BUILDING AN IN-HOUSE MATLAB-BASED SOFTWARE FOR GRAPHICAL ANALYSIS OF FTICR MS DATA OF RIVER WATER

Nguyen Tien Tai[1, *], Nguyen Thi Hong Nhung[2], Nguyen Dac Thanh[3], Akio Miyazato[4], Ngo Van Quang[1], Nguyen Quang Tam[1]

[1]*Institute of Chemistry – Vietnam Academy of Science and Technology (VAST)*

[2]*Water Environment & Ocean Dept. – Hanoi University of Science and Technology*

[3]*Computer Science Dept. – University of Engineering and Technology Hanoi*

[4]*Nano Materials and Technology Center – Japan Advanced Institute of Science and Technology*

[*]Email: *nttai@ich.vast ac.vn*

Received: 3 April 2014 ; Accepted for publication: 12 July 2014

ABSTRACT

Fourier transform ion cyclotron resonance mass spectrometry (FTICR MS) is becoming an efficiently tool for characterization of dissolved organic matter (DOM) at the molecular level. FTICR MS data of DOM samples often involve thousands m/z peaks, therefore, visual presentation for extracting structural information from the complicated data are challenge. MATLAB (Matrix Laboratory) is a high-level language and interactive environment for visualizing and analyzing complete data. An in-house Matlab-based software for graphical presentation of the Kendrick and the van Krevelen diagrams, which are popular tools for analysis of FTICR MS data of mixture samples, is built and applied for graphical analysis of two DOM water samples from Suwannee river (USA) and Cuakhau lake (Northern Vietnam) as demonstration examples. With this software, complicated FTICR MS data of DOM samples can be visualized in a way that allows identifying DOM type and possible reaction pathways. This in-house user friendly software can be also easy to upgrade with more image diagrams as well as histograms.

Keywords: FTICR MS, DOM, the Kendrick plots, the van Krevelen plots.

1. INTRODUCTION

Dissolved organic matters (DOMs) are products of several biogeochemical processes in water environment. DOMs are important research objects as they play important roles in the movement of nutrients in ecosystems and are one of the largest carbon reservoirs.

DOM is very complex, consisted of thousands of chemically distinct compounds of polyfunctional, heterogeneous, polyelectrolytic complex mixture with varying molecular weight and concentration; therefore characterization of DOM in molecular level has become a primary research objective in environmental and ecological chemistry of water.

For analyzing the molecular composition of numerous highly complex DOM in water, petroleum, aerosols, metabolome, Fourier transform ion cyclotron resonance mass spectrometry (FTICR MS) with ultrahigh resolution of hundreds thousand, mass accuracy of ppm and higher, allowed the simultaneous detection of thousands of chemical components per single sample without chromatographic separations, is superiority [1].

As a single FTICR mass spectrum contains thousands of individual peaks, each of them representing a unique molecular mass, signal magnitude, and a specific molecular formula, visualization and graphical analysis this complex multidimensional spectrum remain major obstacles and a challenge. Some methods of visualizing and for analyzing complex FTICR MS data were developed, included most extensively used as the Kendrick, the van Krevelen diagrams.

The observed mass is change into Kendrick mass (KM) or normalized KD in a CH_2 mass-normalized scale by a simple converting the measured IUPAC mass (based on 12.00000 Da) to a new mass, based on CH_2 = 14.00000 Da [2].

$$KM = \text{Observed m/z} \times (14/14.01565) \quad (1)$$

Kendrick mass defect (KMD) is then calculated as the difference between the KM and the nominal observed mass.

$$KMD = (\text{Nominal observed m/z} - KM) \times 1.000 \quad (2)$$

The van Krevelen diagram [3] is originally used in the geochemistry literature to study the evolution of coals or oil samples and latterly extensively used in the environment and biology field [4]. The van Krevelen diagram is constructed using the molar ratio of hydrogen to carbon (H/C ratio) as the ordinate and the molar oxygen-to-carbon ratio (O/C ratio) as the abscissa. Major biogeochemical classes of compounds (such as lignin compounds, lipids, carbohydrates, etc.) have their own characteristic H/C or O/C ratios. As a result, each class of compounds plots in a specific location on the diagram.

There is only some software for visualization FTICR MS data of mixture samples. They are commercialized [5] or academic [6] software and not easy to access. Some research groups favor to develop in-house software, appropriated to their FTICR MS equipment and DOM subjects.

