МФК МГУ
ХИМИЯ ПРИРОДНОГО ОРГАНИЧЕСКОГО ВЕЩЕСТВА И ПРИРОДОПОДОБНЫЕ ТЕХНОЛОГИИ

Лекция 5.
Масс-спектрометрия сверхвысокого разрешения природного органического вещества: способы обработки данных

И.В. Перминова

Химический факультет МГУ им. М.В. Ломоносова

18.03.2020

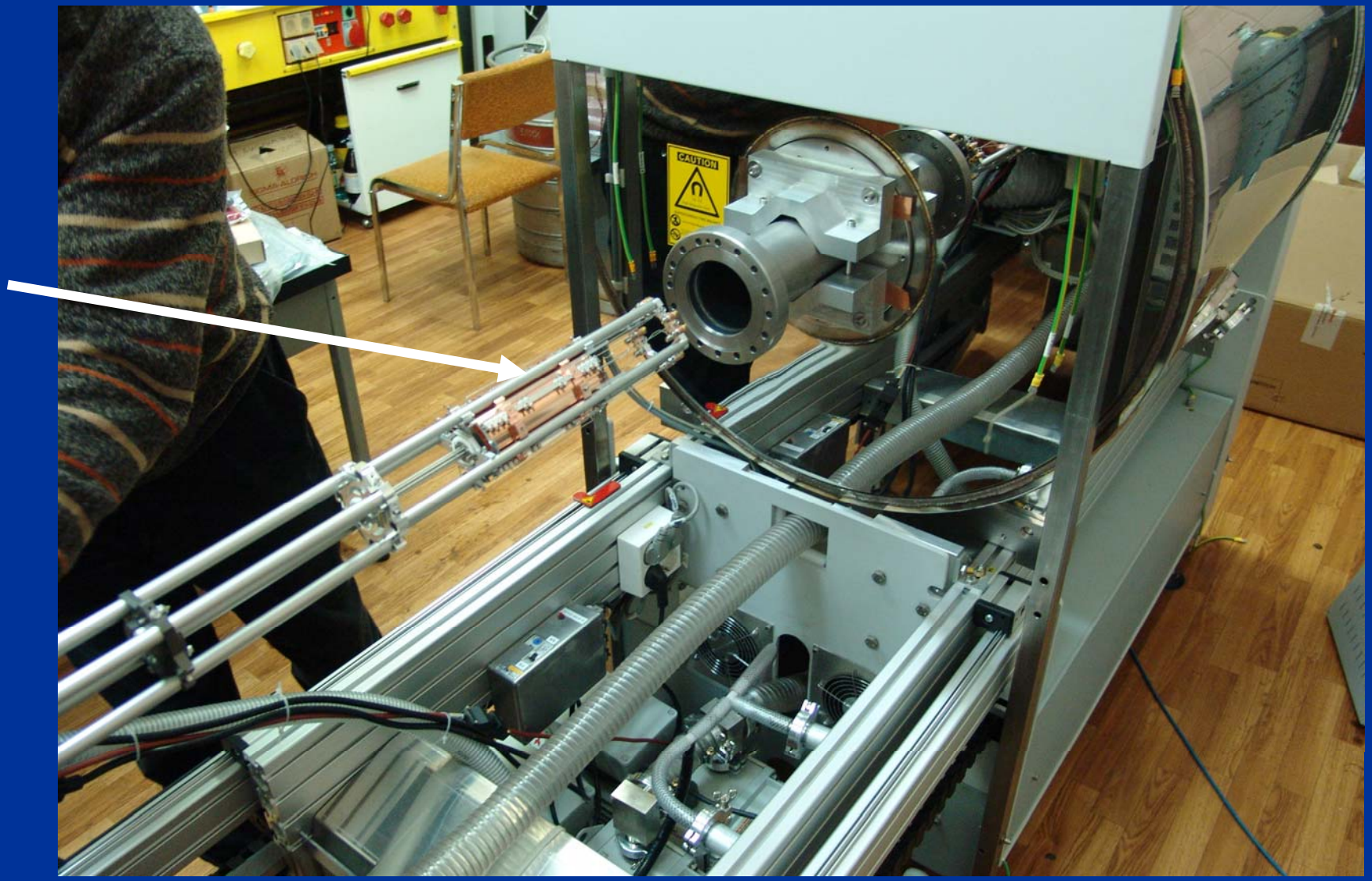
Содержание

- FTICR-MS для анализа сложных смесей: способы интерпретации данных
- Диаграммы Дефекта Масс Кендрика
- Диаграммы Ван Кревелена
- Кластерный и корреляционный многомерный анализ

УСТАНОВКА СВЕРХПРОВОДЯЩЕГО МАГНИТА

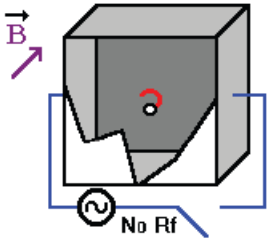
ИОННАЯ ОПТИКА FTICR МАСС СПЕКТРОМЕТРА

УСТАНОВКА ЯЧЕЙКИ ИОННОГО ЦИКЛОТРОННОГО РЕЗОНАНСА



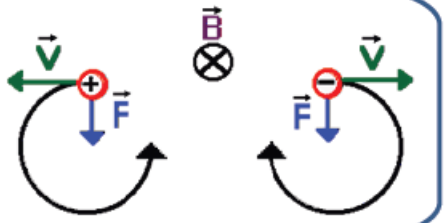
FTICR MS

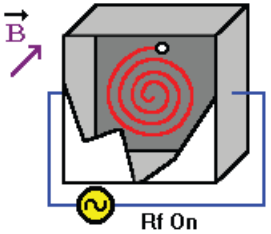
FTMS in 4 Easy Steps



1) Ions before excitation

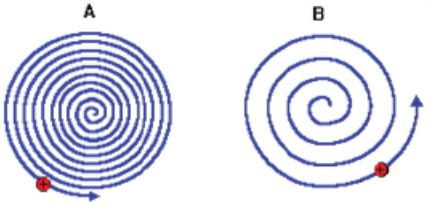
$$F = qv \times B$$

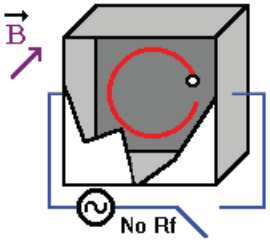




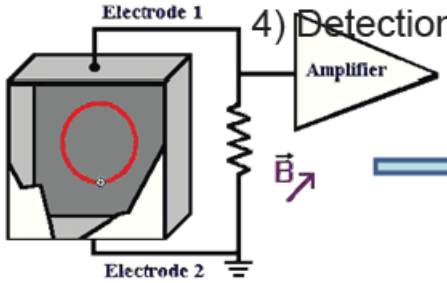
2) Ions during excitation

$$\omega_c = \frac{zB}{2\pi m}$$

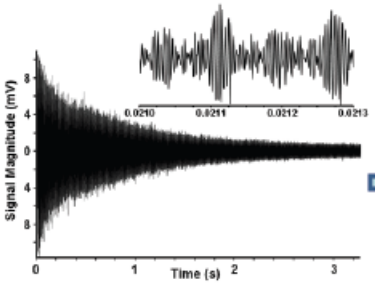




3) Ions after excitation

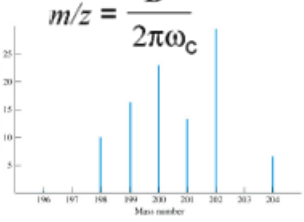


4) Detection



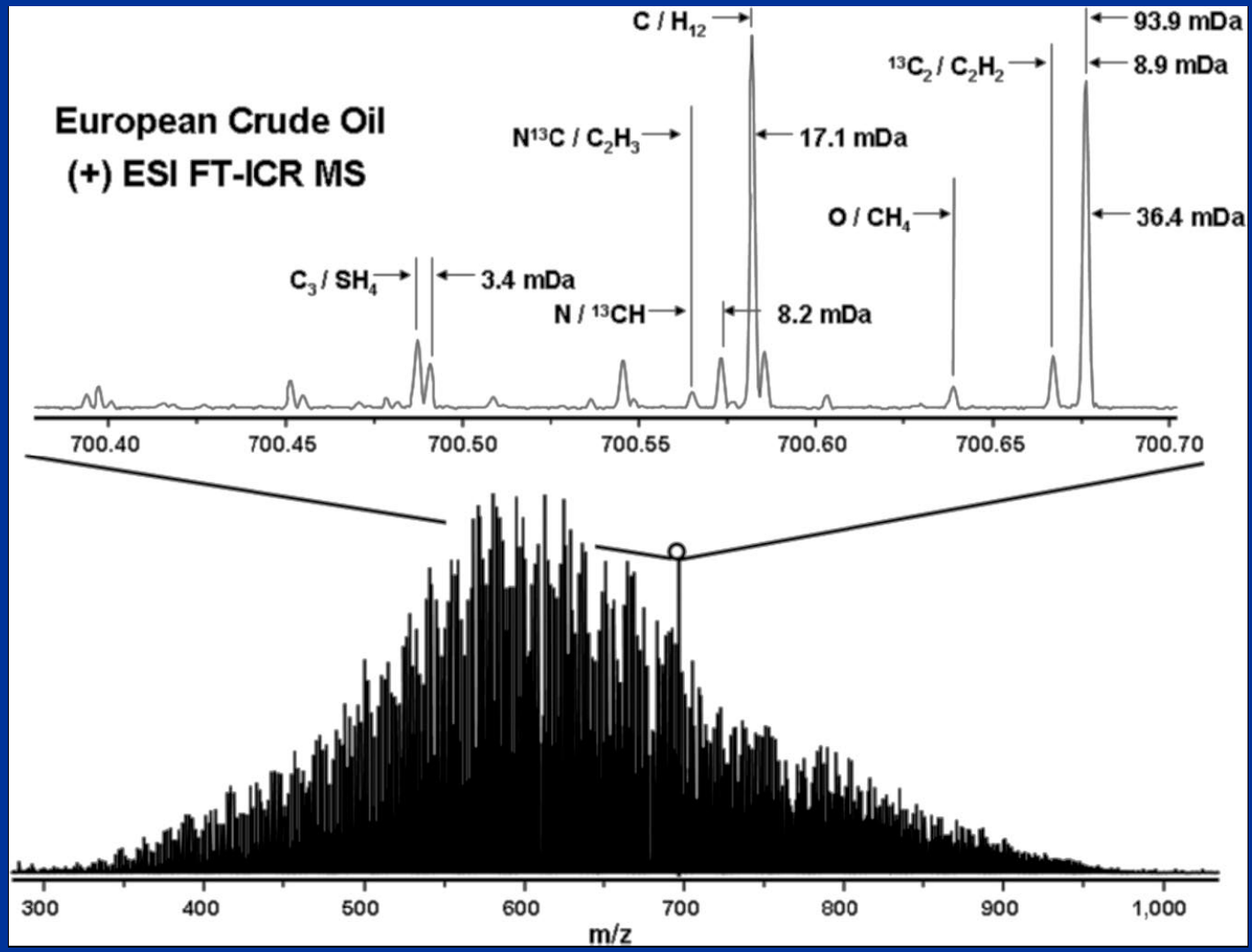
FT

$$m/z = \frac{B}{2\pi\omega_c}$$



11

МАСС СПЕКТР ИЦР ПФ НЕФТИ



МАСС СПЕКТР ИЦР ПФ НЕФТИ

Anal. Chem. **2001,** *73,* 4676−4681

Kendrick Mass Defect Spectrum: A Compact Visual Analysis for Ultrahigh-Resolution Broadband Mass Spectra

Christine A. Hughey,[†] Christopher L. Hendrickson, 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

Kuangnan Qian

ExxonMobil Research and Engineering, 1545 Route 22 East, Annandale, New Jersey 08801

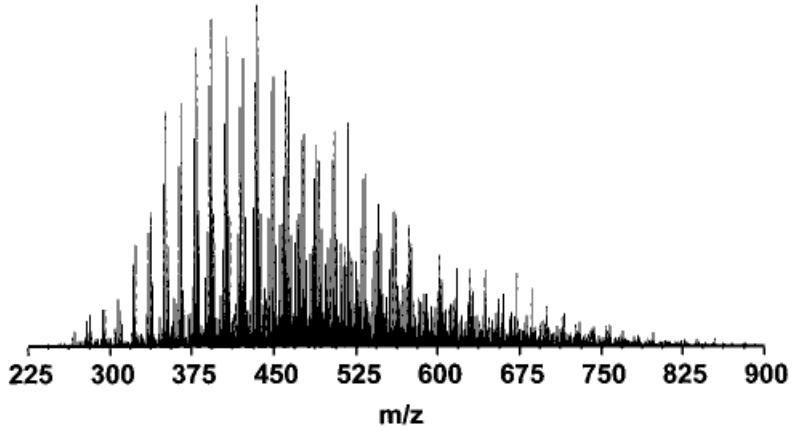


Figure 1. Broadband electrospray ionization FTICR mass spectrum of a heavy crude oil. The average resolving power across the spectrum is $m/\Delta m_{50\%}$ = 390 000 and the signal-to-noise ratio of the base peak is 1480:1.

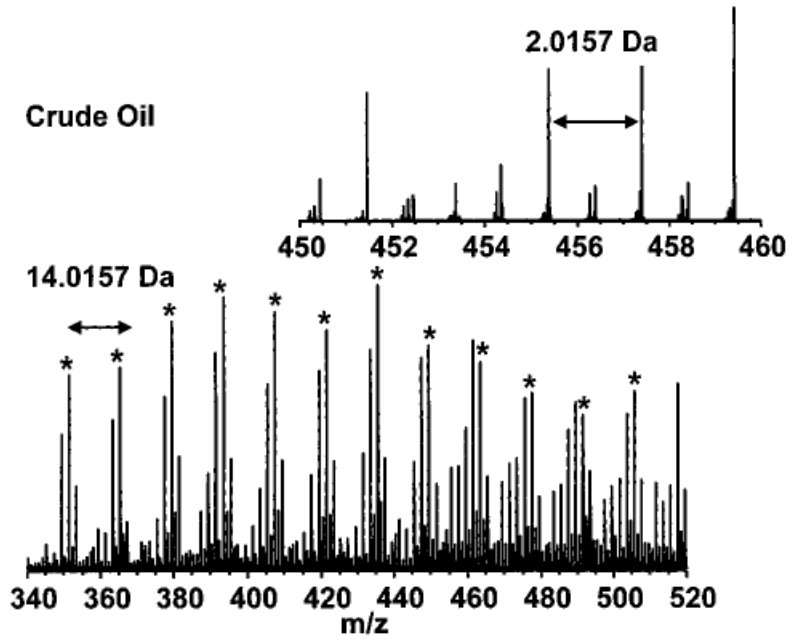


Figure 2. Mass scale-expanded segments of the full range petroleum mass spectrum of Figure 1. Species are present at every nominal mass, with obvious periodicities at every 2 (top) or 14 (bottom) nominal mass units.

