Analytica Chimica Acta 924 (2016) 68–76

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Collision cross section prediction of deprotonated phenolics in a travelling-wave ion mobility spectrometer using molecular descriptors and chemometrics

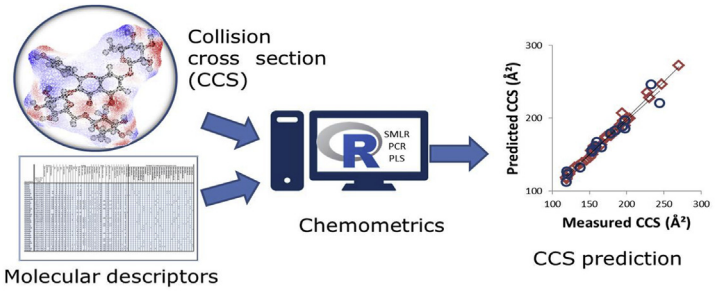
Gerard Bryan Gonzales [a, b, d], Guy Smagghe [b], Sofie Coelus [a], Dieter Adriaenssens [a], Karel De Winter [c], Tom Desmet [c], Katleen Raes [d], John Van Camp [a, *]

[a] Food Chemistry and Human Nutrition (NutriFOODChem), Department of Food Safety and Food Quality, Faculty of Bioscience Engineering, Ghent University, Belgium
[b] Laboratory of Agrozoology, Department of Crop Protection, Faculty of Bioscience Engineering, Ghent University, Belgium
[c] Center for Industrial Biotechnology and Biocatalysis, Faculty of Bioscience Engineering, Ghent University, Belgium
[d] Department of Applied Biological Science, Faculty of Bioscience Engineering, Ghent University, Belgium

HIGHLIGHTS

- CCS for deprotonated phenolics were measured using TWIMS.
- Isomeric phenolics were separated in the IMS based on their CCS.
- SMLR, PLS and PCR models were developed to predict CCS values.
- The generated models yielded high predictive ability and efficiency.
- The generated models could be easily integrated into metabolite ID platforms.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 26 January 2016
Received in revised form
29 March 2016
Accepted 15 April 2016
Available online 25 April 2016

Keywords:
Ion mobility
Mass spectrometry
Phenolics
Collision cross section prediction
Chemometrics
Flavonoids

ABSTRACT

The combination of ion mobility and mass spectrometry (MS) affords significant improvements over conventional MS/MS, especially in the characterization of isomeric metabolites due to the differences in their collision cross sections (CCS). Experimentally obtained CCS values are typically matched with theoretical CCS values from Trajectory Method (TM) and/or Projection Approximation (PA) calculations. In this paper, predictive models for CCS of deprotonated phenolics were developed using molecular descriptors and chemometric tools, stepwise multiple linear regression (SMLR), principal components regression (PCR), and partial least squares regression (PLS). A total of 102 molecular descriptors were generated and reduced to 28 after employing a feature selection tool, composed of mass, topological descriptors, Jurs descriptors and shadow indices. Therefore, the generated models considered the effects of mass, 3D conformation and partial charge distribution on CCS, which are the main parameters for either TM or PA (only 3D conformation) calculations. All three techniques yielded highly predictive models for both the training ($R^2_{SMLR} = 0.9911$; $R^2_{PCR} = 0.9917$; $R^2_{PLS} = 0.9918$) and validation datasets ($R^2_{SMLR} = 0.9489$; $R^2_{PCR} = 0.9761$; $R^2_{PLS} = 0.9760$). Also, the high cross validated R^2 values indicate that the generated models are robust and highly predictive ($Q^2_{SMLR} = 0.9859$; $Q^2_{PCR} = 0.9748$; $Q^2_{PLS} = 0.9760$). The predictions were also very comparable to the results from TM calculations using modified mobcal (N2). Most importantly, this method offered a rapid (<10 min) alternative to TM calculations without

* Corresponding author.
E-mail addresses: gerard.gonzales@ugent.be (G.B. Gonzales), john.vancamp@ugent.be (J. Van Camp).

http://dx.doi.org/10.1016/j.aca.2016.04.020

compromising predictive ability. These methods could therefore be used in routine analysis and could be easily integrated to metabolite identification platforms.

1. Introduction

The increased use of high resolution mass spectrometry (HRMS) in bioanalysis has prompted the need for better metabolite identification tools. For instance, the success of metabolomics, an increasingly popular and useful analytical technique to understand biological systems, relies on the confident identification of metabolites in a high-throughput scale [1]. However, although HRMS can provide elemental composition and many structural clues, it is generally not definitive, and rarely distinguishes isomeric compounds that produce identical MS/MS spectra, such as substituted metabolites [2]. Recently, the integration of ion mobility spectrometry (IMS) to mass spectrometry offered an increased capability to characterize biological mixtures both in terms of metabolite identification [1,3] and increase in peak capacity [4]. IMS has also been previously shown to resolve metabolites that co-elute during liquid chromatographic separation, thus offering a rapid characterization tool for complex matrices [5].

