LIMNOLOGY
and
OCEANOGRAPHY: METHODS

Limnol. Oceanogr.: Methods 7, 2009, 269–276
© 2009, by the American Society of Limnology and Oceanography, Inc.

Expressing biomarker data in stoichiometric terms: shifts in distribution and biogeochemical interpretation

Robert J. Panetta and Yves Gélinas*
GEOTOP and Department of Chemistry and Biochemistry, Concordia University, Montréal, Québec, Canada

Abstract

Quantitative biomarker analysis is an invaluable tool used routinely by organic geochemists to interpret and explain environmental processes. Since the advent of organic geochemistry, all levels of biomarker methodology from wet chemistry to data interpretation have significantly advanced; however, an important aspect of data analysis has remained constant and that is the expression of biomarkers in terms of mass (e.g., mg gOC^{-1}). We argue that biomarkers are more appropriately expressed in terms of moles (e.g., mmol $molOC^{-1}$) to better reflect molecular-level distribution of compounds of interest, as well as introducing a chemical consistency for better comparability and transferability between data sets. Using modeled and real data culled from the literature, we demonstrate that the use of moles is not a trivial conversion and that distribution and relative weighting of biomarker data sets are shifted and affect the total data structure. The shift is sometimes strong enough to exert changes in interpretation (e.g., microbial abundance in an estuarine system), alter proxies (e.g., terrestrial-to-aquatic ratio), and potentially influence the use of multivariate statistical methods (e.g., principal components analysis).

Introduction

Compound-specific identification, or biomarker analysis, in organic geochemistry lends an added dimension to the understanding of the carbon cycle in environmental systems, in many cases highlighting subtle yet important details not revealed by bulk analyses alone. For example, the stable carbon isotopic signature of estuarine organic matter is generally proportional to the fraction of terrestrially derived material; however, the analysis of lignin oxidation products can shed light on its degradation state and provide more specificity on the origins of the terrestrial matter (e.g., Louchouarn et al. 1999). This higher level of understanding is invariably linked to the quantitative analysis of target biomarkers, and in particular to the calculation of the relative abundance of individual or groups of molecules to the total mass of quantified biomarkers. The analysis of source material (without which biomarker interpretation is extremely limited, as explained in Volkman et al. 2008), and its comparison to natural samples, has led to the development of a number of effective, simple, and robust biomarker-derived proxies that have advanced the understanding of carbon dynamics in environmental systems, including the terrestrial to aquatic ratio (TAR), the alkenone paleothermometer ($U_{37}^{K'}$), and average chain length (ACL), among several others (see Meyers 2003 and Eglinton and Eglinton 2008 for two thorough reviews on the topic).

The use of biomarker analysis in organic geochemistry arose from the overlapping interests of petrochemistry, geology, and ecology through the link between current biological production of organic matter to its deposition, diagenesis, and ultimate transformations to petroleum products. Alkanes, fatty acids, amino acids, polycyclic aromatic hydrocarbons, lignin, pigments, etc., were identified in a number of studies attempting to address the source of natural organic matter (e.g., Treibs 1936), its transformations in the environment (e.g., Brown et al. 1972), and how it might be converted to petroleum through geological processes (e.g., Breger 1960).

Historically, the most widely studied group of biomarkers are lipids, broadly defined as any compounds extractable with organic solvents such as dichloromethane or hexane, which comprise hundreds of distinct organic chemicals. For simplicity, lipids can be liberally separated into three branches: (1) hydrocarbons, (2) alkanoic acids, and (3) sterols. Other biomarkers that are routinely reported in biogeochemical studies

*Corresponding author: E-mail: rpanetta@live.concordia.ca

Acknowledgments

This work was funded by the National Science and Engineering Research Council of Canada (NSERC), the Fonds québécois de la recherche sur la nature et les technologies (FQRNT), and the Canadian Foundation for Innovation (CFI). The authors extend thanks to Karine Lalonde, Alexandre Ouellet, and Mina Ibrahim for helpful discussion and comments on the concept presented in this paper. This is GEOTOP publication number 2009-0003.

include lignin phenols (e.g., Hedges and Parker 1976, Louchouarn et al. 1999), pigments (Schubert et al. 2005), and amino compounds (e.g., Kaiser and Benner 2008), among others (Killops and Killops 2004).

