

# From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter

B. P. Koch<sup>1\*</sup> and T. Dittmar<sup>2</sup>

<sup>1</sup>Alfred-Wegener-Institut für Polar- und Meeresforschung, Am Handelshafen 12, 27570 Bremerhaven, Germany

<sup>2</sup>Florida State University, Department of Oceanography, OSB 311, Tallahassee, FL 32306-4320, USA

Received 16 November 2005; Revised 9 January 2006; Accepted 9 January 2006

Recent progress in Fourier transform ion cyclotron resonance mass spectrometry (FTICRMS) has provided extensive molecular mass data for complex natural organic matter (NOM). Structural information can be deduced solely from the molecular masses for ions with extreme molecular element ratios, in particular low H/C ratios, which are abundant in thermally altered NOM (e.g. black carbon). In this communication we propose a general aromaticity index (AI) and two threshold values as unequivocal criteria for the existence of either aromatic (AI > 0.5) or condensed aromatic structures (AI ≥ 0.67) in NOM. AI can be calculated from molecular formulae which are derived from exact molecular masses of naturally occurring compounds containing C, H, O, N, S and P, and is especially useful for substances with aromatic cores and few alkylations. In order to test the validity of our model index, AI is applied to FTICRMS data of a NOM deep-water sample from the Weddell Sea (Antarctica), a fulvic acid standard, and an artificial dataset of all theoretically possible molecular formulae. For graphical evaluation a ternary plot is suggested for four-dimensional data representation. The proposed aromaticity index is a step towards structural identification of NOM and the molecular identification of polyaromatic hydrocarbons in the environment. Copyright © 2006 John Wiley & Sons, Ltd.

Natural organic matter (NOM) is a highly complex and polydisperse mixture of different compounds,<sup>1</sup> the number of which is still unknown but exceeds several thousands. It is abundant in terrestrial, limnic and marine environments. Major contributors to the global NOM pool are humic substances, sedimentary and dissolved organic matter, which are operationally defined. Highly degraded NOM like dissolved organic matter (DOM) in the deep ocean resists further degradation and can reach average ages of several thousand years.<sup>2</sup> The extent of molecular information on NOM is largely restricted due to its complexity which prevents molecular resolution in chromatographic methods. Also chemical degradation methods yield only small amounts of analytically accessible compounds probably because of the refractory character of NOM. The proportion of NOM that can be characterized on the molecular level is therefore small, in deep-sea DOM, e.g., less than 10% of organic carbon can be assigned to molecular structures.<sup>3</sup>

Fourier transform ion cyclotron resonance mass spectrometry (FTICRMS) opened a new analytical window for the

analysis of NOM.<sup>4–7</sup> In combination with electrospray ionization (ESI), extensive molecular elemental information can be achieved without preceding chromatographic separation. Due to its very high resolution (>200 000) and mass accuracy (often < 0.5 ppm), FTICRMS delivers unequivocal molecular formulae. Formula assignment is possible because NOM consists of few abundant elements, mainly of C, H, and O, and to a lower degree of N, P and S. Other elements are rare in NOM and can be disregarded for a general examination of molecular structures in NOM.

NOM is a very heterogeneous mixture of largely different molecules. Molecular O/C and H/C ratios range from 0–0.8 and 0.3–1.8, respectively, in both deep-sea DOM (Weddell Sea, Antarctica)<sup>8</sup> and a fulvic acid standard (Suwannee River).<sup>7,8</sup> Typical molecular masses in DOM determined by FTICRMS range from approximately 300 to 700 Da. Deduction of specific structural configurations exclusively from a given molecular formula<sup>9–11</sup> can be challenging. Even formulae of very small ions can result in a large variety of different structures and functionalities. In most cases structural information on NOM can only be achieved by additional chemical and analytical techniques. However, for molecules with extreme elemental ratios, the number of possible isomers is more restricted. Low H/C ratios, for instance, diminish the number of configurational isomers and can be associated with unsaturations and C–C double bonds. The calculation of the 'double-bond equivalent' (DBE) is a well-established tool in mass spectrometry. DBE represents the sum of unsaturations plus rings in a molecule. Since triple bonds or cumulated double bonds are rare in

\*Correspondence to: B. P. Koch, Alfred-Wegener-Institut für Polar- und Meeresforschung, Am Handelshafen 12, 27570 Bremerhaven, Germany.

E-mail: bkoch@awi-bremerhaven.de

Contract/grant sponsor: The Petroleum Research Fund; contract/grant number: ACS PRF#41515-G2.

