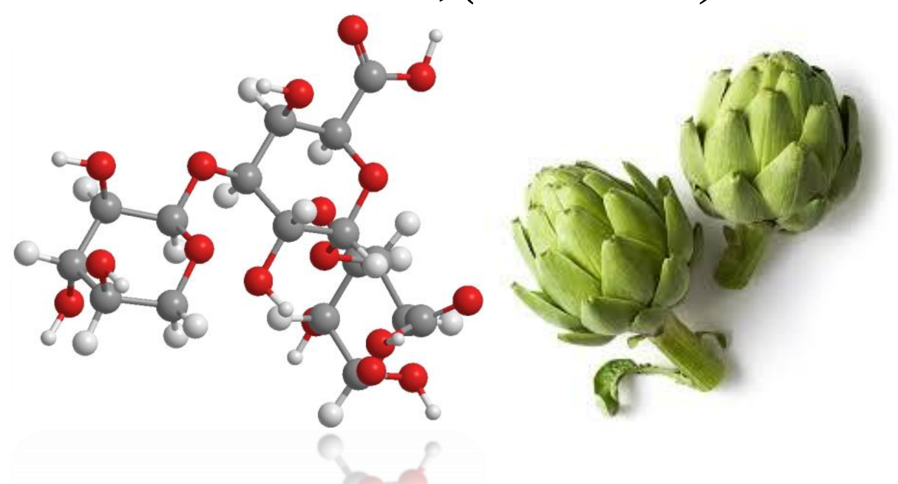
GC-MS characterisation of novel pectic-oligosaccharides derived from artichoke pectin using machine learning and competitive fragmentation modelling

Carlos Sabater* , Antonia Montilla, Nieves Corzo

Departamento Bioactividad y Análisis de Alimentos, Instituto de Investigación en Ciencias de la Alimentación CIAL, (CSIC-UAM)

*carlos.sabater@csic.es

Wednesday 9 October, 2019



CIAL
INSTITUTO DE INVESTIGACIÓN
EN CIENCIAS DE LA ALIMENTACIÓN

INTRODUCTION

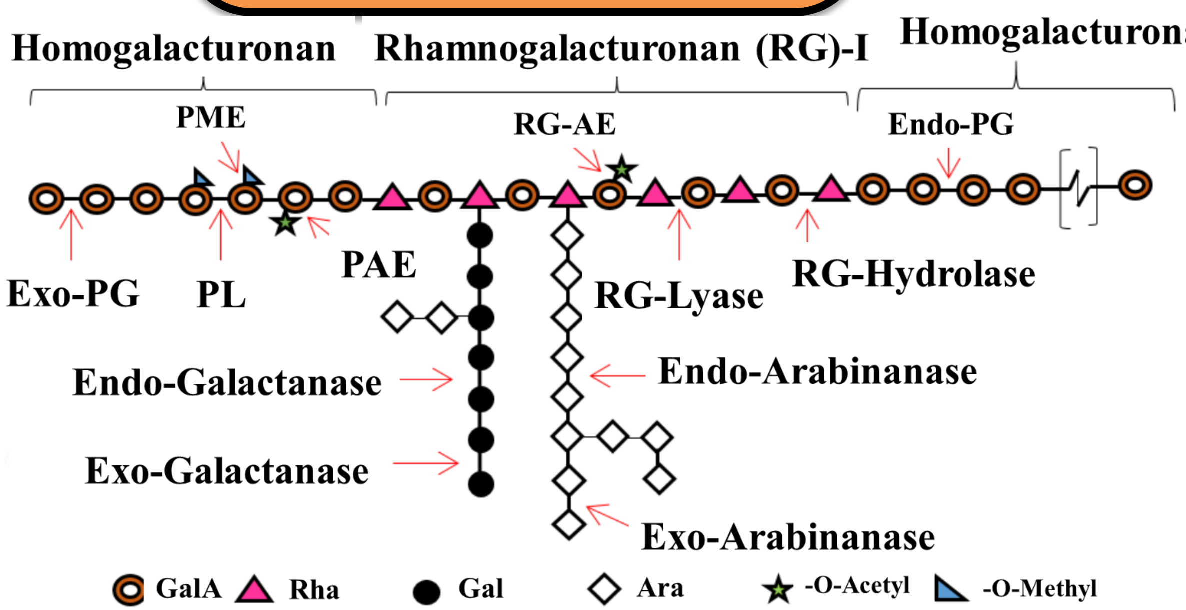
- **Pectin**: HG, RG-I, RG-II
- **Biological** and **technological** properties
- **Alternative sources**:
  different structures and bioactivity
- **Depolymerisation** of **pectin**:
  Pectic oligosaccharides (**POS**)

Structural characteristics:

- Monomeric composition
  - **Galacturonic acid (~ 70%)**
- Distribution of side chains
- DM and DA
- Molar mass
- Charge distribution