In this study, we demonstrate the graphical analysis of FTICR MS data of DOM water samples using the Matlab-based software, built as a stand-alone package. Two DOM samples taken from Suwannee river (USA) and Cuakhau lake (Dong Cao, northern VietNam), denoted as SR and CKL, respectively, were used as testing data. DOM analysis of SR was analyzed by FTICR MS and published elsewhere [7], while nutrition export but not DOM image of CKL water was also characterized recently [8].

2. EXPERIMENTAL SECTION

2.1. FTICR data of DOM water samples

Description of two water samples and experimental condition are shown in Table 1.

Table 1. Information of two FTICR MS data for TTN 01.13 testing.

Description	SR DOM sample	CKL DOM Sample
Place	Suvanee river (USA)	Cuakhau lake (Vietnam)
Extraction	Lyophilized	SPE C18 Disks
FTICR MS equipment	SolariX-JA/9.4T (Bruker Daltonic) at JAIST	910-MS TQFTMS/7T (Agilent/Varian) at VAST
Composition Software	SmartFormula ™	Omega
Number of assigned compounds	4.632	2.176

Despite both FTICR MS data and two DOM samples are experimental, not simulated ones, however attention should be paid that these data are prepared and measured by quite different methods, equipments and condition. So that graphical analysis of these samples below is used only as illustration for TTN 01.13 software application but not as result for environment analysis.

2.2. Matlab-based programming

Matlab is an interactive environment for programming. The Kendrick, the van Krevelen and other statistic algorithms are programmed in Matlab GUI (Graphical User Interface), version 7.14, then the Matlab coded project is compiled by Deployment Tool/Matlab® CompilerTM, created a standalone execute software, named TTN with size of 234 Mb. Input data is FTICR data file in txt or xls formats, while output data is png, bmp, jpg or dbf formats.

As training version, only the Kendrick, the van Krevelen basic rules are programmed in TTN 01.13, but the software is quite opening for upgrading to other modified or expanded graphical tools in next versions.

3. RESULTS AND DISCUSSION

3.1. The in-house software for graphical analysis of FTICR MS data

The TTN 01.13 allows to graphical visualize of FTICR MS data into two basic types: image plots and histograms. As basic image plots for analyzing FTICR MS data of DOM river samples, both the Kendrick, van Krevelen can be plotted by the TTN 01.13. All plots are displaced within a single sample ("Single data") as well as across multiple samples ("Multiple data"), in 2D, 3D or psedo-3D modes. The software also supports for plots of DBE versus chemical classes.

Beside basic plots, the TTN 01.13 also allows to present some histograms, helpful for statistical analysis of FTICR MS data. Typical histograms, designated in TTN 01.13, are distribution of chemical lass (O_x, NO_x), DBE and Carbon number in DOM samples.

English interface of TTN 01.13, which is academic user friendly, is shown in Figure 1.

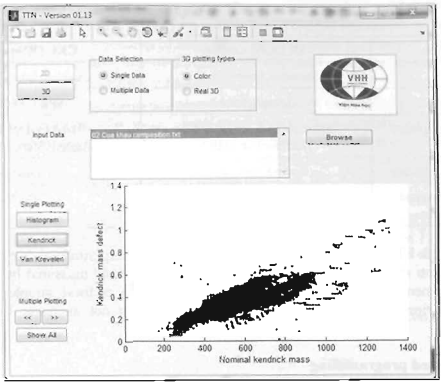

Figure 1. Interface of the TTN 01.13 software.

3.2. The Kendrick plot

Kendrick plot sorts compounds into homologous series according to compound class, compound type and degree of alkylation. In the Kendrick diagram, members of the same homologous series will have KM differing by exactly 14 Da and will have the same KMD value. Additional, the compounds differing by the corresponding masses of H2 or O can be identified also by parallel lines with a specific slope of 6.7 and 1.4, respectively.

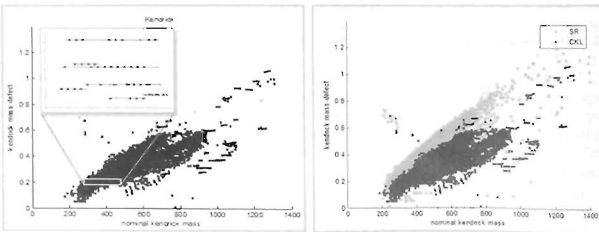

Figure 2. The Kendrick plot of single CKL sample (left) and multi SR+CKL samples (right). Some homologous series are identified as parallel line in zoom plot of CKL sample. Meanwhile, difference between CKL and SR DOM is apparently in overseeing plot, where SR DOM is distributed at higher DBE.