In the time that has passed since rock extracts were first analyzed by gas chromatography, analytical methodology has become more sophisticated and specialized, particularly over the last decade, but the general approach remains the same: extraction, fractionation, chromatographic separation, and quantitation of the compounds of interest. As data analysis has progressed and still evolves, various biomarker proxies identifying the source materials, diagenetic state, and reaction pathways have emerged from the quantitative data, and some common ones are summarized in Table 1. Biomarker-based proxies are invaluable tools, but because they are based on a limited number of compounds they can potentially lead to an oversimplification or generalization of a given system when addressing total organic carbon sources or diagenetic processes. Because a biomarker analysis can result in the identification and quantification of sometimes hundreds of compounds in a single analysis, using all the information obtained (i.e., all biomarkers identified and quantified) would intuitively result in a more powerful interpretation for the system under study (e.g., Oros and Simoneit 2000). Such complex data sets are not easily deciphered, however, and statistical tools more capable of sifting and sorting data of this nature have recently been put to use, such as principal components analysis (Zimmerman and Canuel 2001, Yunker et al. 2005), clustering analysis (Dittmar et al. 2007), and multivariate curve resolution (Salau et al. 1997). These methods all depend on the precise quantitation of individual compounds and their weighting relative to other compounds within their class. These approaches will likely become more prominent in the search for unique or novel biomarker proxies for additional robust and far-reaching interpretations of environmental processes, as shown by the use of PCA in the development of the amino acid degradation index (Dauwe and Middelberg 1998).

Whereas biomarker data analysis has significantly evolved, an important characteristic remains constant, at least for lipids and lignin: the expression of the concentration of individual compounds in mass yields. This can be attributed to habit borne of day-to-day simplicity: a standard solution containing *n* mg per unit volume is prepared to determine the response factor of the detector used. Even today, the concentration of most commercial analytical standards used in biomarker analysis is given in mg per liter (e.g., ASTM Petroleum standards for gas chromatographic analysis of *n*-alkanes). Thus biomarker analysts express concentrations of lipid biomarkers including alkanes, alkanoic acids, sterols, *n*-alcohols, among many others, as either μg g^{-1} of dry weight or mg g of organic carbon $(OC)^{-1}$; lignin phenols as mg 100mgOC^{-1}; polycyclic aromatic hydrocarbons (PAH) as part per billion dry weight or mass percent of total PAH, etc. In contrast, amino acids and amino sugars are conventionally expressed as moles, a

Table 1. Some common proxies and their interpretation.

Proxy	Formula	Interpretation
Terrestrial to aquatic ratio[a]	$TAR_{HC} = \frac{C_{27}+C_{29}+C_{31}}{C_{15}+C_{17}+C_{19}}$ $TAR_{FA} = \frac{C_{24}+C_{26}+C_{28}}{C_{12}+C_{14}+C_{16}}$	Used to interpret terrestrial versus aquatic dominance
Alcohol preservation index[b]	$API = \frac{C_{24}OH+C_{26}OH+C_{28}OH}{\left(C_{24}OH+C_{26}OH+C_{28}OH\right)+(C_{24}+C_{26}+C_{28})}$	Infer TOC diagenesis
Alkenone paleothermometer[c]	$U_{37}^{K'} = \frac{\left[C_{37:2}\right]}{\left[C_{37:2}\right]+\left[C_{37:3}\right]}$	Surface water temperature
Carbon preference index[c]	$CPI_{HC} = \frac{2\times\left[\sum Odd\ nC_{23} to\ nC_{31}\right]}{\sum Even\ nC_{22} to\ nC_{30} + \sum Even\ nC_{24} to\ nC_{32}}$ $CPI_{FA\ or\ OH} = \frac{2\times\left[\sum Even\ nC_{22} to\ nC_{30}\right]}{\sum Odd\ nC_{21} to\ nC_{29} + \sum Odd\ nC_{23} to\ nC_{31}}$	Terrestrial versus aquatic source indicator
Average chain length[d]	$TAR_{FA} = \frac{27\times C_{27}+29\times C_{29}+31\times C_{31}}{C_{27}+C_{29}+C_{31}}$	Differentiate grasses from leaves/temperature

[a]Bourbonniere and Meyers (1996)were the first to formulate the equation as presented.
[b]Poynter and Eglinton (1990).
[c]Bray and Evans (1961) and Kvenvolden (1966) formulated the hydrocarbon and alkanoic acid/alkanol equations, respectively.
[d]Cranwell (1973) first presented the equation.

historical consequence of the fact that these analyses were conducted by biochemists interested in molar ratios of these biochemicals (e.g., Benson and Hare 1975).

Although it may be more practical in the laboratory to use mass units, one basis of biomarker data interpretation is fueled by the idea of comparing the total amount of one molecule present in a sample to another, in other words stoichiometric ratios. Stoichiometric expression also allows direct comparability between different classes by introducing molar consistency. The conversion from mass to stoichiometric units is not trivial and results in shifts of the weighted distribution of some biomarker profiles with potential to alter biomarker interpretations, especially those making use of multiple biomarker "fingerprints" that incorporate full suites of compounds (e.g., Oros and Simoneit 2000, Zimmerman and Canuel 2001, Yunker et al. 2005). In this article, theoretical treatments of gaussian-modeled and randomly generated data sets representing a number of biomarker classes, complemented by recalculated biomarker data from the literature, are used to illustrate the impact of a mass- to a mole-based conversion of biomarker data on their environmental interpretations.