МАСС СПЕКТР ИЦР ПФ НЕФТИ

Anal. Chem. **2001**, *73*, 4676−4681

Kendrick Mass Defect Spectrum: A Compact Visual Analysis for Ultrahigh-Resolution Broadband Mass Spectra

Christine A. Hughey,† Christopher L. Hendrickson, 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

Kuangnan Qian

ExxonMobil Research and Engineering, 1545 Route 22 East, Annandale, New Jersey 08801

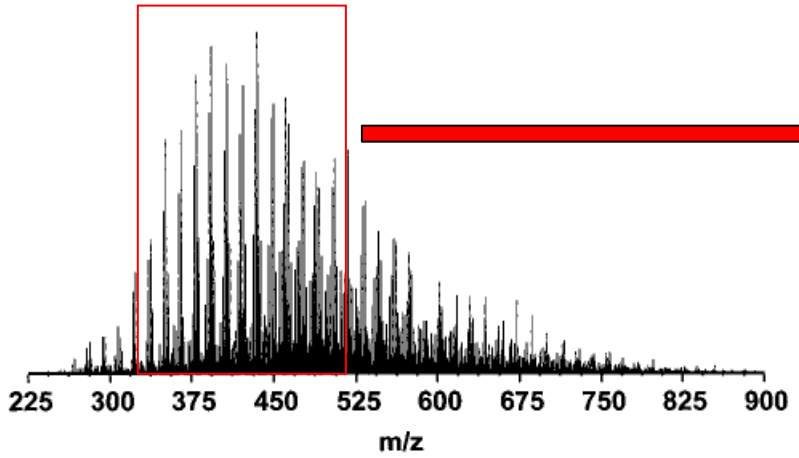


Figure 1. Broadband electrospray ionization FTICR mass spectrum of a heavy crude oil. The average resolving power across the spectrum is $m/\Delta m_{50\%}$ = 390 000 and the signal-to-noise ratio of the base peak is 1480:1.

Масс-спектр тяжелой нефти

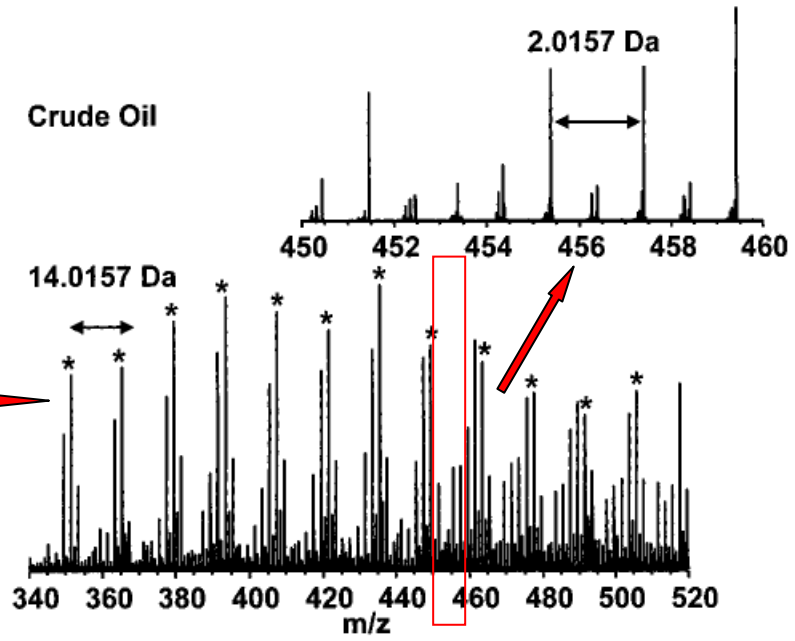


Figure 2. Mass scale-expanded segments of the full range petroleum mass spectrum of Figure 1. Species are present at every nominal mass, with obvious periodicities at every 2 (top) or 14 (bottom) nominal mass units.

Периодичность в масс-спектрах нефти

МАСС СПЕКТР ИЦР ПФ НЕФТИ ДЛЯ ОДНОЙ НОМИНАЛЬНОЙ МАССЫ: 457

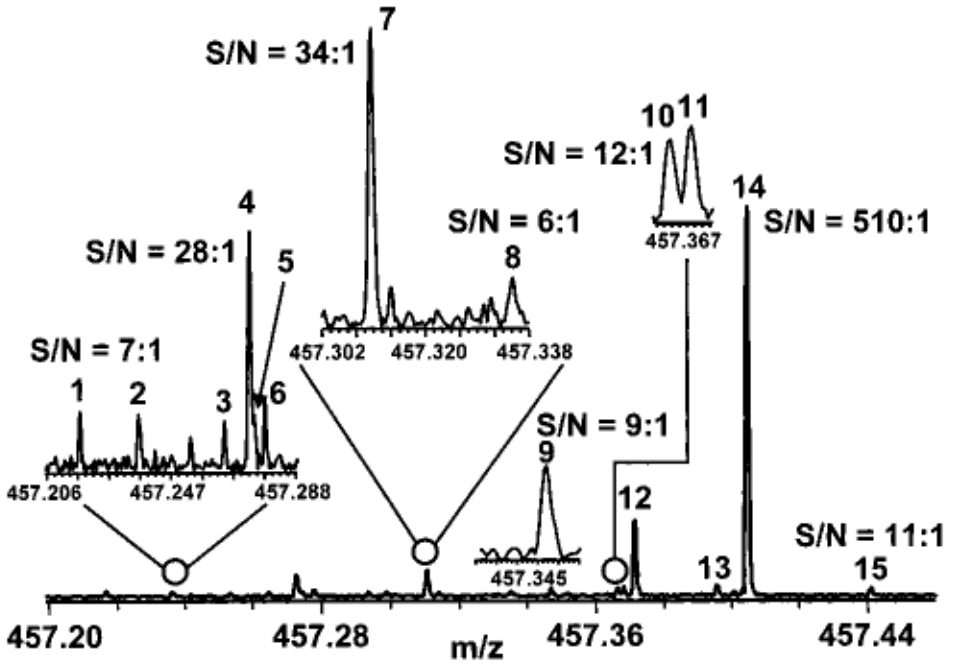


Figure 3. A further mass scale-expanded segment (plotted as IUPAC mass) of the crude oil mass spectrum of Figure 1, allowing for visual resolution and elemental composition assignment (based on accurate mass measurement) of all of the 15 chemically distinct species at a single nominal mass. The tabulated data for the 15 peaks show an average mass error of ~0.0001 Da for the proposed assignments. The elemental composition reveals both the "class" (i.e., number(s) of various heteroatoms) and "type" (i.e., number of rings plus double bonds), expressed according to the "*Z*" value in the chemical formula, $C_cH_{2c+Z}N_nO_oS_s$.

МАСС СПЕКТР ИЦР ПФ НЕФТИ ДЛЯ ОДНОЙ НОМИНАЛЬНОЙ МАССЫ: 457

Table 1. Elemental Composition Assignments for the 15 Resolved Peaks of Nominal Mass, 457 (Figure 3)[a]

peak no.	measd mass	rel abund (%)	theor mass	error (mDa)	$m/\Delta m_{50\%}$	type (Z)	class (NOS)
1	457.2174	0.50	457.2173	+0.1	517 000	−36	O_2
2	457.2367	0.47	457.2366	+0.1	345 000	−35	$NO^{13}C$
3	457.2650	0.42	457.2649	+0.1	413 500	−32	N_2
4	457.2731	2.01	457.2730	+0.1	413 500	−33	$N^{13}C$
5	457.2747	0.47	457.2748	−0.1	138 000	−24	O_3
6	457.2784	0.60	457.2782	+0.2	413 500	−14	O_3S
7	457.3112	2.40	457.3112	0.0	413 500	−22	O_2
8	457.3357	0.40	457.3357	0.0	258 000	−2	O_4S
9	457.3475	0.64	457.3476	−0.1	413 500	−20	O
10	457.3669	0.74	457.3669	0.0	413 500	−19	$N^{13}C$
11	457.3687	0.85	457.3687	0.0	413 500	−10	O_3
12	457.3720	7.13	457.3721	−0.1	413 500	0	O_3S
13	457.3961	1.04	457.3961	0.0	413 500	−10	$O_2{}^{13}C_2$
14	457.4049	36.7	457.4051	−0.2	517 000	−8	O_2
15	457.4416	0.74	457.4415	+0.1	344 500	−6	O

[a] The "Z" value in the chemical formula, $C_cH_{2c+Z}N_nO_oS_s$, is related to the number of rings plus double bonds. Note that the listed masses refer to the observed (deprotonated) species but that the "Z" values are calculated for the corresponding neutral.

$C_cH_{2c+Z}N_nO_oS_s$

ОРТОГОНОЛИЗАЦИЯ МАСС СПЕКТРА ИЦР ПФ НЕФТИ ДЛЯ РАСЧЕТА ДЕФЕКТА МАСС

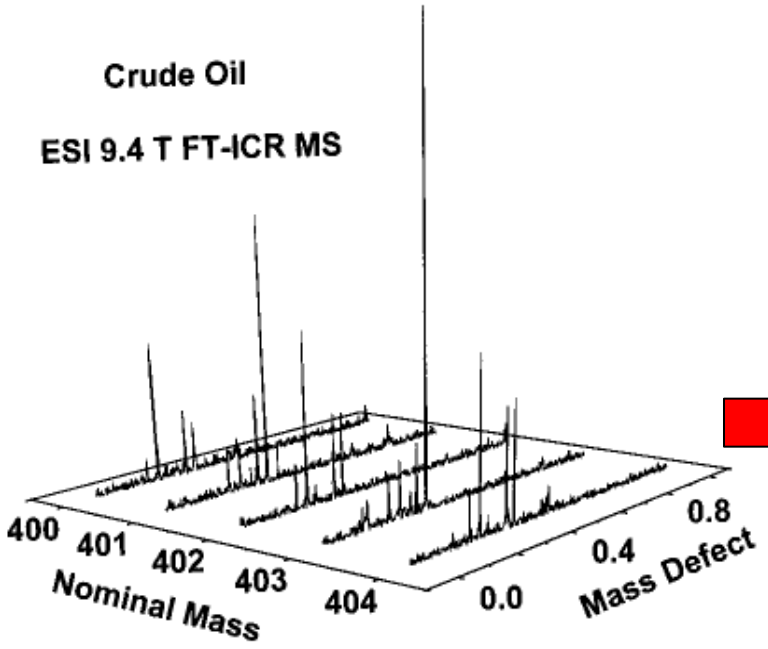


Figure 4. Three-dimensional display of a 5-Da segment of the mass spectrum of Figure 1. The spectrum has been sliced into 1-Da segments, and each segment is then rotated by 90° and scaled according to its "mass defect" (i.e., difference between exact and nominal mass). The segments are then stacked according to their nominal mass as shown.

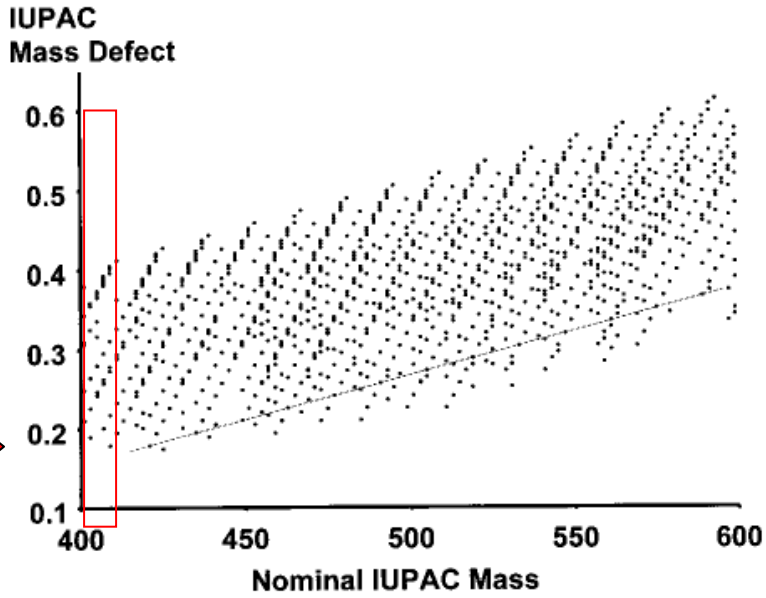
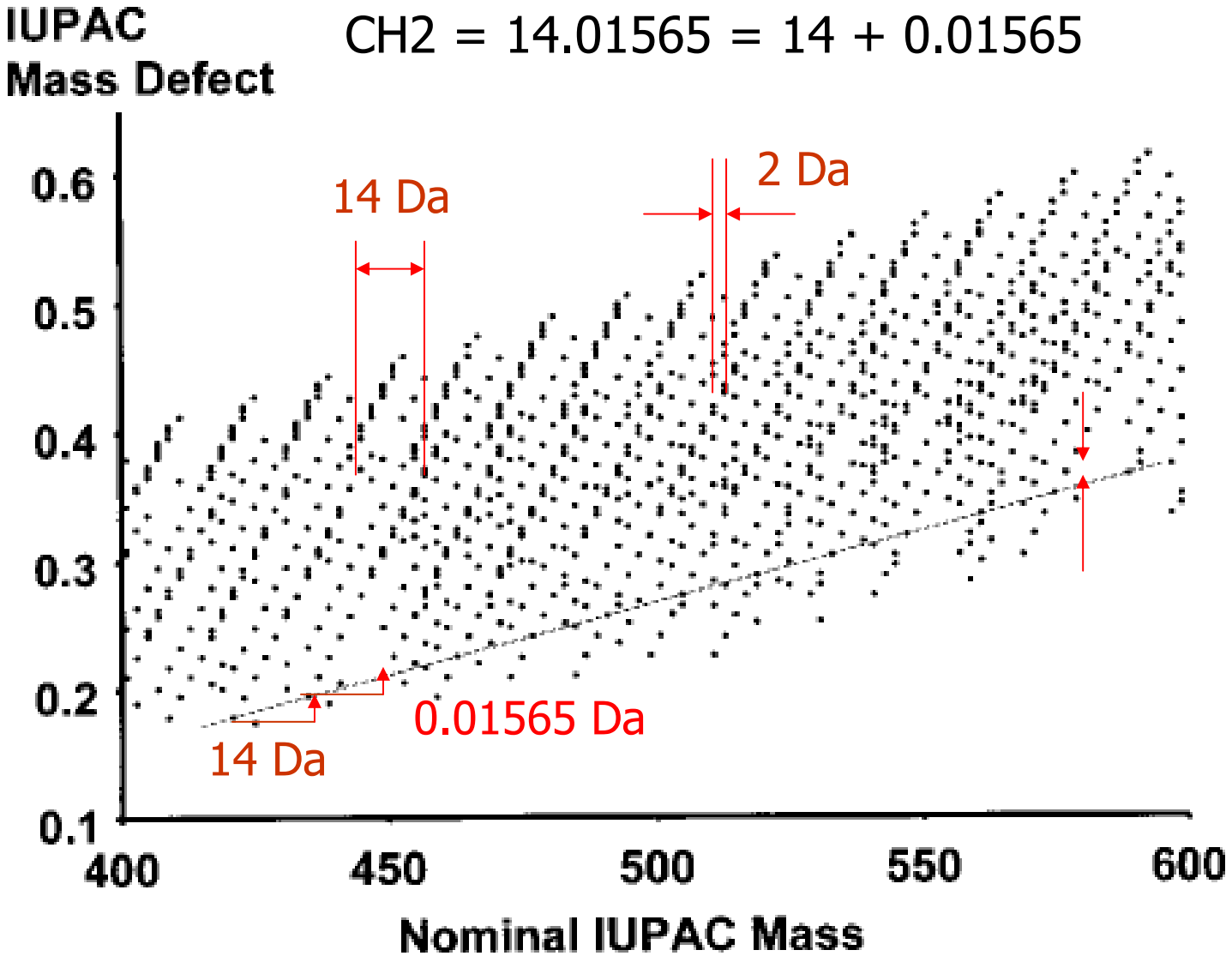


Figure 5. Plot of IUPAC mass defect vs nominal IUPAC mass for ions of *odd* nominal mass within a 200-Da segment of the mass spectrum of Figure 1. (It is convenient to display the even-mass and odd-mass data separately.) Note the periodicities of 2 and 14 Da on the nominal mass axis, as well as 0.015 65 Da (i.e., mass defect for two hydrogens) on the mass defect axis. These spacings make it possible to determine molecular "class" and "type" simultaneously over a wide mass range from a single display (see Figure 6 ff).

