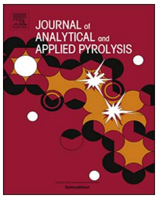
Journal of Analytical and Applied Pyrolysis xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Analytical and Applied Pyrolysis

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



Application of Py-GC/MS coupled with PARAFAC2 and PLS-DA to study fast pyrolysis of genetically engineered poplars

Hilal E. Toraman[a], Victor Abrahamsson[b], Ruben Vanholme[c], Rebecca Van Acker[c], Frederik Ronsse[d], Gilles Pilate[e], Wout Boerjan[c], Kevin M. Van Geem[a,*], Guy B. Marin[a]

[a] Ghent University, Laboratory for Chemical Technology, Technologiepark 914, 9052 Ghent, Belgium
[b] Lund University, Department of Chemistry, Centre for Analysis and Synthesis, 22100 Lund, Sweden
[c] Ghent University, Department of Plant Systems Biology, VIB, Technologiepark 927, 9052 Ghent, Belgium
[d] Ghent University, Department of Biosystems Engineering, Coupure Links 653, 9000 Ghent, Belgium
[e] Unité Amélioration, Génétique et Physiologie Forestières, INRA Orléans, BP 20619, Ardon, 45166 Olivet, France

ARTICLE INFO

Keywords:
Pyrolysis
Field-grown transgenic plants
Phenolic compounds
PARAFAC2
PLS-DA

ABSTRACT

Field-grown genetically engineered and wild-type poplars were pyrolyzed in a micro-pyrolysis (Py-GC/MS) setup under fast pyrolysis conditions. Poplars (*Populus tremula x P. alba*) down-regulated for cinnamoyl-CoA reductase (*CCR*), which catalyzes the first step of the monolignol-specific branch of the phenylpropanoid biosynthetic pathway, were grown in field trials in France and harvested after a full rotation of 2 years. The effect of small compositional differences, specifically small shifts in lignin composition and their impact on the bio-oil composition, could not be identified using principal component analysis (PCA), necessitating the use of more advanced analysis techniques. The combination of parallel factor analysis 2 (PARAFAC2) and partial least squares-discriminant analysis (PLS-DA) for detailed characterization and classification of the pyrolysis data enabled the classification of the poplars with a success rate above 99% using the PARAFAC2 scores. This methodology proved to be extremely valuable to identify subtle information in complex datasets, such as the one used in this study. The obtained PLS-DA models were validated by cross-validation, jackknifing and permutation tests in order to ensure that the model was not overfitting the data. PLS-DA showed that down-regulation of *CCR* disfavored the relative amount of both guaiacyl and syringyl lignin-derived compounds. This study shows that lignin engineering can be a promising strategy to alter the lignin composition of the biomass for the production of high value-added phenolic compounds.

1. Introduction

Lignocellulosic biomass is the most abundant and renewable resource in the world and has great potential as a sustainable source of fuels and chemicals which are presently mostly obtained from finite fossil-based resources [1]. There is an increasing research effort from government, academia and industry to develop technologies which can either be an alternative or complementary to the well-established manufacturing platforms of the petrochemical industry.

There are two main groups of biomass conversion processes, namely bio-chemical and thermochemical. During bio-chemical conversion such as digestion and fermentation, biomass breaks down into molecules of lower molecular weight by enzymes or micro-organisms. Compared to thermochemical conversion, bio-chemical conversion processes suffer from longer reaction times and lengthy reaction steps, but they do not require much external energy input as in the case of thermochemical conversion processes [2]. On the other hand, biomass is converted into smaller molecules at high temperatures during thermochemical conversion processes: combustion, pyrolysis and gasification [3–5].

Among thermochemical conversion methods, fast pyrolysis is a promising route for the production of fuels and chemicals for thermochemical biorefineries [3,6]. In principle, fast pyrolysis requires the use of finely ground biomass to achieve high heating rates and very high heat transfer rates at a temperature of ca. 500 °C. Short vapor residence times (e.g., 1–2 s) and rapid quenching of the pyrolysis vapors are necessary to minimize the secondary vapor phase reactions and to obtain the maximum bio-oil yield (up to 75 wt% on a dry-feed basis) [3,4,7,8].