Analysis

Data sets—Two flavors of modeled data sets comprised of m samples x n variables (compounds) were generated with routines developed in MATLAB (Mathworks; version 7.0, release 14). In all cases, there are 50 samples, whereas the number of compounds varies between biomarker classes (12 for lignin phenols and up to 31 for fatty acids, for example). The first set, expressed as mass percent, is based on a bimodal gaussian distribution, designed to mimic a transition from high molecular weight– to low molecular weight–dominated envelopes of long-chain lipid biomarkers (*n*-alkanes, *n*-alkanols, and *n*-alkanoic acids) as is found in many natural samples. The second is a random data set representing the long-chain lipid biomarkers as well as the commonly reported sterols, polycyclic aromatic hydrocarbons, and lignin oxidation products, expressed as mass percent. The random data are generated as a control to the modeled sets to ensure that any differences resulting from changes in units are not due to trends intrinsic to the ordered distribution of gaussian data. A string of random values was created to represent the summed biomarkers. Each value was then converted in a stepwise fashion from relative mass % to mg gOC^{-1} or mg $100mgOC^{-1}$ for lignin, to μmol $molOC^{-1}$, and finally to a mole fraction (as %). All data are normalized to the amount of organic carbon to eliminate the influence of mineral dilution and because biomarkers are typically used to reflect the origin and fate of organic carbon. Direct comparisons are made between absolute (mg gOC^{-1} versus μmol $molOC^{-1}$) and relative (mass percent versus mole fraction) concentrations.

Standardized data—Because there is no consensus or treatise on how biomarkers should be analyzed using multivariate statistical approaches that exploit factors controlling data structure and variability, we investigated how data distribution patterns within a sample and how overall data structure can be affected. One of the best approaches to render data sets comparable with respect to distribution and variability is to autoscale, a procedure commonly used before multivariate analyses. Known in Microsoft Excel 2002 as the "standardize" function, autoscaling first mean-centers a data point then divides by the standard deviation:

$$Sc = \frac{x - \bar{x}}{\mu}, \quad (1)$$

where x is the data point in question, $\bar{x}$ is the mean of all data for that sample (e.g., the mean biomarker concentration for 31 fatty acids quantified in a single sample), μ is the standard deviation, and Sc is the unitless scaled value. This procedure eliminates the effects of absolute intensity differences and scales the data so that variance is unity and all values occur over the same range (± 1.0 units), maintaining data structure, but allowing direct comparisons of data sets through the same lens.

Calculations and interpretation—We conducted a discussion and interpretation of data based solely on relative distributions for data sets to highlight the effect of the choice of units on the interpretation of biomarker data. Published data that are recalculated and referred to in the text include a coal smoke particulate fingerprinting study (Oros and Simoneit 2000), lipids extracted from Arctic Ocean sediments (Belicka et al. 2002), estuarine dissolved organic matter (McCallister et al. 2006), estuarine sediments (Colombo et al. 1997), and lignin phenols isolated from source materials (Hedges and Parker 1976).

Results and discussion

Weighting redistribution—Figure 1A shows a modeled *n*-alkane profile depicting a typical riverine-influenced system as either mg gOC^{-1} or μmol $molOC^{-1}$. The general bimodal trend exists in both mass and mole-normalized profiles, but it is clear that the abundance of low-molecular-weight components is elevated in mole-normalized data relative to high-molecular-weight components, with the ratio of the major peaks of each envelope, C_{17} and C_{27}, increasing from 0.51 to 0.77, or by 51%. This general trend is true for all biomarkers, regardless of class; however, as shown in Fig. 2, the factor by which the biomarker concentration changes is not constant, thus stoichiometric-based expression may lead to differing interpretation and conclusions compared to mass-based expression. This is especially true for families of biomarkers that cover a wide range of molecular masses (~100–600 amu), such as alkanes, fatty acids, or long-chain alcohols, and potentially other classes that are typically on the low molecular weight end (<300 amu) of the range, such as lignin phenols.

