Resolution of 11 000 Compositionally Distinct Components in a Single Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrum of Crude Oil

Christine A. Hughey,[†] Ryan P. Rodgers,[†] and Alan G. Marshall[†],*

Ion Cyclotron Resonance Program, National High Magnetic Field Laboratory, Florida State University, 1800 East Paul Dirac Drive, Tallahassee, Florida 32310

Mass spectrometry is well-suited for complex mixture analysis, because unlike other types of spectroscopy, the number of mass spectral peaks per analyte is of order one. Here, we extend significantly the upper limit for the number of chemically distinct components resolved and identified in a single step. Specifically, electrospray selectively ionizes only the basic compounds (i.e., a small fraction of the entire chemical composition) in a sample of South American crude oil. Nevertheless, their positive ion Fourier transform ion cyclotron resonance mass spectrum (average mass resolving power of ~350 000 from 225 to 1000 Da) contains more than 11 100 resolved peaks, of which >75% may be assigned to a unique elemental composition ($C_cH_hO_oN_nS_s$). Mass scale expansion and graphical representations reveal increased heteroatom diversity, aromatic rings, and alkyl substitution with increased mass. These results set a new standard and illustrate the potential of high-resolution mass spectrometry for analysis of compositionally complex chemical mixtures.

The history of spectroscopy is the history of increasing resolution. However, in most forms of spectroscopy, a single molecule can generate up to thousands of spectral peaks; e.g., tens of thousands (a little more than one-half of which are assigned) for the UV/vis spectrum of a diatomic molecule. In contrast, mass spectrometry (MS), particularly with electrospray ionization (ESI)[1,2] or matrix assisted laser desorption (MALDI)[3] sources, typically yields ions of one or a few different masses for each analyte. (Electrospray, especially for proteins and other large biomolecules,[4,5] can generate ions of several different charge states for a given mass, but high-resolution mass-to-charge ratio (m/z) data can readily be deconvolved to yield a mass-only spectrum.[6])

Thus, MALDI-MS and ESI-MS are well-suited for resolution and identification of complex mixtures of analytes below ~1000 Da in mass, because the number of mass spectral peaks does not greatly exceed the number of analytes to be distinguished.

One measure of the chemical complexity of a mixture is the number of species with different elemental compositions ($C_cH_hO_oN_nS_s\cdots$). (We consider only different elemental compositions, because one-dimensional mass analysis typically does not distinguish isomers, i.e., molecules of the same elemental composition but different bond arrangement.) If the dynamic range (ratio of highest to lowest abundance species) in the mixture is <10 000 or so, then petroleum heavy crude oils or liquefied coals present the most complex readily available chemical mixtures. Thus, we are led to focus on crude oil as an ultimate benchmark for mass spectrometric (MS) mixture analysis.

Nuclear magnetic resonance, gas chromatography (GC), benchtop GC/MS, and LC/MS can discriminate between different functional groups (aromatic, olefinic, paraffinic, carboxylic acids, thiophenes, etc.) but cannot distinguish individual elemental compositions in complex petrochemical samples.[7] Separation techniques coupled to quadrupole or ion trap mass spectrometers are limited by insufficient chromatographic resolution and low mass resolving power. High-resolution double-focusing sector mass spectrometers can provide mass resolving power ($m/\Delta m_{50\%}$ ~50 000, in which m is ion mass and $\Delta m_{50\%}$ is the mass spectral peak width at half-maximum peak height) high enough to assign molecular formulas to individual compounds, but only after the crude oil has been fractionated into saturates, aromatics, and polars or acidic, basic and neutral fractions (at the very least). Most recently, Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) has shown the greatest promise in molecular characterization of complex mixtures, such as petrochemical samples without prior chromatographic fractionation.

As in other forms of spectroscopy, the history of FT-ICR-MS for mixture analysis is the history of spectral resolution. In 1996, a 3.0 T FT-ICR mass spectrometer equipped with electron ionization (EI) was used to analyze a neutral gas oil fraction. Mass segments of ~20 Da were collected over the range, 100 < *m* < 700 Da, to reduce peak shift, broadening, distortion, and coales-

---

[†] Department of Chemistry and Biochemistry, Florida State University.