3D-представление спектра путем нарезки на участки длиной 1Da и их поворота на 90°

Свертка спектра в виде графика «дефекта масс»

ГРАФИК ДЕФЕКТА МАСС ИЮПАК



ДЕФЕКТ МАСС КЕНДРИКА

Преобразование массы ИЮПАК в массу Кендрика:

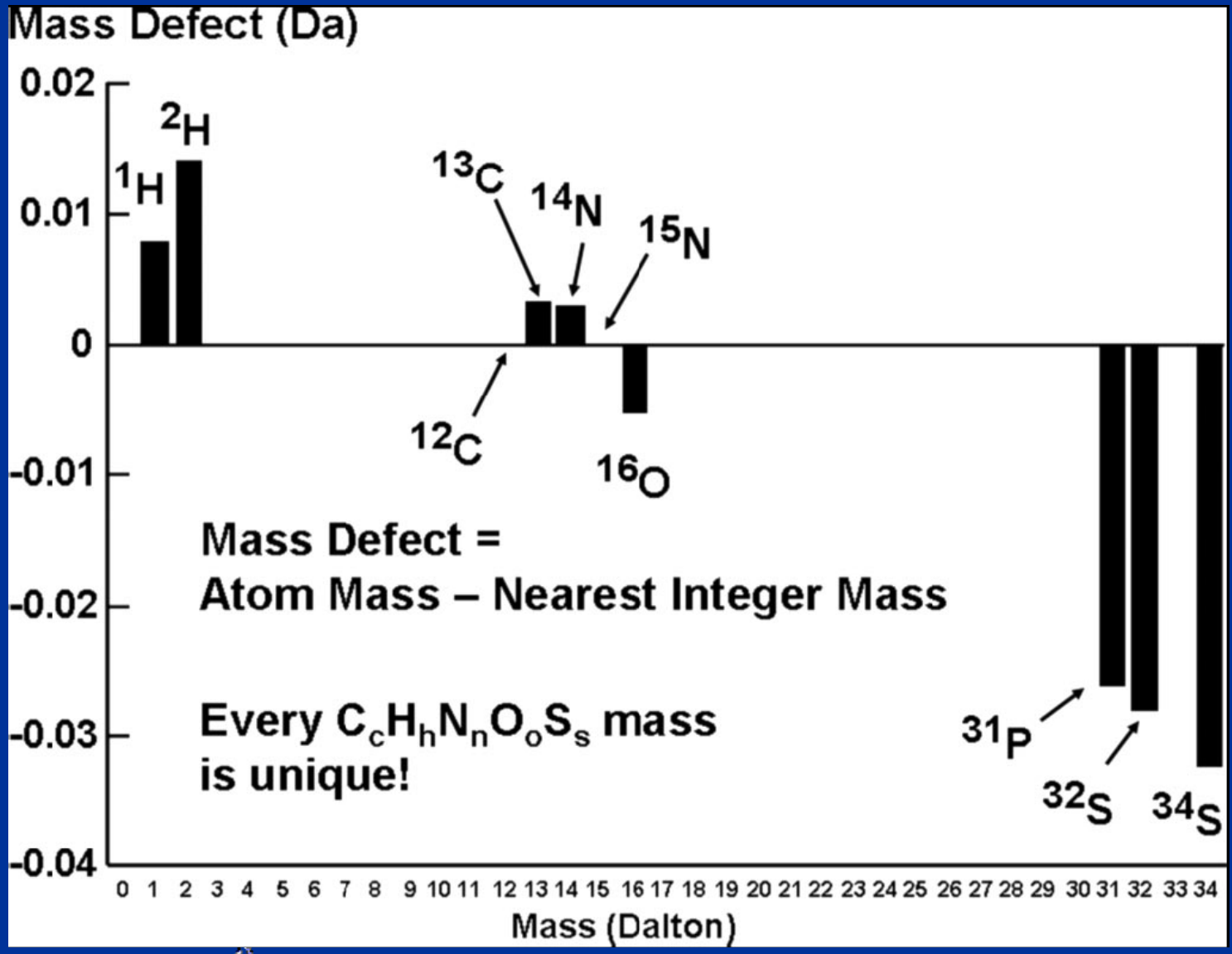
Масса Кендрика = Масса ИЮПАК x (14.00000/14.01565),

где 14.015650 – масса CH2 в шкале ИЮПАК, а
14.00000 – масса CH2 в шкале Кендрика

Массу Кендрика (KM) можно представить как сумму номинальной массы (NMK), представляющей собой ближайшее целое, и разности между этим ближайшим целым и массой Кендрика. Тогда эта разность и будет называться дефектом масс Кендрика (KMD), который можно рассчитать как:

KMD = NKM - KM

ДЕФЕКТ МАСС КЕНДРИКА ДЛЯ РАЗНЫХ ЭЛЕМЕНТОВ



ПРЕОБРАЗОВАНИЕ ГРАФИКА ДЕФЕКТА МАСС ИЮПАК В ДМ КЕНДРИКА

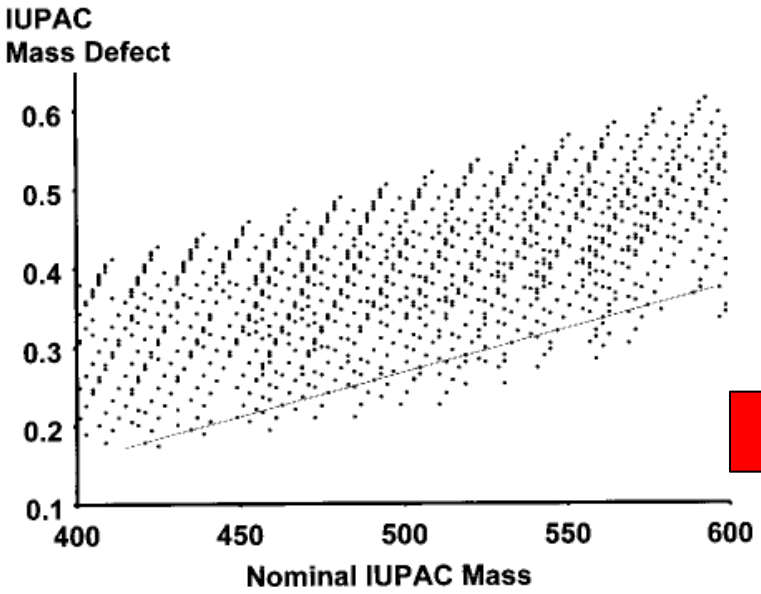


Figure 5. Plot of IUPAC mass defect vs nominal IUPAC mass for ions of *odd* nominal mass within a 200-Da segment of the mass spectrum of Figure 1. (It is convenient to display the even-mass and odd-mass data separately.) Note the periodicities of 2 and 14 Da on the nominal mass axis, as well as 0.015 65 Da (i.e., mass defect for two hydrogens) on the mass defect axis. These spacings make it possible to determine molecular "class" and "type" simultaneously over a wide mass range from a single display (see Figure 6 ff).

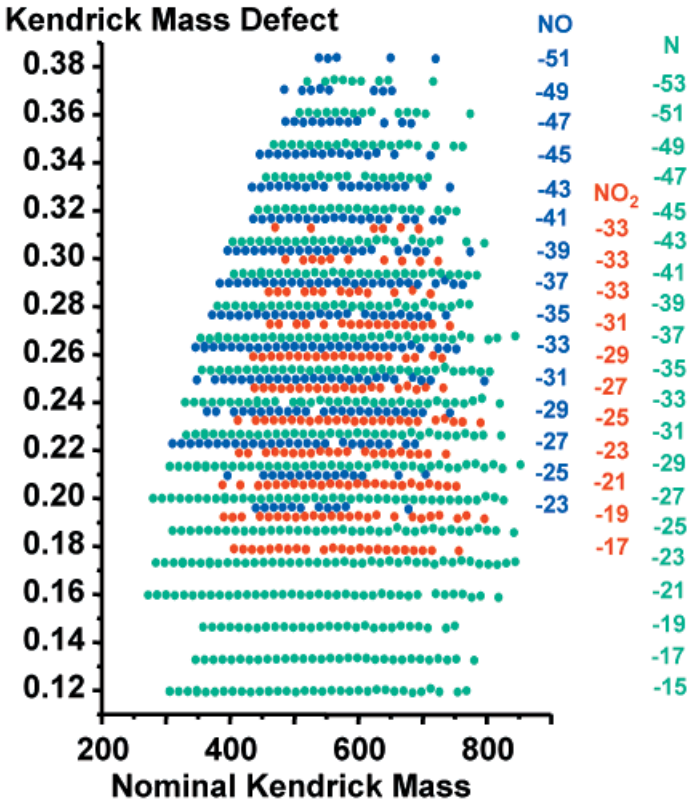


Figure 6. Kendrick mass defect vs nominal Kendrick mass (full mass range) for *even*-mass $^{12}C_c$ ions. The replacement of IUPAC mass by Kendrick mass converts the skewed display (e.g., Figure 5) into a rectilinear display, for simpler visual interpretation. Note the visual *vertical* separation of compound classes (e.g., N vs NO) and types (compounds with different number of rings plus double bonds) based on mass defect and the simultaneous visual *horizontal* distribution of number of CH_2 groups for a given compound class and type.

14.01565x(14.00000/14.01565) - 14 = 0

МАСС-СПЕКТР ИЦР ПФ РАСТВОРЕННОГО ОРГАНИЧЕСКОГО ВЕЩЕСТВА

УНИКАЛЬНО ВЫСОКАЯ РАЗРЕШАЯ СПОСОБНОСТЬ (до 500000 В ДИАПАЗОНЕ 200 – 700 Да)

До 100 тысяч пиков в спектре одного препарата ГВ!

ОПРЕДЕЛЕНИЕ БРУТТО-ФОРМУЛ ИЗ МАСС-СПЕКТРОВ ИЦР ПФ

Линейное Диофантово уравнение:
12.0000 n(C) + 1.0078 n(H) + 15.9950 n(O)+ 14.0030 n(N) = M