Methods
- Chemical
- Enzymatic
- Combined

Structures
- OGalA
- GalOS
- AraOS
- RhaGalAOS
- XylOGalA
- AraGalOS



GalA: Galacturonic acid, Rha: Rhamnose, Gal: Galactose, Ara: Arabinose
PME: Pectin methyl esterase, PAE: Pectin acetyl esterase, PL: Pectin lyase
PG: Polygalactuonase, RG-AE: Rhamnogalacturonan acetyl esterase

INSTITUTO DE INVESTIGACIÓN
EN CIENCIAS DE LA ALIMENTACIÓN

INTRODUCTION

- **POS characterisation** → HPSEC-ELSD, HAPEC-PAD, MALDI-TOF-MS, HILIC-MS, GC-MS
  - **MS**: intricate and high dimensional data (Leijdekkers et al., 2015)
  - **Data modelling**: extract relevant chemical information (*m/z* ions) through machine learning
  - **GC-MS**

✓ **Complex fragmentation** process
✓ **No reference** libraries for **novel oligosaccharides**
✓ How to **interpret individual fragments?**

***In silico* fragmentation (Allen et al., 2016)**

OBJECTIVE

To characterise **novel POS** obtained from hydrolysis of **artichoke pectin** using different types of enzymes. With this aim, a **GC-EI-MS** data analysis strategy based on the **combination** of machine **learning** and ***in silico* fragmentation** is presented.

MATERIALS AND METHODS

POS Obtainment

2% (w/v) **artichoke pecin**

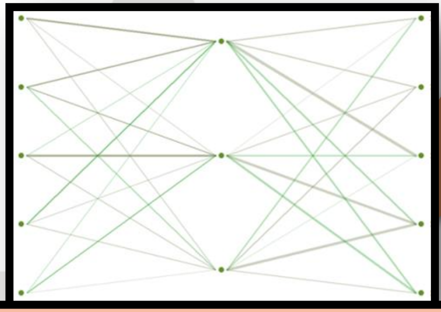
pH 5.0; 50 ºC
750 rpm
0-4 h

0.1-44.6 U/mL
Commercial enzymes

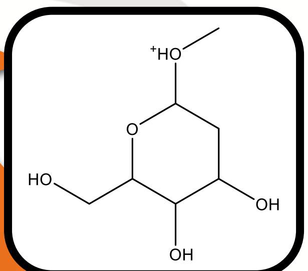
POS Mixtures

Pentopan®Mono-BG
Cellulase from *Aspergillus niger*
Pectinex®Ultra-Olio

GC-MS spectra



Machine learning models
(get characteristic *m/z* ions)



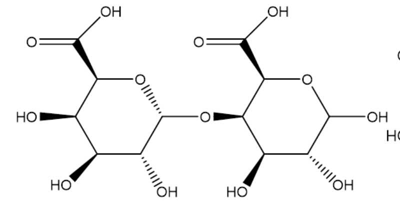
***In silico* fragmentation**
(calculate structures of characteristic *m/z* ions)

Determine structural differences among POS obtained with different enzymes

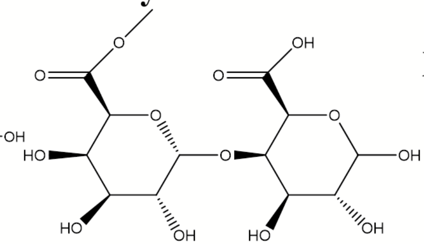
MATERIALS AND METHODS

Structures contained in *in silico* fragmentation library

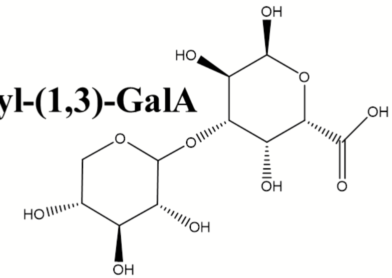
DiGalA



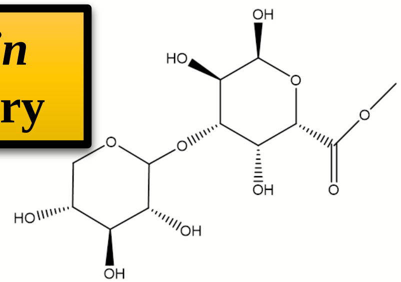
Methyl-esterified DiGalA



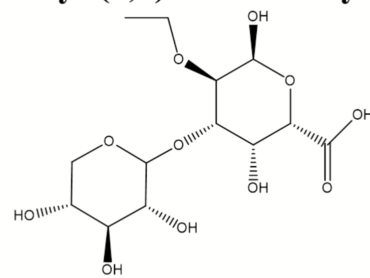
Xyl-(1,3)-GalA



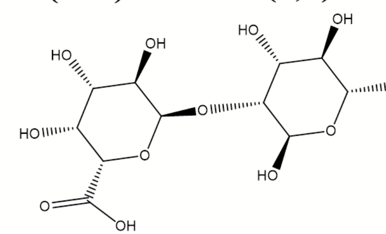
Xyl-(1,3)-GalA Methyl-esterified



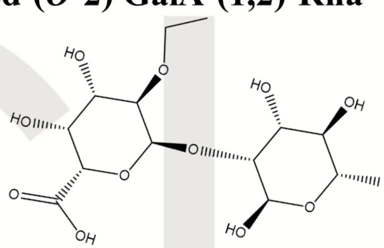
Xyl-(1,3)-GalA Acetylated (*O*-2)



