

IUPAC Technical Report

Russell J. Molyneux*, John J. Beck, Steven M. Colegate, John A. Edgar, William Gaffield, John Gilbert, Thomas Hofmann, Laura L. McConnell and Peter Schieberle

Guidelines for unequivocal structural identification of compounds with biological activity of significance in food chemistry (IUPAC Technical Report)

<https://doi.org/10.1515/pac-2017-1204>

Received December 21, 2017; accepted May 5, 2019

Abstract: Isolation of endogenous constituents of foods is generally performed in order to elucidate the biological activity of individual compounds and their role with respect to factors such as organoleptic qualities, health and nutritional benefits, plant protection against herbivores, pathogens and competition, and presence of toxic constituents. However, unless such compounds are unequivocally defined with respect to structure and purity, any biological activity data will be compromised. Procedures are therefore proposed for comprehensive elucidation of food-based organic structures using modern spectroscopic and spectrometric techniques. Also included are guidelines for the experimental details and types of data that should be reported in order for subsequent investigators to repeat and validate the work. Because food chemistry usually involves interdisciplinary collaboration, the purpose is to inform chemists and scientists from different fields, such as biological sciences, of common standards for the type and quality of data to be presented in elucidating and reporting structures of biologically active food constituents. The guidelines are designed to be understandable to chemists and non-chemists alike. This will enable unambiguous identification of compounds and ensure that the biological activity is based on a secure structural chemistry foundation.

Keywords: bioactivity; chiral recognition; circular dichroism; FT infrared spectroscopy; guidelines; mass spectrometry; natural products; NMR; stereochemistry; structure-function; UV-visible spectroscopy; X-ray structure.

CONTENTS

1. Introduction.....	1418
2. Physical properties	1420
3. Ultraviolet/visible (UV/Vis) spectroscopy.....	1422

Article note: Sponsoring body: IUPAC Chemistry and the Environment Division: see more details on page 1435.

***Corresponding author: Russell J. Molyneux,** Daniel K. Inouye College of Pharmacy, University of Hawaii at Hilo, Hilo, HI 96720, USA, e-mail: molyneux@hawaii.edu. <https://orcid.org/0000-0001-6285-8449>

John J. Beck: Chemistry Research Unit, Center for Medical Agricultural and Veterinary Entomology, ARS/USDA, Gainesville, FL 32608, USA

Steven M. Colegate: Poisonous Plant Research Laboratory, ARS/USDA, Logan, UT 84321, USA

John A. Edgar: CSIRO Agriculture and Food, PO Box 52, North Ryde, NSW 1670, Australia


William Gaffield: Western Regional Research Center, ARS/USDA, Albany, CA 94710, USA

John Gilbert: FoodLife International Ltd., York, UK

Thomas Hofmann: Chair for Food Chemistry and Molecular Sensory Science, Technical University of Munich, D-85354 Freising, Germany

Laura L. McConnell: Bayer CropScience, Research Triangle Park, NC 27709, USA

Peter Schieberle: Faculty of Chemistry, Technical University of Munich, D-85354 Freising, Germany

 © 2019 IUPAC & De Gruyter. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. For more information, please visit: <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Bereitgestellt von | Technische Universität München
Angemeldet
Heruntergeladen am | 19.03.20 16:05

4. Infrared (IR) spectroscopy.....	1423
5. Optical rotation.....	1424
6. Optical rotatory dispersion/circular dichroism (ORD/CD).....	1425
7. Nuclear magnetic resonance (NMR) spectroscopy.....	1427
8. Mass spectrometry (stand-alone)	1429
9. Gas chromatography–mass spectrometry (GC-MS).....	1431
10. Liquid chromatography–mass spectrometry (LC-MS).....	1433
11. X-ray crystallography	1435
Membership of the sponsoring body.....	1435
References.....	1436

1 Introduction

Advances in food chemistry are frequently dependent on determination of biological activity of specific components. Such activities define, for example, odour and taste, quality descriptors, resistance of plants to herbivory and fungal attack, inter-plant competition, toxicity to man and animals due to the presence of exogenous microbial toxins or endogenous phytotoxins, and health-promoting and nutritional benefits. However, without adequate compound identification, reports of biological activity for specific components are highly suspect or even meaningless. It should be noted that in the context of the following discussion “biological activity” refers to discrete organic chemical constituents in foods that affect the well-being, positive or negative, of the consuming organism. Natural macromolecules, such as lignin, polysaccharides, *etc.*, that are responsible for the structural matrix of foodstuffs are not generally included in this category because they are biopolymers encompassing a range of molecular masses. Furthermore, while biological activity of a given compound is obviously dependent on the amount ingested, quantitation is not addressed herein, since it warrants a comprehensive separate treatment.

Unfortunately, increasing reliance on instrumental methods to establish the structure of compounds of relevance in food chemistry has often resulted in improper or incomplete characterization of chemical constituents, with a consequent low level of confidence in the significance of the work. For example, many reports exist in the literature where compounds have been defined solely by comparison with mass spectrometric databases. Similarly, compounds isolated by preparative chromatography are frequently described as “amorphous solid” with no attempt being made to establish purity criteria. Another common problem is the failure to report optical activity data or to define the stereochemistry of compounds that possess chiral centres. A classic example of the importance of this property is the odour of *R*-(-)- and *S*-(+)-carvone, which smell of spearmint and caraway, respectively, because olfactory receptors can distinguish between these enantiomers. If a biological detection system is capable of such discrimination, then it is essential that the structural elucidation power of chemistry be applied in a way that matches this capability. In the context of food chemistry, these deficiencies are a serious problem since the rationale for isolating and identifying constituents is almost always to elucidate their biological activity. The dependency of many biological activities on often subtle changes in structure requires that structural elucidation and reporting is sufficiently comprehensive and unambiguous to differentiate the differences in bioactivity. These guidelines are designed to identify the **type** and **quality** of data that are essential to provide a secure foundation for structural elucidation and identification of food constituents, so that subsequent studies on biological properties are consequential and capable of replication by other investigators. The manuscript has been prepared in accordance with IUPAC standards for quantities, units and symbols [1], and terminology [2]. However, it is recognized that other journals may specify different expressions which might lead to ambiguities.

Most constituents isolated in food chemistry studies are a subset of natural products. Multi-volume compendia exist [3, 4] that comprehensively describe aspects of small molecule and biopolymeric natural products, organized by structural types integrated on a biosynthetic basis. Many of the topics covered include compounds of direct interest in food chemistry, for example, polyketides (including fatty acids, jasmonoids,

and polyphenolics such as the flavonoids), isoprenoids (including carotenoids and terpenoids), carbohydrates (including celluloses, tannins and lignins), amino acids and peptides, pheromones, and plant, insect and microbial hormones. Other chapters deal with food composition (chemistry of tea, coffee and wine) or food quality (beer flavour, and flavours and fragrances in general). While the chemistry is presented from a biosynthetic perspective, many of these chapters encompass sections dealing with structural elucidation. The more recent version of Mander and Liu [4] includes a particular volume (Volume 9) with chapters that specifically address modern methods for structure elucidation, including X-ray crystallography, circular dichroism, NMR spectroscopy and mass spectrometry.

A number of textbooks exist that deal specifically with the subject matter of food chemistry. One volume that has been consistently revised and updated is that by Belitz *et al.* [5], organized both by compound class and food type. The tabular listing of compounds and structures, together with their organoleptic properties, is particularly useful in representing the diverse nature of food constituents and the changes in properties in relation to often quite minor structural modifications.

Structural elucidation of organic compounds in general is covered both by a number of textbooks and in Web-based learning modules. A volume that illustrates structure determination with particular use of examples from natural products chemistry is that of Crews *et al.* [6]. As with most such texts, there is a major focus on nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS), but there are also chapters on infra-red (IR) and ultraviolet/visible (UV/Vis) spectroscopy, and a brief treatment of chiroptical techniques, such as optical rotatory dispersion-circular dichroism (ORD-CD) and the exciton chirality method. A particularly useful chapter is that dealing with strategies for integrating data from such techniques to elucidate the structure of an unknown compound by combining molecular formula, functional groups and substructures to obtain hypothetical 2D structures and refinement of these into a unique 2D structure and finally a 3D structure. Worked examples are presented that illustrate this process. While structural elucidation is common in published papers on constituents of foods, there is often inconsistency in its application, with relevant compounds not adequately purified or only partially characterized. Even when structures are confidently described, supporting or even critical physical and spectroscopic information may be missing, either because it has never been obtained or is simply unreported because it is deemed unnecessary. This issue has been briefly addressed in a short review [7] but this was designed for manuscripts submitted to a specific journal. The current paper is aimed at establishing consistent standards for structural identification of biologically active compounds relevant to all aspects of food chemistry, irrespective of the site of publication.