Contract/grant sponsor: The National Oceanic and Atmospheric Administration; contract/grant number: NOAA GC05-099.

Contract/grant sponsor: Deutsche Forschungsgemeinschaft; contract/grant number: DFG KO 2164/3-1.

Contract/grant sponsor: The German Academic Exchange Service; contract/grant number: DAAD PPP USA 315/ab.

NOM, a high density of double bonds (or high DBE/C ratios) can be indicative for aromatic or even condensed aromatic structures. Recent results in NOM research show that natural samples can contain significant amounts of thermogenic carbon, i.e. mainly condensed functionalized polyaromatic hydrocarbons (PAHs)<sup>8,12–16</sup> with H/C ratios of less than 0.5.

The potential to unambiguously identify condensed aromatic structures in NOM from ultrahigh-resolution mass spectrometry data is intriguing because it provides a molecular tool to simultaneously identify a large number of polyaromatic compounds of thermogenic origin.<sup>14</sup> To date all available methods to determine this fraction of NOM are based on operationally defined parameters associated with significant artefacts.<sup>17</sup> Thermogenic carbon is believed to play a major role in the global carbon cycle. It can originate from burning of biomass and fossil fuels, generally termed black carbon,<sup>14</sup> and from petrogenic sources such as coals and hydrothermal vents.<sup>15</sup>

The objective of this study was to develop an unambiguous parameter for the identification of aromatic and condensed aromatic structures from ultrahigh-resolution mass spectrometry data. We introduce an aromaticity index (AI) which reflects C–C double-bond 'density' in a molecule and which can be calculated solely from the exact molecular masses of single NOM compounds. In order to test the validity of this new index, we applied AI to FTICRMS data from deep-sea DOM (Weddell Sea, Antarctica),<sup>8</sup> a fulvic acid standard (Suwannee River), and to a calculated matrix of all theoretically possible molecular formulae in a mass range from 400–500 Da.

## EXPERIMENTAL

### Sampling and mass spectrometry

In order to establish and validate an aromaticity index (AI), we used the extensive ultrahigh-resolution mass spectrometry data set on marine and terrigenous DOM published by Koch *et al.*<sup>8</sup> For details on sampling and mass spectrometry refer to this publication. In brief, samples were taken from the abyssal ocean (3500–4600 m) of the Weddell Sea (Antarctica). DOM was isolated from the saline aqueous matrix via C18 solid-phase extraction (Varian Mega Bond Elut) at pH 2, and eluted with methanol. The Suwannee River fulvic acid standard was obtained from the International Humic Substances Society (IHSS). For ESI, an aliquot of the DOM methanol extract, or a methanolic solution of the IHSS standard, was mixed with Milli-Q water (50:50 v/v) and formic acid (0.2% final concentration). All analyses were performed on an APEX-Q FTICR mass spectrometer equipped with a 7 Tesla superconducting magnet (Bruker Daltonics, Bremen, Germany). The spectra were internally calibrated with a poly(ethylene glycol) standard and measured in positive ionization mode. To increase peak resolution the ions were mass-selected with a quadrupole filter between the ion source and the FTICR analyzer using a setting for ion selection of about 100 mass units. Thus, several mass spectra were acquired for each sample and subsequently merged into one continuous peak list. Ion accumulation time was set to 3 s for each scan and 200 scans were added for each mass spectrum.

All detected ions ( $m/z$  window 300–600) were singly charged. Once the exact masses of the molecules had been determined, their molecular formulae were calculated by arbitrarily combining any possible combination of atoms. For each detected mass every possible chemical formula in a 0.001 Da mass window was computed. The following elements (and number of atoms of each element) were considered in the calculation: <sup>12</sup>C (1–100), <sup>1</sup>H (1–200), <sup>14</sup>N (0–10), <sup>16</sup>O (0–50), <sup>23</sup>Na (0–1), and <sup>13</sup>C (0–1). The rules and assumptions described in Koch *et al.*<sup>8</sup> were applied to exclude formulae which do not occur abundantly in NOM, in particular: DBE must be an integer value, H/C ≤ 2.2, O/C ≤ 1.2 and N/C ≤ 0.5. After applying these rules, all detected masses could be assigned to one unambiguous chemical formula. The most abundant ions (signal-to-noise ratio >20) were all nitrogen-free, consisting of <sup>12</sup>C, <sup>1</sup>H, and <sup>16</sup>O.