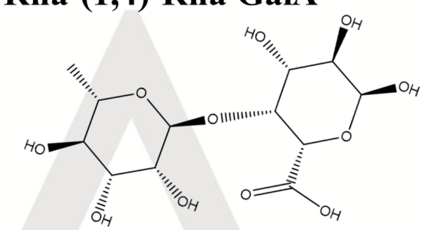
GalA-(1,2)-Rha



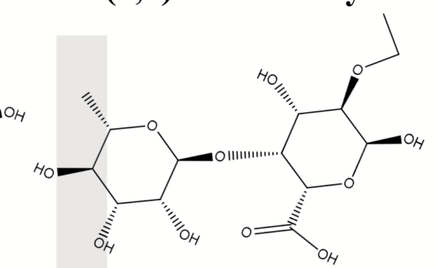
Acetylated (*O*-2) GalA-(1,2)-Rha



Rha-(1,4)-Rha GalA

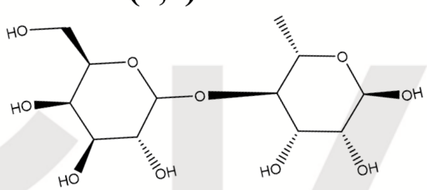


Rha-(1,4)-GalA Acetylated (*O*-2)

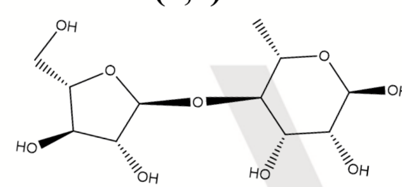


Structures previously described in the bibliography for other substraes different than artichoke: citrus, apple (Atmodjo et al., 2013; Mohnen, 2008)

Gal-(1,4)-Rha



Ara-(1,4)-Rha



MATERIALS AND METHODS

Pectin monomers and **most abundant POS structures** found in the bibliography

In silico **fragmentation (CFM-ID)**

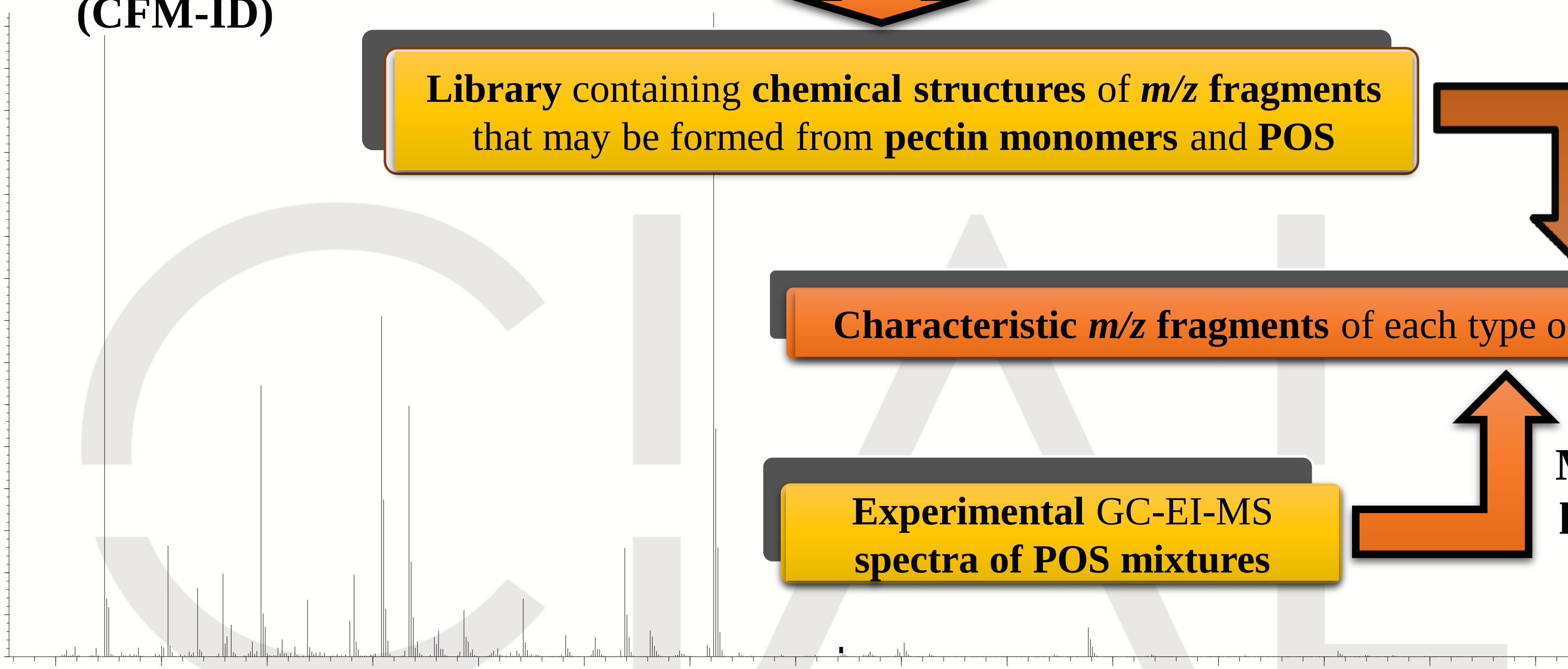
GC-EI-MS **Theoretical fragmentation** patterns