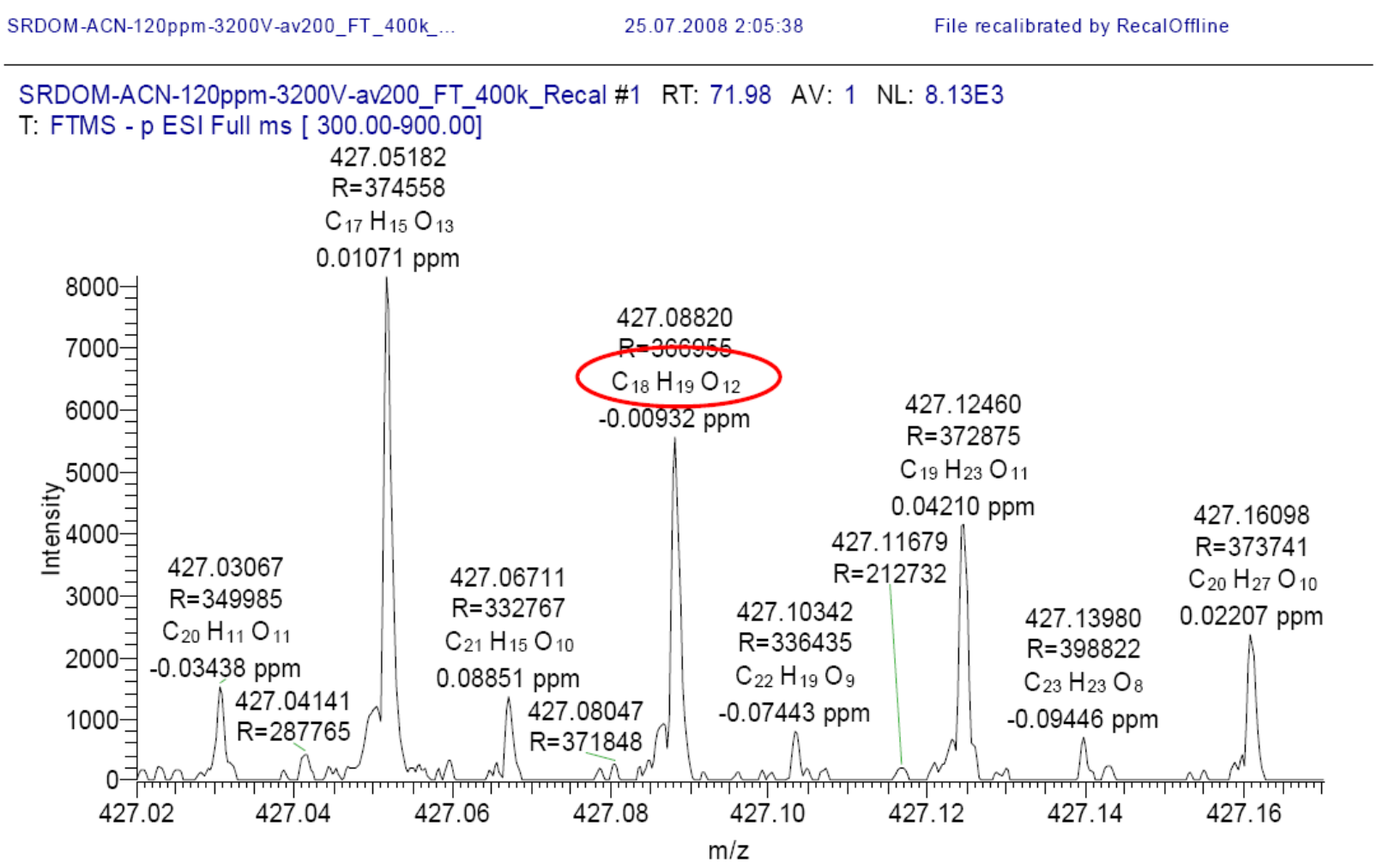
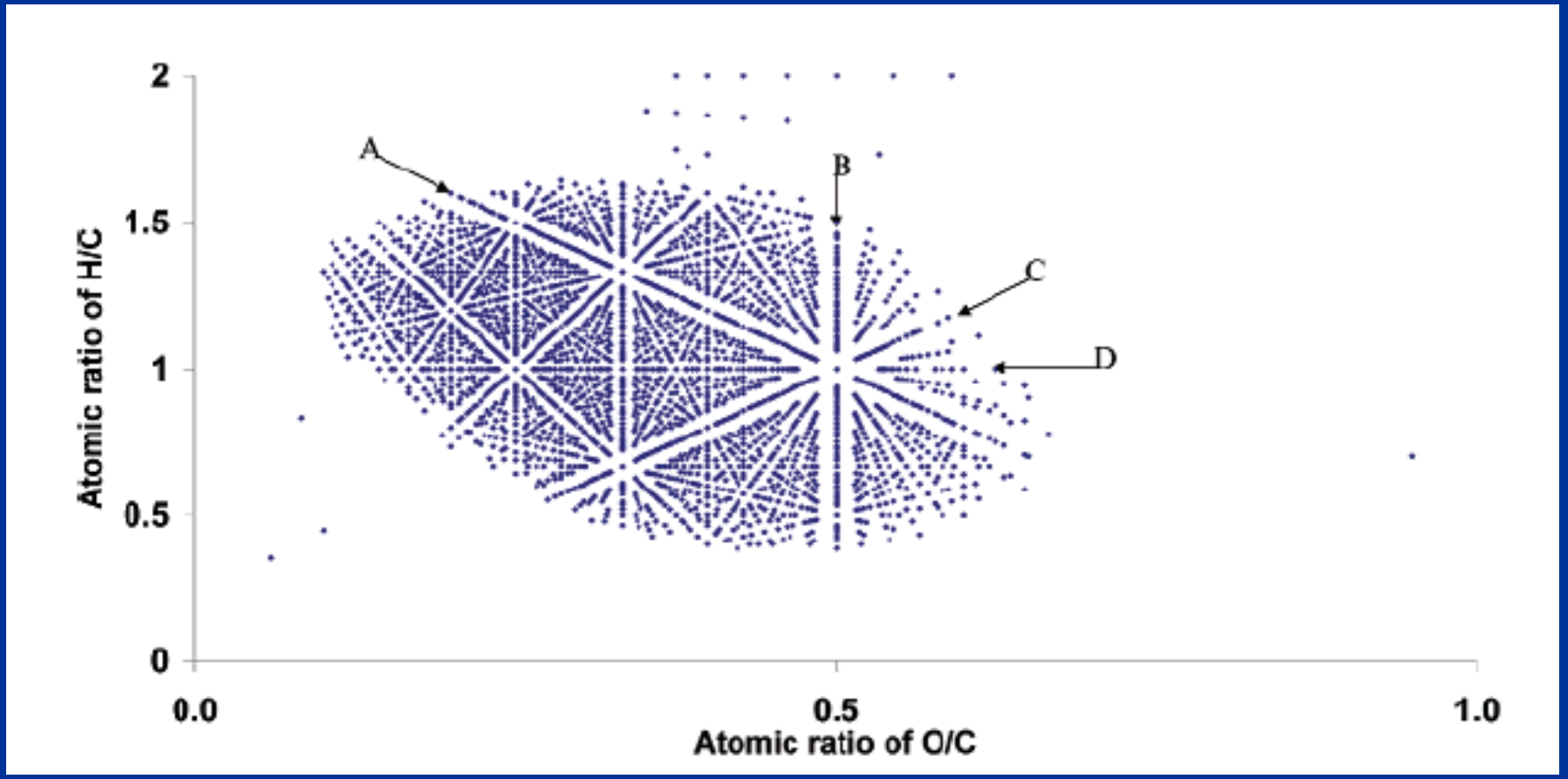


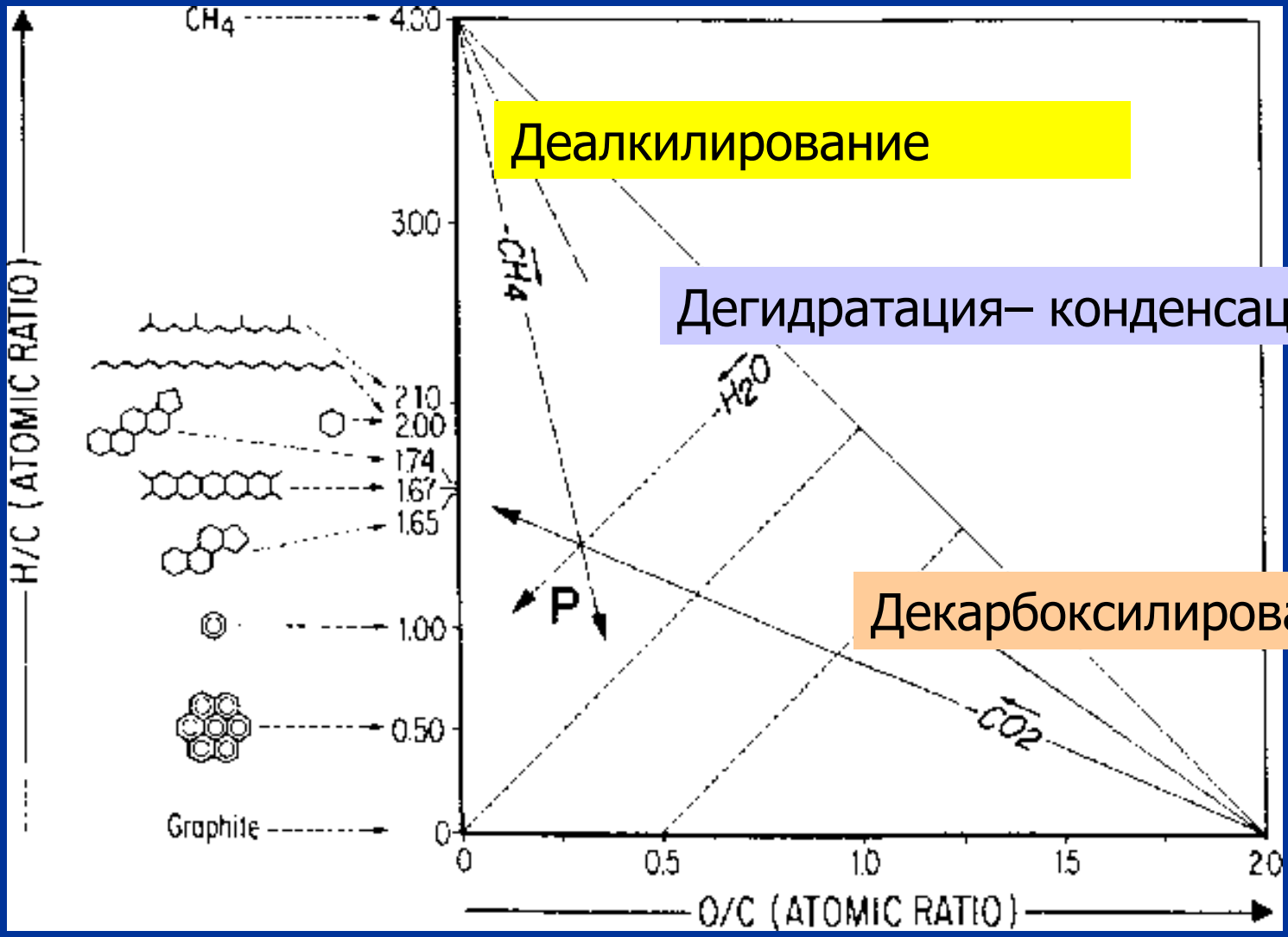
ДИАГРАММА ВАН КРЕВЕЛЕНА



(A) methylation, demethylation, or alkyl chain elongation; (B) hydrogenation or dehydrogenation; (C) hydration or condensation; and (D) oxidation or reduction.

Analytical Chemistry, Vol. 75, No. 20, October 15, 2003

ДИАГРАММА ВАН КРЕВЕЛЕНА - ПРОЦЕССЫ



Деалкилирование

Дегидратация– конденсация

Декарбоксилирование

МЕТОДЫ ГРАФИЧЕСКОГО ПРЕДСТАВЛЕНИЯ ДАННЫХ МС ИЦРПФ ДЛЯ ГВ

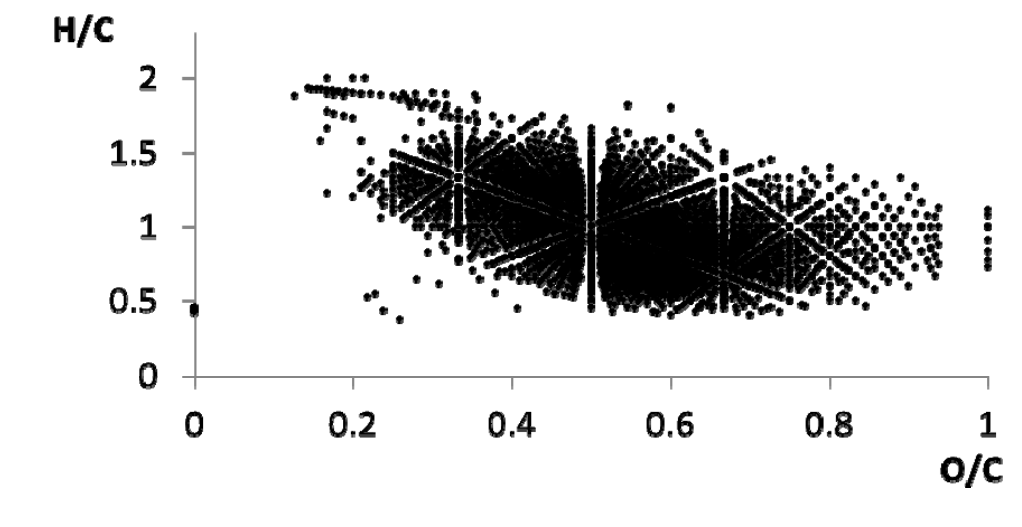
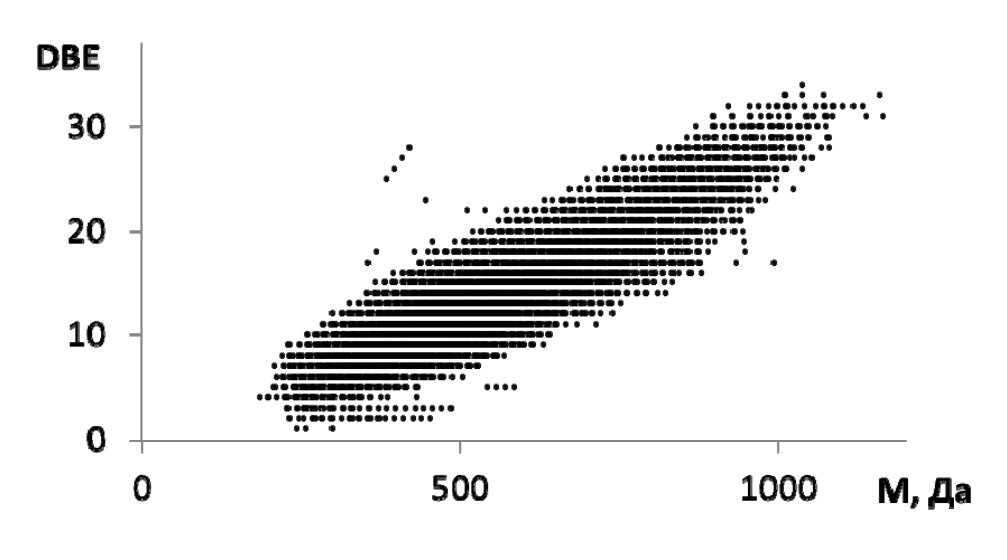
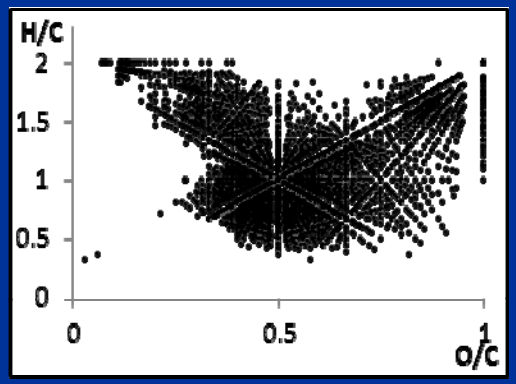


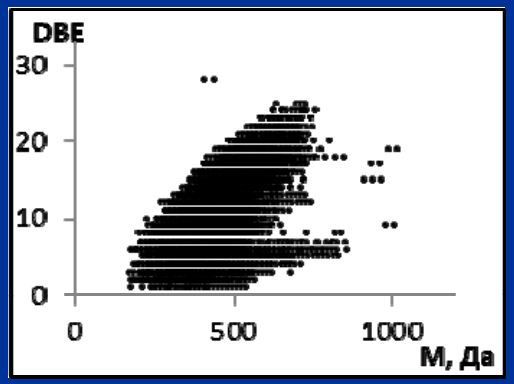
Диаграмма ван Кревелена



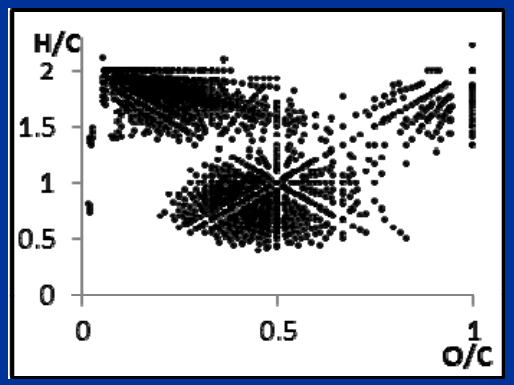
DBE=C-H/2+N/2+1
(степень ненасыщенности)