Gaussian-modeled alkanoic acid data in the form of mg gOC^{-1} and μmol $molOC^{-1}$ were autoscaled, and we calculated the difference between the two standardized ratios (henceforth "residual"). If the distribution between data sets had remained identical, the residual would be equal to zero, but if the weighting shifted, residuals would be nonzero values. Figure 3 shows

Fig. 1. Redistribution of relative weighting of biomarker profiles when converting mass-based units to stoichiometric-based units. Numbers below *x*-axis in (A–C) denote carbon chain length. (A) Gaussian modeled *n*-alkane profile of individual biomarkers as mg gOC^{-1} or µmol $molOC^{-1}$. (B) Fatty acid profile, including *n*-alkanoic and alkenoic acids, reported for St. Lawrence Estuary sediments as ng gOC^{-1} or nmole $moleOC^{-1}$ (Colombo et al. 1997). (C) *n*-Alkane homologous series identified in lignite coal smoke particulate matter expressed as mg $Kg(d.w.)^{-1}$ or µmol $Kg(d.w.)^{-1}$ (Oros and Simoneit 2000). (D) Lignin phenols resulting from CuO oxidation of *Pinus Caribeae* expressed as mg $100mgOC^{-1}$ and nmol $molOC^{-1}$ (pOH, Syr, and Van are *p*-hydroxybenzoic, syringyl, and vanillyl derivatives, respectively) (Hedges and Parker 1976). In all profiles, the relative proportion of low-molecular-mass components increases, and in the case of (C) and (D), the major compound is changed.

a plot of the residuals as a function of molecular mass for three modeled samples (A and B are dominated by low- and high-mass alkanoic acid envelopes, respectively, and M is a mixture of the two), and clearly the scaled values are not identical. Generally, the residuals are positive on the low molecular mass side and decrease with molecular weight, reflecting a shift to higher relative abundances for low-mass molecules. The overall trend is for greater deviation between normalizations when the relative abundance of both the high- and the low-mass compounds are not negligible, further emphasizing how individual data structures can be altered in different manners.

Biomarkers are often expressed as the percent contribution to the total summed biomarkers, or relative to the major compound in a set, allowing easy comparison between data sets that differ by orders of magnitude in their absolute amounts of target biomarker class. Such relative abundances also lay the foundation for quantitative biomarker interpretation, and an

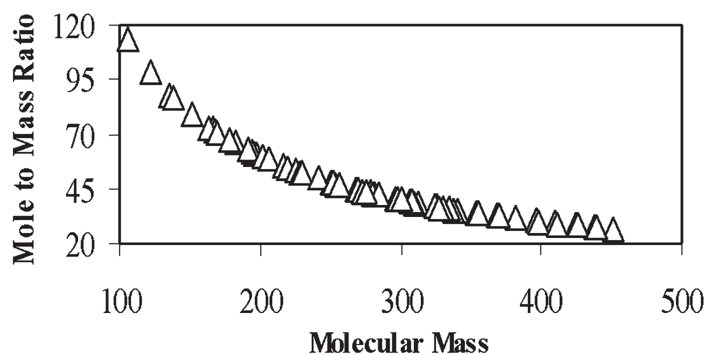

Fig. 2. The ratio of molar (µmol $molOC^{-1}$) to mass (mg gOC^{-1}) normalized data for individual gaussian and randomly modeled alkyl (*n*-alkanes, *n*-alkanoic acids, and *n*-alkanols) and ring (sterols, polycyclic aromatic hydrocarbons, lignin phenols) biomarkers. Low-molecular-weight components are more strongly altered than the higher-molecular-weight components.

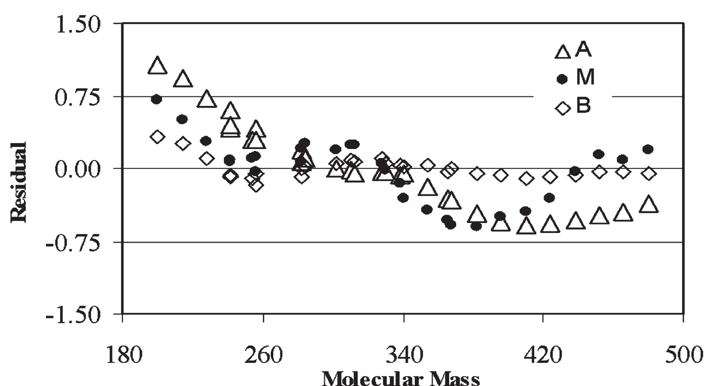

Fig. 3. Modeled *n*-alkanoic acid data expressed as mg gOC^{-1} and µmol molOC^{-1} are autoscaled. Residuals (difference between the scaled mass- and molar-based values) are plotted as a function of molar mass. Shown are three samples (two end-members and a mixture).

elevation in the relative amount of a compound from one sample to another implies an increasing contribution of the source of that biomarker to the total organic carbon pool of a sample. The expression of biomarkers as a stoichiometric fraction results in the same trends as the absolute concentration, as discussed above.