Library containing **chemical structures** of ***m/z*** **fragments** that may be formed from **pectin monomers** and **POS**

Characteristic *m/z* fragments of each type of **POS**

Machine learning

Experimental GC-EI-MS **spectra of POS mixtures**



INSTITUTO DE INVESTIGACIÓN
EN CIENCIAS DE LA ALIMENTACIÓN

RESULTS AND DISCUSSION

Most **POS** remain **unidentified**

Computational models
to **interpret spectra**

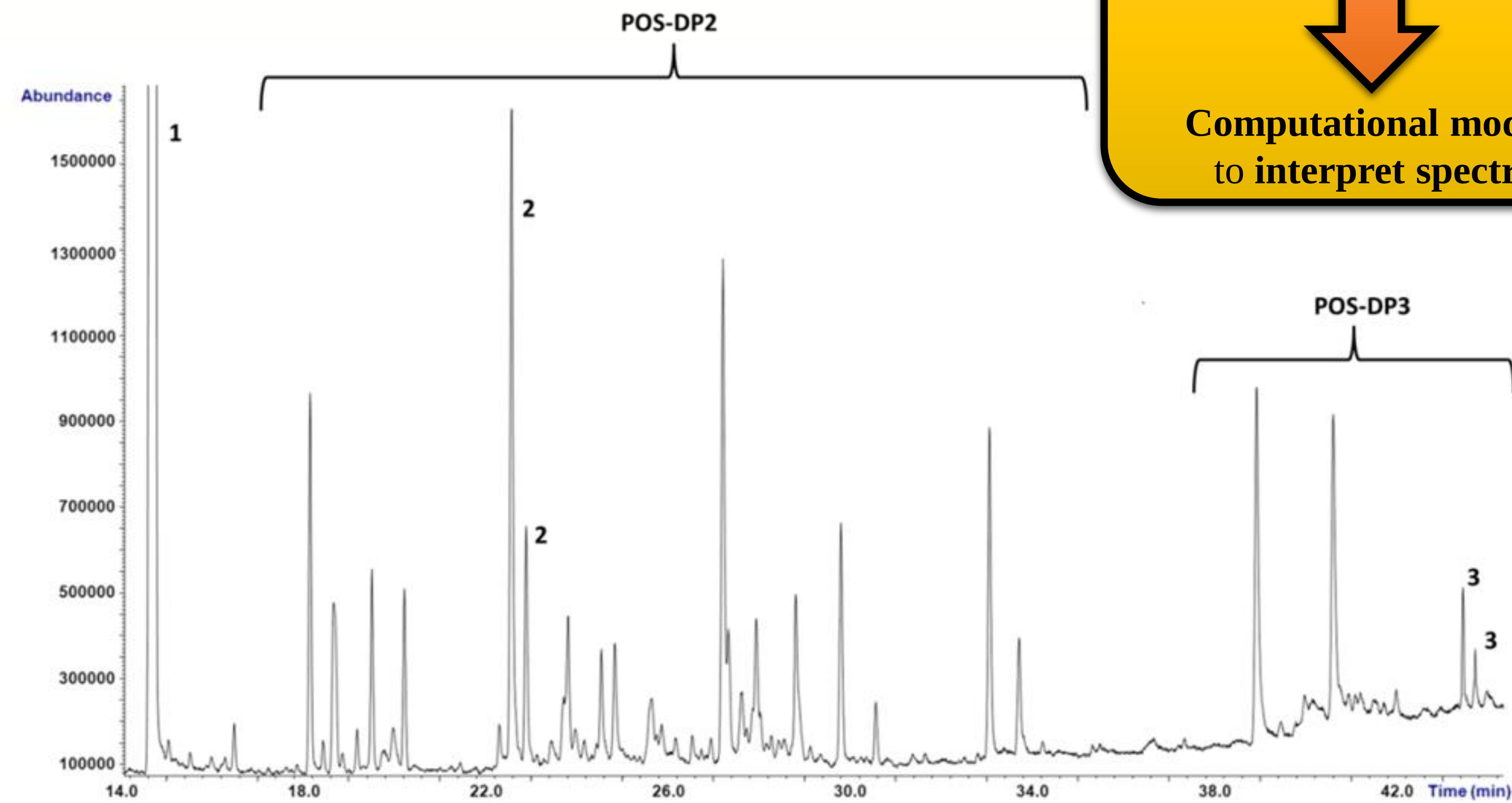
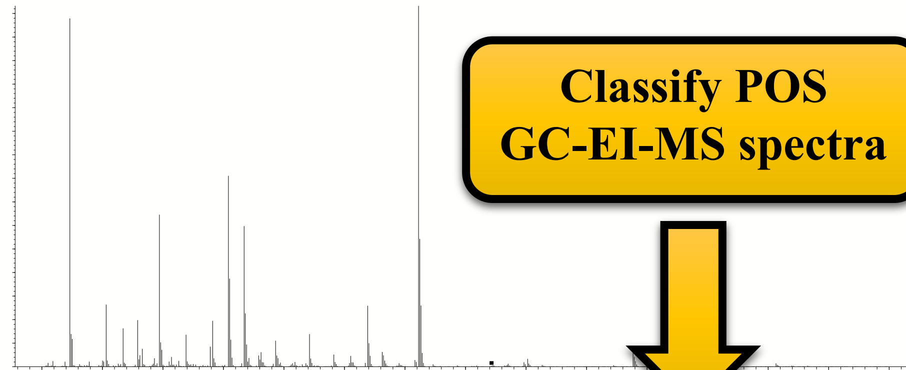