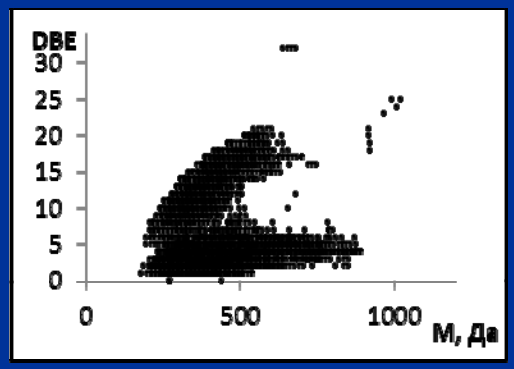
ДАННЫЕ МС ИЦР ПФ ДЛЯ РАЗЛИЧНЫХ ФРАКЦИЙ ГУМИНОВЫХ ВЕЩЕСТВ ТОРФА



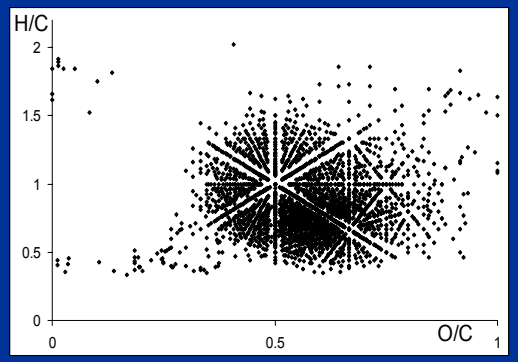


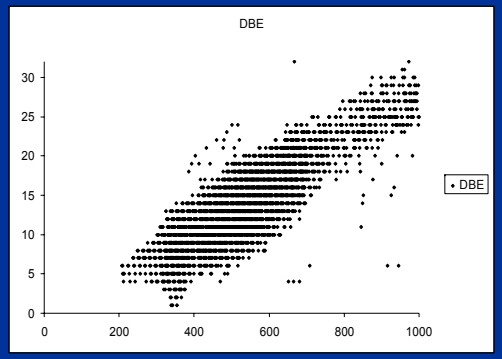
PHF (ГК+ФК торфа)





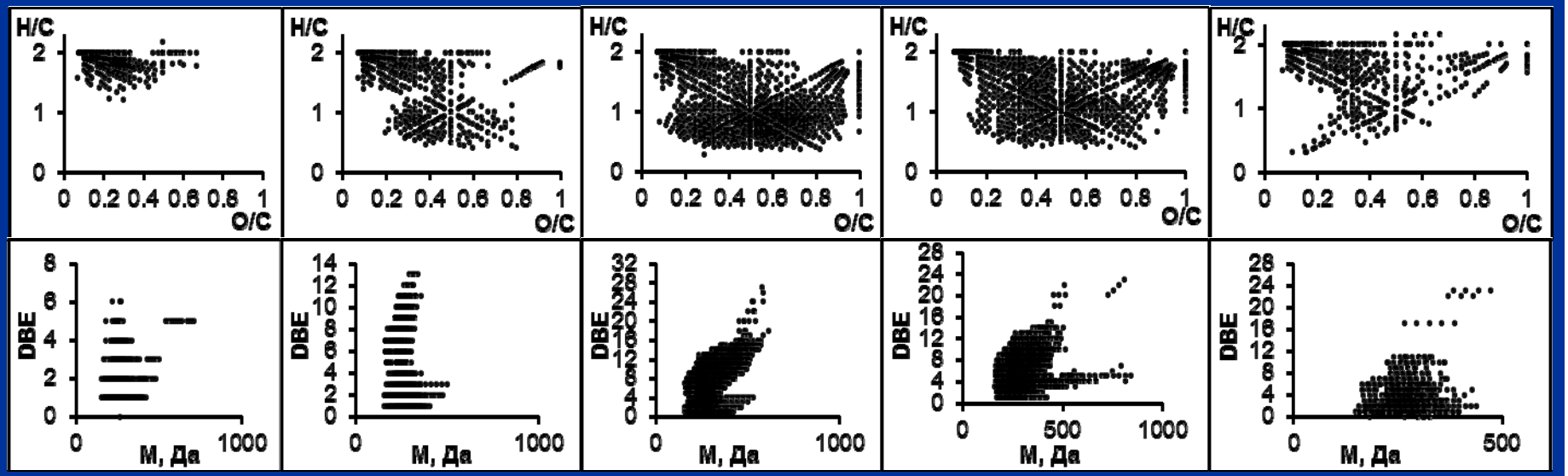
PHA (ГК торфа)

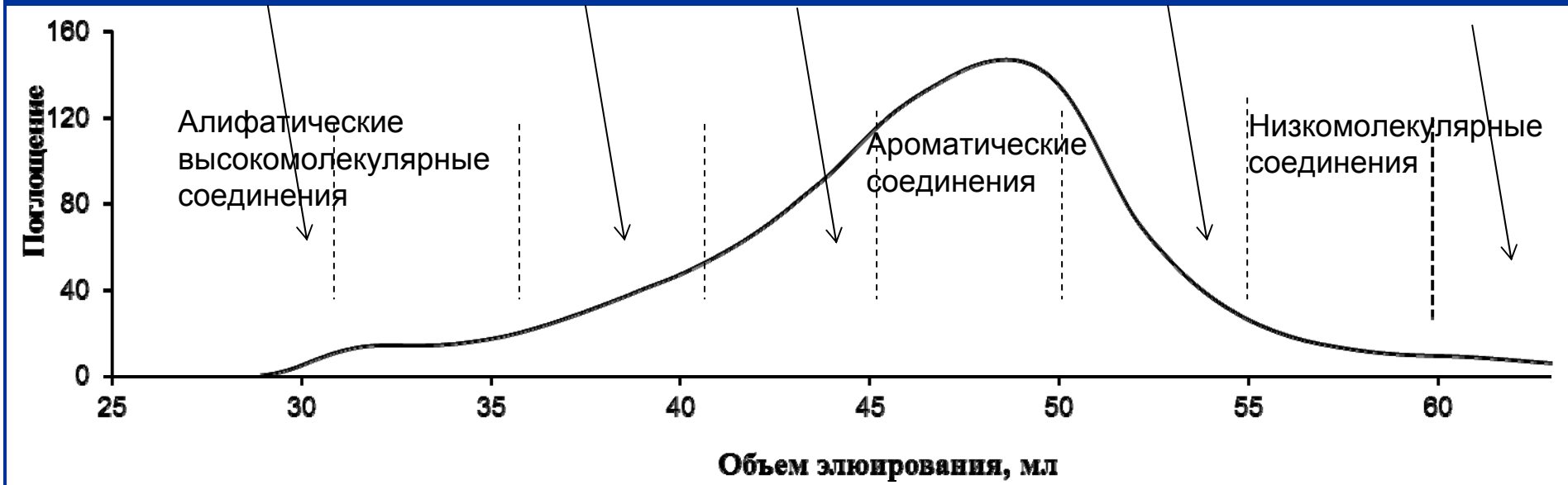




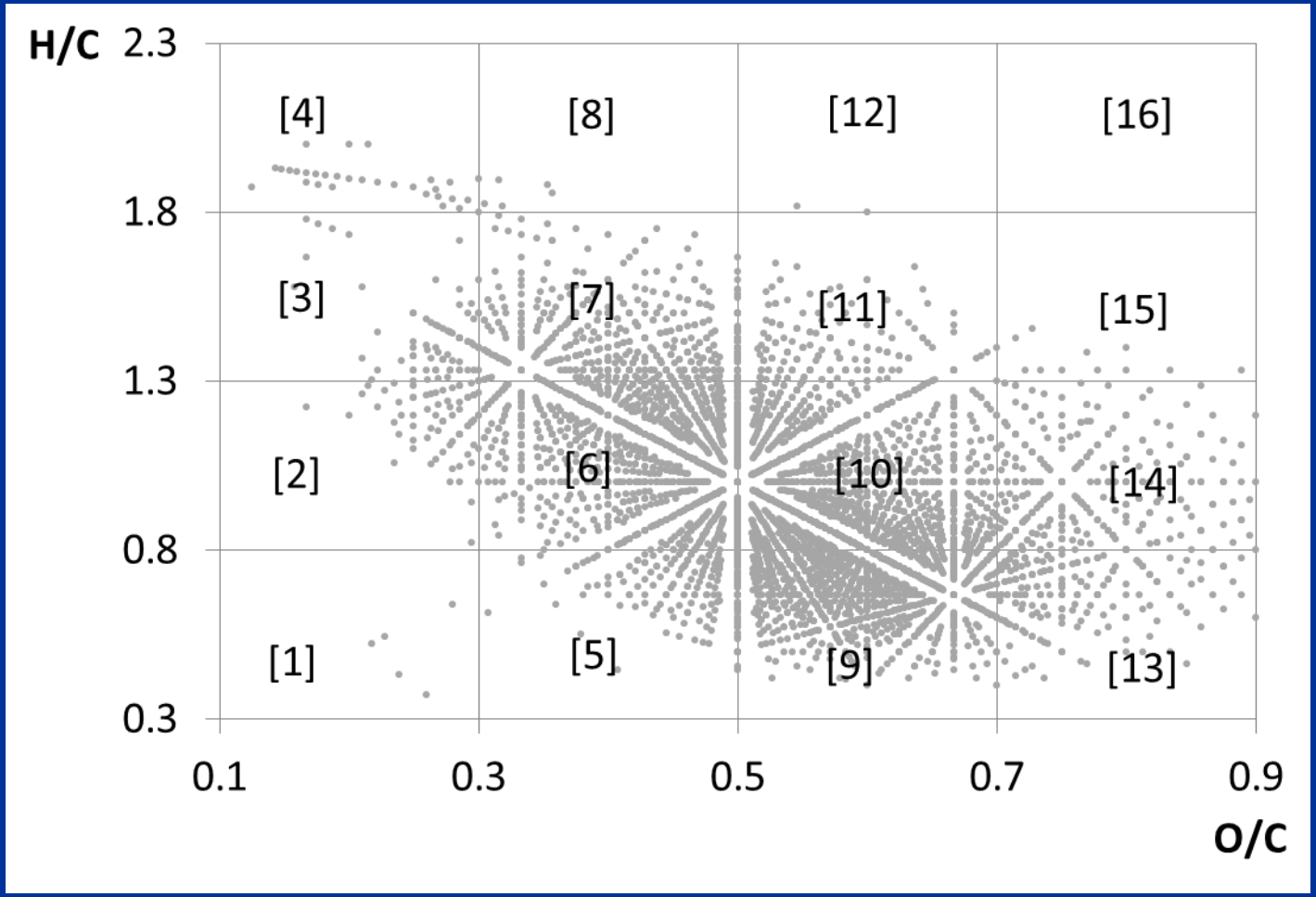
PFA (ФК торфа)

МАСС-СПЕКТРЫ ИЦР ПФ ММР-ФРАКЦИЙ ГУМИНОВЫХ ВЕЩЕСТВ ТОРФА





ПОЛУЧЕНИЕ ДЕСКРИПТОРОВ СТРОЕНИЯ ИЗ ДИАГРАММ ВАН КРЕВЕЛЕНА

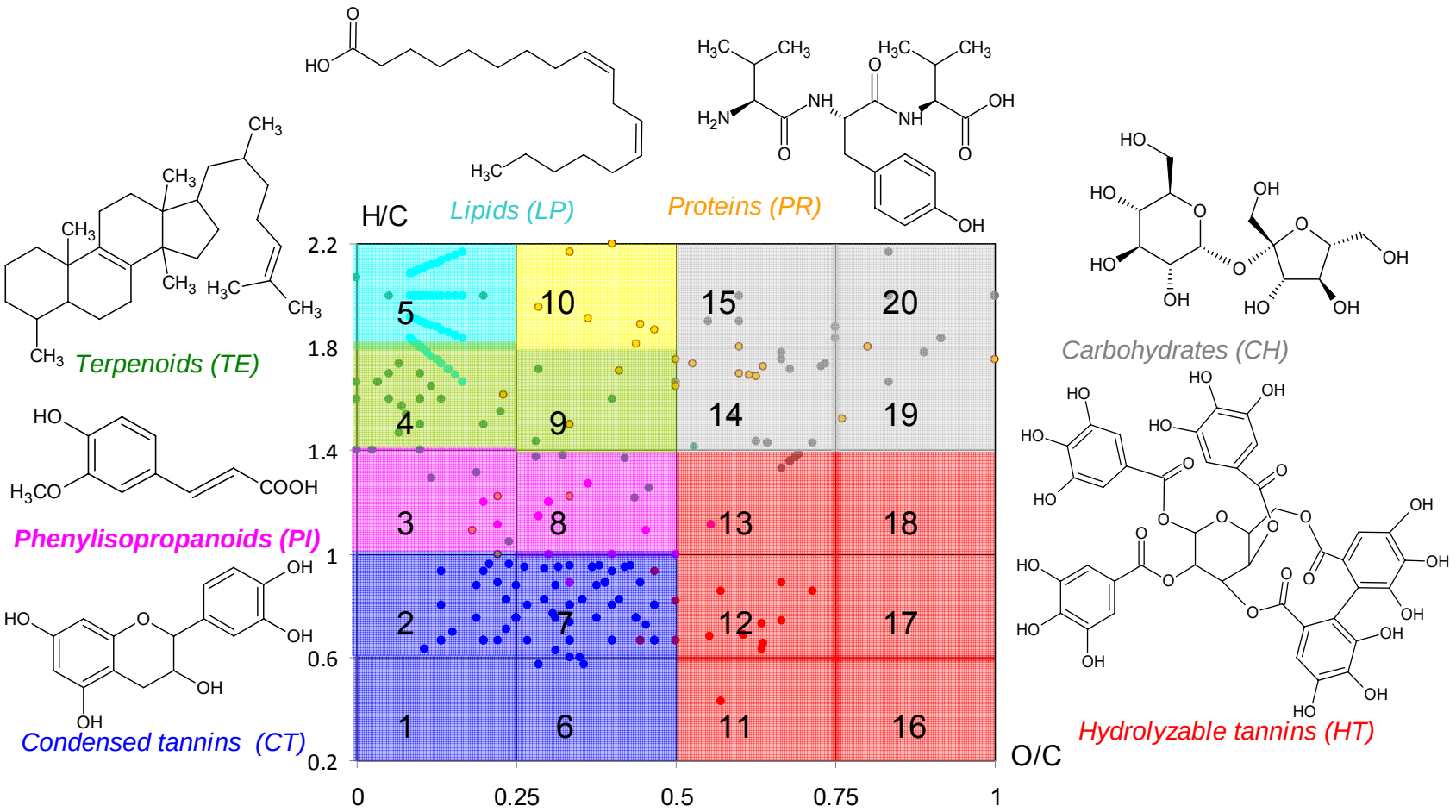


N=16

[k] – дескриптор, соответствующий k-ой области,
N – общее число соединений на диаграмме
N_k – число соединений, принадлежащих k-ой области