Figure 1(B–D) shows how the overall distribution of biomarkers in real samples is altered when concentrations are mole-normalized. For *n*-alkanoic acids isolated from the St. Lawrence Estuary (Fig. 1B), the dominant compound in the series clearly remains the 7-hexadecenoic acid (C_{16}), but the relative abundances of acids greater than 21 carbons in length are significantly reduced relative to the smaller compounds. For example, hexacosanoic acid (C_{26}) is lowered from 78% to 55% of the relative abundance of 7-hexadecenoic acid. The *n*-alkanes of lignite (coal) combustion products (Fig. 1C) show significant changes in their entire distribution, the most notable being the dominant *n*-alkane changing from *n*-C_{29} to *n*-C_{23}. Interesting differences in the shape of the main envelope are also incurred, where a smooth gaussian shape is no longer rounded at the apex but plateaus between the C_{23} and C_{25}. The shift in relative distribution is not limited to alkyl chain–based biomarkers, as the lignin phenol profile of *Pinus Caribeae* wood shows (Fig. 1D). When expressed as mass (mg 100mgOC^{-1}), syringaldehyde is dominant over vanillin and *p*-hydroxybenzoic acid (100:82:79 ratios, respectively), but as moles (µmol molOC^{-1}), the ratio is significantly altered, so that all three are almost equally proportioned (100:104:98).

These examples show how the relative distribution may change when biomarkers are expressed as moles and that these differences are not limited to a single class or family of compounds. They also highlight the subtlety by which different samples can be affected, including the "shape" of the data distribution, as well as the abundance order of a biomarker profile.

Extending the average chain length calculation (Table 1) to include all alkyl functionality compounds identified (ACL_{TOT}) provides an additional measure of data redistribution upon relative weighting. Again looking at the data found in Oros and Simoneit (2001), the recalculated ACL_{TOT} of some components remain invariant (e.g., *n*-alkenoic acids of subbituminous coal), whereas for *n*-alkanoic acids the change varies from a decrease of 0.5 units to as many as 2 full units of chain length (Table 2). In McCallister et al. (2006), the majority of estuarine fatty acids identified are <22 carbons in length, and for 12 samples reported, the average ACL very slightly decreases from 16.8 to 16.6 when switching from mass-based to mole-based concentrations; thus in this case redistribution of abundance does not occur. In another example, the ACL_{TOT} of saturated *n*-alkanoic acids isolated from nine locations in the Arctic Ocean (Belicka et al. 2002) decrease on average 0.8 units, from 20.2 to 19.4, with the largest decrease by 1.1 units. It is clear that the changes do not affect all samples in a similar manner, and generally, but not as a rule, samples with higher ACL_{TOT} values are affected most strongly.

Biomarker interpretation—Because the relative abundance of low-molecular-mass compounds generally increases when expressed in molar fractions, interpretation based on their relative contributions may change as well. Biomarker classes in which the lowest molecular mass is >350 amu, such as sterols, hopanoids, and long-chain alkenones, are not significantly altered. Although individual relative abundances change significantly, the interpretation of PAHs and lignin

Table 2. Shifts in the ACL_{TOT} of alkyl substituent biomarker classes of four types of coal smoke particulate matter based on either mass (mg kg^{-1}) or moles (nmol kg^{-1}).

	Lignite			Brown			Subbituminous			Bituminous		
	Mass	Moles	Difference	Mass	Moles	Difference	Mass	Moles	Difference	Mass	Moles	Difference
Alkan-2-ones	26.1	25.0	–1.1	26.0	25.1	–0.9	24.5	23.4	–1.1	21.8	20.8	–0.9
Alkanals	24.2	22.9	–1.3	22.4	21.4	–1.0	—	—	—	—	—	—
Alkanes	26.0	25.3	–0.7	25.9	25.1	–0.8	25.6	25.0	–0.6	23.5	22.8	–0.7
Alkanoic acids	28.5	28.0	–0.5	24.6	23.8	–0.8	22.9	20.9	–2.0	16.5	15.7	–0.8
Alkanols	26.9	26.3	–0.6	25.7	25.5	–0.2	—	—	—	—	—	—
Alkenes	24.1	23.6	–0.5	24.7	24.4	–0.3	20.8	20.7	–0.1	20.3	19.9	–0.5
Alkenoic	17.9	17.5	–0.4	—	—	—	17.5	17.1	–0.4	18.0	18.0	0.0
Alkylbenzenes	26.0	25.1	–0.9	23.6	22.6	–1.0	23.8	23.1	–0.7	21.2	20.4	–0.7

Data recalculated from Oros and Simoneit 2000.

phenols are very little affected, but this is because of the manner in which they are treated, with most proxies comparing structural isomers or compounds of relatively close molecular mass (see Louchouarn et al. 1999 and Lima et al. 2005 for examples of lignin and PAH proxies, respectively). A brief treatment of published lignin data (Hedges and Parker 1976) illustrating changes in the relative distribution of individual phenols, but little change in interpretation of proxies, is found in Web Appendix 1.