3.3. The van Krevelen plot

The van Krevelen allows elucidating what compound classes are present as well as identifying what reaction pathways are taking place.

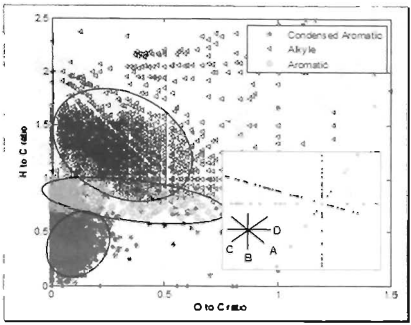

Figure 3. The van Krevelen plot of SR sample Classification of alkyle, aromatic and condensed aromatic compounds as well as trend lines of methylation, hydrogenation, hydration and oxidation are visualized.

In the van Krevelen plot, the dots, represented the molecular formula of detected molecules, can be distributed by clusters, related to families of similar compounds. Determination of clusters can be based on H/C and O/C ranges or can be applied the aromaticity index (AI) limitation [9].

It is apparent from Fig. 3, shown DOM CKL, that numerous trend lines in DOM can be clearly discerned. These lines represent chemical reactions methylation, demethylation, or alkyl chain elongation (A); hydrogenation or dehydrogenation (B); hydration or condensation (C); and oxidation or reduction (D).

3.4. The 3D-plots

The Kendrick and, in particular, the van Krevelen diagrams can be expanded to a 3D plot by using interested ratio, such as N/C, S/C or peak intensities as third axis (Oz). In most 3D plot, relative significance of compound classes of individual sample or compositional differentiation among samples is displaced, provided an indication of which compound class is in highest abundance.

In the so call pseudo-3D plot, colors of points were varied in certain order according to relative peak magnitudes. This color mode is more intuitive for presentation of results in power point or color printed reports.

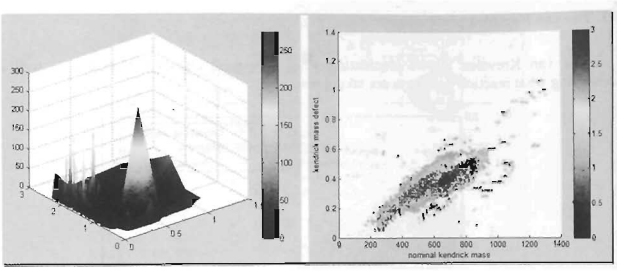

Figure 4. The 3D- van Krevelen plot of SR sample (left), and the pseudo 3D Kendrick plot of CKL sample (right). They are 3D and pseudo 3D plot of the same data, presented in Fig.3 and Fig. 2 (left), respectively, color codded by m/z peak intensities.

4. CONCLUSION

The in-house software for graphical analysis of FTICR MS data was built as standalone Matlab application. In the first version, the software can present basic the Kendrick and the van Krevelen diagrams. With illustration of analysis of FTICR MS data of two DOM samples, we demonstrate how this under-completed in-house user-friendly software can be used to extract structural information from complex FT-ICR-MS data of DOM sample.

In short term, this in-house software will be completed and integrated to FTICR-MS system at ICH-VAST for analysis DOM data from Cuakhau lake and Red river water samples in frame of two ICH projects on environment water. For long term, the software will be upgraded further with more graphical function for DOM research and education in environment chemistry and metabolomics.

Acknowledgment. This work was supported by the VAST project VAST07 on "Evaluation of polution water of Red river by advanced analytical methods". The manuscript was developed after lengthy discussions with other member of VAST07 project. We thank Dr. Emma Rochelle-Newall at Institut de recherche pour le développement (IRD) and LOTUS project for supply the SR and CKL DOM samples.

REFERENCES

1. Antonio Nebbioso and Alessandro Piccolo - Molecular characterization of dissolved organic matter (DOM): a critical review, Anal. Bioanal. Chem. **405** (2013) 109–124.
2. Kendrick E. - A mass scale based on CH_2=14.0000 for high resolution mass spectrometry of organic compounds, Anal. Chem. **35** (1963) 2146-2154.
3. Van Krevelen D W. - Graphical-statistical method for the study of structure and reaction processes of coal, Fuel. **29** (1950) 269-284.