### Generation of an artificial data matrix

To allow validation on a more general basis, an artificial dataset was constructed which comprised all theoretical molecular formulae in the mass range from 400–500 Da containing the elements C, H, O and N. For this purpose, all masses between 400 and 500 Da in 0.001 Da steps were fed into the same software we used for molecular formula computation from real mass spectrometry data. The same assumptions and rules described above were applied to exclude rare or impossible formulae. The final dataset comprised a total of 25 130 possible molecular formulae.

## RESULTS AND DISCUSSION

The first step towards the establishment of an unambiguous index for the existence of aromatic and condensed aromatic moieties in a molecule is to assess the degree of unsaturation or the density of C–C double bonds in a molecule. The sum of rings and double bonds in each molecule or 'double-bond equivalent' (DBE) can be calculated from the number of atoms ( $N_i$ ) and the valence ( $V_i$ ) of each element ( $i$ ) according to Eqn. (1). Uncharged molecules have a DBE with an integer value.

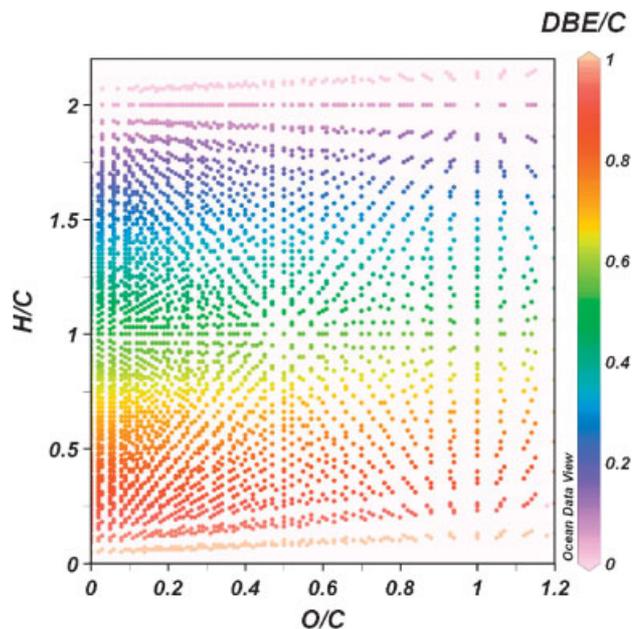
$$\text{DBE} = 1 + \frac{\sum_i^{\text{imax}} N_i(V_i - 2)}{2} \quad (1)$$

By including common elements of NOM (C, H, O, N, S, P) and their prevailing covalences, the DBE for NOM can be expressed as:

$$\text{DBE} = 1 + \frac{1}{2}(2C - H + N + P) \quad (2)$$

A decreasing number of H atoms in a molecule increases unsaturation and hence leads to higher DBE values. However, DBE is independent of the number of O and S (Eqn. (2)).

Since large molecules can potentially contain more double bonds and rings than small molecules, the maximum number of DBEs increases with the number of C, N and P atoms. A way to assess the degree of unsaturation or double-bond density in a molecule is to normalize DBE to the total



**Figure 1.** Molecular H/C vs. O/C ratios for all possible molecular formulae of the artificial data matrix containing C, H, and O in the mass range of 400–500 Da. Color scale represents the DBE/C ratio.

number of C atoms in a given molecule (DBE/C). Consequently, the DBE/C ratio increases with decreasing H/C ratio and remains unchanged with respect to O/C (Fig. 1). For hydrocarbons that exclusively contain C and H, DBE/C can reach values from 0 to 1 and is proportional to the H/C ratio which ranges between 0 and 2 for large molecules (Fig. 2(a)). The DBE/C ratio was recently proposed as an empirical criterion to identify condensed aromatic structures in charcoal from FTICRMS data.<sup>18</sup> DBE/C > 0.7 was used as a threshold to identify molecules which contain condensed aromatic rings. Benzene (see Supplementary table, A), e.g., has a DBE/C ratio of 0.67 (4/6), and any higher ratio must involve condensed ring systems (e.g. naphthalene, see

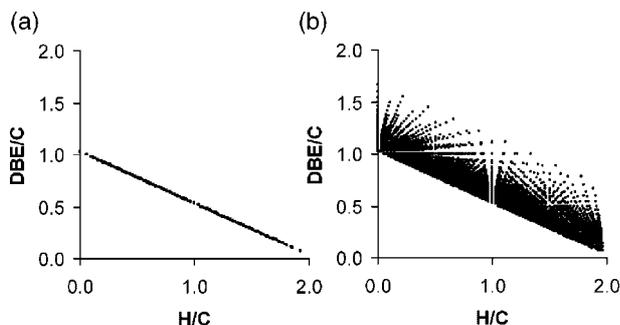
Supplementary table, B). Bi- or polycyclic ring systems (see Supplementary table, C) and cycle sizes with less than five C atoms could also explain high DBE/C ratios but are considered to be minor components in NOM or black carbon.