The vast majority of biomarker work is derived from lipids; thus consistency in their reporting is of pivotal importance. Using a combination of relative contributions and biomarker proxies, the quantitative analysis of this most widely used class of compounds has been used to infer source apportionment in recent (e.g., Colombo et al. 1997, Yunker et al. 2005, McCallister et al. 2006) and ancient (e.g., Feakins et al. 2007) samples, to estimate microbial biomass (Parkes and Taylor 1983), to evaluate diagenetic alterations within settling particles (e.g., Wakeham et al. 1997) and subsequent sedimentary processes (e.g., Madureira et al. 1995), and for paleoecological reconstruction (e.g., Zimmerman and Canuel 2001, Sluijs et al. 2006), among many other environmental and ecosystem conditions and processes examined (see Eglinton and Eglinton 2008 for a recent review). Any changes in the data distribution within this class of compounds could potentially impact the interpretations of a wide swath of biogeochemical processes.

Proxies based on compounds within a narrow mass window—such as $U_{37}^{K'}$, in which molecules differing by 2 amu are compared—or those that make use of the full range of compounds, well distributed over both the numerator and denominator—such as the CPI (see Table 1 for full equations of some lipid biomarker proxies)—are the least affected. The proxies that are most altered are those comparing high- to low-molecular-weight components, such as the TAR ratios.

The TAR ratio is defined as the quotient of the concentrations of either odd chain length C_{27}–C_{31} over C_{15}–C_{19} *n*-alkanes (TAR_{HC}) or even chain length C_{26}–C_{30} over C_{12}–C_{16} *n*-alkanoic acids (TAR_{FA}) (Bourbonniere and Meyers 1996). Higher TAR values are generally attributed to terrestrial sources and lower values as aquatic. Because the relative contribution of low-molecular-weight components (the denominators in TAR equations, Table 1) tends to increase when mass-based concentrations are converted to mole-based ones, both stoichiometric TAR ratios give considerably lower values compared to their mass-based TAR counterparts. This is demonstrated with modeled and random data sets where the molar-based TAR ratios are only 56% to 60% that of the mass-based ratios, generally trending toward a greater disparity between the differently normalized sets as the TAR decreases. A similar trend is found for Arctic Ocean sedimentary fatty acids (Belicka et al. 2002), where the TAR_{FA} decreases by a factor of 0.64 ± 0.01 for nine samples upon conversion to molar units. As long as the TAR value is greater than 0.1, this decrease by a factor of ~0.6 holds true for all literature data tested so far. Although the change in this proxy is relatively constant, it highlights an important concept: the abundance of molecules considered of terrestrial origin is overestimated relative to that of planktonic/microbial compounds when mass-based units are used.

Besides molecular proxies, a very common approach is to report the relative contribution from a subclass of compounds to a total fraction, such as the relative proportion of phospholipid fatty acids to the total pool of identified fatty acids. In some cases, this relative proportion remains more or less impartial to mass or mole normalizations, but in others, important processes can be overlooked with mass-based concentrations. For example, in an estuary with very little terrestrial biomarker influence (McCallister et al. 2006), mass-normalized data clearly show that the relative contribution of branched fatty acids in the dissolved phase remains invariant at ~10% of total fatty acids across a salinity gradient, implying constant bacterial contributions to this pool of organic matter. Upon conversion of the dataset to moles, however, a clear spike in the relative contribution of these bacterial markers is found in the marine end-member during the fall sampling season, with about 16% of the total fatty acids pool. This implies increased microbial biomass at this location that was not initially detected in the mass-based data set.

We recalculated the relative abundance of major alkyl substituent compound classes from the extensive coal smoke fingerprinting study of Oros and Simoneit (2000) (Table 3). In lignite, the most abundant compound class remains the *n*-alkanoic acids; however, their relative abundance decreases from 35.5% to 31.5%, whereas *n*-alkanols, *n*-alkanes, and *n*-alkenes, originally separated within a range of 3.1% (15.6%–18.7%) become almost equal, separated by only 0.5% in stoichiometric abundance (17.7%–18.2%). In brown coal, the order of most abundant compound classes remains the same, but a noticeable change occurs in the difference between *n*-alkanoic acids and *n*-alkenes shrinking from 2.3% to 0.8%, making them almost equal as the most abundant class. In subbituminous coals, the relative abundances of alkyl class compounds is not significantly altered. The most interesting and significant change occurs with bituminous coal, where as a function of mass, the relative abundance of *n*-alkanes and *n*-alkanoic acids are equal (30.3% versus 30.1%), but as a function of moles, the *n*-alkanoic acids are 7.1% more abundant than the *n*-alkanes, owing to the high abundance of relatively shorter-chain components present in the acid fraction, as attested by the ACL_{TOT} (Table 1). There is a general trend of increasing relative abundance with decreasing chain length, but again it is neither a constant nor easily modeled change across all samples.