$$[k] = \frac{N_k}{N}, \qquad k = 1 \div N$$

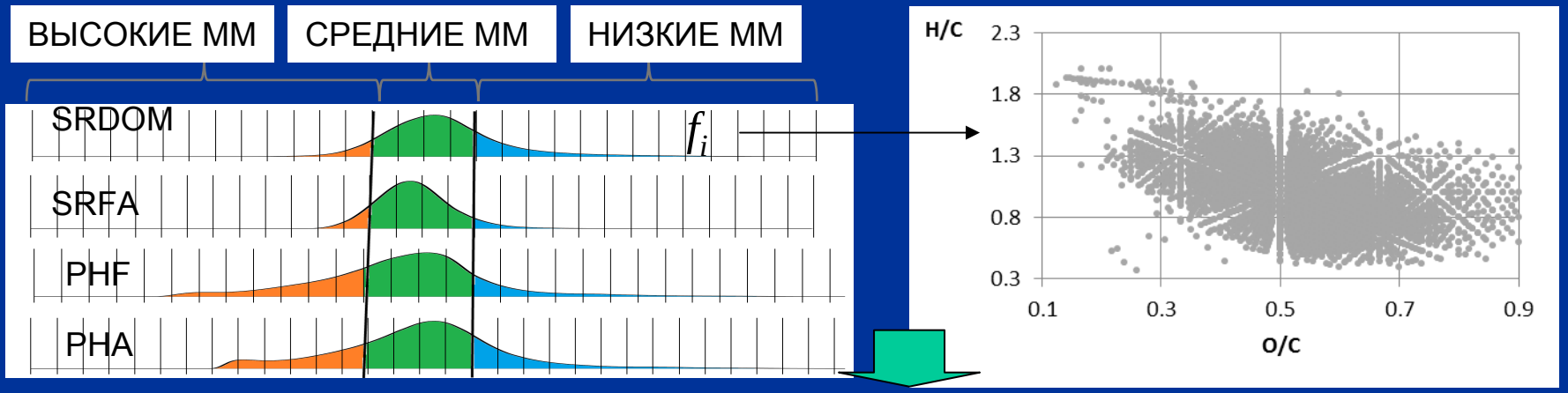
МОДЕЛЬНАЯ ДИАГРАММА ВАН КРЕВЕЛЕНА



Condensed Tannins [CT]	Phenyl-Isopropanoids [PI]	Terpenoids [TE]	Lipids [LP]	Proteins [PR]	Hydrolyzable tannins [HT]	Carbohydrates [CH]
D1, D2, D6, D7	D3, D8	D4, D9	D5	D10	D11-13, D16-D18	D14, D15, D19, D20

КЛАССИФИКАЦИЯ ПО ПРИЗНАКУ «МОЛЕКУЛЯРНАЯ МАССА (ММ)»

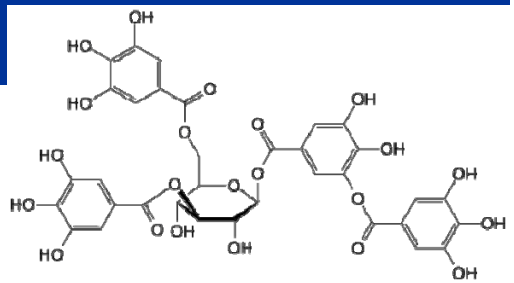
Объекты классификации: 57 ММ фракций (f_i) 4-х препаратов ГВ

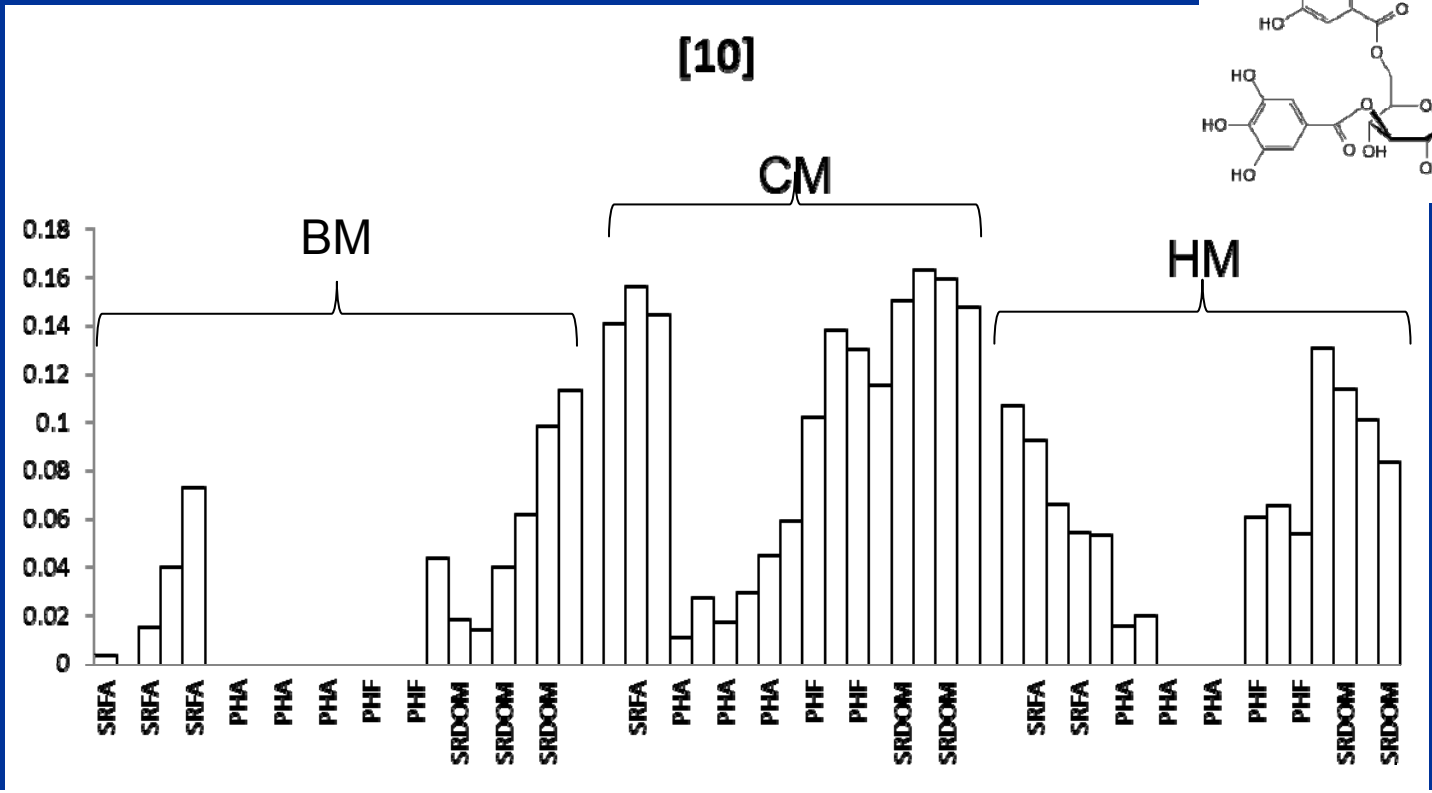


Дескрипторы	Правильно классифицировано, %			
	ВМ	СМ	НМ	Общее
[3, 15]	91	53	78	75
[7, 10, 15]	91	71	94	86
[2, 5, 10, 15]	95	88	83	89
[2, 5, 10, 13, 15]	95	88	86	90

СТРУКТУРНЫЕ ДЕСКРИПТОРЫ В КЛАССИФИКАЦИИ «СТРОЕНИЕ-МОЛЕКУЛЯРНАЯ МАССА»

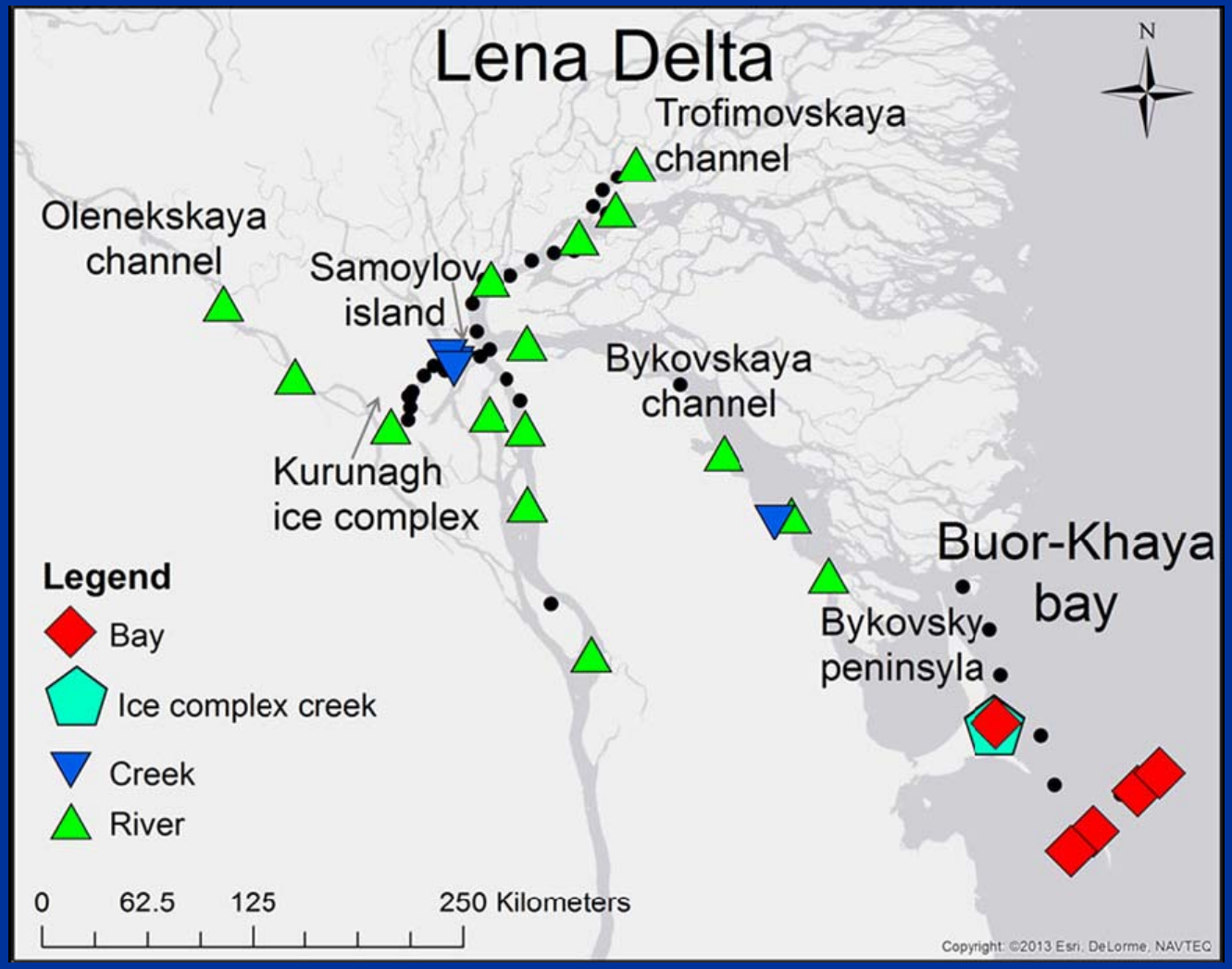
[2]	**лигнин+ флавоноиды**
[5]	**флавоноиды**
[10]	**таннины**
[15]	**Углеводы**





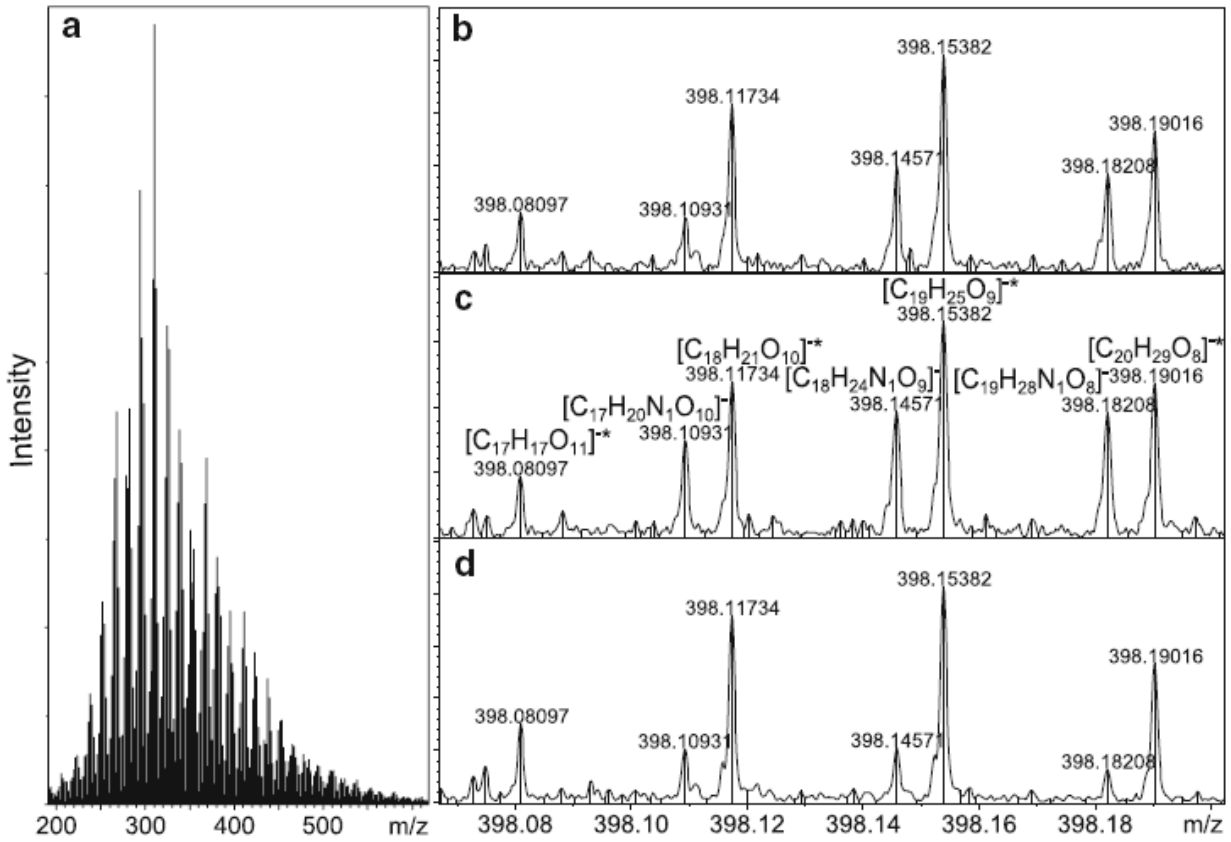
Для сопоставления больших блоков данных необходимо привлечение методов многомерной статистики – очень интенсивное направление современных исследований