The DBE, however, is only a measure for C–C double bonds (and rings) in pure hydrocarbons. The introduction of an O atom to the molecular formula changes neither the remainder of the molecular formula nor the number of DBEs (Table 1, (a)–(c)). However, the addition of oxygen can, because of carbonyl unsaturations, reduce the potential number of C–C double bonds at a given DBE. The introduction of nitrogen into a molecular formula includes the possibility to form N–C  $\pi$ -bonds (Table 1, (a) and (d)), which again reduces the potential number of C–C double bonds at a given DBE. In contrast to oxygen, the addition of an N–C  $\pi$ -bond requires an additional H atom to fill the additional valence of nitrogen. For example, if one N atom is introduced into the molecular formula  $C_3H_6$ , an additional H atom is necessary to form an uncharged molecule ( $C_3H_7N$ ). As a result, the introduction of a heteroatom with an odd valence changes the H/C ratio and results in additional values for possible H/C ratios ( $7/3 = 2.33$ ). Therefore, the correlation between DBE/C and H/C ratios is weaker than for molecules consisting only of H and C (Fig. 2). Consequently, the introduction of nitrogen also weakens the significance of DBE/C as a measure for C–C double bonds which makes the DBE/C ratio insufficient as a sole criterion to define a threshold for condensed aromates.

The aromaticity index (AI) proposed in this communication includes the possibility that heteroatoms can form double bonds which are not contributing to aromaticity, ring formation or condensation. We address the potential contribution of heteroatoms to the DBE by calculating AI from an alternative  $DBE_{AI}/C_{AI}$  ratio. For this, C in Eqn. (2) is reduced by the total number of heteroatoms (Eqn. (3)). Each heteroatom (in particular O in NOM) can potentially contribute to DBE by forming double bonds with C atoms (Table 1). These unsaturations do not necessarily contribute to aromaticity. The number of H atoms in Eqn. (2) is reduced

**Table 1.** Example structures and their corresponding molecular formulae and DBE values. For the calculation of AI all functional groups which potentially contribute DBEs are eliminated from the original formula

|     | Possible structure | Molecular formula, DBE | Molecular formula after conversion |
|-----|--------------------|------------------------|------------------------------------|
| (a) |                    | $C_3H_6$ , DBE = 1     | $C_3H_6$ , DBE = 1                 |
| (b) |                    | $C_3H_6O$ , DBE = 1    | $C_2H_6$ , DBE = 0                 |
| (c) |                    | $C_3H_4O$ , DBE = 1    | $C_2H_6$ , DBE = 0                 |
| (d) |                    | $C_3H_7N$ , DBE = 1    | $C_2H_6$ , DBE = 0                 |



**Figure 2.** DBE/C vs. H/C ratios for all theoretical data including elements (a) C and H, (b) C, H, O and N.

by the number of N and P atoms, because the addition of one P or N atom, e.g. in the form of phosphines or imines, necessarily requires the addition of one H atom (Eqn. (3)). For DBE calculation we only considered P with 3 valences. P with 5 valences (e.g. organophosphates and phosphonates), which would increase DBE by 1 in comparison to 3-valent P, is considered implicitly in the AI because DBE is reduced by the total (including P-bonded) number of oxygens.  $DBE_{AI}$  is the minimum number of C–C double bonds plus rings in a common molecular structure containing heteroatoms (Eq. (4)).

$$DBE_{AI} = 1 + \frac{1}{2}(2(C - O - N - S - P) - (H - N - P) + N + P) \quad (3)$$

$$DBE_{AI} = 1 + C - O - S - \frac{1}{2}H \quad (4)$$

In the same way we calculate  $C_{AI}$  for the new  $DBE_{AI}/C_{AI}$  ratio. The number of C atoms is again reduced by the number of potential double bonds contributed by heteroatoms (Eqn. (5)):

$$C_{AI} = C - O - N - S - P \quad (5)$$

Subsequently, the new aromaticity index (AI, Eqn. (6)) can be calculated from the  $DBE_{AI}/C_{AI}$  ratio (Eqns. (4) and (5)):