(1) Lorenz, S. A.; Maziarz, E. P. I.; Wood, T. D. *Appl. Spectrosc.* **1999**, *53*, 18A–36A.

(2) Hendrickson, C. L.; Emmett, M. R. *Annu. Rev. Phys. Chem.* **1999**, *50*, 517–536.

(3) Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. *Anal. Chem.* **1991**, *63*, 1193A–1203A.

(4) McLafferty, F. W.; Fridriksson, E. K.; Horn, D. M.; Lewis, M. A.; Zubarev, R. A. *Science* **1999**, *284*, 1289–1290.

(5) Palmblad, M.; Wetterhall, M.; Markides, K.; Hakansson, P.; Bergquist, H. J. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1029–1034.

(6) Zhang, Z.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 225–233.

(7) Altgelt, K. H.; Boduszynski, M. M. *Composition and Anlysis of Heavy Petroleum Fractions*; Marcel Dekker: New York, 1994.

cence due to space charge and to increase resolving power. In this way, mass resolving power of ∼200 000 sufficed to assign more than 500 different molecular formulas.[8] With a higher-field (5.6 T) EI FT-ICR instrument,[9] it became possible to detect ions over a mass range of ∼300 Da simultaneously at a mass resolving power of ∼100 000 and mass accuracy <1 ppm for identification of hundreds of individual elemental compositions in processed and unprocessed diesel fuel,[10] weathered transportation fuels,[11,12] and arson accelerants[13] without prior chromatographic isolation. More recently, implementation of a 9.4 T electrospray (ESI) FT-ICR mass spectrometer[14] provides mass resolving power >300 000 from 225 < m <1000 Da[15-20] for identification of thousands of polar species per sample without prior chromatographic isolation.

In this paper, we report the successful resolution of more than 11 000 singly charged ions, corresponding to ∼8000 different elemental compositions (the remaining 3000 constitute the same molecules but with one ^{13}C or ^{34}S instead of ^{12}C or ^{32}S) from a single one-dimensional positive ion ESI FT-ICR mass spectrum with no prior chromatographic separation. That result sets a new standard for the most complex chemical mixture ever analyzed in a single step and sets a standard for analysis of similarly complex mixtures (biological, environmental, etc.).

EXPERIMENTAL METHODS

Sample Preparation. Twenty mg of South American crude oil was first dissolved in 3 mL of toluene and then diluted with 17 mL of methanol to give a final volume of 20 mL. The final solution was spiked with 0.5% acetic acid (to facilitate protonation of the basic nitrogen compounds found in the crude).

Mass Analysis. Mass analyses were carried out with a home-built FT-ICR mass spectrometer[14] equipped with a 22-cm horizontal room-temperature bore 9.4 T magnet (Oxford Instruments America, Inc., Concord, MA). Data were collected and processed by a modular ICR data acquisition system (MIDAS).[21,22] Positive ions were generated from a microelectrospray source equipped with

a 50-μm- i.d. fused-silica micro-ESI needle.[23] Samples were infused at a flow rate of 500 nL/min. Typical ESI conditions were needle voltage, 2 kV; tube lens, 370 V; and heated capillary current, 4 A. Ions were accumulated externally in a linear octopole ion trap for 15 s and transferred through rf-only multipoles to a 10-cm-diameter, 30-cm-long open cylindrical Penning ion trap.[24] Multipole ion guides[25] were operated at 1.8 MHz at 170 V$_{p-p}$ rf amplitude. Broadband frequency-sweep (chirp)[26,27] dipolar excitation (∼70−641 kHz at a sweep rate of 150 Hz/μs and peak-to-peak amplitude of 190 V) was followed by direct-mode image current detection that yielded 4 Mword time-domain data. Two hundred co-added time domain data sets were Hanning-apodized, followed by a single zero-fill before fast Fourier transformation and magnitude calculation.