1.1 Structural elucidation strategy

There are various levels of confidence associated with reporting identifications of natural products, including food constituents [8] and these should be recognized at the outset of any studies. It is understandable that modern structural elucidation relies substantially on NMR and various MS methods. These two techniques are so powerful that structures can often be derived without recourse to other methods. There is now a vast array of ^1H and ^{13}C NMR correlation experiments that can establish direct and indirect connectivities between atoms in a molecule, so that unequivocal structures can be derived. Mass spectrometric (MS) techniques, either stand-alone or in combination with chromatographic separation, are particularly useful for establishing the presence of specific structural moieties. Possible structures for subsequent identification can initially be made by careful matching with large MS databases of known compounds. High resolution MS is particularly useful for establishing the elemental composition of a compound, especially when the amount is too limited for combustion analysis. In addition, the ratio of hydrogen to carbon atoms from high resolution MS allows the so-called “double bond equivalents” to be determined, which is not only an indication of the number of double bonds but also the presence of a carbonyl group or ring system.

In spite of these capabilities, both NMR and MS have some limitations. For example, neither the purity of a compound nor its absolute configuration can be directly determined, although for compounds that have substituents capable of reacting and forming derivatives with a chiral reagent, Mosher's method [9] can be applied. Furthermore, the instrumentation is relatively expensive and often available only with significant

time constraints. Therefore, a potentially useful fundamental strategy for structural elucidation would be to acquire data from simpler, rapid techniques (UV/Vis, IR, ORD/CD, *etc.*) that can be used to guide the selection of the most appropriate NMR or MS experiments. For example, IR will usually suggest the presence and type of carbonyl groups that often give only weak signals in the NMR spectrum. Other techniques, such as various forms of chromatography using a wide range of detection methods, can be used to determine the purity of the isolated compound. A calculation of the difference in retention indices on two stationary phases of different polarity can give important hints on the presence of functional groups in an unknown compound, such as phenolic hydroxy groups, double bonds, oxo groups, *etc.* UV/Vis spectra can indicate the presence of extended chromophores that can be helpful in linking structural moieties that are deduced from the NMR and MS. Similarly, optical rotation measurements and ORD/CD will establish the presence of chiral centers that need to be accounted for in deriving a complete structure. With few exceptions the compound can be recovered from these spectroscopic techniques and used for subsequent experiments that may be destructive (*e.g.* MS) and is therefore an economical procedure with compounds that may be limited in availability and/or needed for bioassay experiments. Comprehensive physical, spectroscopic, and spectrometric data is also beneficial to other researchers who may subsequently isolate the same compound. It is generally understood that, when available, samples of compounds will be provided to other scientists who may request them for direct comparison. Finally, the identification of previously known compounds by applying the principle of dereplication [10] requires the use of as many different techniques as possible in order to provide multiple comparison points.

The following sections provide a detailed description of the types of experiments and reporting of experimental conditions and acquired data that are essential for unequivocally establishing the structures of compounds that are involved in the many aspects of food chemistry. Obviously, a great deal of data can be generated from the application of multiple techniques, often too much for publication within a single paper. Most journals now provide a mechanism to provide this data, in the form of Supporting Information that can be accessed electronically by readers. However, it is important that authors provide the information most important to the structural elucidation within the body of the manuscript.

2 Physical properties

Physical properties are important criteria in describing a compound because they give an immediate sense of its nature without recourse to sophisticated instrumentation. For example, a colourless compound will eliminate any possibility of the compound being one with an extensive chromophore, such as a carotenoid or anthocyanin. Within the specific context of food constituents, a compilation of compounds has been published with a companion searchable CD-ROM, listing physical properties such as melting and boiling points, together with UV spectra [11].

The following factors should be presented when describing the physical properties of a compound:

2.1 Description

A physical description of the compound must be given. This should include its form, *e.g.* solid (crystalline/amorphous), liquid, viscous oil, *etc.*, and observed colour. It must be recognized that colour can often be conferred on the sample by very small amounts of intensely pigmented contaminants. Recrystallization in the presence of charcoal can often remove such contaminants and associated colour.

For identification of compounds with biological activity in food chemistry, odour and taste are important, even essential, physical descriptors. It is very important that such tests be performed only after establishing thoroughly the provenance of the original sample, a lack of contaminants (*e.g.* mycotoxins, pesticides, *etc.*), and review of the various extraction steps (*e.g.* toxic solvents) such that no residues are present. Since the use of trained panels is common in evaluating the contribution of individual components to the characteristic flavour and aroma of various foods, beverages and spices, evidence must be presented that each individual

has given informed consent before participation and that relevant approvals have been documented with respect to proper authorities.

The degree of solubility in solvents of differing polarity will indicate the hydrophilic or hydrophobic nature of the compound. With an awareness of potential artefact formation and consequent confounding of observations, such tests can be supplemented by checking the solubility in alkaline (*e.g.* aqueous ammonium hydroxide, sodium hydrogen carbonate/sodium carbonate, or sodium hydroxide) or acidic (*e.g.* aqueous hydrochloric acid) solutions, thereby establishing whether the compound itself is acidic (*e.g.* fatty acid; phenolic) or basic (*e.g.* amine; alkaloid) in nature, respectively.

2.2 Melting point

The melting point (m.p.) of a compound is an important descriptor because it provides a simple and inexpensive method for compound comparison and evaluation of purity. However, the m.p. of amorphous compounds may not be consistent and attempts should always be made to recrystallize them. If the compound is crystalline the solvent(s) from which it has been recrystallized must be reported and whether or not it has been recrystallized to a constant m.p. If the compound is believed to be identical with a known compound, the literature m.p. of that compound and corresponding citation(s) must be reported for comparison. Whenever possible an authentic sample should be obtained and a mixed melting point (m.m.p.) determination performed in order to establish whether or not the compounds are identical.

A description of the type of melting point apparatus and model must be given, together with whether the thermometer is corrected or uncorrected. Typical apparatus are oil-bath or hot-stage types and the temperature of melting may be determined either by manual observation or automatically. While the automatic type is convenient and time-saving, manual observation of a suitably slow temperature gradient (to avoid overshooting the actual m.p.) can provide invaluable information such as evolution of gases, colour change, or sublimation prior to melting. If melting is accompanied by decomposition or charring this should be reported, *e.g.* 123 °C (decomp.). An approximate value can be established by using a fairly rapid rate of heating followed by a more careful determination for which the rate of heating in the range of the m.p. should be slow enough (*ca.* 1 to 2 °C/min) so that the actual value is not overshoot.

2.3 Elemental analysis

The molecular formula of the unknown compound should be established as soon as possible, not only to provide fundamental elemental composition but also because it is the first step in dereplication, namely establishing whether the compound under investigation has been previously identified [10]. Combustion analysis can be used to determine the empirical formula and mass spectrometry (MS) to determine the molar mass. The amount of sample available may limit combustion analysis to determination of content of carbon and hydrogen only. Judicious use of high resolution MS (HRMS) can establish the elemental composition and is particularly useful for establishing the presence of bromine, chlorine, silicon and sulfur, which result in parent ion clusters due to these elements having more than one isotope of relatively high abundance. A combustion analysis that does not fit an elemental composition within the accepted parameters indicates that the sample is impure. On the other hand, HRMS should be used with caution because if the compound of interest does not ionize readily but contains an impurity that does so, the data may establish the elemental composition of the contaminant, rather than the compound of interest.

2.4 Thin-layer chromatography (TLC)

TLC [12] is a useful indicator of whether or not trace impurities are present in a sample. Description of the method used should include the substrate (silica gel, polyamide, C18, *etc.*) and its thickness, plate backing

(glass, plastic, aluminum, *etc.*) and developing solvent or solvent mixture. Incorporation of a fluorescent compound into the adsorbent also enables visualization of spots under UV light and the detection wavelength used and whether the spot fluoresces or quenches should be stated. Any subsequent visualization methods must be reported. These can be a general test, such as charring with sulfuric acid, or color tests specific for a particular compound or class of compounds. The retardation factor (R_f) [2] (n.b. not “retention factor”), defined as the ratio of the distance travelled by the centre of a spot to the distance travelled by the solvent front, must be reported. If the spot is relatively immobile (close to the origin) or too mobile (close to the solvent front) the accuracy and reproducibility of the R_f value will be compromised and the solvent should be changed to obtain a more acceptable value, typically within the R_f range 0.2 to 0.8.

If the identity of the compound under investigation is suspected, an authentic sample should be obtained and both compounds analysed, separately and together, on the same TLC plate. Experiments should be conducted with a variety of substrates and solvents. Correspondence of R_f values in several different experiments will ensure the identity of the unknown and authentic compounds. If trace impurities are detected in an unknown, it is good practice to run a blank of the solvents and/or reagents used in its isolation to establish whether or not these are the source of the contamination.

2.5 Colour tests

Numerous specific colour tests or spot tests have been devised for various classes of compounds [13]. These can be performed in a test tube or by spraying TLC plates with a solution of the reagent. Such tests are extremely useful for categorizing the type of compound (*e.g.* phenolic) and even for subclassification (*e.g.* flavonoid) [14]. The development of different colours with the same reagent can also be useful for differentiating classes of compound such as indole alkaloids, β -carboline and tyramines [15].

