JOURNAL OF
AGRICULTURAL AND FOOD CHEMISTRY

PERSPECTIVE

Compound Identification: A *Journal of Agricultural and Food Chemistry* Perspective

RUSSELL J. MOLYNEUX* AND PETER SCHIEBERLE

U.S. Department of Agriculture, 800 Buchanan Street, Albany, California 94710, and
Lehrstuhl für Lebensmittelchemie, Technischen Universität München, Lichtenbergstrasse 4,
D-85748 Garching, Germany

This perspective is designed to summarize the standards that authors of manuscripts submitted to the *Journal of Agricultural and Food Chemistry* are expected to follow in establishing the structures of either new or unknown compounds identified in the course of a study. It is especially important that the molecular formulas of new compounds be determined by either high-resolution mass spectrometry or combustion analysis. All relevant physical, spectroscopic, and spectrometric data should also be reported, so that other research workers have criteria for comparison with compounds that may be isolated in the future. In the case of flavor and aroma constituents, it is not sufficient to depend upon mass spectrometric identifications based solely on comparison with commercial databases. Mass spectra and retention indices on GC stationary phases of different polarities must be determined and the results compared to data for reference compounds and with commercial standards, when available. If geometric or positional isomers may be present, or for chiral compounds, the retention indices of all isomers or enantiomers must be determined. Odor properties or odor thresholds determined by GC−olfactometry may also serve as appropriate tools for compound identification. Adherence to these standards will ensure that processing of manuscripts proceeds expeditiously and that the high standards of the *Journal* are maintained.

INTRODUCTION

The *Journal of Agricultural and Food Chemistry* deals primarily with chemistry topics, as its name implies. A significant number of manuscripts describe the isolation and identification of natural products from agricultural sources. This is particularly true for topics that fall within the "Bioactive Constituents" and "Flavors and Aromas" categories, although many other sections of the *Journal* touch, at least to some extent, on structural identification. The ultimate aim of most such research is to describe the biological and sensory activities of food and agricultural products, as constituted by a complex matrix of individual components. The credibility of the resulting papers is absolutely dependent on the rigorous and unequivocal characterization of, at the very least, those compounds (new or known) that contribute to the observed bioactivity. This does not mean that noncontributory compounds should be neglected, because it is also important to identify constituents, sometimes major, that have little or no effect on the raw material or product. Such constituents may serve the purpose of authenticating the source of the material under investigation or indicating potential contamination.

Why do we need to discuss this topic in a perspective? The answer is that modern separation techniques, especially when interfaced with exceptionally powerful spectroscopic methods, appear to have made the determination of structure a routine procedure. Unfortunately, dependence on such procedures, in combination with commercial compound libraries or databases that may contain errors, can give a false sense of security, unless they are applied with extreme care and informed judgment. Our purpose in writing this perspective is to set down the standards that authors are expected to follow in establishing the structures of compounds, so that processing of the manuscript proceeds in an orderly manner through review, revision, and editing to eventual acceptance with a minimum of time lost. It is important to note that these standards should be an inherent part of the initial experimental design, not requirements subsequently demanded by the Reviewers or Editors as a condition of publication.

Recent advances in instrumentation have led to the interfacing of separation techniques with modern spectroscopic techniques, not only UV and IR but also nuclear magnetic resonance spectroscopy (NMR) and mass spectrometry (MS). This exceptionally powerful combination has enabled the structures of compounds to be determined without the necessity of physical isolation, and has changed the traditional approach to natural products chemistry (*1*). It is generally now possible to separate

10.1021/jf070242j CCC: $37.00 © 2007 American Chemical Society
Published on Web 05/10/2007

virtually all components present in any extract of a particular organism and to determine the structures of each and every one, even though these may be present in extraordinarily low amounts. Minor constituents may exhibit exceptional potency in selected bioassays, whereas, at the same time, the contribution of natural products to bioactivity, flavor, or aroma is likely to depend upon complex interactions. Studies of such interactions involve much greater complexity than previously involved in natural product research; however, the sophisticated separation techniques and spectroscopic methods now available are capable of such tasks. In many cases the experimental approach is reduced to selecting the most economical, available, or elegant choice of techniques and devising suitable experiments within these constraints. Determination of structure has, therefore, turned into a relatively routine procedure, to be replaced by investigations of biological properties or of mode of action. Nevertheless, it is important that the standards that have traditionally been applied to the identification of compounds and structural elucidation be maintained to ensure that the results reported are fully accepted by the research community.