Figure. GC-MS profile of a hydrolysate of artichoke pectin obtained by incubation with cellulase from *A. niger*. Peaks: (**1**) Internal standard, (**2**) Digalacturonic acid (Di-GalA), (**3**) Trigalacturonic acid (Tri-GalA). **POS-DP2**: unknown pectic disaccharides, **POS-DP3**: unknown pectic trisaccharides.

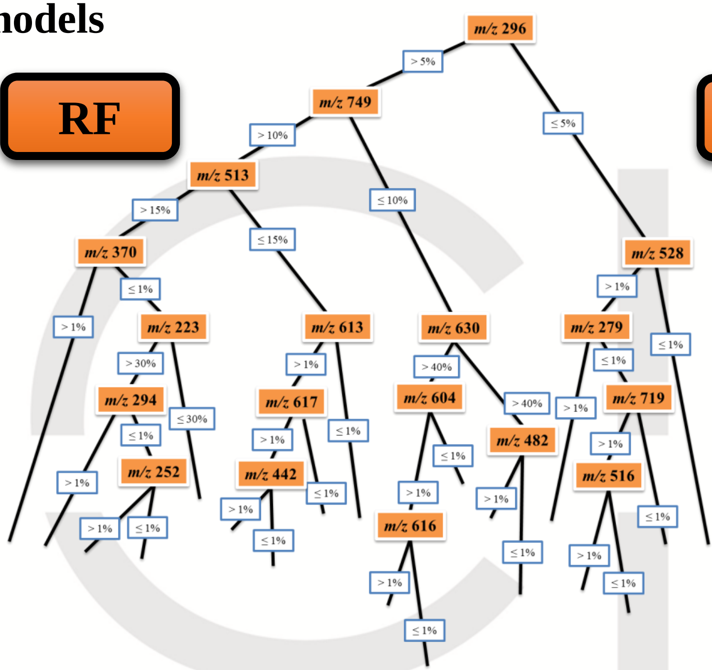
RESULTS AND DISCUSSION

- **RF**: multiple **decision trees**, each node is split, outputting different classes (i.e. enzyme used). using a **subset of predictors** (i.e. *m/z* ions)
- **ANN**: formed by an **input layer** (i.e. *m/z* ions), an **output layer** (i.e. enzyme used) and several **neurons** connected through functions
- **BLR**: strong model **from numerous simple models**