3 Ultraviolet/visible (UV/Vis) spectroscopy

UV/Vis spectroscopy encompasses electromagnetic radiation absorbed by a compound in the range 200 to 400 nm (UV) and 400 to 800 nm (Vis). The value of UV/Vis spectroscopy lies in its ability to indicate the presence of specific chromophores in a molecule and their existence either in isolation or in combination to form extended conjugated systems. The presence (or absence) of particular functional groups (or combinations thereof) can be inferred by the chromophores observed. In certain cases, such as acyclic dienes and cyclic polyenes, rules have been formulated to calculate the wavelength of maximum absorption for particular substituents, thereby permitting discrimination between a number of hypothetical structural moieties [16].

The large databases that exist for certain classes of compounds enable UV/Vis spectroscopy to be a convenient and inexpensive technique to assist dereplication, allowing known compounds to be distinguished from possible new natural products [17]. For some classes of compounds (*e.g.* flavonoids) addition of specific reagents to the solution being measured results in shifts in the λ_{\max} value which is diagnostic for a particular class of compound or a distribution of substituents [18, 19].

UV/Vis spectroscopy is frequently used in tandem with high-performance liquid chromatography (HPLC-UV) to detect classes of compounds and quantitate individual components of a mixture. In order to do so it is essential that appropriate standards be selected and their λ_{\max} and molar absorption coefficients be determined on compounds of established purity.

3.1 Experimental conditions

The manufacturer and model of the UV spectrometer used must be indicated. In addition, the type of dispersion (prism or diffraction grating) and detector (photomultiplier or photodiode array), the cell path-

length (typically 1 cm), together with the resolution of the instrument, should be given. The solvent used and concentration of the analysed compound must be reported.

When shift reagents such as AlCl_3 , NaOMe or H_3BO_3 are used, the solvent and concentration must be given. The solvent may differ from that of the original sample (*e.g.* AlCl_3 in aqueous acidic methanol).

3.2 Reporting data

The presentation of data will vary slightly between journals but as a general rule the format should be as follows: UV (solvent): λ_{max} (ϵ) or λ_{max} ($\log_{10} \epsilon$). Significant shoulders (sh) or points of inflexion (infl.) should be indicated in parentheses after the corresponding wavelength. Changes in spectrum maxima on addition of shift reagents should be indicated as bathochromic (to longer wavelength) or hypsochromic (to shorter wavelength) [2].

While it is common for simple spectra to be reported in the textual form indicated above, it may be helpful to readers for the actual UV curves to be published, particularly if the spectra are complex, when a series of compounds are being compared, or to illustrate changes in the spectrum produced by addition of a shift reagent. When a journal has page space restrictions but offers the opportunity to submit a supporting information file, all spectra should at least be submitted in the latter form.

4 Infrared (IR) spectroscopy

IR spectroscopy applied to organic structure elucidation is confined to the mid-IR wavenumber range covering 400 to 4000 cm^{-1} (wavenumber $\tilde{\nu}$, in cm^{-1}). In this region of the IR spectrum stretching, bending and twisting vibrations of individual chemical bonds result in a series of absorption bands. For all but the simplest molecules, the number of potential vibrations is large, resulting in numerous, often overlapping bands [20].

In contrast to UV/Vis spectroscopy, which is generally most indicative of combinations of functional groups that generate chromophoric systems, IR spectroscopy is particularly useful in giving rise to characteristic frequencies corresponding to individual functional groups within the molecule. The large number of absorption bands observed in an IR spectrum means that it can often be highly characteristic of a particular structure and be used as a “fingerprint” for that molecule by comparison with database libraries [21, 22]. Furthermore, near-infra red (NIR) spectroscopy can also be used for determining the chemical profiles of plant matrices when combined with chemometric models such as principal component analysis (PCA) [23].

4.1 Experimental conditions

The manufacturer and model of the IR spectrometer used is required. In addition, the type of dispersion (diffraction grating or Fourier transform (FT-IR)) must be stated, together with the temperature of measurement and the experimental resolution. To ensure accuracy of the recorded frequencies, the spectrometer should be calibrated periodically with a polystyrene film. The physical form in which the sample is analysed must be specified, *i.e.* liquid (neat or in solution), solid (KBr pellet), Nujol mull, or gas. For liquid samples, the solvent used and the cell path-length should be stated. For instruments that have the capability, especially FT-IR spectrometers, the spectrum of the solvent can be subtracted from the spectrum of the sample. For samples that are incompatible for analysis by transmission IR, an instrument with an attenuated total reflectance (ATR-IR) attachment can be used. This technique is particularly applicable to compounds that are water-soluble and for solid samples.

4.2 Reporting data

IR data is reported in wavenumbers ($\tilde{\nu}$) in units of cm^{-1} . Typically, only frequencies that can be unequivocally assigned to specific functional groups are listed, as these comprise only a small proportion of the observed bands. Optionally, the bands may be annotated with respect to their intensity as strong (s), medium (m) or weak (w), *e.g.* IR (Nujol) $\tilde{\nu}$ (cm^{-1}): 3500 (m, OH stretching), 1730 (s, C=O stretching). It is not common to publish the entire spectrum but in journals that provide the capability to submit supporting information, spectra should be provided in the specified format for direct comparison. When structural elucidation is supported by a “fingerprint” comparison it is essential that the sample and reference spectra be provided, recorded under identical conditions.

5 Optical rotation

Optical activity is detected in certain organic compounds when passage of a beam of plane polarized light through a solution of such compounds results in rotation of the plane of polarization of the beam. The angle of optical rotation, $[\alpha]$, (often abbreviated to “specific rotation”) is a physical property of the compound and is constant for a given wavelength, mass concentration, solvent and temperature. It is calculated from the angle of optical rotation, α ; at pathlength l , mass concentration, γ , Celsius temperature, θ , (typically *ca.* 20 °C), and wavelength, λ , (usually the sodium D line at 589.3 nm) according to:

$$[\alpha]_{\lambda}^{\theta} = \alpha / (\gamma l) \quad (1)$$

When the defining quantities used are in SI units, the corresponding unit of the specific optical rotatory power, $[\alpha]_{\lambda}^{\theta}$, is given in $^{\circ} \text{m}^2 \text{kg}$. A more convenient unit is obtained when the pathlength is 10 cm and the mass concentration is given in g/100 mL:

$$[\alpha]_{\lambda}^{\theta} = (\alpha / ^{\circ}) / ((l / 10 \text{ cm}) (\gamma / (\text{g} / 100 \text{ mL}))) \cdot 10^2 \text{ } ^{\circ} \text{g cm}^2 \quad (2)$$

Enantiomers are chiral compounds (non-superimposable mirror images of each other) that can be different in biological activity but are usually identical to each other in other physical properties (melting point, boiling point, refractive index, density, solubility, *etc.*), except that their specific rotation are equal, but opposite in sign. Diastereomers are optically active molecules (*e.g.* sugars) that are not mirror images of each other but do contain chiral centres that are designated *R* or *S*, based upon substituent priority rules. Compounds producing a clockwise rotation of the light beam generate positive specific rotation values and are described as dextrorotatory, and conversely those with negative (counterclockwise) values as laevorotatory. Molecules with two or more chiral centres, for which the mirror images are superimposable are referred to as meso compounds and are achiral. For terminology and concepts in stereochemistry, optical rotation, and polarimetry refer to Eliel *et al.* [24].

The development of manual polarimeters allowed the use of polarimetry to study the rate of hydrolysis of sugar. Semi- and fully-automatic instruments were subsequently introduced, and polarimetry is now routinely employed as a quality attribute for raw ingredients and finished products in process control in the food and other industries. Although polarimetry is a mature technique, modern polarimeters feature benefits that were not available with purely manual models. Current polarimeters provide the option of automatic data capture, variable wavelength and readouts accurate to 0.0001 degrees. To restrict or eliminate environmental influences on the optical rotation, modern polarimeters are equipped with temperature control systems that allow an accurate monitoring and control of the sample temperature with an accuracy of only a few hundredths of a °C, since minimal changes in temperature can often cause changes in the optical rotation of a substance. Thus, extremely narrow quality standards can be set based solely on optical rotation.

Specific rotation values are only useful in comparison with known standards, or inter-sample evaluation, if the isolated compounds are pure. Reports of specific rotations of compounds that have not been scrupulously purified are effectively meaningless and should be avoided. For example, contamination with an achiral compound will reduce the specific rotation value, whereas the presence of even a small quantity of a chiral compound with a particularly high specific rotation will cause the value to be significantly enhanced or reduced, depending on its sign relative to the compound of interest. If the compound is to be used for biological evaluation, it is especially important that the enantiomeric purity be established, as enantiomers can differ greatly in properties such as odour, taste, toxicity, allelopathy and enzymatic activity, or as semiochemicals, pesticides, herbicides, fungicides, *etc.* Fortunately, chiral analytical or (semi)preparative stationary phases have been developed for gas- and liquid-chromatography that can be used not only to establish chiral purity but also to separate enantiomers [25].

5.1 Experimental conditions

The manufacturer and model of the polarimeter used must be given. To ensure that the solvent does not contain optically active impurities or that the cell is not contaminated from previous samples, a measurement must be made with neat solvent prior to introduction of the sample solution.