For the purposes of this discussion we have divided the compounds under consideration into two broad classes, conforming to the subcategories of the *Journal* mentioned above, namely, "Bioactive Constituents" and "Flavors and Aromas". This is not to imply that the standards to be applied are substantially different, but rather to recognize the technical difficulties that may exist in isolating and characterizing individual components. Bioactive constituents are generally isolated on the basis of a directed fractionation focused upon a particular bioassay of interest, for example, toxicity, antifungal, anti-insecticidal, or herbicidal activity. Because the bioassay is an inherently selective process, the number of compounds involved may not be great, numbering in the tens at the most, even though many other compounds may exist in the organism being studied. On the other hand, volatile constituents often number in the hundreds, a majority of which need to be identified before the specific flavor- or odor-active components can be identified by a combination of odor-activity thresholds and concentration. Furthermore, bioactive constituents are frequently solids of moderate to high molecular weight, whereas aroma constituents are generally relatively low molecular weight, volatile oils. In either case, the identity of the source organism must be comprehensive, with description to the subspecies, variety, or cultivar level if appropriate. Reference samples should be deposited at a specific location (herbarium, etc.) where they can be accessed by other scientists for comparison purposes.

BIOACTIVE CONSTITUENTS

Physical Data. It is the essence of organic chemistry that compounds need to be characterized as completely as possible, so that other research workers can compare directly data for compounds which may subsequently be isolated or retrospectively with data already in the literature. Such data include color, physical form, melting point if solid or boiling point if liquid, and optical rotation when compounds potentially have chiral centers. Melting points are particularly useful in this respect, especially when determined as mixed-melting points with reference compounds. They not only provide a convenient and easily determined point of comparison but are also a criterion of purity. With the advent of preparative HPLC there has been an unfortunate tendency to report isolates as "amorphous solids". In many cases, amounts obtained are insufficient to recrystallize to constant melting point, but there are a significant number of reports where tens of milligrams or even gram quantities have been isolated and no attempt has been made to recrystallize the sample.

The importance of melting points was recently illustrated by the NOVA television documentary *Forgotten Genius*, describing the life of the African-American chemist Dr. Percy Julian, a program partially funded by the American Chemical Society (*2*). Julian, working at DePauw University, a small liberal arts college, was in direct competition with F. E. King and Sir Robert Robinson at Oxford University, England, to complete the first formal synthesis of the indole alkaloid physostigmine, used for the treatment of glaucoma. When Robinson published the synthesis of eserethrole (*3*), easily convertible into physostigmine, Julian noticed that the reported melting point was almost 50 °C higher than the literature value for the compound and that the structure of Robinson's synthetic compound therefore had to be in error. This enabled Julian and Pikl (*4*) to publish the correct synthesis shortly thereafter.

When appropriate from examination of the structure, it is also essential to measure optical rotation, so that it can be identified as a particular enantiomer or a racemic mixture. This is particularly important in relation to bioactivity as different enantiomers often have entirely different biological properties. Obviously, unless purity has been established, optical rotation measurements are of little value.

Yields from the natural source should be reported, preferably on a dry weight basis. It is also expected that, whenever possible, reference samples will be retained and made freely available to other research groups on request, to provide comparisons with compounds that may subsequently be isolated.

The principle of dereplication, namely, the rapid identification of previously identified natural products (*5*, *6*), should be applied to all "novel" isolates and a thorough literature search conducted for structural analogies with known compounds from related organisms at the genus and family level. On many occasions the full power of modern instrumental techniques has been applied to compounds isolated from a new source only to have their structures established as identical to that of a natural product previously isolated from a different source. Dereplication can avoid such waste of time, money, and resources.

In particular, the standards of the *Journal* require that the molecular formula of a new compound **must** be determined either by high-resolution mass spectrometry (HRMS) or by combustion analysis. The latter is preferable as it provides confirmation of purity. If HRMS is used, it is essential to ensure that the ion under examination is the molecular ion. It should be noted that the molecular formula cannot be **established** from the low-resolution MS molecular ion and NMR carbon count, it can only be **inferred**. Such an approach is fraught with problems; the ion observed may be only a pseudo-molecular ion, and the signals of carbon atoms with long relaxation times may be so weak that they are not observed. There are many instances in the literature where this approach has led to unfortunate errors in structural identification.