Conclusions

The underlying goal driving quantitative biomarker analysis is the determination of the quantity of a given compound originating from one source compared to another, with the aim of deriving information on past or present environmental conditions and/or processes. We argue here that quantitative

Table 3. Shifts in the relative proportion (% of total composition) of alkyl substituent biomarker classes of four types of coal smoke particulate matter based on either mass (mg gOC^{-1}) or moles (nmol $molOC^{-1}$).

	Lignite			Brown			Subbituminous			Bituminous		
	Mass	Moles	Difference	Mass	Moles	Difference	Mass	Moles	Difference	Mass	Moles	Difference
Alkan-2-ones	3.7	3.8	0.1	15.6	15.0	–0.6	5.3	5.2	–0.1	22.1	20.6	–1.5
Alkanals	1.3	1.4	0.2	6.2	7.0	0.8	—	—	—	—	—	—
Alkanes	17.0	17.9	1.0	20.8	20.7	–0.1	60.7	57.8	–2.9	30.3	27.2	–3.2
Alkanoic acids	35.5	31.5	–4.0	25.9	25.1	–0.8	11.3	11.6	0.4	30.1	34.2	4.1
Alkanols	18.7	18.2	–0.5	5.4	5.1	–0.3	—	—	—	—	—	—
Alkenes	15.6	17.7	2.1	23.6	24.3	0.8	9.4	10.9	1.5	11.6	12.0	0.4
Alkenoic	1.2	1.7	0.4	—	—	—	2.4	3.0	0.6	0.7	0.7	0.0
Alkylbenzenes	7.1	7.8	0.6	2.6	2.9	0.3	10.9	11.5	0.6	5.2	5.3	0.1

Data recalculated from Oros and Simoneit 2000.

values should be expressed as moles to best reflect these molecular relationships, because moles, rather than grams, are the currency of molecules. Furthermore, the emergence and development of multivariate statistical analyses of biomarker data sets calls for consistency in the expression of all biomarkers to allow better comparability and transferability between not just different samples, but also different molecules within one sample. The aim of this work was to show that the conversion from mass-based units to mole-based units in biomarker analysis is not a trivial matter, but a calculation with potential to affect the way biomarker data are shaped and interpreted. This is an especially important point, as the drive to develop novel molecular proxies through better understanding of the relationship between biochemistry and environmental conditions has never been stronger (Eglinton and Eglinton 2008).

References

Belicka, L. L., R. W. Macdonald, and H. R. Harvey. 2002. Sources and transport of organic carbon to shelf, slope, and basin surface sediments of the Arctic Ocean. Deep-Sea Res. I 49:1463-1483.

Benson, J. R., and P. E. Hare. 1975. *o*-Phthalaldehyde: fluorogenic detection of primary amines in the picomole range: comparison with fluorescamine and ninhydrin. Proc. Nat. Acad. Sci. U. S. A. 72:619-622.

Bourbonniere, R. A., and P. A. Meyers. 1996. Sedimentary geolipid records of historical changes in the watersheds and productivities of Lakes Ontario and Erie. Limnol. Oceanogr. 41:352-359.

Bray, E. E., and E. D. Evans. 1961. Distributions of *n*-paraffins as a clue to the recognition of source beds. Geochim. Cosmochim. Acta. 27:1113-1127.

Breger, I. A. 1960. Diagenesis of metabolites and a discussion on the origin of petroleum hydrocarbons. Geochim. Cosmochim. Acta. 19:297-308.

Brown, F. S., M. J. Baedecker, A. Nissenbaum, and I. R. Kaplan. 1972. Early diagenesis in a reducing fjord, Saanich Inlet, British Columbia—III. Changes in organic constituents of sediment. Geochim. Cosmochim. Acta. 36:1185-1208.

Colombo, J. C., N. Silverberg, and J. N. Gearing. 1997. Lipid biogeochemistry in the Laurentian Trough—II: Changes in composition of fatty acids, sterols and aliphatic hydrocarbons during early diagenesis. Org. Geochem. 26:257-274.

Cranwell, P. A. 1973. Chain-length distribution of *n*-alkanes from lake sediments in relation to post-glacial environmental change. Freshwater Biol. 3:259-265.

Dauwe, B., and J. J. Middelburg. 1998. Amino acids and hexosamines as indicators of organic matter degradation state in North Sea sediments. Limnol. Oceanogr. 43:782-798.