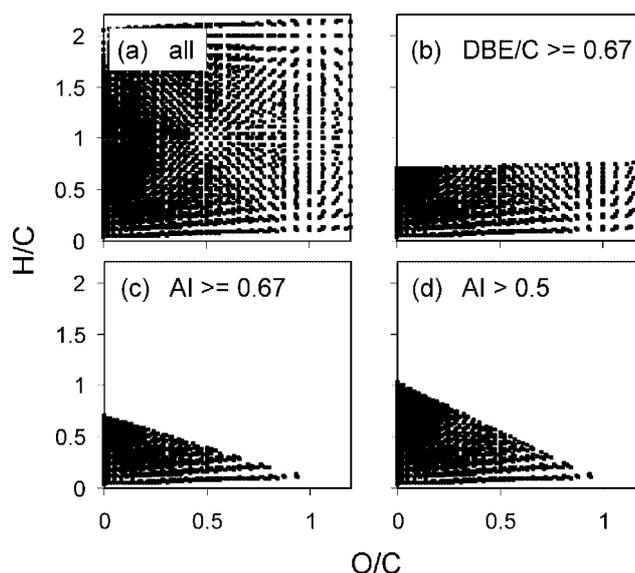
$$AI = \frac{DBE_{AI}}{C_{AI}} = \frac{1 + C - O - S - 0.5H}{C - O - S - N - P} \quad (6)$$

and if  $DBE_{AI} \leq 0$  or  $C_{AI} \leq 0$ , then  $AI = 0$ .

AI is a measure for C–C double-bond density and considers the contribution of  $\pi$ -bonds by heteroatoms. Since all heteroatoms are taken into account as potential contributors, the actual aromaticity in a molecule can be higher and AI must be regarded as the most conservative approach.

If the number of heteroatoms exceeds the number of  $\pi$ -bonds in a molecule, the calculated AI would be negative.  $C_{23}H_{36}O_{12}$ , e.g., would have an AI of  $-0.54$  because the high number of O atoms reduces the number of C atoms to a value below 1. Since it is ineligible to calculate C–C double-bond density for a molecular formula without C atoms, AI is defined as 0 for those cases (see boundary conditions, Eqn. (6)).

A threshold value of  $AI \geq 0.67$  provides an unambiguous minimum criterion for the presence of condensed aromatic structures in a molecule. In comparison to  $DBE/C \geq 0.67$ , AI

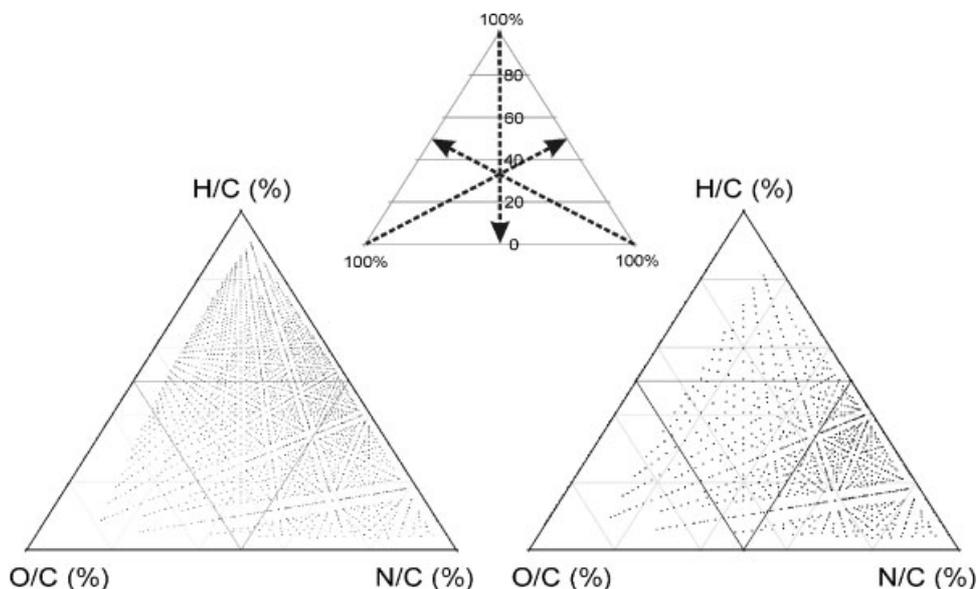


**Figure 3.** (a) All possible formulae ( $n = 3385$ ) for compounds containing C (1–100 atoms), H (1–200) and O (0–50) and threshold criteria of (b)  $DBE/C \geq 0.67$  ( $n = 1296$ ), (c)  $AI \geq 0.67$  ( $n = 903$ ), and (d)  $AI > 0.5$  ( $n = 1351$ ).

is more conservative. However, by considering unsaturations of heteroatoms, AI provides more reliable information. Figure 3 illustrates how the amount of condensed aromatic compounds might be overestimated by applying  $DBE/C \geq 0.67$  as a sole criterion. At a ratio of  $O/C = 1$  several compounds are identified as condensed aromates (Fig. 3(b)). For  $AI \geq 0.67$  the number of molecules that necessarily contain condensed aromatic structures decreases to 0 at  $O/C = 1$  (Fig. 3(c)).