Spectroscopic and Spectrometric Data. UV−visible and IR spectra should be provided whenever necessary for structural elucidation. These physical techniques can provide important information at a fraction of the time, effort, and expense involved with more sophisticated techniques. Furthermore, information such as the "fingerprint" spectrum from IR can provide an unequivocal comparison with authentic samples. Similarly, the addition of specific reagents to UV sample

solutions can induce spectroscopic shifts that are very diagnostic for particular structural features, as, for example, with flavonoids (*7*).

X-ray Crystallography. Recrystallization of solid samples may provide crystals suitable for establishing structure and stereochemistry by X-ray crystallography. It is important to ascertain that the crystalline material analyzed by this technique has spectroscopic data in accord with that of the bulk sample. On occasion, crystals have been selected for X-ray structure elucidation that represent minor impurities of the sample as a whole.

FLAVOR AND AROMA CONSTITUENTS

The term "flavor" comprises in particular the aroma and taste impression perceived when food is eaten. It is generally accepted that the aroma compounds causing the smell or odor are part of the volatile fraction, whereas taste-active compounds usually belong to the nonvolatile food constituents. Thus, compounds having the typical aromas of herbs or other plant materials have been isolated by steam distillation, for example, since ancient times. The steam volatile fractions, the so-called essential oils, have been used up to the present time either as fragrances or to modify or improve the aroma of foods. The majority of essential oil volatiles are terpenes, and, besides the tremendous number of terpenes occurring in nature, these compounds often contain two or more chiral centers and, thus, require certain efforts for structural elucidation. Besides these "natural" sources, biochemical or chemical degradation reactions of nonvolatile constituents during food processing or storage can also lead to the formation of additional volatile compounds.

The development of GC-MS coupling about 50 years ago has led to the identification of thousands of food volatiles. However, a detailed study of the literature indicates that many of these "identifications" are based just on assumptions made on the basis of fragmentation patterns obtained by mass spectrometric measurements, and quite often no syntheses or NMR measurements were performed. In addition, the development of headspace isolation techniques, such as solid-phase microextraction (SPME), has facilitated the process of volatile isolation. However, the combination of SPME isolation with mass spectrometric identifications based simply on the use of commercial databases, as performed by many authors, may be a source of errors.

It is without doubt that the exact identification of aroma compounds is a challenging task, because in a given food hundreds of volatiles are quite often present, additionally differing by many magnitudes in their concentrations. However, often a certain "bioactivity" of the volatiles, such as antimicrobial, antifungal, and, of course, sensorial properties, are the focus of a study. In such cases, the application of bioactivity-guided screening procedures may help to reduce the efforts needed in the identification of inactive compounds and, thus, avoid waste of time and resources. For example, it is now well-accepted in the literature that only those volatiles able to interact with the receptor proteins in the human olfactory bulb are responsible for a certain aroma. Therefore, if the aroma or aroma changes are the purpose of a study, the identification experiments should be focused on the odor-active volatiles. These compounds are usually present in a lower number as compared to the bulk of odorless volatiles and can be evaluated by screening techniques based on GC–olfactometry, such as the aroma extract dilution analysis (AEDA), the CHARM analysis, or others (*8*).

After the odor-active compounds have been located, the identification procedure usually starts with mass spectrometric measurements and the determination of retention indices relative to a series of homologues, such as *n*-alkanes. However, many authors compare the data obtained only with data available in commercial databases, but, as shown in the following two examples, this procedure may not be sufficient to unequivocally identify a volatile constituent or aroma compound. Thus, 5-methyl-2-furfurylthiol (**1**) and 2-(1-mercaptoethyl)furan (**2**) show very similar mass spectra (**Figure 1**). In this case, simply comparing mass spectra with data taken from the literature will obviously lead to erroneous results. However, on the basis of retention indices, which should at least be determined on two columns of different polarities, both compounds can easily be differentiated (**Table 1**) (*9*). Retention times, still used by many authors, are, however, not appropriate for identification purposes, because these cannot be standardized, for example, between different laboratories. On the other hand, odor qualities determined by GC–olfactometry can serve as an additional parameter in the identification experiments. For example, 5-methyl-2-furfurylthiol elicits a coffee-like odor, whereas 2-(1-mercaptoethyl)furan smells sulfury and burnt (**Table 1**).