5.2 Reporting data

The solvent used (methanol, ethanol, dimethylsulfoxide, chloroform, acetone, water, *etc.*) must be specified. Occasionally, different solvents will yield an opposite sign of rotation for the same substance. If a pure liquid compound is measured, the term “neat” should be used for description. Use of wavelengths lower than 589 nm may provide advantages in sensitivity. Other wavelengths available include 578, 546, 436, 405, and 365 nm. Often the observed optical rotatory power at 436 nm is approximately double and at 365 nm about triple that measured at 589 nm. Reporting rotation for different wavelengths enables multiple data point comparison with standards or other samples.

The optical rotatory power due to a solute in solution may be specified by a statement of the type:

$$\alpha(589.3 \text{ nm}, 20^\circ\text{C}, \text{ sucrose}, 10 \text{ g dm}^{-3} \text{ in H}_2\text{O}, 10 \text{ cm path}) = +0.6647^\circ$$

The same information may be conveyed by quoting either the specific optical rotatory power $\alpha/\gamma l$, or the molar optical rotatory power α/cl , where γ is the mass concentration, c is the amount (of substance) concentration, and l is the path length [1].

6 Optical rotatory dispersion/circular dichroism (ORD/CD)

Studies of optical activity using radiation near to the wavelength at which resonance absorption occurs in an optically active sample allow detection not only of anomalous optical activity or optical rotatory dispersion (ORD) but also a difference in the absorption indices at these wavelengths for right and left circularly polarized rays, an effect known as circular dichroism (CD) [26]. Circular dichroism also results in conversion of plane polarized light into elliptically polarized light. The bell-shaped CD curves and the S-shaped ORD curves observed in regions of UV absorption for optically active molecules are known as Cotton effects.

Circular dichrometers used for such measurements are commercially available instruments equipped with temperature control systems that allow accurate monitoring and control of the sample temperature with an accuracy of only a few hundredths of a degree centigrade, since minimal changes in temperature can often

cause changes in the optical rotation of a substance. These advanced instruments allow ORD measurements from *ca.* 190 nm to 600 nm and they possess high resolving power and high tolerance of sample absorption. ORD and CD measurements can thus be recorded through the entire region of absorption for a compound, which provides valuable structural and stereochemical information on chiral organic compounds, including various small molecule food components, as well as peptides, proteins, and carbohydrates. For example, long-chain hydroxy acids that do not contain an accessible absorption band provide plane ORD curves from *ca.* 200 nm to 600 nm, which yields useful structural information.

A particular advantage of ORD/CD measurements is that they can be used to determine absolute configuration of a compound by application of the octant rule or other sector rules [27], either from the known relative configuration or by comparison with compounds having similar chromophores for which the absolute configuration is already established. However, application of the octant rule and other semiempirical rules must be applied judiciously. For compounds with two independent chromophores, the exciton chirality method can be used to establish their disposition to each other in space [28]. Derivatization of molecules lacking a suitable chromophore can be used to generate chromophoric systems to which the exciton chirality method can be applied. Application of the method specifically to bioactive natural products has been reviewed by Humpf [29].

6.1 Experimental conditions

As with conventional optical rotations measurements, the manufacturer and model of the polarimeter used for ORD/CD measurements must be specified, together with the solvent used, temperature of the experiment, and wavelength range over which the spectrum is recorded. A blank run with pure solvent prior to the sample solution is essential.

6.2 Reporting data

Molar rotation $[\Phi]_{\lambda}^{\theta}$ is the most suitable experimental quantity for computing rotations of different substances due to comparisons being made on the basis of molar mass. Thus, $[\Phi]_{\lambda}^{\theta}$ is generally used in reporting ORD data. The molar optical rotatory power is defined as follows:

$$[\Phi]_{\lambda}^{\theta} = \alpha / cl = [\alpha]_{\lambda}^{\theta} \cdot M \quad (3)$$

where $[\alpha]$ is the specific optical rotatory power, c is the amount concentration and M is the molar mass of the optically active substance. When the defining quantities used are in SI units, the corresponding unit of $[\Phi]_{\lambda}^{\theta}$ is given in $^{\circ} \text{m}^2 \text{mol}^{-1}$. A more convenient unit is obtained when the pathlength is 10 cm, the amount concentration is given in mol/100 mL and the molar mass in g/mol:

$$[\Phi]_{\lambda}^{\theta} = [\alpha]_{\lambda}^{\theta} (M / (\text{g mol}^{-1})) / 10^2 \quad (4)$$

Circular dichroism is reported in one of two ways, either as molar ellipticity, $[\theta]$ or as molar circular dichroism, $\Delta\epsilon$. The two terms are directly convertible by the equation:

$$\Delta\epsilon = [\theta] / 3298 \quad (5)$$

where the value 3298 arises from the conversion of radians to degrees [26].

$$\text{Molar ellipticity: } [\theta] = 10^2 \cdot \theta / (cl) \quad (6)$$

in $^{\circ} \text{cm}^2/\text{dmol}$, where θ is the degree of ellipticity, c is the amount concentration, and l is the cell pathlength in cm.

$$\text{Molar circular dichroism: } \Delta\varepsilon = \Delta A / (cl) \quad (7)$$

where ΔA is the difference in absorbance of left and right circularly polarized light, c is the amount concentration and l is the pathlength in cm.

Determination of $[\Phi]$ values over a particular wavelength range permits visual presentation of optical rotatory dispersion and is referred to as an ORD curve whereas recording values of either $[\theta]$ or $\Delta\varepsilon$ versus wavelength provides a CD curve. While values for peak positions can be reported in text form, it is generally advantageous to provide graphical ORD/CD curves so that the observed positive and negative Cotton effects can be directly compared for a pair of enantiomers, or for determination of absolute configuration by application of the octant rule or other sector rules.

7 Nuclear magnetic resonance (NMR) spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy provides molecular structural identification with a high degree of confidence. With the rapid development in column technology leading to exceptional resolution and detection sensitivity of components within a mixture, there is an increasing tendency to identify chromatographic peaks based upon similar, or presumed, characteristic retention time (t_r) indices; IR or UV/Vis absorption spectra; or various forms of mass spectrometric data including HRMS and tandem mass spectrometry (MSⁿ). However, without recourse to detailed comparison with authenticated standards, such “identifications” must be considered tentative.

NMR spectroscopic data should be acquired for any isolated compounds of sufficient purity for which proposed structures are not unambiguous. In many cases, relatively simple one-dimensional (1D) proton (¹H) or proton-decoupled carbon (¹³C) NMR spectra will suffice to confirm a proposed structure if the data conform to published data for an unambiguously defined compound. If the published structure was not unambiguously defined, or if the isolated compound is previously unreported, then more extensive NMR spectroscopy experiments, including selective 1D experiments and homonuclear and heteronuclear 2D experiments, are required with the aim of comprehensively mapping the presence and connectivities of all nuclei [6, 30, 31]. A precautionary principle requires that sweep widths be extended beyond the typical range in both the ¹H and ¹³C 1D NMR experiments (0 to 8 and 0 to 180 ppm, respectively) to ensure that all resonance peaks are observed. There are many variants of NMR experiments, however, a useful sequence for the unambiguous determination of structures might be:

1. ¹H spectrum;
2. Proton-decoupled ¹³C spectrum;
3. 2D ¹H–¹H correlation spectroscopy (COSY) to establish proton connectivities;
4. 2D ¹³C–¹H heteronuclear single quantum coherence (HSQC) spectroscopy to establish direct connectivity of protons to carbons, thereby establishing carbon types (*i.e.* quaternary, methine, methylene or methyl). If carbon type is ambiguous, then 1D distortionless enhancement by polarization transfer (DEPT) experiments or attached proton test (APT) experiments will confirm the character;
5. 2D ¹³C–¹H heteronuclear multiple bond correlation (HMBC) experiments will provide carbon-proton connectivity over 2 to 4 bonds.

It should be recognized that certain resonances may be relatively weak, especially ¹³C signals (*e.g.* carbonyl groups) due to long relaxation times. Acquisition of data by other spectroscopic methods such as IR may indicate the presence of such groups, so that extra effort is made to ensure inclusion of such entities in the NMR spectrum. Although software for assigning resonances and generating potential conforming structures is becoming more sophisticated and more widely used, it is essential that such structure elucidation be evaluated and verified by direct inspection and interpretation.

7.1 Experimental conditions

In all cases, the NMR spectrometer attributes (manufacturer, instrument model and magnet strength for the observed nuclei), the sample solvent, the field-frequency lock signal and chemical shift reference must be provided. The temperature of the sample and use of any shift reagents should be reported. When multi-pulse sequences are applied the meanings of the acronyms must be given together with a reference to sequence and its source (*i.e.* either literature or proprietary manufacturer-supplied).