Furthermore, an additional AI-threshold value can be deduced: Any additional unsaturation in a conjugated unsaturated aliphate, e.g. in hexatriene (see Supplementary table, D;  $AI = 0.5$ ), necessarily leads to the formation of an aromatic ring (benzene) if cumulated  $\pi$ -bonds (as present in, e.g., allene) and triple bonds are excluded. For such a conjugated  $\pi$ -system, AI is 0.5 and any higher value must involve aromatic structures. We therefore propose a threshold of  $AI > 0.5$  as a necessary minimum criterion for the presence of aromatic cores in general, including all combinations with heteroatoms.  $AI > 0.5$  is the most conservative calculation for the existence of aromatic structures and, of course, also includes condensed aromates (Fig. 3(d)).

Ultrahigh-resolution MS data of complex NOM require suitable ways for data presentation. For N-containing compounds, O/C versus H/C plots are less appropriate, because both ratios are affected by the number of N atoms present in a molecule. A ternary plot is suitable to reflect the effect of AI thresholds for CHON compounds (Fig. 4). Each corner of the plot of the artificial dataset represents 100% of the assigned ratio decreasing to 0% on the opposite line of the isosceles triangle. 100% values for H/C, O/C and N/C were set to 2.2, 1.2 and 0.5, respectively. As stated earlier, high O/C and high H/C ratios result in low values for AI. Applying the  $AI \geq 0.67$  criterion demonstrates that especially H- and O-poor molecules can be reliably identified as

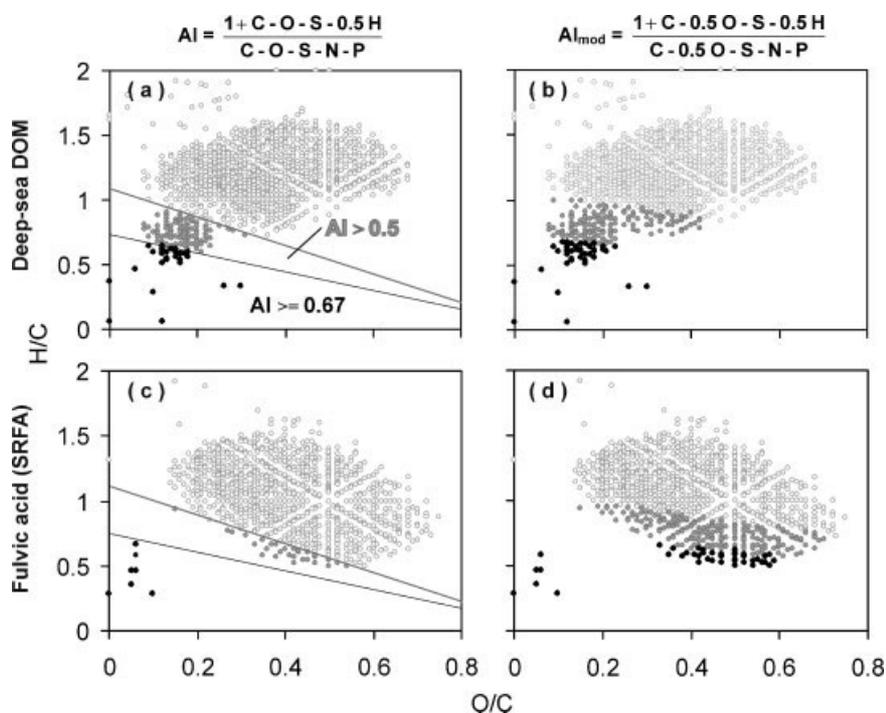


**Figure 4.** (a) All possible formulae ( $n = 25130$ ) for compounds containing C (1–100), H (1–200), O (0–50) and N (0–10) and a threshold criterion of (b)  $AI \geq 0.67$  ( $n = 9235$ ). 100% values (corners) for H/C, O/C and N/C are defined as 2.2, 1.2 and 0.5, respectively.

aromatic compounds and that compounds in the upper (high H/C) and lower left corner (high O/C) disappear from the ternary plot due to their low AI values (Fig. 4(b)).