Methoxyalkylpyrazines have been identified as important aroma compounds, for example, in peas, raw cocoa beans, or bell peppers, and 2-methoxy-3,5-dimethylpyrazine (**3**) (**Figure 2A**) has previously been characterized as an important odorant

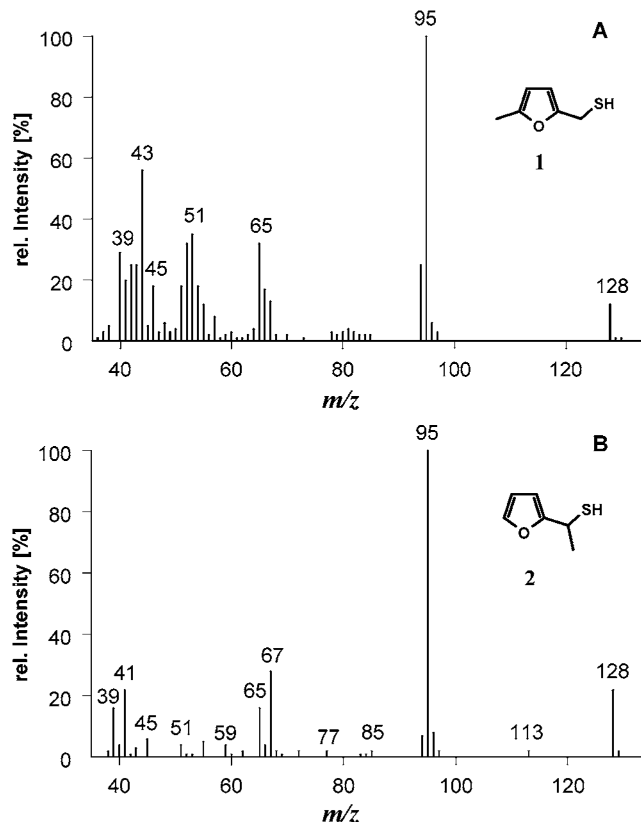

Figure 1. Mass spectra (MS-EI) of (**A**) 5-methyl-2-furfurylthiol (**1**) and (**B**) 2-(1-mercaptoethyl)furan (**2**).

Table 1. Analytical Data[a] for 5-Methyl-2-furfurylthiol (**1**) and 2-(1-Mercaptoethyl)furan (**2**)

odorant	odor quality	retention index DB-5	retention index FFAP	odor threshold (ng/L in air)
1	coffee-like	995	1500	0.006
2	sulfury, burnt	951	1396	0.009

[a] Hofmann and Schieberle (*9*).