Mass Calibration and High Mass Accuracy. Frequency was converted to mass-to-charge ratio according to the quadrupolar electric trapping potential approximation[28,29] to generate the m/z spectra shown in Figures 1, 3, and 4. Mass spectra were initially frequency-to-m/z-calibrated internally with respect to PEG 400 and then recalibrated from the homologous series −15 N (see below for notation), over the range, 225 < m/z < 1000. Nineteen mass values from the −15 N homologous series gave a calibration rms m/z error of 0.254 ppm. All species in the present mass spectra are singly charged, as evidenced by the ∼1 Da spacing between each monoisotopic species and its corresponding nuclide containing one ^{13}C atom.[30,31] Thus, from here on, we shall denote peaks by their mass in Da rather than as mass-to-charge ratio, m/z.

Sorting and Assignment of Elemental Compositions. The mass values for (singly charged) ions of 225−1000 Da and >0.45% relative abundance (corresponding to peak height >3 SD of baseline noise) were imported into an Excel spreadsheet. Measured masses were converted from the IUPAC mass scale (CH_2 = 14.015 65 Da) to the Kendrick mass scale (CH_2 = 14.000 00 Kendrick mass units) as previously described.[32-34] Even and odd nominal Kendrick masses were then sorted by an Excel macro, and the Kendrick mass defect (i.e., difference between exact and nominal (nearest-integer) Kendrick mass) was calculated. Even

(8) Guan, S.; Marshall, A. G.; Scheppele, E. S. *Anal. Chem.* **1996**, *68*, 46−71.

(9) Rodgers, R. P.; White, F. M.; McIntosh, D. G.; Marshall, A. G. *Rev. Sci. Instrum.* **1998**, *69*, 2278−2284.

(10) Rodgers, R. P.; White, F. M.; Hendrickson, C. L.; Marshall, A. G.; Andersen, C. V. *Anal. Chem.* **1998**, *70*, 4743−4750.

(11) Rodgers, R. P.; Blumer, E. N.; Freitas, M. A.; Marshall, A. G. *Anal. Chem.* **1999**, *71*, 5171−5176.

(12) Rodgers, R. P.; Blumer, E. N.; Freitas, M. A.; Marshall, A. G. *Environ. Sci. Technol.* **2000**, *34*, 1671−1678.

(13) Rodgers, R. P.; Blumer, E. N.; Freitas, M. A.; Marshall, A. G. *J. Forensic Sci.* **2001**, *46*, 268−279.

(14) Senko, M. W.; Hendrickson, C. L.; Pasa-Tolic, L.; Marto, J. A.; White, F. M.; Guan, S.; Marshall, A. G. *Rapid Commun. Mass. Spectrom.* **1996**, *10*, 1824−1828.

(15) Hughey, C. A.; Rodgers, R. P.; Marshall, A. G.; Qian, K.; Robbins, W. K. *Org. Geochem.* **2002**, *33*, in press.

(16) Hughey, C. A.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G. *Energy Fuels* **2001**, *15*, 1186−1193.

(17) Hughey, C. A.; Rodgers, R. P.; Hendrickson, C. L.; Marshall, A. G.; Qian, K.; Robbins, W. K. *Selective ionization, resolution and chemical identification of naphthenic acids and neutral nitrogen compounds in heavy petroleum crudes by negative ion electrospray FT-ICR mass spectrometry*, In Proceedings of the 49th ASMS Conference on Mass Spectrom. and Allied Topics, Chicago, IL, 2001.

(18) Qian, K.; Robbins, W. K.; Hughey, C. A.; Cooper, H. J.; Rodgers, R. P.; Marshall, A. G. *Energy Fuels* **2001**, *15*, 1505−1511.

(19) Qian, K.; Rodgers, R. P.; Hendrickson, C. L.; Emmett, M. R.; Marshall, A. G. *Energy Fuels* **2001**, *15*, 492−498.

(20) Rodgers, R. P.; Hendrickson, C. L.; Emmett, M. R.; Marshall, A. G.; Greaney, M.; Qian, K. *Can. J. Chem.* **2001**, *79*, 546−551.

(21) Senko, M. W.; Canterbury, J. D.; Guan, S.; Marshall, A. G. *Rapid Commun. Mass. Spectrom.* **1996**, *10*, 1839−1844.

(22) Blakney, G. T.; van der Rest, G.; Johnson, J. R.; Freitas, M. A.; Drader, J. J.; Shi, S. D.-H.; Hendrickson, C. L.; Kelleher, N. L.; Marshall, A. G. *Further improvements to the MIDAS data station for FT-ICR mass spectrometry*, In Proceedings of the 49th ASMS Conference on Mass Spectrom. and Allied Topics, Chicago, IL, 2001.