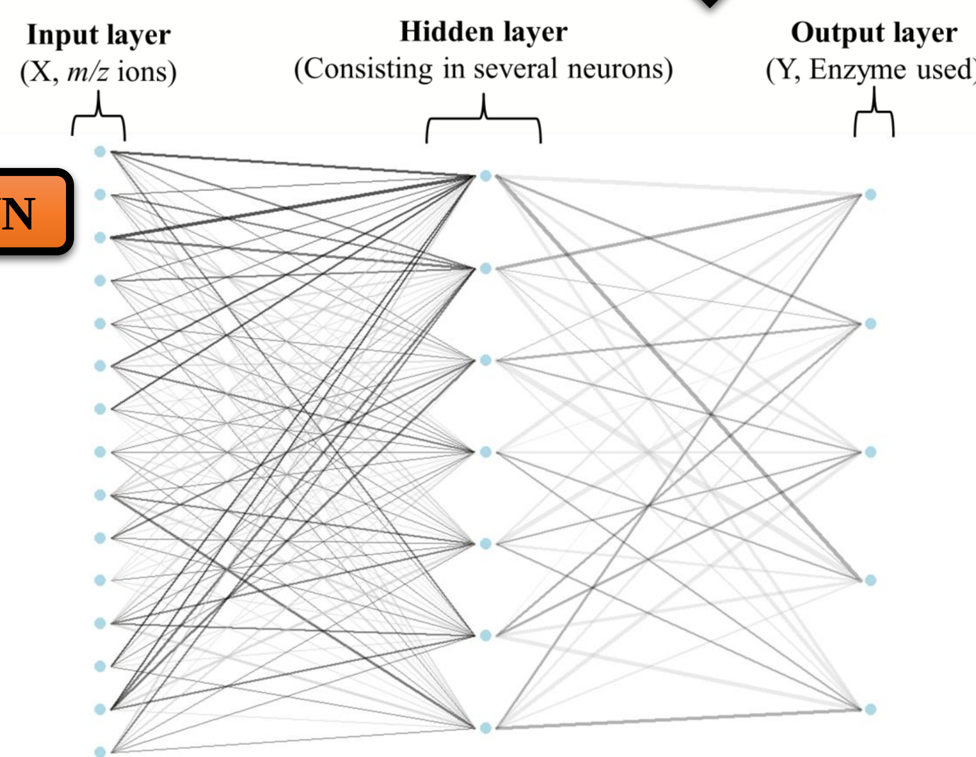


Classify POS GC-EI-MS spectra

RF



ANN



RESULTS AND DISCUSSION

Machine learning models

Characteristic fragments

Compare to *in silico* POS structure library (Competitive fragmentation modelling)

Structures of Characteristic fragments

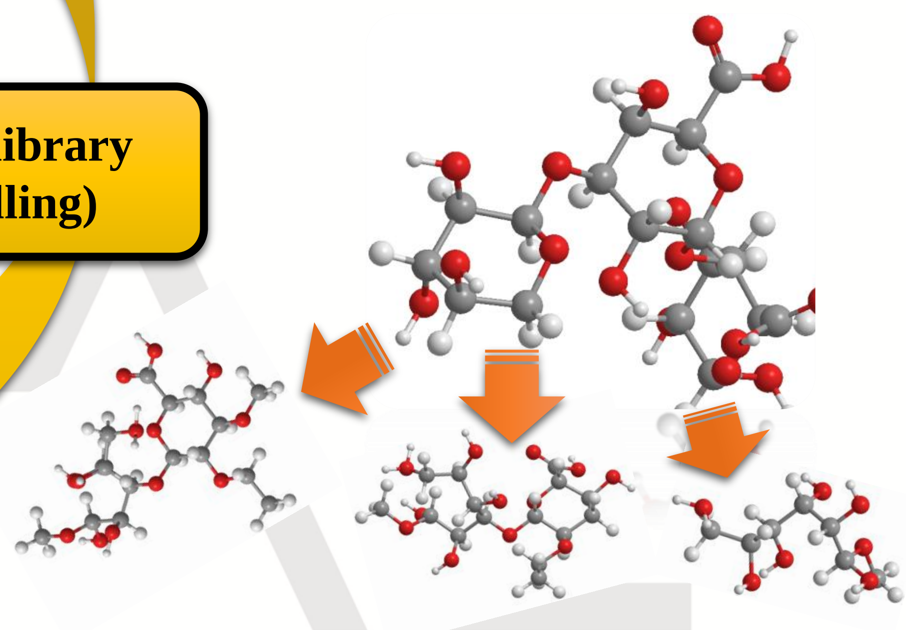
Classify POS according to the enzyme used

- Pentopan-POS
- Cellulase from *Aspergillus niger*-POS
- Pectinex Olio-POS

High accuracy rates (95-100%)

- Trained on 70% spectra and tested on 30% (new samples)

RF outperformed ANN

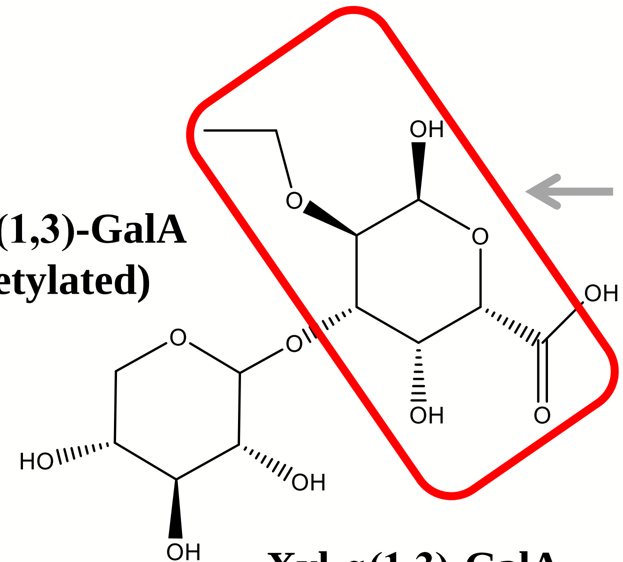


GC-MS characterisation of artichoke POS using machine learning and CFM
XIX SECyTA