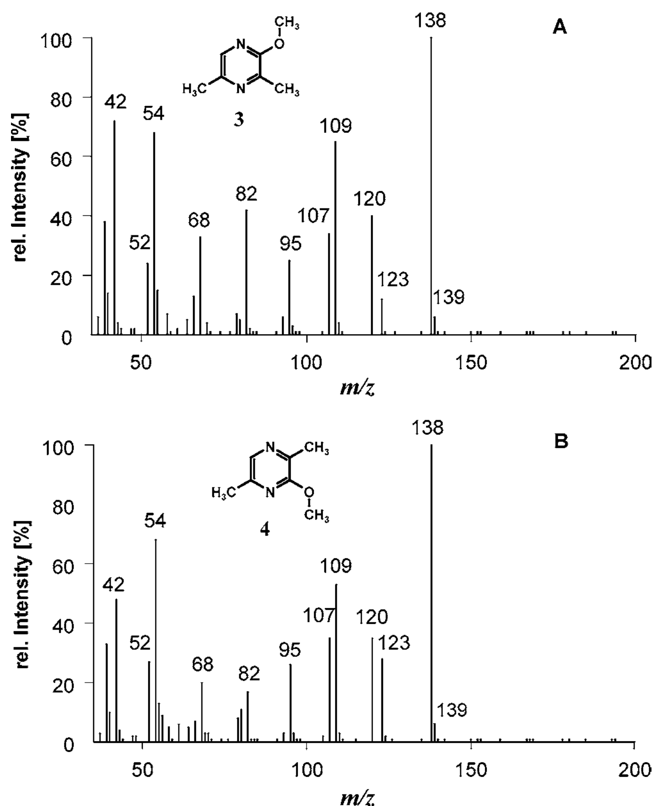

Figure 2. Mass spectra (MS-EI) of (**A**) 2-methoxy-3,5-dimethylpyrazine (**3**) and (**B**) 3-methoxy-2,5-dimethylpyrazine (**4**).

Table 2. Analytical Data for 2-Methoxy-3,5-dimethylpyrazine (**3**) and 3-Methoxy-2,5-dimethylpyrazine (**4**)

		retention index		
odorant	odor quality	DB-5	FFAP	odor threshold (ng/L in air)
3[a]	earthy	1055	1417	0.001
4[b]	pea-like	1057	1421	56

[a] Czerny and Grosch (*10*). [b] Michael Czerny, Ph.D. thesis, Technical University Munich, 1999.

in raw coffee beans (*10*). However, because several positional isomers of this compound might exist, synthesis of all possible isomers was the prerequisite to unequivocal identification of the correct structure. For example, 3-methoxy-2,5-dimethylpyrazine (**4**) (**Figure 2B**) shows a nearly identical mass spectrum. Also, for both pyrazines, the retention indices on two columns of different polarities (**Table 2**) are so close that an unequivocal identification on this basis is not possible. Because the odor qualities of both isomers are also similar, a determination of the odor thresholds by GC−olfactometry is the only way to differentiate between both compounds. Because pyrazine **3** is more odor-active by a factor of 56000 than pyrazine **4**, such identifications must be carefully performed and strictly require the use of reference compounds.

To avoid false identifications, the following procedure is recommended to unequivocally identify volatiles or aroma compounds:

(i) First, mass spectra (MS-EI; MS-CI) and retention indices must be determined on at least two different GC columns that have stationary phases of different polarities.

(ii) The results must be compared to spectra and retention indices of reference compounds. If literature data are used, it must be ascertained that mass spectrometric data and retention indices have been checked by the use of synthesized reference compounds. In this case the reference must be cited. If a compound is commercially available, the data must be checked in the author's laboratory, particularly if geometrical or positional isomers may be present.

(iii) For chiral compounds the retention indices of both enantiomers must be determined by using the respective enantiomers.

(iv) Odor qualities or odor thresholds determined by GC−olfactometry may serve as appropriate tools to aid in compound identification.

(v) If a new structure is detected (no data are available in the literature), two possibilities do exist: either enough material can be isolated, for example, by preparative GC for NMR measurements followed by synthesis, or, if the preparation of enough material is not possible, HRMS should be applied to obtain the elemental composition. A proposal for the structure can then be derived from all mass spectrometric data, but the suggested structure must be synthesized and confirmed by NMR measurements. If the target molecule and the synthesized reference agree in all mass spectrometric data, retention indices, and the respective bioactivity, NMR measurements of the analyte can be omitted.

Quite recently, such screening procedures have also been developed to elucidate compounds responsible for taste (*11*). However, the determination of the exact chemical structure is also the key step in this area of "bioactive" food constituents.

REPORTING AND PRESENTATION OF DATA

It is not the purpose of this perspective to deal with the specifics of reporting physical, spectroscopic, or spectrometric data. The *Journal* has adopted a format that conforms in general to that of other ACS journals. Examples of the proper description of such properties are given in the Guidelines for Authors, published in the first issue of each calendar year and also available online (*12*).

It should be noted that the physical and spectroscopic data must be grouped together for each specific compound under the Materials and Methods section of the manuscript. As a general rule, the presentation of data in the form of tables consumes valuable page space and should not be done unless it provides a convenient comparison of a series of closely related compounds. Similarly, figures of spectra should be included only if they are essential to illustrate points that would be difficult for the reader to visualize. Many submissions contain mass spectra that show, for example, the molecular ion and one or two fragment ions; such data are better represented in textual form. Even when the data are presented as text, there is no rule against providing spectra as "Supporting Information", available electronically at no charge. Constraints in figures do not apply to structural representations as these are the primary "shorthand" for chemical communication (*13*).