IMS has the ability to rapidly separate (microseconds to milliseconds) ions based on their mobilities in a gas-filled chamber under the influence of a weak electric field [6]. This separation is based on their size and shape (collision cross section, CCS, Ω), as well as their ionic interaction with the buffer gas, typically nitrogen or helium, and their charge state (z) [3,6]. This technique therefore provides crucial structural and ionic information. which is essential especially in characterizing isomeric compounds [3]. For this purpose, IMS has been used in various bioanalytical research domains such as lipidomics [7,8], proteomics [9], metabolomics [1,10], analysis of large protein complexes [11] and metabolic transformations of drugs [2,12]. However, the main challenge lies on the interpretation of the data and deciphering structural information from the obtained CCS values.

In order to infer structural and conformational information from the obtained CCS values, the gold standard for IMS data interpretation is the satisfactory correlation between the measured and the theoretical CCS values calculated using computational chemistry techniques that model the interaction of the ion with the buffer gas [3,13]. The most widely used CCS prediction methods include trajectory method (TM) [14,15], the exact hard sphere scattering (EHSS) [16] and the projection approximation (PA) [14,15] calculated using the mobcal software developed by the Jarrold Group of the University of Indiana. A comprehensive discussion of these techniques has previously been reported elsewhere [13]. TM is generally accepted as the most reliable prediction method [17]. However, the original mobcal was parametrized to predict CCS values obtained using helium as buffer gas. A modification was later introduced to parametrize CCS calculations using nitrogen as buffer gas, which is more commonly used in commercial instrumentations [3,18].

Although a widely used and reliable method, TM calculations are highly computationally expensive [17] and hence not applicable to larger molecules, bigger sets of molecules and routine laboratory analysis. This also prevents the easy integration of CCS calculation in metabolite identification platforms especially in the absence of high-powered computational facilities. To circumvent this problem, researchers turned to chemometrics to develop accurate yet computationally efficient methods to predict CCS values. Compared to atomistic models derived from mobcal, chemometric analysis requires a database of experimentally-derived CCS values from where a predictive model is developed using molecular descriptors (such as in quantitative-structure property relationship studies). This approach has been successfully employed for the prediction of the CCS of peptides in positive mode ionization commonly using amino acid parameters and sequence information [19]. The success of using these descriptors owes to the fact that peptides are composed of a limited number of known and repeating units of amino acids.

However, small molecules such as non-peptide metabolites do not have a certain predictable sequence. Also, most of the chemometric predictive models in literature use positive ionization, for instance a chemometric tool for prediction of drift times in positive mode has been previously described [20], and relatively few CCS measurements have been done for negatively charged ions [21]. This is unfortunate especially because many small metabolites such as phenolics ionize better in negative ionization mode. Chemometrics-based prediction models for small molecules are thus lacking in the current literature. Therefore, identification of these metabolites remains an analytical challenge due to their wide chemical and structural diversity. Developing a CCS prediction method for negatively charged small molecules could therefore increase the capacity of IMS for confident metabolite identification.

Thus, in this study, we explored the use of chemometrics tools, stepwise multiple linear regression (SMLR), principal components analysis regression (PCR), and partial least squares regression (PLS) in predicting the CCS of 56 deprotonated phenolics using molecular descriptors. Initially, the CCS values of these compounds were measured in a travelling-wave ion mobility instrument, and the measured CCS values were correlated with their predicted CCS values. Also, the predictive performance of these models was compared to the conventional TM prediction using the modified mobcal (N_2).

2. Materials and methods

2.1. Chemical reagents

Poly DL-alanine (P9003) and most of the small molecule standards used in this paper were purchased from Sigma–Aldrich. Many of the glycosylated derivatives were synthesized at the Center for Industrial Biotechnology and Biocatalysis of the Ghent University (Ghent, Belgium) (see Table 1) using an enzymatic glycosylation method [22]. The poly DL-alanine (10 μM) and all the standards were prepared in acetonitrile/water/formic acid (50/50/0.1) and filtered through a 0.4 μm syringe filter prior to analysis.

2.2. Collision cross section (CCS) measurement in TWIMS

All analytical analyses were performed using a Waters Synapt HDMS instrument (Waters Corp., Milford. MA. USA). Mass calibration of the mass spectrometer was achieved using sodium formate cluster ions and was assessed by analyzing the mass of leucine-enkephalin (m/z = 554.2615). All working standards and the poly-alanine calibrant were infused directly to electrospray ionization at a flow rate of 5 μL/min using a 250 μL glass syringe. Data were