHs will be extracted at a greater or lesser rate by different organic solvent mixtures, and at different temperatures and pressures. Another example is the measurement of the fat content of foodstuffs. This measurand is often referred to as the 'extractable' fat content, since the amount of fat removed from the foodstuff prior to its measurement depends on the method of chemical extraction used. Therefore the measurement process defines the measurand.

In the case of empirically defined measurands it is most likely to be the case that the measurement processes are not transparent or completely understood metrologically. In the future it is hoped that the concept of coherence can provide a basis for overcoming this type of difficulty.

5 Accuracy and traceability

In this review we have discussed how the principle of traceability underpins measurement results that are stable, comparable and coherent. In this section we consider the relationship between the traceability of a measurement result and its accuracy. These two concepts are sometimes confused, despite the fact that their definitions elaborate the difference between them. The accuracy of a measurement is defined as the "closeness of the agreement between the result of a measurement and the true value of the measurand".[19] The difference between the result and the true value is known as the 'error of measurement'. Accuracy (or 'trueness') is a qualitative concept. It is not the same as 'precision', which refers to the repeatability or reproducibility of measurement results ("the closeness of the agreement between successive measurements of the same measurand carried out under the same or changed conditions of measurement"[19]). Precision indicates the spread of the observed response of an instrument (or different instruments) as a result of variations of the measurement parameters from one measurement to another, it gives no indication of the agreement between the result of a measurement and the true value of the measurand, and therefore does not relate directly to accuracy.

One way to simplify the confusion sometimes caused by these terms is to recognise that the uncertainty of a measurement result has many components. These originate from the different input quantities within the measurement equation that describes the measurement process. The GUM describes the internationally accepted method of calculating and combining uncertainties.[26] The possible range of the measured value is bounded by the measurement uncertainty and will cover the 'true value' (the value that would be obtained from a perfect measurement) if full traceability has been established for the measurement procedure. Estimates of 'accuracy' incorrectly evaluated without reference to traceability actually only represent the precision. Similarly, the implementation of traceability does not necessarily ensure accuracy—it is possible for a measurement result to be fully traceable and appear to be less accurate than a measurement result that has not taken full account of traceability.

6 The organisation of international metrology

In order to make traceable chemical measurements, an infrastructure is required within which the traceability described above can exist. This infrastructure is usually overseen by a National Measurement Institute (NMI). As the requirements for standards have grown, many countries have increased the effectiveness of their dissemination activities by linking working standards to their NMI through chains of direct measurement comparisons. As these mechanisms have become more complex, quality and accreditation systems have been developed to ensure that they deliver the required levels of accuracy and traceability. One consequence of this approach has been the emergence of the concept of 'traceability to an organisation' *e.g.* traceability to a stated NMI. This is an unhelpful description as it obfuscates the fact that the measurement result should be traceable to the SI system of units; which are maintained globally and not solely at a particular organisation.

International measurement standards play an important role in metrology, they are ‟standards recognized by an international agreement to serve internationally as the basis for assigning values to other standards of the quantity concerned”.[19] They are closely linked to national measurement standards, defined as ‟a standard, often a primary standard, recognized by national law to serve in a country as the basis for assigning values to other standards of the quantity concerned”.[19] Although this infrastructure is well defined for physical measurements, there is a dearth of recognised chemical measurement standards. These primarily exist in the form of certified reference materials, such as mono-elemental solutions and environmental samples certified for their content of specific analytes.

The origin of the effort towards the international harmonisation of measurement standards was the signing by 17 nations of the Convention of the Metre in 1875. It is a diplomatic treaty which gives authority to the International Committee for Weights and Measures (Comité International des Poids et Mesures, CIPM) to act in matters of world metrology, particularly concerning the demand for measurement standards of ever increasing accuracy, range and diversity, and the need to demonstrate equivalence between national measurement standards. It established a permanent organizational structure for member governments to act in common accord on all matters relating to units of measurement.

The CCQM (Consultative Committee for Amount of Substance—Metrology in Chemistry),[30] is the part of this structure charged with dealing with matters relating to amount of substance and was established in 1993. Its present activities concern the development of primary methods for measuring amount of substance and the conduct of international comparisons (so called 'CCQM Key Comparisons'[31]), which lead to the establishment of international equivalence between national laboratories. The CCQM currently has 6 technical working groups concerned with different branches of chemical measurement: Electrochemical Analysis, Inorganic Analysis, Organic Analysis, Gas Analysis, Surface Analysis and Bioanalysis. These working groups discuss detailed technical matters, improvements to methodologies and procedures, and work towards improved international comparability by organising, and participating in measurement comparison exercises. A country's performance in these comparisons provides the technical basis for the Mutual Recognition Arrangements drawn up by the CIPM in 1999. This is a mechanism that is used to establish the comparability of national measurement standards maintained by the world's NMIs.

7 Summary and conclusions

In the last ten years there has been a considerable increase in interest in the application of the principles of measurement science to chemistry. This has led to the recognition of 'metrology in chemistry' as an area of importance in analytical chemistry. In this review we have shown how chemical measurement is concerned with quantities that measure composition. There are subtle and important differences between these quantities. It has been demonstrated that traceability is a central concept in chemical measurement, leading to the comparability, stability and coherence of measurements, and providing confidence in measurement results. It is increasingly becoming a requirement of international standards.[32,33] Whilst traceability and accuracy are related, this review has shown that they are not the same thing. This review has also discussed why primary measurement methods are needed in order to establish traceability to the SI, although other classifications of measurement methods are in use. Furthermore, the way in which the operation of a primary measurement method can be described mathematically and how the concept of traceability can be quantified has been demonstrated. The final part of the review has discussed the structure that exists both internationally and nationally to produce and disseminate reference materials and standards of the highest quality.