(23) Emmett, M. R.; White, F. M.; Hendrickson, C. L.; Shi, S. D.-H.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 333−340.

(24) Senko, M. W.; Hendrickson, C. L.; Emmett, M. R.; Shi, S. D.-H.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 970−976.

(25) Hendrickson, C. L.; Quinn, J. P.; Emmett, M. R.; Marshall, A. G. *Quadrupole mass filtered external accumulation for Fourier transform ion cyclotron resonance mass spectrometry*, In Proceedings of the 48th ASMS Conference on Mass Spectrom. and Allied Topics, Long Beach, CA, 2000.

(26) Comisarow, M. B.; Marshall, A. G. *Chem. Phys. Lett.* **1974**, *26*, 489−490.

(27) Marshall, A. G.; Roe, D. C. *J. Chem. Phys.* **1980**, *73*, 1581−1590.

(28) Ledford, E. B. J.; Rempel, D. L.; Gross, M. L. *Anal. Chem.* **1984**, *56*, 2744−2748.

(29) Shi, S. D.-H.; Drader, J. J.; Freitas, M. A.; Hendrickson, C. L.; Marshall, A. G. *Int. J. Mass Spectrom.* **2000**, *195/196*, 591−598.

(30) Henry, K. D.; McLafferty, F. W. *Org. Mass Spectrom.* **1990**, *25*, 490−492.

(31) Marshall, A. G.; Hendrickson, C. L. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 232−235.

(32) Kendrick, E. *Anal. Chem.* **1963**, *35*, 2146−2154.

(33) Hsu, C. S.; Qian, K.; Chen, Y. C. *Anal. Chim. Acta* **1992**, *264*, 79−89.

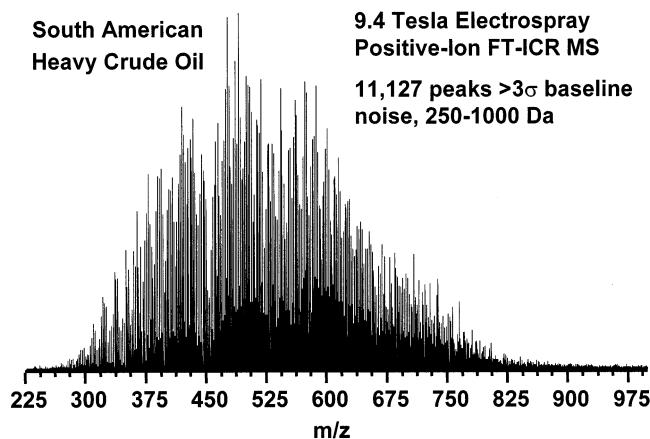(34) Hughey, C. A.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G. *Anal. Chem.* **2001**, *73*, 4676−4681.

Figure 1. Broadband positive ion electrospray ionization FT-ICR mass spectrum of a South American crude oil. The 11 127 peaks (each exceeding 3σ baseline noise) represent the most complex chemical mixture ever resolved and identified in a single mass spectrum. The average mass resolving power, $m/\Delta m_{50\%}$, is $\sim$350 000 from 250 to 950 Da.

Figure 2. Plot of Kendrick mass defect vs nominal Kendrick mass for *all* resolved peaks in the broadband ESI FT-ICR positive ion mass spectrum of Figure 1. This single-page display not only visually displays all of the 11 127 peaks in the original mass spectrum, but also reveals each ion's class and type (by projection to the vertical axis), and carbon distribution (by projection to the horizontal axis), as explained in the text.

and odd nominal Kendrick masses were then sorted[33] into homologous series on the basis of their identical Kendrick mass defects. Molecular formulas of species less than $\sim$400 Da in mass could be assigned solely on the basis of mass measurement to $\pm$1 ppm. Elemental compositions were assigned by use of a mass calculator program limited to molecular formulas consisting of up to 100 ^{12}C atoms, 2 ^{13}C, 200 ^{1}H, 5 ^{14}N, 5 ^{16}O, 5 ^{32}S, and 1 ^{34}S. If two (or more) elemental compositions were found within the mass tolerance of $\pm$1 ppm, one formula could usually be confirmed/eliminated unequivocally by the presence/absence of the corresponding nuclide containing one ^{13}C. Because members of a homologous series differ only by integer multiples of CH_2, assignment of a single member of such a series usually sufficed to identify all higher-mass members.[34]