7.2 Reporting data

7.2.1 Presentation of NMR spectra

To facilitate comparisons and assessments by other researchers, it is preferable to make available clear scans of all 1D ^1H and ^{13}C NMR spectra, together with any homonuclear and heteronuclear 2D spectra used to elucidate the structure. This will provide readers not only with a way to evaluate the overall quality of the spectra but also with a means to directly compare similarities and/or differences in the spectra with those for any unidentified compound that they may isolate. Depending on the policy of the particular journal, such scanned spectra can be placed within the body of the text or in the supplementary/supporting information section. In some cases, it may only be necessary to illustrate portions of the spectrum, or on the other hand, to enlarge specific regions for clarity purposes.

Even when scanned spectra are provided, each resonance signal must be described. For individual compounds it is usually most economical in terms of page space to provide the data in text form. A typical format is: *Compound X*. ^1H NMR (solvent, field strength (MHz)): chemical shift (δ), integrated relative intensity, multiplicity, coupling constant between specified protons (*e.g.* $J_{x,y} = 6.5$ Hz) and signal assignment for numbered atoms in a chemical structure. Multiplicities are represented by the abbreviations: singlet (s), doublet (d), triplet (t), quartet (q), multiplet (m), and may be modified by the term broad (br.), for example, (br.s).

If relative stereochemistry can be assigned then protons should be labelled as “ α ” or “ β ” but otherwise simply as “a” or “b”, or “u” and “d” for the upfield and downfield partners in a geminal proton pair. Such designations are presented as subscripts to the atom number (*e.g.* H-1_α or H-1_u).

For previously described compounds where the spectra concur, it is generally not necessary to present the data but merely to insert a statement such as: “NMR data consistent with published report(s)”, each identified by a reference number. However, if significant differences are observed in any chemical shift, multiplicity, or coupling constant value then a reason must be provided. If no rational explanation is forthcoming, the complete spectrum must be re-evaluated *a priori*.

When NMR data are obtained for a series of closely related compounds it is often advantageous to present the information in tabular form so that direct comparison of similarities and differences between structural moieties can be made. The same conventions for abbreviations as with textual presentation are used, with explanations appended to the Table as footnotes.

7.3 NMR spectroscopy data for structure elucidation

All NMR spectroscopy data, including heteronuclear correlations that allow for a complete and unambiguous mapping of all nuclei should be presented. The preferred method is tabulation with associated footnotes that confirm the ^{13}C assignments (*e.g.* ^a all assignments confirmed using a gradient enhanced HSQC experiment). In addition to the abbreviations defined above for all NMR data, others such as: not applicable (n.a.); not observed (n.o.); weak (w); *etc.* may be used. However, provided that a comprehensive NMR spectroscopy data set is presented, any combination of text and Tables is acceptable, depending upon journal requirements and personal preference. Key correlations should be explained in the text. Other key structural elucidation data,

such as from 1D selective or 2D NOESY NMR experiments, should be described in the text or, if extensive, also be tabulated. Such key structural correlations can often be represented graphically to advantage on a structure with numbered positions by a series of arrows showing through-space correlations. However, this format should not be provided as an alternative to the presentation of the complete data set, but rather in order to facilitate comprehension on the part of the reader.

8 Mass spectrometry (stand-alone)

Modern mass spectrometers capable of producing tandem mass spectra (MS/MS), multiple-stage mass spectra (MSⁿ) and high resolution, exact mass measurement, combined with computer systems for comprehensive data interpretation, can provide considerable structural information [32, 33]. However, using MS data alone is likely to leave some questions and doubts in regard to certain aspects of molecular structures, so that unequivocal identification of a compound by MS alone is not generally achievable [33]. This is especially the case in regard to determining the structures of previously unknown molecules and also when authentic samples of known substances are not available for direct MS comparison. Future developments in MS equipment, computer-aided data interpretation and cheminformatics, may eventually approach the point where reasonably secure identification can be cautiously claimed. At present MS data needs to be accompanied by other data to achieve an acceptable level of information necessary for unequivocal identification. While some laboratories have up-to-date MS equipment with data systems capable of producing detailed identification data, this is not generally the case [33]. Thus, the availability of authentic standards for comparison, combined with independently obtained chromatographic and spectroscopic data (*e.g.* NMR data) are also required before claims of unequivocal identification based primarily on MS data meet expectations. When a claim of “unequivocal” identification is based primarily on MS data alone, as many of these factors as possible need to be considered and the results reported along with the MS data.

Definitions and recommendations for terms in mass spectrometry have been published by Murray *et al.* [34].

8.1 Experimental conditions

It is essential that the types of MS equipment (manufacturer, model, *etc.*), software used, and the performance capability be described in detail, as they determine the accuracy and reliability of the data reported and reflect on the confidence in the structure identification. All operating conditions such as voltage potentials, gas flows, temperatures, the mass range scanned, *etc.* must also be reported. The instrument must be carefully calibrated with an appropriate standard prior to use, and the frequency of calibration and the ions monitored to determine stability reported. A blank run with the solvent used to dissolve the sample, if it is not introduced into the probe as a solid, should be performed in order to subtract impurity peaks and background from the mass spectrum of the compound of interest. It should also be recognized that even instruments of the same model may give slightly different mass spectra so that between-instrument or day-to-day samples are rarely 100 % comparable. To satisfy expectations of unequivocal structural identity will generally require additional data. Prior recognition of the existence of chiral compounds through optical rotation information that differentiates between diastereomers and especially enantiomers is essential, as these will usually not be distinguishable from the mass spectrum alone.

The most commonly used ionization techniques are electron ionization (EI) and chemical ionization (CI). The specific details of the particular experiment must be provided, for example, the ionization energy and in CI the reagent gas (*e.g.* methanol, ammonia, *etc.*). Matrix-assisted laser desorption ionization (MALDI), requires a completely different type of instrument and is only employed for analysis of relatively high mass compounds. For MALDI experiments, the type and energy of the laser must be specified, and it is essential

that details of the matrix and sample preparation be given. Details should be sufficient for others to repeat the experiment as closely as possible, within individual instrumental limitations. Other methods such as direct analysis in real time (DART) [35] and desorption electrospray ionization (DESI) are ambient ionization techniques using an atmospheric pressure ion source that permits MS analysis of samples in their native state. Ionization occurs on the sample surface, making the method applicable to food samples such as fruits and vegetables. However, these only offer the possibility of screening for compounds of interest, although tentative identification might be claimed.

Ideally a compound being identified solely or largely on the basis of MS data should be shown to be essentially pure prior to investigation of its structure because interfering background noise and co-occurring compounds can adversely affect aspects such as molar mass and elemental composition determination. While MS data alone has considerable ability to determine the level of purity of a substance, the compound to be identified should also be shown to be essentially pure by other means, *e.g.* chromatographic characteristics and physical properties, before being subject to stand-alone MS structure determination. Individual compounds have different ionization potentials and on occasion an impurity may give a deceptively much stronger mass spectrum than the major constituent. If the presence of impurities is not predetermined, this may mislead the operator into believing that the impurity is the compound of interest.

The concept of “stand-alone” MS identification implies that the substance being identified has been inserted directly into the ion source of an MS and it has not entered the ion source via coupled gas or liquid chromatography (see below). Attempted identification of a substance by stand-alone MS data may therefore, in the worst-case scenario, involve MS detection and identification of an individual substance in very complex mixtures or, more likely, following chemical fractionation into a smaller number of components having similar characteristics, for example in terms of solubility, volatility, basicity, acidity or selective adsorption properties on solid phase materials. The details of how the substance being identified by MS was obtained therefore need to be provided as they will add vital information on the chemical properties and likely level of purity of the sample and help to support and substantiate the chemical identification claimed. If chromatographic methods were employed in purifying the substance of interest prior to MS but not employed directly in introducing the substance into the MS, these too need to be described in detail as an indication of chemical properties and sample purity. Ideally the substance subject to identification based on MS data alone will have been purified to a single component and there will be both MS data and other information provided to substantiate its level of purity.

8.2 Reporting data

The physical properties of a “purified” substance subject to identification by MS data alone should be reported as evidence of purity and as characteristics of the substance for future reference.

Molar mass and elemental composition information are basic requirements for structure identification and MS is ideally suited to such determination [36]. Elemental composition determination requires that the MS has an appropriate soft ionization source and is capable of measuring the mass of the molecular ion or protonated molecule and ideally all fragment ions, with an accuracy below 2 ppm difference between the measured accurate mass (observed) and calculated mass. Verification of this capability and the results obtained should be reported. High resolution MS elemental composition and molar mass data should be reported in comparison with the calculated values and the error indicated as ppm. A discussion of terminology and treatment of data with respect to accurate mass measurement has been published [37].