The requirements on chemistry and analytical science are becoming increasingly demanding as the boundaries of analysis are constantly being pushed downwards into the 'trace' and 'ultra-trace' regions.[34] The current challenges in analytical chemistry surround the need to quantify accurately and repeatably, analytes with mass fractions of 10^{-8} or less. The dominant source of uncertainty in very low level analytical measurements is often the systematic bias of the analytical methodology itself. This is usually because the different techniques are measuring subtly different measurands, because of either matrix effects or the different principles on which the analytical methodologies themselves operate.

Future challenges will include the robust application of these principles to areas of bioanalysis, medicine,[35] and nanoscience, where the requirement for accuracy and traceability is just as great, but the principles of metrology are arguably even harder to establish. One of the ultimate goals in chemical analysis is reproducible quantification at the level of single molecules. Surface enhanced Raman spectroscopy is one technique that shows the potential to reach this goal[36] although issues surrounding the reproducibility of detection need to be addressed before it will be a suitable technique for use in quantitative analysis.

Acknowledgements

This work was funded by the UK DTI National Measurement System's Valid Analytical Measurement Programme.

References

1 T. J. Quinn and J. Kovalevsky, *C. R. Phys.*, 2004, **5**, 791.
2 M. Burns, *Trends Anal. Chem.*, 2004, **23**, 393.
3 R. Kaarls and M. J. T. Milton, *C. R. Phys.*, 2004, **5**, 907.
4 M. L. McGlashan, *Metrologia*, 1994/95, **31**, 447.
5 M. L. McGlashan, *Metrologia*, 1997, **34**, 7.
6 The International System of Units, 7th edn, 1998, BIPM, Paris.
7 I. M. Mills, B. N. Taylor and A. J. Thor, *Metrologia*, 2001, **38**, 353.
8 I. M. Mills, P. J. Mohr, T. J. Quinn, B. N. Taylor and E. R. Williams, *Metrologia*, 2006, **43**, 227.
9 P. Becker, *Metrologia*, 2003, **40**, 366.
10 T. Cvitas, *Metrologia*, 1996, **33**, 35.
11 M. J. T. Milton, P. T. Woods and P. E. Holland, *Metrologia*, 2002, **39**, 97.
12 ISO 31:1992, Quantities and Units Part 8: Physical Chemistry and Molecular Physics, 1992, ISO, Geneva.
13 I. M. Mills, *Metrologia*, 1997, **34**, 101.
14 E. A. Guggenheim, *Philos. Mag.*, 1942, **33**, 21.
15 I. M. Mills, *Metrologia*, 1994/95, **31**, 537.
16 R. Dybkaer, *Metrologia*, 2004, **41**, 69.
17 W. H. Emerson, *Metrologia*, 2004, **41**, L26.
18 T. J. Quinn and I. M. Mills, *Metrologia*, 1998, **35**, 807.
19 International Vocabulary of Basic and General Terms in Metrology, BIPM, IEC, IFCC, ISO, IUPAC, IUPAP and OIML, 1993, ISO, Geneva.
20 ISO 17511:2003 In vitro diagnostic medical devices. Measurement of quantities in biological samples. Metrological traceability of values assigned to calibrators and control materials, 2003, ISO, Geneva.
21 M. J. T. Milton and T. J. Quinn, *Metrologia*, 2001, **38**, 289.
22 Analytical Chemists' Committee of ICI Ltd., *Analyst*, 1950, **75**, 577.
23 R. J. C. Brown, M. J. T. Milton and P. J. Brewer, NPL Report COAM 18, January 2003.
24 M. J. T. Milton and R. I. Wielgosz, *Metrologia*, 2000, **37**, 199.
25 A. Wallard, *Accred. Qual. Assur.*, 2003, **8**, 319.
26 Guide to the Expression of Uncertainty in Measurement, BIPM, IEC, IFCC, ISO, IUPAC, IUPAP and OIML, 1995, ISO, Geneva.
27 *IUPAC Compendium of Chemical Terminology*, ed. A. D. McNaught and A. Wilkinson, 2nd edn, Royal Society of Chemistry, Cambridge, 1997.
28 A. Hulanicki, *Pure Appl. Chem.*, 1995, **67**, 1905.
29 K. Danzer and L. A. Currie, *Pure Appl. Chem.*, 1998, **70**, 993.
30 http://www.bipm.org/en/committees/cc/ccqm/.
31 http://kcdb.bipm.org/.
32 ISO/IEC 17025:2005 General requirements for the competence of testing and calibration laboratories, 2005, ISO, Geneva.
33 EURACHEM/CITAC Guide: Traceability in Chemical Measurement, EURACHEM, CITAC, 2003.
34 R. J. C. Brown and M. J. T. Milton, *Trends Anal. Chem.*, 2004, **24**, 266.
35 R. Dybkaer, U. Ornemark, A. Udall and W. Richter, *Accredit. Qual. Assur.*, 1999, **4**, 349.
36 P. Etchegoin, R. C. Maher, L. F. Cohen, H. Hartigan, R. J. C. Brown, M. J. T. Milton and J. C. Gallop, *Chem. Phys. Lett.*, 2003, **375**, 84.