Figure 5 shows a comparison of the results obtained for two natural samples by applying  $AI > 0.5$  versus  $AI \geq 0.67$ . A group of compounds in the deep-sea DOM sample is

characterized by very low O/C and H/C ratios (lower left corner, Fig. 5(a)). These substances were identified as condensed aromatic ring systems.<sup>15</sup> If we calculate AI for all of these masses, they can be characterized as compounds which necessarily contain aromatic structures ( $AI > 0.5$ , grey dots, Fig. 5(a)). However, only compounds with the lowest



**Figure 5.** (a, b) Element ratio plot for FTICRMS data for a deep-sea (3500 m) DOM sample from the Weddel Sea and (c, d) for the Suwannee River fulvic acid standard. (a, c) Aromatic and condensed aromatic structures can be identified with threshold criteria of the aromaticity index  $AI > 0.5$  (grey dots) and  $AI \geq 0.67$  (black dots). (b, d) Based on the assumption that half of the oxygen is  $\sigma$ -bound a modified aromaticity index ( $AI_{mod}$ ) can be calculated.

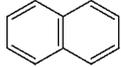
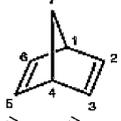
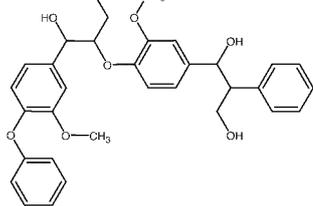
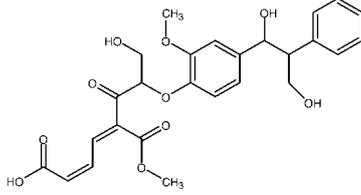
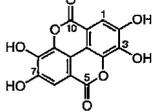
H/C ratios are classified as condensed aromates ( $AI \geq 0.67$ , black dots, Fig. 5(a)). Underestimation of PAHs is a result of the conservative approach for AI since we assumed that all oxygen is bound as carbonyl oxygen. This effect also becomes obvious in the fulvic acid sample (SRFA, Fig. 5(c)). This material contains a large fraction of lignin and tannin degradation products (see Supplementary table, E–G),<sup>7</sup> and hence numerous aromatic compounds. The conservative AI only identifies compounds with specifically low H/C ratios to be aromatic.

According to published NMR data approximately half of the oxygen in marine DOM is bound with  $\sigma$ -bonds, especially as carboxyl oxygen, rather than bound as carbonyl oxygen.<sup>3</sup> Consequently, only half of the oxygen is bound with  $\pi$ -bonds in carboxyl groups. In this case, a modified AI can be calculated ( $AI_{mod}$ , Figs. 5(b) and 5(d)) by considering only half of the oxygen being present in carbonyl functional groups. By reducing the number of possible carbonyl unsaturations, a larger number of compounds was identified by  $AI_{mod}$  as aromatic and condensed aromatic components, respectively (Figs. 5(c) and 5(d)). However, this procedure assigns bulk NMR information to a specific individual molecule which is highly speculative. The proposed AI is the most conservative case; any modifications ( $AI_{mod}$ ) where bulk chemical information are assigned to individual molecules introduce uncertainties.

## CONCLUSIONS

The aromaticity index (AI) provides a most conservative criterion for the unequivocal identification of aromates and condensed aromates in natural organic matter (NOM) solely from molecular formulae. AI was developed for the purpose of a general structural characterization of NOM. Based on existing knowledge on the structure of NOM the following assumptions were made: NOM consists mainly of C, H, O, N, P, and S. Triple and cumulated double bonds, heterocycles, cycle sizes with less than five carbon atoms and bi- or polycyclic compounds are not common. Applying these assumptions, AI is a suitable parameter for the identification of aromatic structures in any other sample material as well. Since AI provides a necessary minimum criterion for the existence of aromatic and condensed structures, several natural compounds containing aromatic structures like lignin degradation products may not be recognized due to their aliphatic side chains. Lignin compounds, e.g., would also not be identified as an aromatic structure by applying our model, because non-aromatic molecules can be constructed with the same molecular formula. However, for other NOM compounds like tannin and especially black carbon, AI is a helpful tool to unequivocally identify structural subunits exclusively on the basis of exact mass determination. For practical use, additional assumptions