MS data is typically reported in textual form. A common convention is to report up to 10 of the most intense ions (including the molecular ion or protonated molecule), followed by the % intensity (in parentheses) of each, relative to the most intense ion (100 %). However, low mass ions (m/z below 50) are rarely reported because they are virtually ubiquitous. When polyisotopic elements (Br, Cl, *etc.*) are present the values and relative intensities for each isotope should be reported. It is often not necessary to provide the

mass spectra in graphic form but for the convenience of readers these can be submitted as supplementary/supporting information in accordance with the policy of the publishing journal. Submission of data to MS databases is encouraged. When putative identification of ions is made, this can be indicated after each reported value. However, it must be recognized that such assignments can only be unequivocally confirmed by high-resolution MS or labelling experiments. Many ionization techniques will result in adduct formation and these should be indicated in a form such as: $[M+H]^+$, $[M+NH_4]^+$, $[M+Na]^+$, *etc.*

While small molecules generate singly-charged ions some, such as symmetrical dimers, can produce both single- and double-charged ions. Most proprietary mass spectrometers include software for charge state discrimination. Such doubly-charged ions must be identified in the dataset as M^{2+} , *etc.*

Large MS library databases are available [38] together with specialist libraries of interest to food chemists, such as flavours and fragrances, pesticides, *etc.* Identification based on library search data alone is not sufficient for “unequivocal” identification of a previously unknown substance or a substance without an authentic sample for comparison but can be greatly enhanced/improved using multiple-stage MS (MS^n) with associated libraries, and precursor ion fingerprinting data comparison with related compounds [39]. Details of the MS, MS/MS and MS^n libraries and search algorithms employed should be reported.

9 Gas chromatography–mass spectrometry (GC-MS)

The GC-MS technique combines the separating power of chromatography for volatile compounds with a versatile identification technique based on ionization to produce molar mass information and fragment ions indicative of structure. It is unlikely that an entirely novel organic compound can be unequivocally identified based only on GC-MS data, and other supporting spectroscopic data such as NMR is normally required. However, unknown compounds can be matched by GC-MS against known standards from databases or their structures can be extrapolated from closely related known compounds. In all cases obtaining a standard and conducting comparative GC-MS analysis is the most rigorous way to support identification.

In GC-MS analysis, the chromatographic relative retention time (or retention index, RI) provides an insight into the polarity of the compound of interest and its volatility. The retention time (or index) must be measured on at least two different GC columns with different stationary phases, and it should be demonstrated that identical mass spectra are obtained on both columns thereby showing effective separation has been achieved. The minimum acceptable retention time is twice the retention time corresponding to the void volume of the column. A calculation of the difference in retention indices on two different stationary phases such as a nonpolar silicone SE-54 and a polar free fatty acid phase (FFAP) is a powerful tool to get information on the structure of an unknown compound, as demonstrated by the following example. Decanal is characterized by an RI on SE-54 of 1207 and of 1497 on FFAP, a difference of 290. The more polar (2*E*)-dec-2-enal (1262/1635) shows a difference of 373, caused by the conjugated double bond, and the (2*E*,4*E*)-dec-2,4-dienal is characterized by 1318/1804 with a difference of 486, due to the second double bond (higher polarity). Even a weak difference in polarity caused by a (*Z*)-double bond can be derived. (2*E*,4*Z*)-dec-2,4-dienal is characterized by 1294/1752, *i.e.* a difference of 458 compared to 486 in the (*E,E*) isomer. Thus, in combination with high resolution MS, IR or UV, a proposal for a structure can be obtained. More examples can be found in Rychlik *et al.* [40].

High resolution MS not only supplies the elemental composition, but also gives an indication on the so-called “double bond equivalents”, which may also indicate the presence of a ring system or the presence of an aldehyde function. For example, in an alkane such as nonane the C:H ratio is 9:20, whereas in the corresponding saturated aldehyde it is 9:18. The high resolution data of the previously identified (2*E*,4*E*,6*Z*)-nona-2,4,6-trienal gives $C_9H_{12}O$, which clearly indicates four double bond equivalents, *i.e.* the presence of an aldehyde function and three double bonds [41]. An aromatic ring system, like benzene, would give C_6H_6 , *i.e.* four double bond equivalents compared to the open chain hexane with C_6H_{14} .

The full scan mass spectrum from a GC-MS run can provide a possible indication of molar mass, whilst the fragmentation pattern in electron ionization (EI) mode can provide some insights into certain structural elements of the compound of interest. That the highest m/z ion (M^+) observed in EI mode is in fact the molecular ion, can be confirmed if an $(M + X)^+$ ion is found in the chemical ionization spectrum, where X is determined by the nature of the reagent gas (e.g. $X = 18$ when the reagent gas is ammonia). Identification relies on proper selection of diagnostic (characteristic) ions. The (quasi) molecular ion is a diagnostic ion that should be included in the measurement and identification procedure whenever possible. In general, and especially in single-stage MS, high m/z ions are more specific than low m/z ions (e.g. $m/z < 100$). However, high mass m/z ions arising from loss of water or loss of common moieties may be of little use, although characteristic isotopic ions, especially Cl or Br clusters, may be particularly useful.

It is important to recognize that the extent and efficiency of sample purification (clean-up) and the quality of the chromatographic separation will both affect the ability to obtain a full scan spectrum of the unknown which is free of interfering ions. When unknowns are present at trace levels in foods, this ability to achieve good chromatographic separation prior to MS analysis is particularly critical.

The volatility of some polar compounds can be increased by derivatization of functional groups, making them amenable to GC analysis and additionally increasing the molar mass thereby providing more characteristic diagnostic ions. For example, derivatization of hydroxyl groups to form trimethylsilyl (TMS) ethers is frequently employed, with a wide range of derivatization reagents being commercially available.

9.1 Experimental conditions

The manufacturer and model of the GC-MS instrumentation is required. Where not obviously apparent from the GC-MS model, the configuration of the mass spectrometer (quadrupole, ion-trap, magnetic sector, time-of-flight) should be indicated. For the gas chromatography, the dimensions of the capillary column, its stationary phase and film thickness, type of carrier gas, its flow rate and conditions for the temperature program must be reported. The temperature of the GC injector and the injection mode (split, split/splitless, on-column or headspace) should be stated. For the mass spectrometry, the mode of ionization (electron impact and/or chemical ionization), the m/z scan range and scan speed should be specified. In chemical ionization mode the reagent gas should be specified. The resolution of the mass spectrometer should be stated as nominal mass, or for high resolution accurate mass measurement the resolution and mode of measurement of resolution should be given (e.g. full width at half-maximum (FWHM) at a specified m/z value).

9.2 Reporting data

The relative retention time should be given with respect to a known standard under prescribed GC conditions. The Kovats retention index is by definition system independent and the index should be reported normalized to the retention times of adjacently eluting n -alkanes. Retention time data is available for searching against various databases. For example, the NIST library [42] gives retention index (RI) data for 82 868 compounds and can be a useful starting point in identification.

For a full scan GC-MS EI spectrum, the top ten ions should be tabulated, or the scanned spectrum shown as a plot of m/z against ion intensity. The NIST library gives EI spectra for 242 466 chemical compounds and can be searched electronically to find a possible match. In chemical ionization (CI) mode the m/z value of the quasi-molecular ion should be reported. In high resolution, the accurate mass of the molecular ion should be reported, and a search conducted to establish possible empirical formulae that might be a reasonable match for the compound of interest. The closeness of the match as 10^{-6} (ppm) between the measured m/z and the theoretical m/z of the suspected molecular formulae should be reported. In some situations, tandem mass spectrometry (GC-MS/MS) may be employed to provide additional structural indications, in which case the m/z value of the selected precursor ion and the m/z values of the product ions should be reported as nominal (low resolution) or accurate masses.

9.3 Identification criteria [43, 44]

1. Where identification is reported on the basis of the GC retention time of a known compound, the relative retention time should match that of the standard at a tolerance of $\pm 0.5\%$.
2. When full scan mass spectrometric identification is confirmed by analysis of a standard, the presence of all measured diagnostic ions (the molecular ion, characteristic adducts of the molecular ion, characteristic fragment ions and isotope ions) with a relative intensity of more than 10 % should be present for the unknown and the reference standard.
3. The intensities of detected ions, expressed relative to the intensity of the most intense ion or transition, must correspond to those of the standard, at a comparable concentration, measured under the same conditions, within the following tolerances: $\pm 10\%$ for relative intensity greater than 50 %, $\pm 15\%$ for relative intensity between 20 and 50 %, $\pm 20\%$ for relative intensity between 10 and 20 % and $\pm 50\%$ for relative intensity *equal to or lower than* 10 %.
4. In high resolution mass spectrometry, for at least two diagnostic ions, preferably including the (quasi) molecular ion and at least one fragment ion, the agreement with the calculated theoretical mass or with measured ions from the analytical standard should be below 2 ppm.

10 Liquid chromatography–mass spectrometry (LC-MS)

The LC-MS technique combines the separating power of chromatography for non-volatile compounds with a versatile identification technique based on ionization to produce molar mass and structural information. It is unlikely that an entirely novel organic compound can be unequivocally identified based only on LC-MS data, and other supporting spectroscopic data such as NMR is normally required. However, unknown compounds can be matched by LC-MS against known standards from databases or their structures can be extrapolated from closely related known compounds. In all cases, obtaining a standard and conducting comparative LC-MS analysis is the most rigorous way to prove unequivocal identification.