Compact Notation for Elemental Compositions. The elemental composition of a petroleum molecule is conveniently expressed by its chemical formula, $C_cH_{2c+z}X$, in which c is the carbon number, Z is the deficiency in number of hydrogen atoms relative to a saturated hydrocarbon with the same number of carbons (e.g., $Z = -6$ for benzene, C_6H_6, relative to hexane, C_6H_{14}), and X denotes the constituent heteroatoms (N, S, O) in the molecule. The hydrogen deficiency index or Z value is determined by the number of rings and double bonds in a molecule. Each additional double bond or ring makes the Z value more negative by 2 units; therefore, the more negative the Z value, the more aromatic the molecule. For convenience, we abbreviate molecular formulas according to their Z value ("type") and heteroatoms ("class"). For example, $C_{12}H_9N$ would be abbreviated as -15 N and always refers to the neutral molecule. Finally, because additional CH_2 groups do not change the "class" or "type", it is also possible to determine the "carbon distribution" of all alkyl substituents for each class and type.

RESULTS AND DISCUSSION

ESI FT-ICR Mass Spectrometry of Crude Oils. We recently employed ESI FT-ICR mass spectrometry to selectively ionize, resolve, and identify polar compounds (acidic and basic NSO-containing compounds) in crude oils of varying geochemical origin
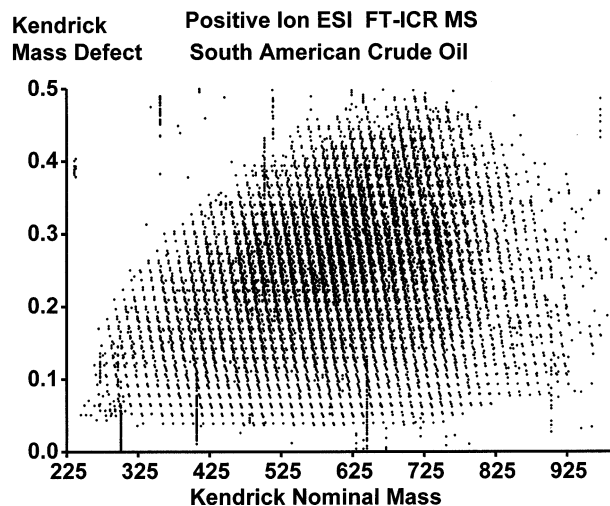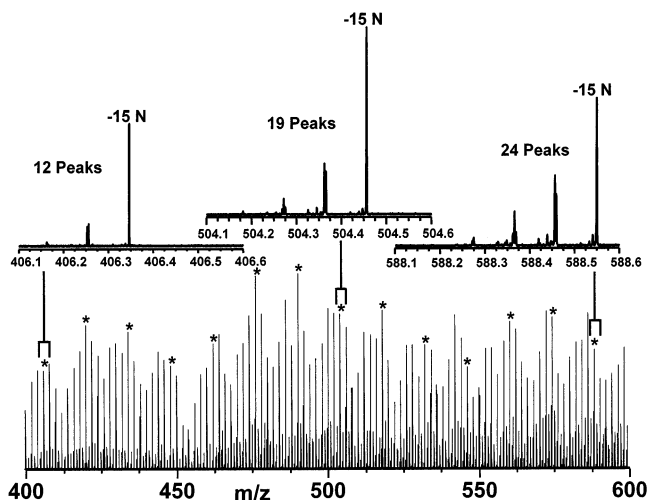

Figure 3. Successive mass scale expanded segments of Figure 1. The 200 Da mass segment (bottom) shows species at every nominal mass, with periodicities at 2 and 14 nominal mass units. Starred peaks (*) are 14 Da apart and constitute members of the same homologous series, the most abundant of which is -15 N (see text). The most expanded insets at 406, 504, and 588 Da (top) demonstrate the increase in spectral complexity with increased mass: 12, 19, and 24 peaks/Da, respectively. The segments were chosen to differ in mass by integral multiples of 14 Da to illustrate homologous type series assignments (see Table 1).