Dittmar, T., K. Whitehead, E. C. Minor, and B. P. Koch. 2007. Tracing terrigenous dissolved organic matter and its photochemical decay in the ocean by using liquid chromatography/mass spectrometry. Mar. Chem. 107:378-387.

Eglinton, T. I., and G. Eglinton. 2008. Molecular proxies for paleoclimatology. EPSL 275:1-16.

Feakins, S. J., T. I. Eglinton, and P. B. deMenocal. 2007. A comparison of biomarker records of northeast African vegetation from lacustrine and marine sediments ca. 3.40 Ma. Org. Geochem. 38:1607-1624.

Hedges, J., and P. L. Parker. 1976. Land-derived organic matter in surface sediments from the Gulf of Mexico. Geochim. Cosmochim. Acta. 40:1019-1029.

Kaiser, K., and R. Benner. 2008. Major bacterial contribution to the ocean reservoir of detrital organic carbon and nitrogen. Limnol. Oceanogr. 53:99-112.

Killops, S. D., and V. J. Killops. 2004. An introduction to organic geochemistry. New York: Wiley-Blackwell.

Kvenvolden, K. A. 1966. Molecular distributions of normal fatty acids and paraffins in some lower cretaceous sediments. Nature 209:573-577.

Lima, A. L. C., J. W. Farrington, and C. M. Reddy. 2005. Combustion-derived polycyclic aromatic hydrocarbons in the environment: a review. Environ. Forensics 6:109-131.

Louchouarn, P., M. Lucotte, and N. Farella. 1999. Historical and geographical variations of sources and transport of terrigenous organic matter within a large-scale coastal environment. Org. Geochem. 30:675-699.

Madureira, L. A. S., M. H. Conte, and G. Eglinton. 1995. Early diagenesis of lipid biomarker compounds in North Atlantic Sediments. Paleoceanogr. 10:627-642.

McCallister, S. L., J. E. Bauer, H. W. Ducklow, and E. A. Canuel. 2006. Sources of estuarine dissolved and particulate organic matter: a multi-tracer approach. Org. Geochem. 37:454-468.

Meyers, P. A. 2003. Applications of organic geochemistry to paleolimnological reconstructions: a summary of examples from the Laurentian Great Lakes. Org. Geochem. 34:261-289.

Oros, D. R., and B. R. T. Simoneit. 2000. Identification and emission rates of molecular tracers in coal smoke particulate matter. Fuel 79:515-536.

Parkes, R. J., and J. Taylor. 1983. The relationship between fatty acid distributions and bacterial respiratory types in contemporary marine sediments. Estuar. Coast. Shelf Sci. 16:173-189.

Poynter, J., and G. Eglinton. 1990. Molecular composition of three sediments from hole 717C: the Bengal Fan. In J. R. Cochran and D. A. V. Stow (eds.), Proc. ODP Sci. Results 116:155-161.

Salau, J. S. I., R. Tauler, J. M. Bayona, and I. Tolosa. 1997. Input characterization of sedimentary organic contaminants and molecular markers in the northwestern Mediterranean Sea by exploratory data analysis. Environ. Sci. Tech. 31:3482-3490.

Schubert, C., J. Niggemann, G. Klockgether, and T. Fedelman. 2005. Chlorin index: a new parameter for organic matter freshness in sediments. Geochem. Geophys. Geosyst. 6, Q03005, doi: 10.1029/2004GC000837.

Sluijs, A., and others, and Expedition 302 Scientists 2006. Subtropical Arctic Ocean temperatures during the Palaeocene/Eocene thermal maximum. Nature 44:610–613.

Treibs, A. E. 1936. Chlorophyll-und Häminderivate in organischen mineralstoffen. Angewandte Chem. 49:682-686

Volkman, J. K., A. T. Revill, D. G. Holdsworth, and D. Fredricks. 2008 Organic matter sources in an enclosed coastal inlet assessed using lipid biomarkers and stable isotopes. Org. Geochem. 39:689-710.

Wakeham, S. G., C. Lee, J. I. Hedges, P. J. Hernes, and M. L. Peterson. 1997. Molecular indicators of diagenetic status in marine organic matter. Geochim. Cosmochim. Acta 61:5363-5369.

Yunker, M. B., L. L. Belicka, H. R. Harvey, and R. W. Macdonald. 2005. Tracing the inputs and fate of marine and terrigenous organic matter in Arctic Ocean sediments: a multivariate analysis of lipid biomarkers. Deep Sea Res. II 52:3478-3508.

Zimmerman, A. R., and E. A. Canuel. 2001. Bulk organic matter and lipid biomarker composition of Chesapeake Bay surficial sediments as indicators of environmental processes. Est. Coast. Shelf Sci. 53:319-341.

Submitted 19 November 2008
Revised 9 February 2009
Accepted 18 February 2009