In LC-MS analysis, the chromatographic retention time provides an insight into the polarity of the compound of interest depending on the LC column and mobile phase. The minimum acceptable retention time is twice the retention time corresponding to the void volume of the column. The full scan mass spectrum from an LC-MS run can provide a possible indication of molar mass, although compared to GC-MS there is generally far less fragmentation to provide structural indications. The highest m/z ion observed in LC-MS will usually be an adduct ion, and with knowledge of the mobile phase the molar mass can be deduced. However, it can also be $[M - H]^-$, or in the case of naturally charged molecules (*e.g.* quaternary alkaloids) even the molecular ion.

Identification relies on proper selection of diagnostic (characteristic) ions. The (quasi) molecular ion is a diagnostic ion that should be included in the measurement and identification procedure whenever possible. In general, and especially in single-stage MS, high m/z ions are more specific than low m/z ions (*e.g.* $m/z < 100$). However, high mass m/z ions arising from loss of water or loss of common moieties may be of little use. Characteristic isotopic ions, especially Cl or Br clusters, may be particularly useful.

It is important to recognize that both the extent and efficiency of sample purification (clean-up) and the quality of the LC separation will affect the ability to obtain a mass spectrum of the unknown which is free of interfering ions. When unknowns are present at trace levels in foods, this ability to achieve good LC separation prior to MS analysis is particularly critical.

10.1 Experimental conditions

The manufacturer and model of the LC-MS instrumentation is required. Where not obviously apparent from the LC-MS model, the configuration of the mass spectrometer (quadrupole, ion-trap, time-of-flight, *etc.*)

should be indicated. For the liquid chromatography, the dimensions of the LC column, its stationary phase and particle size, mobile phase, together with mobile phase flow rate and whether isocratic or programmed gradients, should be specified. The temperature of the LC column should be stated. For the mass spectrometry, the mode of ionization (ESI, APCI), positive or negative ion mode, the m/z scan range and scan speed should be specified. The resolution of the mass spectrometer should be stated as nominal mass or for high resolution accurate mass measurement the resolution and mode of measurement of resolution should be given (e.g. FWHM at a specified m/z). Appropriate instrumental settings should be provided depending on the mode of operation and the instrument type, but might include for example sheath and auxiliary gas flow rates; spray voltage, capillary temperature, capillary voltage, tube lens and skimmer voltage and heater temperature.

10.2 Reporting data

The relative retention time should be given with respect to a known standard under prescribed LC conditions. LC retention time data is less useful than in GC, but nevertheless is a good indicator if matched against an analytical standard.

For a full scan LC-MS spectrum the m/z value of the molecular ion or adducted molecule and any fragment ions if present should be reported. In LC-MS, spectra produced on different instrumentation types from a variety of instrument vendors can vary significantly in terms of ion abundance and content. LC-MS databases are less well developed than GC-MS and proprietary databases from instrument manufacturers provide the best resource for identification.

In high resolution, the accurate mass of the molecular ion should be reported, and a search conducted to establish possible empirical formulae that might be a reasonable match for the compound of interest. The closeness of the match as 10^{-6} (ppm) between the measured m/z and the theoretical m/z based on the hypothetical molecular formulae should be reported.

In LC-MS where there is little fragmentation, tandem MS (LC-MS/MS) may be employed to provide additional structural information. In LC-MS/MS the m/z value of the selected precursor ion and the m/z values of the product ions should be reported as nominal (low resolution) or accurate masses.

10.3 Identification criteria [43, 44]

1. Where identification is reported on the basis of the LC retention time of a known compound acquired in the same system, the relative retention time should match that of the standard at a tolerance of $\pm 2.5\%$.
2. When full scan mass spectrometric identification is confirmed by analysis of a standard, the presence of all measured diagnostic ions (the molecular ion, characteristic adducts of the molecular ion, characteristic fragment ions and isotope ions) with a relative intensity of more than 10 % should be present for the unknown and the reference standard.
3. The intensities of detected ions, expressed relative to the intensity of the most intense ion or transition, must correspond to those of the standard, at a comparable concentration, measured under the same conditions, within the following tolerances: $\pm 10\%$ for relative intensity greater than 50 %, $\pm 15\%$ for relative intensity between 20 and 50 %, $\pm 20\%$ for relative intensity between 10 and 20 % and $\pm 50\%$ for relative intensity *equal to or lower than* 10 %.
4. In high resolution mass spectrometry for at least two diagnostic ions, preferably including the (quasi) molecular ion and at least one fragment ion, the agreement with the calculated theoretical mass or with measured ions from the analytical standard should be below 2 ppm.

11 X-ray crystallography

X-ray crystallography is becoming increasingly available and is perhaps the ultimate technique for defining molecular structure and configuration. It can also be used as an alternative to elemental analysis or mass spectrometry to establish the elemental composition of a compound. However, it should be recognized that such data is only representative of the crystal analysed; it does not establish the purity of the bulk sample. Under ideal experimental conditions and with careful interpretation it can also establish the absolute configuration of the molecule.

11.1 Experimental conditions

The validity of the derived structure is primarily a function of the crystal quality. Care should therefore be devoted to producing crystals that are not only of high quality but that are also representative of the sample as a whole. It should be recognized that impurities may co-crystallize from the solution and selection of such artefacts must be avoided. Crystals may also incorporate the solvent from which they have been recrystallized and subsequent evaporation may disrupt the crystal lattice. For these reasons, the crystal chosen for analysis should be selected by microscopy. Experimental information should include a description of the crystal, the make and model of the diffractometer, and the type and wavelength of the radiation. The particular software programs employed in deriving the structure must be indicated.

11.2 Reporting data

Specialist crystallography journals have specific requirements for reporting the data and the derived structure. However, when crystallographic structures are obtained as a component of the structural elucidation in non-specialist journals, it is common to report the relevant information in an abbreviated form. Typically, this consists of an ORTEP diagram showing the structure and packing of the molecule, including the location of any solvent molecules and counter-ions, together with a discussion of bond lengths, hydrogen bonding, stacking interactions, *etc.*, that may be relevant to the biological activity of the compound. The underlying electronic Crystallographic Information Files (CIFs) are typically appended as Supporting Information. The format and standards for CIF files are maintained by the International Union of Crystallography [45] and must be strictly adhered to. It is strongly recommended that CIF files be deposited in an appropriate database such as the Cambridge Structural Database [46] for organic small molecules or the Protein Data Bank for biologically active macromolecules such as proteins and nucleic acids [47]. The curated CIF files for previously deposited structures can be searched for and freely accessed at these sites.

Acknowledgements: This manuscript was prepared within the framework of IUPAC, Funder Id: <http://dx.doi.org/10.13039/100006987>, project 2013-024-2-600.

Membership of the sponsoring body

Membership of the IUPAC Chemistry and the Environment Division Committee for the period 2016–2017 was as follows:

President: Dr. Petr Fedotov (Russia); **Past President:** Dr. Laura McConnell (USA); **Vice President:** Dr. Rai Kookana (Australia); **Secretary:** Prof. Hemda Garelick (UK); **Titular Members:** Dr. Manos Dassenakis (Greece); Dr. Philippe Garrigues (France); Dr. Irina Perminova (Russia); Dr. Heinz Rüdell (Germany); Dr. John

B. Unsworth (UK); Dr. Baoshan Xing (USA); **Associate Members:** Prof. Guibin Jiang (China); Prof. Nadia G. Kandile (Egypt); Prof. Gijs A. Kleter (Netherlands); Dr. Bradley W. Miller (USA); Prof. Stefka Tepavitcharova (Bulgaria); Prof. Roberto Terzano (Italy); **National Representatives:** Dr. Annemieke Farenhorst (Canada); Prof. Yong-Chien Ling (China/Taipei); Dr. Anna-Lea Rantalainen (Finland); Dr. Pradeep Kumar, FNA (India); Prof. Doo Soo Chung (Korea); Dr. Din Mohammad (Pakistan); Prof. Ana Aguiar-Ricardo (Portugal); Prof. Edgard Resto (Puerto Rico); Prof. Luke Chimuka (South Africa); Dr. Nelly Mañay (Uruguay).

For the period 2018–2019 was as follows: **President:** Prof. Rai Kookana (Australia); **Past President:** Prof. Petr S. Fedotov (Russia); **vice President:** Prof. Hemda Garelick (UK); **Secretary:** Prof. Roberto Terzano (Italy); **Titular Members:** Prof. Doo Soo Chung (South Korea), Prof. Annemieke Farenhorst (Canada), Prof. Nadia G. Kandile (Egypt), Dr. Laura L. McConnell (USA), Prof. Irina Perminova (Russia), Prof. Fani L. Sakellariadou (Greece); **Associate Members:** Dr. Wenlin Chen (USA), Dr. Bradley Miller (USA), Prof. Diane Purchase (UK), Prof. Edgard Resto (Puesto Rico), Dr. John B. Unsworth (UK), Prof. Baoshan Xing (USA); **National Representatives:** Prof. Cristina Delerue Alvim de Matos (Portugal), Prof. Michal Galamboš (Slovakia), Prof. Ester Heath (Slovenia), Prof. Yong-Chien Ling (China/Tapei), Prof. Bulent Mertoglu (Turkey), Prof. Gloria Obuzor (Nigeria), Dr. Bipulbehari Saha (India), Dr. Tiina Sikanen (Finland), Prof. Weiguo Song (China), Prof. Stefka Tepavitcharova (Bulgaria).