(i.e., North America, China, South America, and the Middle East).[15] In the course of differentiating compositions among these crude oils to establish their geochemical significance, we happened upon a South American crude oil that exhibited a complexity greater than any previously analyzed crude oil, namely, >11 000 positive ion ESI-FT-ICR mass spectral peaks, each with peak height exceeding 3σ of baseline noise (Figure 1). That positive ion mass spectrum is, to our knowledge, the most complex mass spectrum ever reported and is rivaled only by the corresponding negative ion ESI-FT-ICR mass spectrum (not shown) of the same sample. On the basis of an average mass resolving power,

Table 1. Assignment of Peaks at 406, 504, 588, and 644 Da

meas mass at 588 Da	theor mass	error (ppm)	homologous series	identified at 406 Da	504 Da	644 Da
588.271 93	588.271 95	−0.03	−47 NS			X
588.275 19	588.275 32	−0.22	−37 NS$_2$		X	X
588.326 15	588.326 09	+0.10	−45 NO		X	X
588.329 41	588.329 46	−0.08	−35 NOS		X	X
588.345 29	588.345 43	−0.24	−44 N$_2$[a]			X
588.348 61	588.348 80	−0.34	−34 N$_2$S[a]		X	X
588.356 81	588.356 91	−0.17	−35 NS[b]			X
588.362 45	588.362 48	−0.05	−43 N		X	X
588.365 70	588.365 85	−0.25	−33 NS	X	X	X
588.369 05	588.369 22	−0.34	−23 NS$_2$	X	X	X
588.419 88	588.419 99	−0.19	−31 NO	X	X	X
588.423 27	588.423 36	−0.15	−21 NOS		X	
588.439 18	588.439 33	−0.25	−30 N$_2$[a]	X	X	X
588.442 60	588.442 70	−0.17	−20 N$_2$S[a]			X
588.447 33	588.447 44	−0.19	−31 N[b]		X	X
588.450 79	588.450 81	−0.03	−21 NS[b]		X	X
588.452 17	2 possibilities					
588.456 25	588.456 38	−0.22	−29 N	X	X	X
588.459 56	588.459 75	−0.32	−19 NS	X	X	X
588.513 74	588.513 89	−0.25	−17 NO	X	X	X
588.533 25	588.533 23	+0.03	−16 N$_2$[a]	X	X	X
588.541 22	588.541 34	−0.20	−17 N[b]	X	X	X
588.545 70	not identified					
588.550 11	588.550 28	−0.29	−15 N	X	X	X
peaks ID/total			22/24	10/12	18/19	23/29

[a] Nuclides containing one ^{13}C. [b] Nuclides containing two ^{13}C.

$m/\Delta m_{50\%} \approx 350\,000$, and a mass accuracy <1 ppm, molecular formulas may be assigned (with the aid of Kendrick mass analysis, see Experimental section) from 225 to 950 Da. Molecular formula assignments correspond predominately to basic nitrogen-containing compounds (i.e., pyridine homologues). Previous work has demonstrated that basic compounds in petrochemical samples are selectively ionized by positive ion ESI,[16,19] whereas acidic compounds are selectively ionized in negative ion ESI.[15,18]

Mass Spectral Complexity of a South American Crude Oil: The Positive Ion ESI Mass Spectrum. The full compositional complexity of the crude oil is revealed by a two-dimensional display of Kendrick mass defect versus nominal Kendrick mass[34] (Figure 2), in which each point represents a resolved mass spectral peak, including species that contain one or two ^{13}C atoms. Most of the 11 000 peaks are visible in that (single) display. Moreover, compound types and classes are visually distinguished as different vertically displaced horizontal rows, whereas carbon distributions appear as points spaced at intervals of 14 Da in each horizontal row.