ПРИМЕР: АНАЛИЗ РОВ ИЗ РАЗЛИЧНЫХ ИСТОЧНИКОВ В РАЙНОЕ РЕКИ Лена



Dubinenkov et al. Biogeochemistry (2015) 123:1–14

МАСС СПЕКТР ИЦР ПФ РОВ реки Лена



РОВ р. Лена

Ручей из мерзлоты

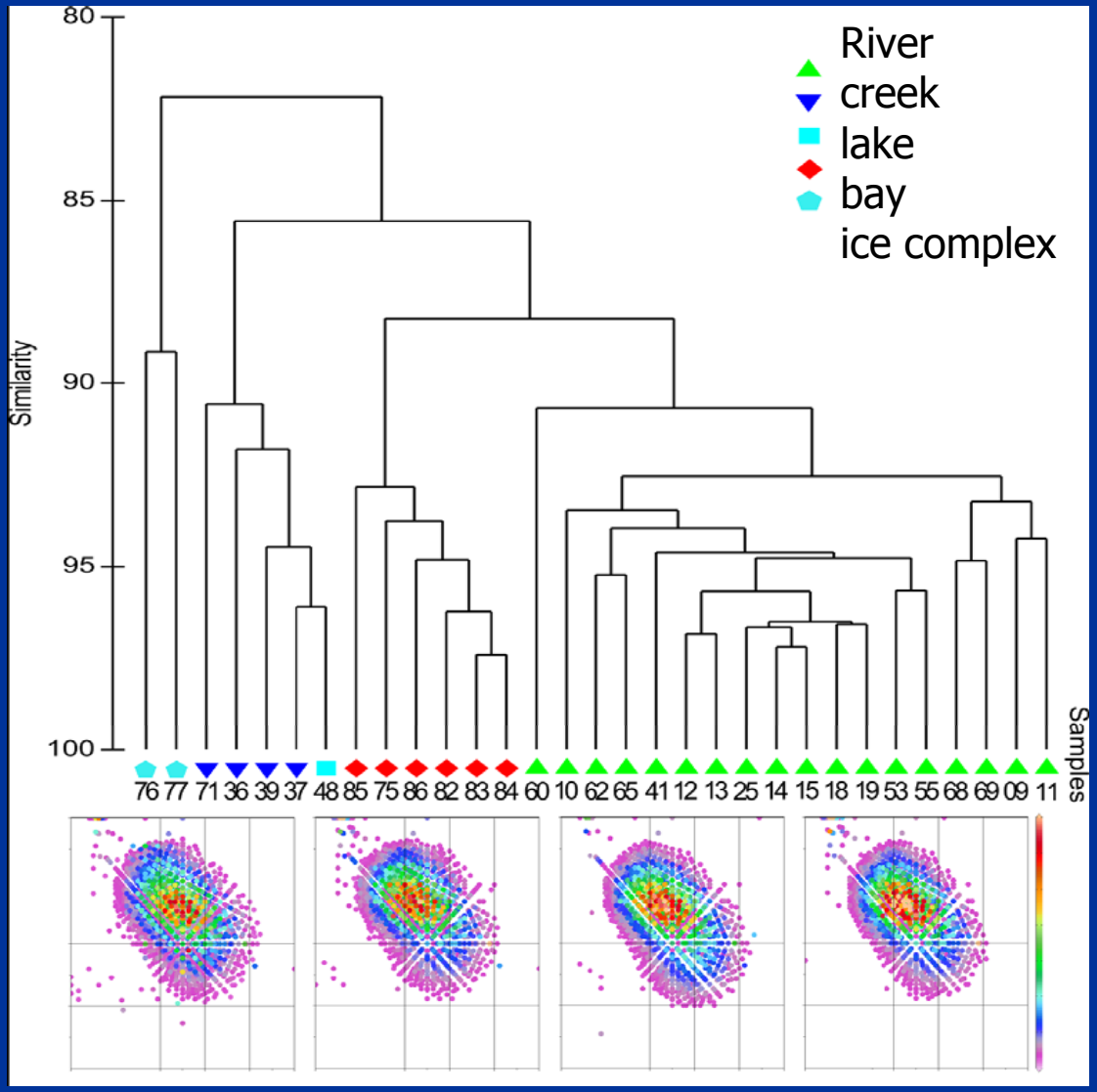
РОВ залива

Fig. 2 Characteristic ESI negative FT-ICR mass spectrum for Lena Delta DOM with **a** mass range of 200–600 m/z, **b** mass spectrum at nominal mass 398 for a typical Lena River sample, **c** melt water creek sample, and **d** Buor-Khaya Bay sample. *Asterisks* indicate the presence of one ^{13}C isotope in the assigned formula

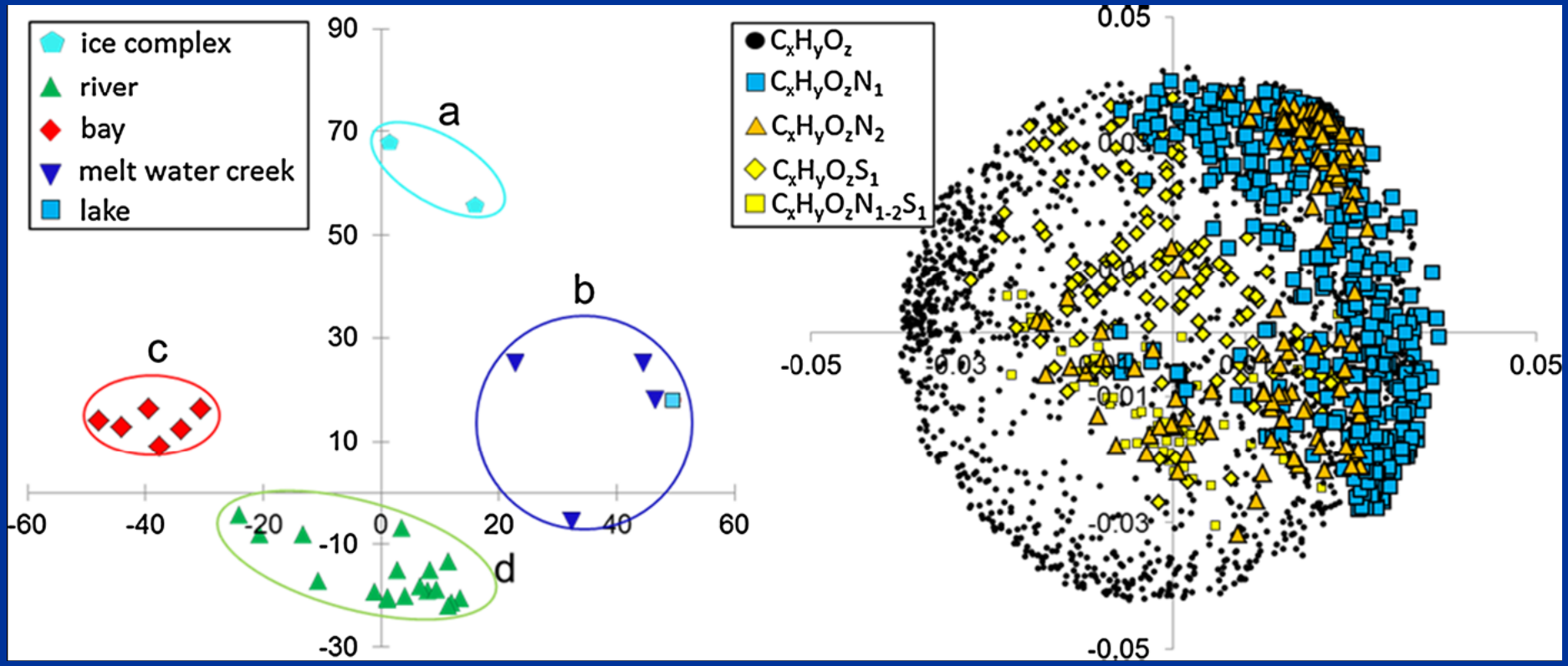
Dubinenkov et al. Biogeochemistry (2015) 123:1–14

КЛАСТЕРНЫЙ АНАЛИЗ ПО ИСТОЧНИКУ ПРОИСХОЖДЕНИЯ

Hierarchical cluster analysis (Bray–Curtis similarity) and van Krevelen diagrams for "average samples". Molecular differences were observed between a ice complex melt water creeks (pentagons, n = 2), creek (triangles, n = 4) plus lake samples (square, n = 1), Buor-Khaya Bay samples (diamonds, n = 6) and Lena River samples (green triangles, n = 18).



АНАЛИЗ ГЛАВНЫХ КОМПОНЕНТ (principal copmonent analysis)



Principal component analysis. PC1 and PC2 explained 56.7 % of the variance in the biplot (left panel): a) ice complex creek DOM samples (pentagons), b) permafrost creeks and lake DOM samples (upside down triangles and square), c Buor-Khaya Bay samples (diamonds), d) Lena River DOM samples

МЕТАДАННЫЕ

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms4804

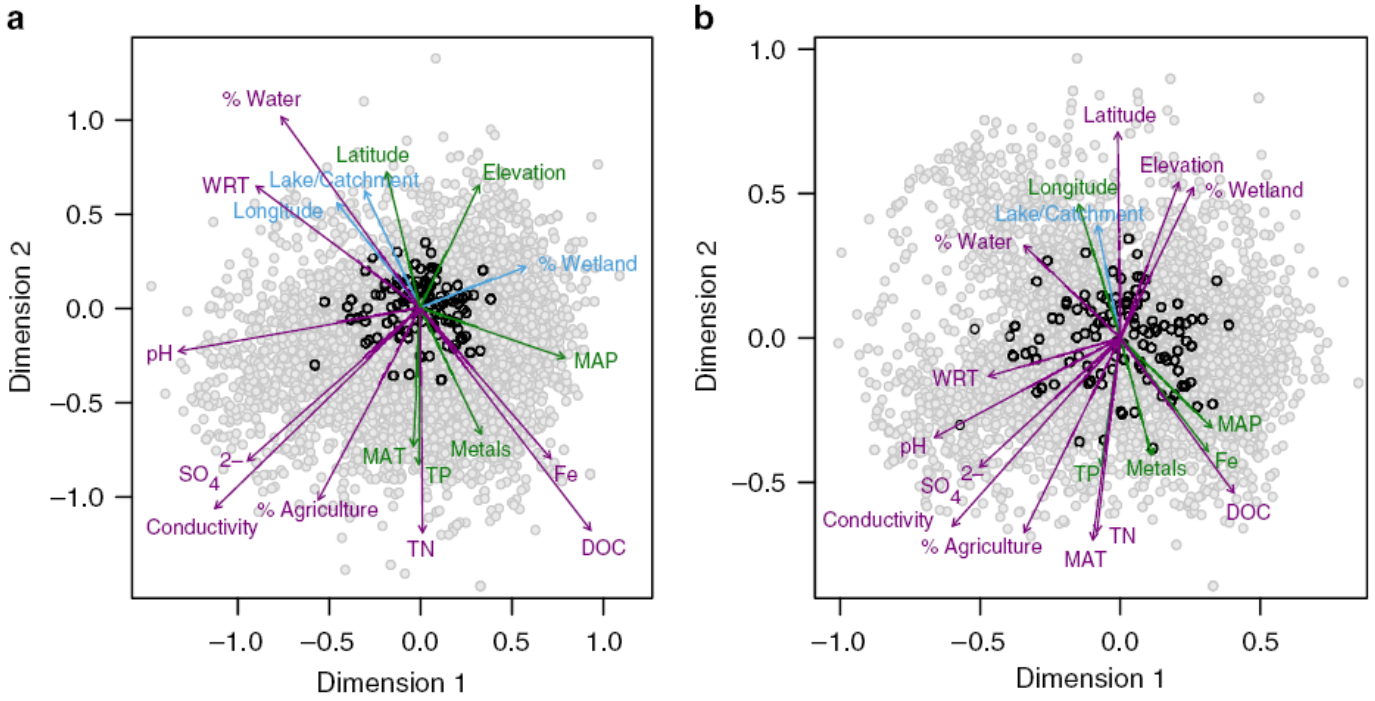


Figure 2 | Multivariate analysis of molecular data and drivers using non-metric multidimensional scaling. Ordinations are based on either Bray–Curtis (**a**, stress = 0.0966), which utilizes relative compound abundances, or Jaccard distances (**b**, stress = 0.1143), which utilizes presence/absence information. Landscape, climate and in-lake chemistry variables were fit to the ordination. Grey-shaded circles are DOM compounds, whereas black circles indicate the site. Variables with a significance level of <0.05 (light blue), <0.01 (green) and <0.001 (purple) are shown. DOC, dissolved organic carbon; MAP, mean annual precipitation; MAT, mean annual temperature; TN, total nitrogen; TP, total phosphorus; WRT, water residence time.