References

- [1] IUPAC. *Quantities, Units and Symbols in Physical Chemistry, 3rd ed.* (the “Green Book”). Prepared by E. R. Cohen, T. Cvitaš, J. G. Frey, B. Holmstrom, K. Kuchitsu, R. Marquardt, I. Mills, F. Pavese, M. Quack, J. Stohner, H. Strauss, M. Takami, A. J. Thor, The Royal Society of Chemistry, Cambridge, UK (2007). Online version available; see <https://iupac.org/project/110-2-81>.
- [2] IUPAC. *Compendium of Chemical Terminology, 2nd ed.* (the “Gold Book”). Compiled by A. D. McNaught and A. Wilkinson. Blackwell Scientific Publications, Oxford, UK (1997). XML on-line advanced version (2006–2014), created by M. Nic, J. Jirat, B. Kosata: <https://goldbook.iupac.org/pages/about>.
- [3] D. Barton, K. Nakanishi, O. Meth-Cohn. (Eds.). *Comprehensive Natural Products Chemistry (9 Vols.)*, Elsevier, Oxford, UK (1999).
- [4] L. Mander, H.-W. Liu. (Eds.). *Comprehensive Natural Products II: Chemistry and Biology (10 Vols.)*, Elsevier, Oxford, UK (2010).
- [5] H.-D. Belitz, W. Grosch, P. Schieberle. *Food Chemistry*, 4th ed., Springer, Berlin (2009).
- [6] P. Crews, J. Rodríguez, M. Jaspars. *Organic Structure Analysis*, Oxford University Press, New York, NY (1998).
- [7] R. J. Molyneux, P. Schieberle. *J. Agric. Food Chem.* **55**, 4625 (2007).
- [8] B. L. Milman. *Chemical Identification and its Quality Assurance*. Springer-Verlag GmbH, Berlin, Heidelberg (2011).
- [9] J. A. Dale, H. S. Mosher. *J. Am. Chem. Soc.* **95**, 512 (1973).
- [10] G. A. Cordell, Y. G. Shin. *Pure Appl. Chem.* **71**, 1089 (1999).
- [11] S. Yannai. *Dictionary of Food Compounds with CD-ROM*, 2nd ed., CRC Press, Boca Raton, FL (2012).
- [12] E. Hahn-Deinstrop. *Applied Thin-Layer Chromatography. Best Practice and Avoidance of Mistakes*, 2nd ed., Wiley-VCH. (2006).
- [13] F. Feigl, V. Anger. *Spot Tests in Organic Analysis*, 7th ed., pp. 796, Elsevier Science, 2012. (Available as an eBook).
- [14] H. Wagner, S. Bladt. *Plant Drug Analysis: A Thin Layer Chromatography Atlas*, 2nd ed., Springer, Berlin (1996).
- [15] N. A. Anderton, P. A. Cockrum, S. M. Colegate, J. A. Edgar, K. Flower. *Phytochem. Anal.* **10**, 113 (1999).
- [16] A. I. Scott. *Interpretation of the Ultraviolet Spectra of Natural Products*, Pergamon, New York, NY (1964). (Available as an eBook at: <https://www.sciencedirect.com/science/book/9780080136158>).
- [17] T. O. Larsen, M. A. E. Hansen. Dereplication and discovery of natural products by UV spectroscopy, in: *Bioactive Natural Products: Detection, Isolation, and Structural Determination, Second Edition*, S. M. Colegate, R. J. Molyneux (Eds.), pp. 221–244, CRC Press, Boca Raton, FL (2007).
- [18] T. J. Mabry, K. R. Markham, M. B. Thomas. Reagents and procedures for the ultraviolet spectral analysis of flavonoids, in: *The Systematic Identification of Flavonoids*, T. J. Mabry, J. B. Harborne (Eds.), pp. 35–40, Springer, New York, Heidelberg, Berlin (1970).
- [19] O. M. Andersen, K. R. Markham. *Flavonoids: Chemistry, Biochemistry and Applications*, CRC Press, Boca Raton, FL (2005).
- [20] P. Larkin. *Infrared and Raman Spectroscopy. Principles and Spectral Interpretation*, Elsevier, Amsterdam, the Netherlands (2011).
- [21] Aldrich® Spectral Viewer™ FT-IR Library, 2002. (Available on CD-ROM) (>11,000 compounds).

- [22] Aldrich® Spectral Viewer™ ATR-IR Library, 2002. (Available on CD-ROM) (>18,000 compounds).
- [23] N. S. Mokgalaka, S. P. Lepule, T. Regnier, S. Combrinck. *Pure Appl. Chem.* **85**, 2197 (2013).
- [24] E. L. Eliel, S. H. Wilen, M. P. Doyle. *Basic Organic Stereochemistry*, Wiley-Interscience, New York, NY (2001).
- [25] M. M. Caja, G. P. Blanch, M. Herraiz, M. L. Ruiz del Castillo. *J. Chromatogr. A* **1054**, 81 (2004).
- [26] N. Berova, K. Nakanishi, R. W. Woody. (Eds.). *Circular Dichroism: Principles and Applications*, 2nd ed., Wiley-VCH, New York, NY (2000).
- [27] D. A. Lightner. The Octant Rule, Chapter 10, in: *Circular Dichroism: Principles and Applications*, 2nd ed., N. Berova, K. Nakanishi, R. W. Woody (Eds.), pp. 261–304, Wiley-VCH, New York, NY (2000).
- [28] N. Berova, K. Nakanishi. Exciton chirality method: principles and applications, Chapter 12, in: *Circular Dichroism: Principles and Applications*, 2nd ed., N. Berova, K. Nakanishi, R. W. Woody (Eds.), pp. 337–382, Wiley-VCH, New York, NY (2000).
- [29] H.-U. Humpf. Determination of the absolute configuration of bioactive natural products using exciton chirality circular dichroism. Chapter 6, in *Bioactive Natural Products: Detection, Isolation, and Structural Determination*, S. M. Colegate, R. J. Molyneux (Eds.), pp. 191–207, CRC Press, Boca Raton, FL (2008).
- [30] L. T. Byrne. Nuclear magnetic resonance spectroscopy: strategies for structural determination. Chapter 3, in *Bioactive Natural Products: Detection, Isolation, and Structural Determination*, S. M. Colegate, R. J. Molyneux (Eds.), pp. 77–111, CRC Press, Boca Raton, FL (2008).
- [31] H. Friebolin. *Basic One- and Two-Dimensional NMR Spectroscopy*, 5th ed., Wiley-VCH, Weinheim, Germany (2010).
- [32] J. T. Watson, O. D. Sparkman. *Introduction to Mass Spectrometry: Instrumentation, Applications, and Strategies for Data Interpretation*, 4th ed., Wiley, Chichester, UK (2009).
- [33] T. Kind, O. Fiehn. *Bioanal. Rev.* **2**, 23 (2010).
- [34] K. K. Murray, R. K. Boyd, M. N. Eberlin, G. J. Langley, L. Li, Y. Naito. *Pure Appl. Chem.* **85**, 1515 (2013).
- [35] J. H. Gross. *Anal. Bioanal. Chem.* **406**, 63 (2014).
- [36] T. Kind, O. Fiehn. *BMC Bioinform.* **8**, 105 (2007).
- [37] A. G. Brenton, A. R. Godfrey. *J. Am. Mass Spectrom.* **21**, 1821 (2010).
- [38] Wiley 11th Edition/NIST 2014 MS Library. <https://www.sisweb.com/software/wiley-registry.htm>.
- [39] M. T. Sheldon, R. Mistrik, T. R. Croley. *J. Am. Soc. Mass Spectrom.* **20**, 370 (2009).
- [40] M. Rychlik, P. Schieberle, W. Grosch. *Compilation of Odor Thresholds, Odor Qualities and Retention Indices of Key Food Odorants*, Deutsche Forschungsanstalt für Lebensmittelchemie, ISBN 3-9803426-5-4 (1998).
- [41] C. Schuh, P. Schieberle. *J. Agric. Food Chem.* **54**, 916 (2006).
- [42] NIST 17 Mass Spectral Library & Search Software (NIST 2017/2014/EPA/NIH). <https://www.sisweb.com/software/ms/nist.htm>.
- [43] European Commission. Commission Decision of 12th August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results. *Off. J. Eur. Commun.* **L221**, 8 (2002).
- [44] European Commission. Guidance document on analytical quality control and method validation procedures for pesticides residues and analysis in food and feed. SANTE/11813/2017. https://ec.europa.eu/food/sites/food/files/plant/docs/pesticides_mrl_guidelines_wrkdoc_2017-11813.pdf.
- [45] International Union of Crystallography. Crystallographic Information Framework. <https://www.iucr.org/resources/cif>.
- [46] The Cambridge Structural Database. <https://www.ccdc.cam.ac.uk/solutions/csd-system/components/csd/>.
- [47] RCSB Protein Data Bank. <https://www.rcsb.org/pdb/home/home.do>.