Successively more detail appears in mass scale-expanded segments (Figure 3) of the full-range mass spectrum of Figure 1. For example, the 400−600 Da segment (Figure 3, bottom) shows peaks at each nominal mass. Starred peaks represent members of the same homologous series (−15 N), differing by integer multiples of ∼14 Da (CH$_2$). Further mass-scale expansion of the regions near the starred masses, 406, 504, and 588 Da (Figure 3, top), illustrates the increase in number of peaks per Dalton with increasing molecular weight (from 12 peaks at nominal 406 Da to 24 peaks at nominal 588 Da). The increased number of possible (and actual) elemental compositions with increased mass is documented in Table 1, which lists elemental compositions, types, and classes assigned at those representative masses.

Table 1 also demonstrates the utility of Kendrick mass analysis. For example, elemental compositions assigned to peaks at 406 Da suffice to identify members of the same homologous series at masses differing by integer multiples of 14.000 00 Da in Kendrick mass (e.g., 322, 504, 588, and 644 Da), because those compounds differ only in their number of CH$_2$ groups. Kendrick mass analysis thereby eliminates the need to identify peaks individually and allows identification of higher-mass species that would otherwise be impossible to assign on the basis of accurate mass alone. For example, at nominal 588 Da (Table 1), molecular formulas could be assigned to within ±0.4 ppm of their theoretical masses by use of Kendrick mass analysis and internal mass calibration.

A further-expanded 0.4 Da-wide mass segment at nominal 644 Da (Figure 4) exhibits 29 peaks, the heights of which exceed a threshold (shown as a dashed line) of 3σ of baseline noise. Of those 29 peaks, 23 could be assigned molecular formulas (Table 1) with an average of ∼45 carbons. (Figure 5, left). Additional periodicity at nominal 644 Da (and other nominal masses) results from compounds of the same class but different number of CH$_2$ groups and different multiples of seven rings plus double bonds, for example, classes N (−15 N, −29 N), NS (−47 NS, −33 NS, −19 NS), NS$_2$ (−37 NS$_2$, −23 NS$_2$), and NO (−45 NO, −31 NO, −17 NO). (Note that an addition of seven rings plus double bonds eliminates 14 hydrogens (i.e., reduces the IUPAC mass by 14.109 55 Da), whereas adding one CH$_2$ group increases the IUPAC mass by 14.015 65 Da); the combined effect is to produce a molecule whose mass is lower by 0.0939 Da.

Another issue is the presence of unresolved doublets and triplets. Figure 5 (right) graphically demonstrates that the number of mass-resolved peaks in the positive ion ESI mass spectrum reaches a maximum of ∼30 peaks at ∼600 Da for three reasons: (a) FT-ICR mass resolving power varies approximately inversely
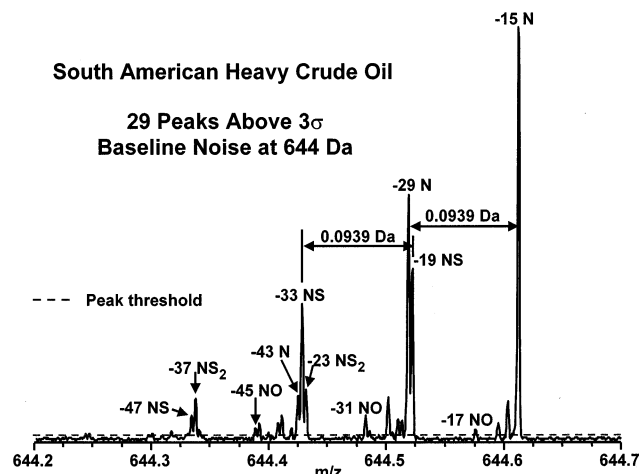
Figure 4. Mass scale-expanded segment from Figure 1 at 644 Da. Twenty-three of 29 peaks are identified and summarized in Table 1. Periodicities observed within a nominal mass unit represent multiple compounds of the same class (e.g., containing a single nitrogen atom) but of different type; e.g., −15 N (9 rings plus double bonds), −29 N (16 rings plus double bonds), −43 N (23 rings plus double bonds). Similar periodicities are denoted for classes NS, NS$_2$, and NO. There are some unresolved doublets and triplets as well.
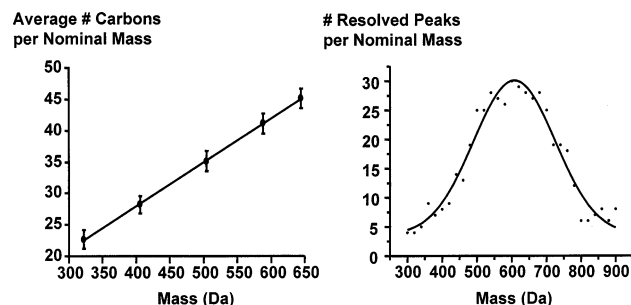
Figure 5. Two graphical measures of compositional complexity as a function of mass. The average number of carbons per Dalton (left) increases linearly ($R^2 = 0.9989$) from $300 < m/z < 650$, whereas the number of resolved peaks per Dalton (right) exhibits a Gaussian distribution over the range, $300 < m/z < 900$, with a maximum of ~30 resolved peaks at ~607 Da. Above 600 Da, mass resolution of peaks becomes more difficult as a result of the increasing number of peaks per Dalton and the decreasing relative abundance of high-molecular-weight compounds.