**Supplementary Table.** Molecular formulae and example structures with corresponding elemental ratios, double-bond equivalents (DBE) and values for the aromaticity index (AI). Structures for lignin degradation products were taken from Stenson *et al.*<sup>7</sup>

|   | Molecular formula (possible structure)                            | H/C  | O/C  | DBE | DBE/C | AI   | Structure   |
|---|---|------|------|-----|-------|------|---|
| A | C <sub>6</sub> H <sub>6</sub> (benzene)                           | 1.00 | 0    | 4   | 0.67  | 0.67 |  |
| B | C <sub>10</sub> H <sub>8</sub> (naphthalene)                      | 0.80 | 0    | 7   | 0.70  | 0.70 |  |
| C | C <sub>7</sub> H <sub>6</sub> (bicyclo[2.2.1] hepta-2,5-diene)    | 0.86 | 0    | 5   | 0.71  | 0.71 |  |
| D | C <sub>6</sub> H <sub>8</sub> (hexatriene)                        | 1.33 | 0    | 3   | 0.5   | 0.5  |  |
| E | C <sub>32</sub> H <sub>34</sub> O <sub>8</sub> (lignin subunit)   | 1.06 | 0.31 | 16  | 0.50  | 0.33 |  |
| F | C <sub>26</sub> H <sub>28</sub> O <sub>10</sub> (degraded lignin) | 1.08 | 0.31 | 13  | 0.50  | 0.19 |  |
| G | C <sub>14</sub> H <sub>6</sub> O <sub>8</sub> (ellagic acid)      | 0.5  | 0.67 | 10  | 0.83  | 0.5  |  |

could be made. For instance, it is very conservative to assume that all oxygen is present in carbonyl functional groups. In NOM, carboxyl groups are far more common, thus it would be legitimate to consider all oxygen to be present as carboxyl. This (less conservative) assumption for a modified aromaticity index ( $AI_{\text{mod}}$ ) increases the number of identified aromatic compounds in a realistic matter. However, if additional structural information on a molecular level is not available, this modified approach can only provide a probability for the existence of aromatic structures. Considering the immense number of possible configurational isomers even for small molecules, the proposed aromaticity index is an important step towards structural identification of NOM.

### Acknowledgements

We would like to thank Matthias Witt (Bruker Daltonics) for FTICRMS analysis. Gerhard Kattner and the anonymous reviewers are acknowledged for helpful comments on the manuscript. This work was financially supported by the Petroleum Research Fund (ACSPRF#41515-G2), the National Oceanic and Atmospheric Administration (NOAA GC05-099), Deutsche Forschungsgemeinschaft (DFG KO 2164/3-1), and the German Academic Exchange Service (DAAD PPP USA 315/ab).

### REFERENCES

- Hertkorn N, Kettrup A. In *Earth and Environmental Sciences*, vol. 52, Perminova IV, Hatfield K, Hertkorn N (eds). Springer: Heidelberg, 2005; 391–435.
- Williams PM, Druffel ERM. *Nature* 1987; **330**: 246.
- Benner R. In *Biogeochemistry of Marine Dissolved Organic Matter*, Hansell DA, Carlson CA (eds). Academic Press: San Diego, 2002; 71–80.
- Stenson AC, Landing WM, Marshall AG, Cooper WT. *Anal. Chem.* 2002; **74**: 4397.
- Kujawinski EB, Hatcher PG, Freitas MA. *Anal. Chem.* 2002; **74**: 413.
- Kujawinski EB, Freitas MA, Zang X, Hatcher PG, Green-Church KB, Jones RB. *Org. Geochem.* 2002; **33**: 171.
- Stenson AC, Marshall AG, Cooper WT. *Anal. Chem.* 2003; **75**: 1275.
- Koch BP, Witt M, Engbrodt R, Dittmar T, Kattner G. *Geochim. Cosmochim. Acta* 2005; **69**: 3299.
- Hu C-Y, Xu L. *Chem. Inform. Computer Sci.* 1997; **37**: 311.
- Robinson RW, Harry F, Balaban AT. *Tetrahedron* 1976; **32**: 355.
- Benecke C, Grund R, Hohberger R, Kerber A, Laue R, Wieland T. *Anal. Chim. Acta* 1995; **314**: 141.
- Kim S, Kramer RW, Hatcher PG. *Anal. Chem.* 2003; **75**: 5336.
- Kim SW, Kaplan LA, Benner R, Hatcher PG. *Mar. Chem.* 2004; **92**: 225.
- Kramer RW, Kujawinski EB, Hatcher PG. *Environ. Sci. Technol.* 2004; **38**: 3387.
- Dittmar T, Koch BP. *Mar. Chem.* 2006; submitted.
- Simpson MJ, Hatcher PG. *Org. Geochem.* 2004; **35**: 923.
- Simpson MJ, Hatcher PG. *Naturwissenschaften* 2004; **91**: 436.
- Hatcher PG, Hockaday WC, Grannas AM, Kim S. *American Chemical Society, Abstracts of the 230th National Meeting*, 2005.