Метаданные – это независимые переменные: температура, содержание биогенных элементов, кислород и т.д. В данном случае речь идет о 240 озерах в Швеции

ПОСТРОЕНИЕ КОРРЕЛЯЦИОННЫХ ДИАГРАММ С МЕТАДАННЫМИ

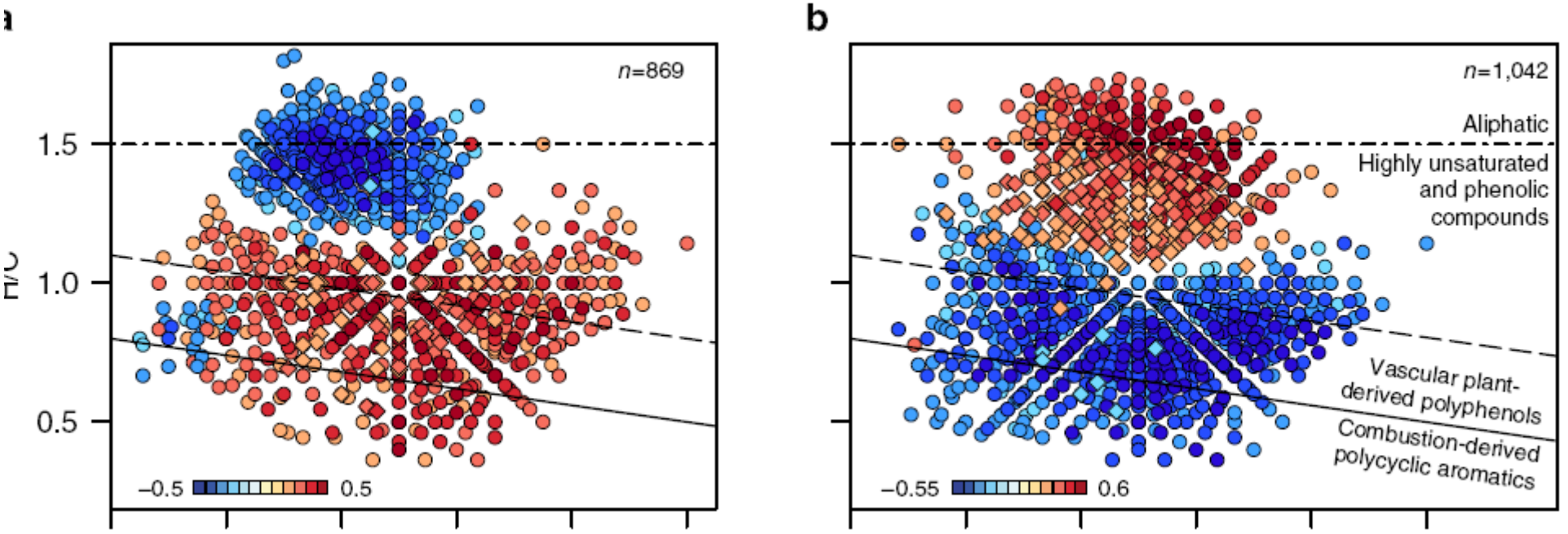
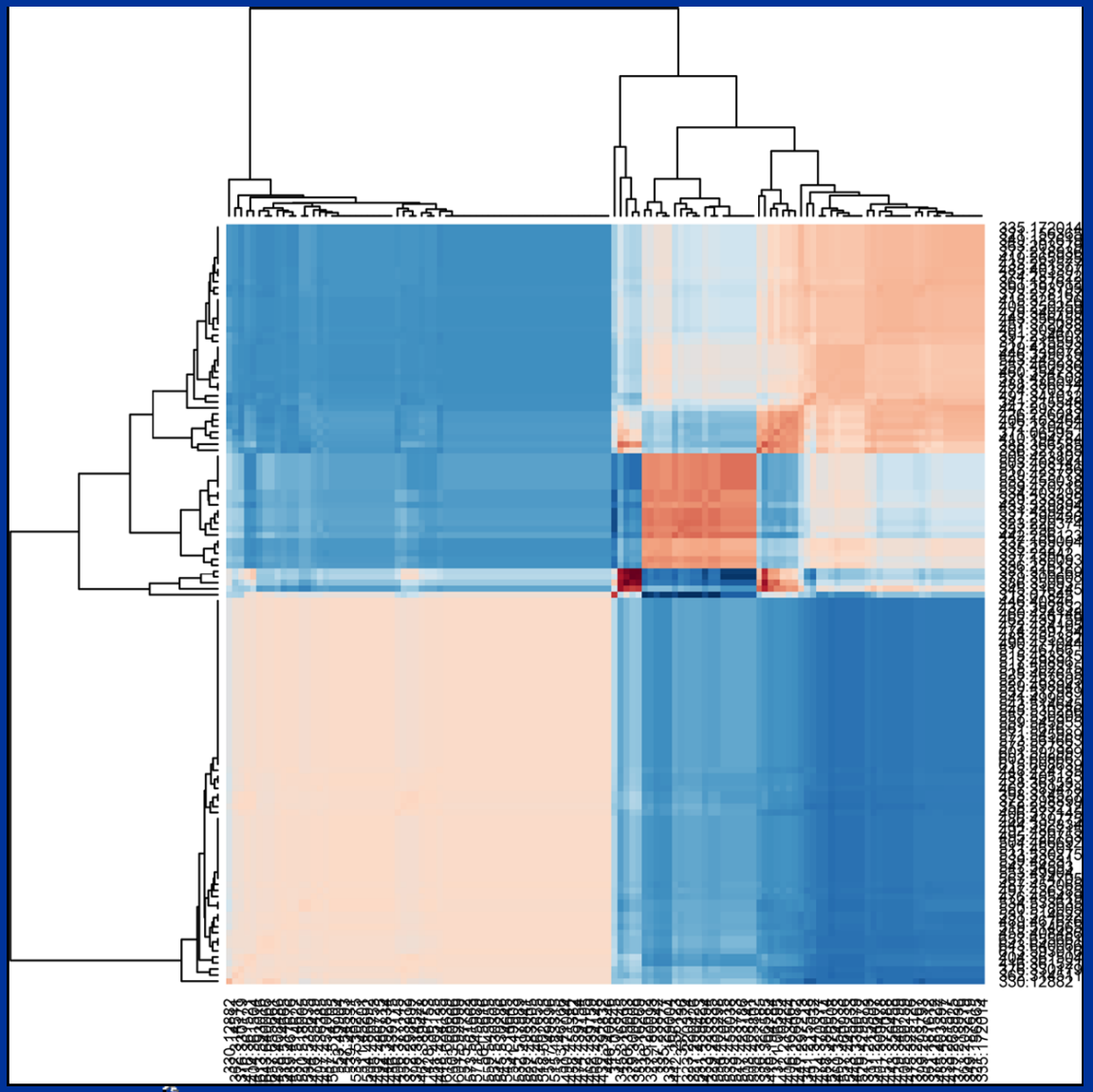
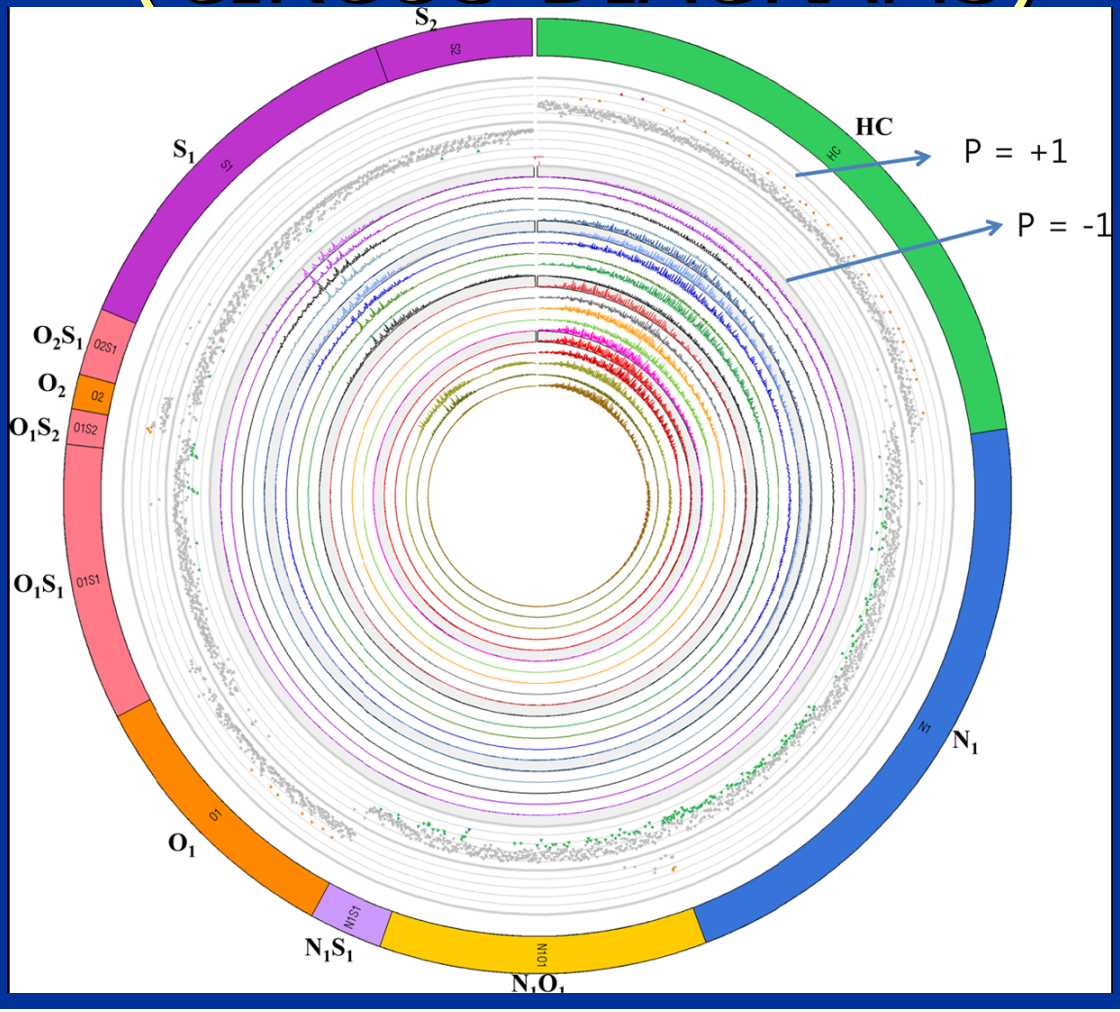


Figure 3 Molecular-level DOM patterns across 120 Swedish boreal lakes. Significant Spearman rank correlation coefficients (P-value <0.02674) molecules with (a) mean annual precipitation, (b) water residence time, The colour scale indicates Spearman correlations

КЛАСТЕРНО-КОРРЕЛЯЦИОННЫЙ АНАЛИЗ (HEATMAPs)

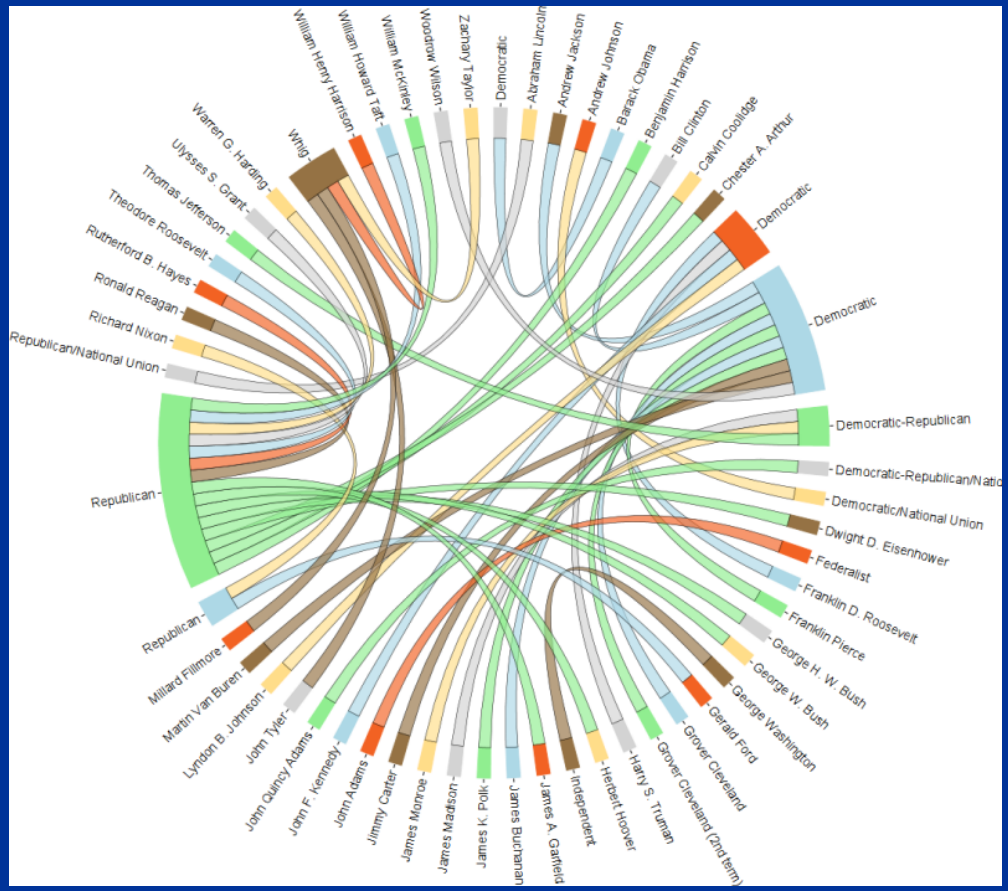


КРУГОВЫЕ СРАВНИТЕЛЬНЫЕ ДИАГРАММЫ (CIRCOS-DIAGRAMS)



Circos diagram showing correlational relationship between high resolution mass spectral peak information and physical property of crude oil.

КРУГОВЫЕ СРАВНИТЕЛЬНЫЕ ДИАГРАММЫ (CIRCOS-DIAGRAMS)



Рекомендуемая литература

ОБЪЯСНЕНИЕ РАСЧЕТА ДЕФЕКТА КЕНДРИКА

- **Hughey et al.** Kendrick Mass Defect Spectrum: A Compact Visual Analysis for Ultrahigh-Resolution Broadband Mass Spectra. *Anal. Chem. 2001, 73, 4676-4681*

САМАЯ ПЕРВАЯ ФУНДАМЕНТАЛЬНАЯ РАБОТА ПО МС ИЦРПФ ГУМИНОВЫХ В-В

- **Stenson et al.** Exact masses and chemical formulas of individual Suwannee River fulvic acids from ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectra. *Anal.Chem. 2003, 75, 1275-1284*

ПРИМЕНЕНИЕ ДИАГРАММ ВАН КРЕВЕЛЕНА ДЛЯ ДАННЫХ МС ЦИР ПФ ГВ

- **Kim et al.** Graphical Method for Analysis of Ultrahigh-Resolution Broadband Mass Spectra of Natural Organic Matter, the Van Krevelen Diagram. *Anal. Chem. 2003, 75 (20), 5336–5344*

ПРОДВИНУТЫЕ СТАТИСТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ ДАННЫХ МС ИЦРПФ

- **Hur et al.** Combination of Statistical Methods and Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for More Comprehensive, Molecular-Level Interpretations of Petroleum Samples. *Anal. Chem. 2010, 82, 211–218*
- **Cho et al.** Developments in FT-ICR MS instrumentation, ionization techniques, and data interpretation methods for petroleomics. *Mass Spectrometry Reviews 2015, 34, 248–263*
- **Kellerman et al.,** Chemodiversity of dissolved organic matter in lakes driven by climate and hydrology. *Nature communications*, 2014, DOI: 10.1038/ncomms4804

СПАСИБО ВСЕМ!