with mass;[35] (b) mass accuracy is proportional to peak height-to-noise ratio,[36] whereas crude oil compound relative abundance (and thus FT-ICR mass spectral peak height) decreases above ~600 Da; and (c) the tendency for closely spaced ion cyclotron resonances to coalesce increases with increasing mass and with decreasing spacing between the resonances.[37]

Estimating the Mass Spectral Complexity of a South American Crude Oil: Ionization of All Components. The compositional complexity of crude oil revealed by FT-ICR MS is truly staggering. In addition to the 11 000+ (basic) compounds reported here by positive ion ESI-FT-ICR-MS, it seems likely that most of the comparable number of (acidic) compounds observed by negative ion ESI-FT-ICR-MS will be chemically different, as well. Moreover, those ~20 000 acidic or basic polar compounds make up <15 wt % of all compounds in crude oil,[7] so that the entire crude oil "petroleome" (hydrocarbons included) should contain a substantially higher number. Although the above estimate includes nuclides containing one or more ^{13}C or ^{34}S atoms, the number of compositionally distinct compounds could still exceed 30 000.

Future Prospects. The present example sets a record for resolved chemical complexity that should soon be broken. First, electrospray ionization (as in this study) fails to generate ions from nonpolar species. (It's worth noting that crude oil contains >85% hydrocarbons.) Therefore, we are currently designing and building a field desorption ionization source for a 9.4 T actively shielded FT-ICR mass spectrometer to gain access to nonpolar species in the same mixture. Second, FT-ICR-MS resolving power increases linearly with magnetic field strength, B, and dynamic range increases as B^2.[38] One can therefore expect to resolve proportionately more species at higher magnetic field (we expect to install a 15 T magnet by the end of 2002). Third, by dividing the mass spectrum into several segments, it is possible to improve the dynamic range for each individual segment, thereby exposing even more components—preliminary results suggest an improvement by a factor of at least two for a spectrum divided into 10 segments. With such improvements, it seems safe to predict that it will soon be possible, in a single one-dimensional mass spectrum, to resolve tens of thousands of chemically distinct components from a complex mixture.

A closely related issue is the ability to *identify* those thousands of components according to their elemental compositions. At the present mass resolving power ($m/\Delta m_{50\%} > 300\ 000$) and mass accuracy (subparts-per-million), we can typically assign unique elemental compositions to >75% of the resolved peaks in an FT-ICR mass spectrum of singly charged ions ranging from 225 to 1000 Da. Although mass accuracy alone is insufficient for species above ~400 Da, the patterns revealed by a Kendrick mass defect spectrum[34] allow for sorting of ions of different class, type, and carbon distribution, thereby extending the range of unique chemical formula assignment to ~900 Da. Automation of that process is under way.

(35) Marshall, A. G.; Comisarow, M. B.; Parisod, G. *J. Chem. Phys.* **1979**, *71*, 4434−4444.
(36) Chen, L.; Cottrell, C. E.; Marshall, A. G. *Chemom. Intell. Lab. Syst.* **1986**, *1*, 51−58.
(37) Mitchell, D. W.; Smith, R. D. *Phys. Rev. E* **1995**, *52*, 4366−4386.
(38) Marshall, A. G.; Guan, S. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1819−1823.