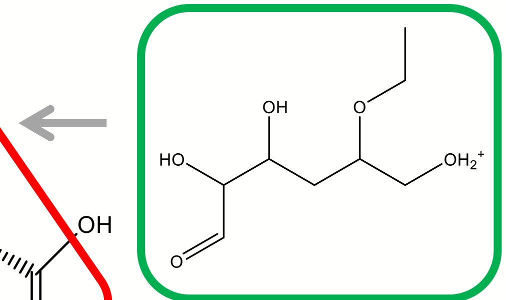
RESULTS AND DISCUSSION

Characteristic fragments of **POS** obtained with **Pectinex®Ultra-Olio**

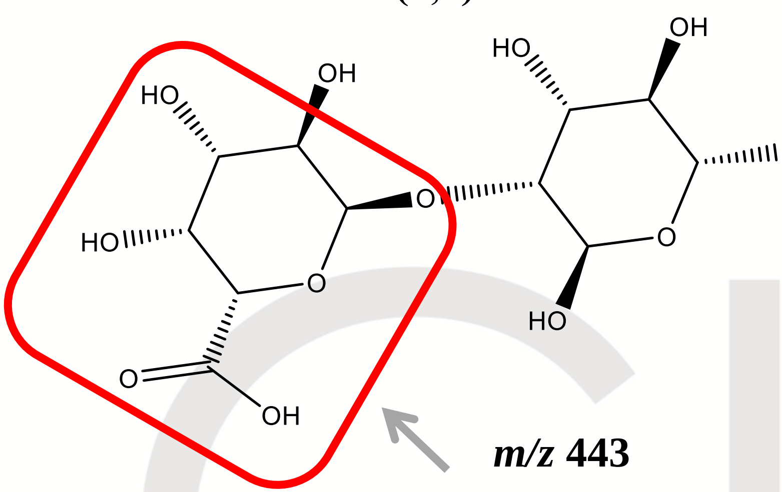
Xyl-α(1,3)-GalA (acetylated)



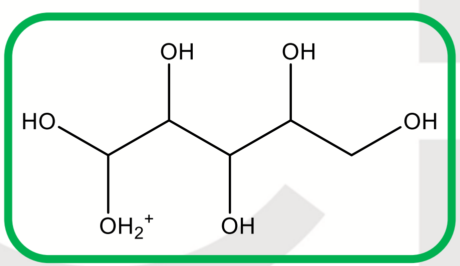
***m/z* 424**



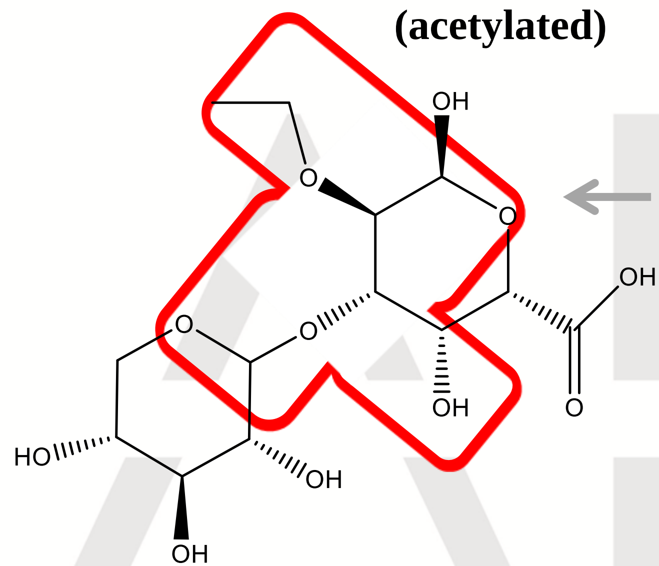
GalA-α(1,2)-Rha



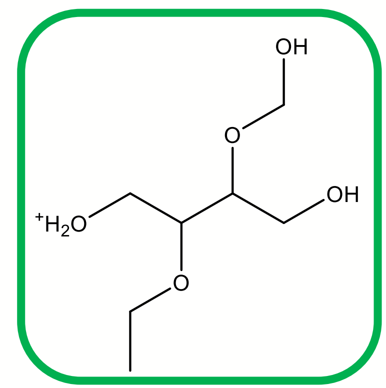
***m/z* 443**



Xyl-α(1,3)-GalA (acetylated)