The data should not be listed again under Results and Discussion, but the relevant NMR signals, MS fragmentations, spectroscopic peaks, etc., should be clearly and concisely discussed in this section so that the reader can follow the arguments for derivation of the structure. If the compound is known and the data have been reported previously, the data should not be listed again unless there are serious discrepancies in values or there is evidence that assignments are in error. It is sufficient to state, for example, "Compound **X**: ^{1}H and ^{13}C NMR data consistent with the literature (*reference no.*)."

The editorial staff of the *Journal of Agricultural and Food Chemistry* is committed to ensuring the pre-eminent position

of the *Journal* in the publication of all aspects of chemistry dealing with agriculture and food. It is therefore important that such endeavors be supported by the most rigorous standards of chemistry. Adherence to the requirements provided in this perspective, and strict attention to the Author Guidelines, will ensure that both the quality and speed of publication, for which the *Journal* has an enviable reputation, will be maintained.

LITERATURE CITED

(1) Nakanishi, K. An historical perspective of natural products chemistry. In *Comprehensive Natural Products Chemistry*; Barton, D., Nakanishi, K., Meth-Cohn, O., Eds.; Elsevier: Amsterdam, The Netherlands, 1999; Vol. 1, pp xxiii−xl.

(2) Public Broadcasting System. *Forgotten Genius*, broadcast Feb 6, 2007; available at http://www.pbs.org/wgbh/nova/julian/ (accessed April 1, 2007).

(3) King, F. E.; Liguori, M.; Robinson, R. Experiments on the synthesis of physostigmine (eserine). Part X. *dl*-Noreseremethole and crystalline *dl*-eserethole. *J. Chem. Soc.* **1934**, 1416−1419.

(4) Julian, P. L.; Pikl, J. Studies in the indole series. IV. The synthesis of *d*,*l*-eserethole. *J. Am. Chem. Soc.* **1935**, *57*, 563−566.

(5) Cordell, G. A.; Shin, Y. G. Finding the needle in the haystack. The dereplication of natural product extracts. *Pure Appl. Chem.* **1999**, *71*, 1089−1094.

(6) Bradshaw, J.; Butina, D.; Dunn, A. J.; Green, R. H.; Hajek, M.; Jones, M. M.; Lindon, J. C.; Sidebottom, P. J. A rapid and facile method for the dereplication of purified natural products. *J. Nat. Prod.* **2001**, *64*, 1541−1544.

(7) Mabry, T. J.; Markham, K. R.; Barrie, T. M. *The Systematic Identification of Flavonoids*; Springer-Verlag: New York, 1970.

(8) Leland, J. V., Schieberle, P., Buettner, A., Acree, T. E., Eds. *Gas Chromatography−Olfactometry. The State of the Art*; ACSSymposium Series 782; American Chemical Society: Washington, DC, 2001.

(9) Hofmann, T.; Schieberle, P. Identification of potent aroma compounds in thermally treated mixtures of glucose/cysteine and rhamnose/cysteine using aroma extract dilution analysis. *J. Agric. Food Chem.* **1997**, *45*, 898−906.

(10) Czerny, M.; Grosch, W. Potent odorants of raw Arabica coffee. Their changes during roasting. *J. Agric. Food Chem.* **2000**, *48*, 868−872.

(11) Ottinger, H.; Soldo, T.; Hofmann, T. Discovery and structure determination of a novel Maillard-derived sweetness enhancer by application of the comparative taste dilution analysis (cTDA). *J. Agric. Food Chem.* **2003**, *51*, 1035−1041.

(12) Scope, Policy, and Instructions for Authors. *J. Agric. Food Chem.*, **2007**, *55*, 7A−14A; available at https://paragon.acs.org/paragon/ShowDocServlet?contentId=paragon/menu_content/authorchecklist/jf_authguide.pdf (accessed April 4, 2007).

(13) Hoffmann, R.; Laszlo, P. Representation in chemistry. *Angew. Chem., Int. Ed. Engl.* **1991**, *30*, 1−112.

Received for review January 28, 2007. Revised manuscript received April 4, 2007. Accepted April 5, 2007.

JF070242J