4. Kim S., Kramer R. W., Hatcher P. G. - Graphical Method for Analysis of Ultrahigh-Resolution Broadband Mass Spectra of Natural Organic Matter, the Van Krevelen Diagram, Anal. Chem. **75/20** (2003) 5336-5344.
5. Sierra Analytics, Inc.: Composer Software User Manual, version 1.05 – June 2012; http://www.massspec.com/Composer.html;
6. Erast V. Kunenkov, Alexey S. Kononikhin, Irina V. Perminova, Norbert Hertkorn, Andras Gaspar, Philippe Schmitt-Kopplin, Igor A. Popov, Andrew V. Garmash, and Evgeniy N. Nikolaev - Total Mass Difference Statistics Algorithm: A New Approach to Identification of High-Mass Building Blocks in Electrospray Ionization Fourier Transform Ion Cyclotron Mass Spectrometry Data of Natural Organic Matter, Anal. Chem. **81** (2009) 10106–10115.
7. Alexandra C. Stenson, Alan G. Marshall and William T. Cooper - 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. **75** (2003) 1275-1284.
8. Luu Thi Nguyet Minh, Trinh Anh Duc, Nguyen Thi Huong Giang, Ngo Thi Le Trang, Jean-Louis Janeau, Didier Orange - Nutrient export to downstream aquatic systems in small upland, farmed catchment. Reported at Scientific Youth Confference, ICH-VAST, Hanoi, January 2013.
9. Koch, B. P., and T. Dittmar - From mass to structure: An aromaticity index for high-resolution mass data of natural organic matter, Rapid Commun. Mass Spectrom. **20** (2006) 926–932.

TÓM TẮT

XÂY DỰNG PHẦN MỀM TRÊN NỀN MATLAB ĐỂ PHÂN TÍCH ĐỒ HỌA DỮ LIỆU PHỔ KHỐI ION CYCLOTRON BIẾN ĐỔI FOURIER CỦA NƯỚC SÔNG HỒ

Nguyễn Tiến Tài[1,*], Nguyễn Thị Hồng Nhung[2], Nguyễn Đắc Thành[3], Akio Miyazato[4], Ngô Văn Quang[1], Nguyễn Quang Tâm[1]

[1]*Viện Hóa học – Viện Hàn lâm KHCNVN*

[2]*Khoa môi trường nước và đại dương – Đại học Khoa học và Công nghệ Hà Nội*

[3]*Khoa Khoa học máy tính – Đại học Công nghệ Hà Nội*

[4]*Trung tâm Công nghệ và vật liệu nano – Học viện Khoa học công nghệ tiên tiến Nhật Bản*

[*]Email: *nttai@ich.vast.ac.vn*

Phổ khối lượng biến đổi Fourier ion cyclotron (FTICR MS) đang trở thành công cụ hữu hiệu để đặc trưng các chất hữu cơ tan (DOM) ở cấp độ phân tử. Dữ liệu FTICR MS của các mẫu thường chứa hàng nghìn tín hiệu m/z, do vậy việc biểu diễn đồ họa để khai thác thông tin cấu trúc của các hệ DOM phức tạp thật sự là một thách thức. MATLAB (Matrix Laboratory) là một ngôn ngữ lập trình bậc cao với môi trường có tính tương tác tốt để phân tích các dữ liệu phức tạp. Phần mềm tự tạo trên nền MatLab để biểu diễn đồ họa theo 2 mô hình thường được dùng

phân tích dữ liệu FTICR MS mẫu hỗn hợp là Kendrick và van Krevelen đã được xây dựng và phân tích biểu diễn trên 02 mẫu DOM của sông Suwannee (Mỹ) và hồ Cửa Khâu (phía Bắc Việt Nam). Với phần mềm này, dữ liệu FTICR MS phức tạp của mẫu DOM có thể được đồ họa hóa cho phép phân loại và nhận biết xu hướng chuyển hóa. Phần mềm tự tạo thân thiện với người dùng này cũng có thể dễ dàng nâng cấp, bổ sung thêm các chức năng đồ họa hữu dụng khác.

Từ khóa: FTICR MS, DOM, họa đồ Kendrick, họa đồ van Krevelen.