***m/z* 340**

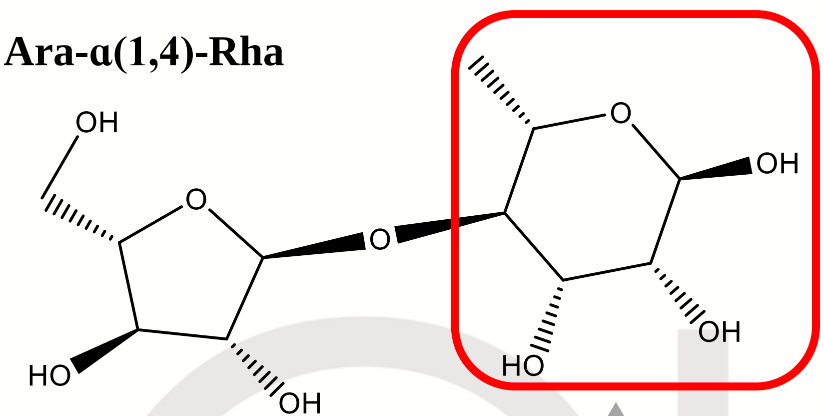


GC-MS characterisation of artichoke POS using machine learning and CFM
XIX SECyTA

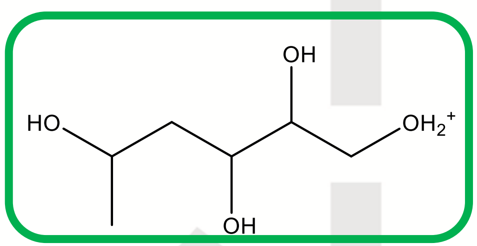
RESULTS AND DISCUSSION

Characteristic fragments of **POS** obtained with **Pentopan®Mono-BG**

Ara-α(1,4)-Rha

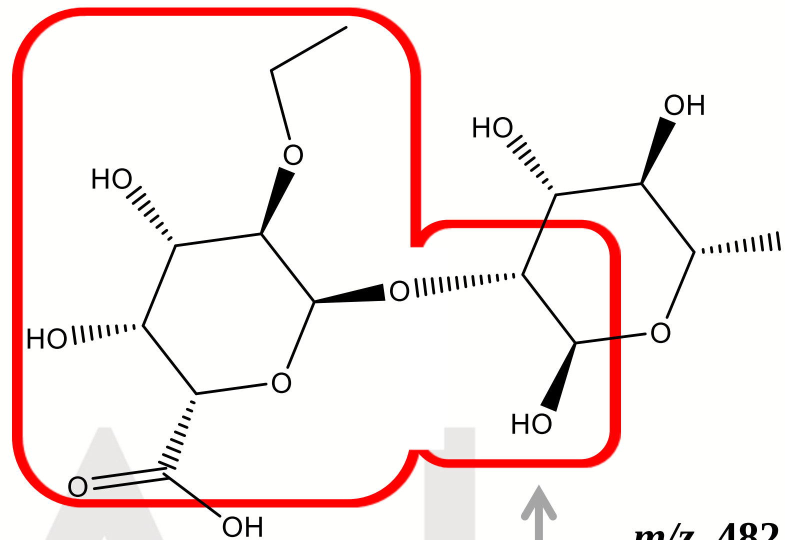


***m/z* 454**

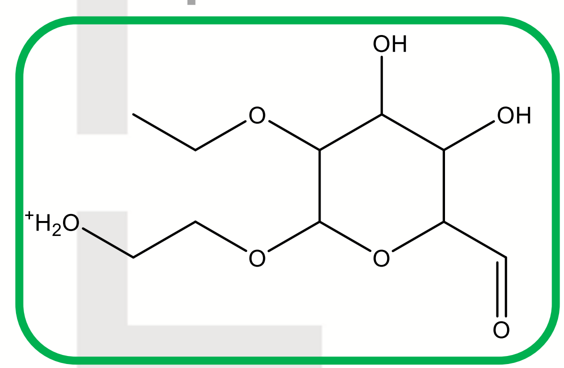


Characteristic fragments of **POS** obtained with **Cellulase from *Aspergillus niger***

(acetylated) GalA-α(1,2)-Rha



***m/z* 482**



CONCLUSIONS

- **Structural profiles** of POS obtained using enzymes with different main activities (pectin-lyase, endoxylanase and cellulase), have been **determined**.
- Oligosaccharide **spectra** can be **classified using machine algorithms**, showing high accuracy rates and determining **characteristic *m/z* ions**.
- ***In silico* GC-EI-MS fragmentation** can be used to calculate the **chemical structures** of characteristic *m/z* ions, and to associate this information to the presence of **specific oligosaccharide** chains
- This methodology allows to **tentatively predict** some of the **most probable POS structures** that may be obtained from artichoke pectin and could be applied to other substrates or novel oligosaccharide mixtures.

INSTITUTO DE INVESTIGACIÓN
EN CIENCIAS DE LA ALIMENTACIÓN

GC-MS characterisation of artichoke POS using machine learning and CFM
XIX SECyTA

ACKNOWLEDGEMENTS

This work has been funded by projects AGL2014-53445-R and AGL2017-84614-C2-1-R.
Carlos Sabater thanks his FPU Predoc contract from Spanish MECD (FPU14/03619).