Bio-oils are complex mixtures of water and hundreds of organic compounds which include acids, aldehydes, ketones, alcohols, esters, anhydrosugars, furans, phenols, guaiacols, syringols, as well as molecular oligomers such as holocellulose-derived anhydro-oligosaccharides

---

* Corresponding author.
E-mail address: Kevin.VanGeem@UGent.be (K.M. Van Geem).

https://doi.org/10.1016/j.jaap.2017.11.022
Received 11 August 2017; Received in revised form 1 November 2017; Accepted 27 November 2017

Please cite this article as: Toraman, H.E., Journal of Analytical and Applied Pyrolysis (2017), https://doi.org/10.1016/j.jaap.2017.11.022

and lignin-derived oligomers [9]. Bio-oils obtained from the pyrolysis of lignocellulosic biomass can be directly used as fuels in furnaces or boilers without any upgrading process. Nevertheless, bio-oil has some undesirable properties such as its water content (15–30 wt.%), high oxygen content (35–40 wt.%), low heating value compared to fossil fuels (higher heating value (HHV) of approximately 17 MJ kg^{-1}), high viscosity (0.025–1 Pa s at 315 K), high acidity (pH of 2–4), and last but not least its instability [8]. Bio-oils cannot directly replace high-grade transportation fuels (e.g. diesel and gasoline) without upgrading by catalytic cracking and/or hydroprocessing [8].

However, it is not necessarily ideal to simply convert bio-oils into fuels since bio-oils contain a wide range of chemicals, such as lignin-derived phenolic compounds, which can be isolated and used as renewable feedstocks for the production of pharmaceuticals, fine chemicals, additives in fertilizing, resin manufacturing and food processing chemicals [1,10–15]. Realizing those potential applications requires having control over the chemical composition of the bio-oil through the selection of the optimal biomass and the identification of the optimal operating conditions for the production of the targeted high value-added specialty chemicals.

Therefore, a lot of research is devoted to understanding the effect of biomass composition on bio-oil composition using whole biomass from different feedstocks or its main constituents cellulose, hemicellulose and lignin [16–20]. However, it is extremely challenging to characterize the relationship between biomass composition and bio-oil composition when the focus is on the comparison of pyrolysis product composition across different taxonomic groups differing in more than one property. The use of genetically modified biomass samples of which the phenotype vary only in one or a few components is one of the most extreme cases which will aid in directly addressing the relationship between the starting material and its product composition in a more systematic way [19,21]. Chemometrics play a crucial role in finding the underlying patterns in complex data sets through the simultaneous assessment of all variables. Principal component analysis (PCA) and partial least squares-discriminant analysis (PLS-DA) are two of the most widely applied multivariate data analysis techniques for chemometrics [22]. They allow graphical visualization and hence better understanding of different data patterns and relations based on scores and loadings for the corresponding components. PCA is an unsupervised dimensionality reduction algorithm which can be utilized to describe the overall variation in the data set, as applied previously for the detailed characterization of the pyrolysis data obtained using caffeic acid *O*-methyltransferase (*COMT*) and caffeoyl-CoA *O*-methyltransferase (*CCoAOMT*) down-regulated poplars [23]. However, PCA attempts to describe the maximum variation with as few principal components as possible and this variation does not necessarily lead to the separation of the two classes from each other. Being a supervised dimensionality reduction algorithm, PLS-DA finds the maximum separation between each class more effectively based on the class membership designation which allows to find patterns in the data, as in this study [22]. Unlike PCA, the risk of overfitting of the PLS-DA models to the data requires extensive validation by using a combination of different techniques such as cross-validation, jackknifing and permutation tests [24–26].

Here, we investigated the effect of biomass composition on bio-oil composition using unique field-grown poplars with an engineered lignin composition. Lignin is a heteropolymer mainly consisting of the monolignols coniferyl and sinapyl alcohol and to a lesser extent of *p*-coumaryl alcohol, which result in guaiacyl, syringyl and *p*-hydroxyphenyl units when incorporated into the lignin structure [27,28]. The monolignols are produced from phenylalanine via the phenylpropanoid and monolignol biosynthetic pathways [29,30]. Cinnamoyl-CoA reductase (*CCR*) catalyzes the first step of the monolignol-specific branch of the phenylpropanoid biosynthetic pathway (Fig. 1). It is responsible for the conversion of *p*-coumaroyl-CoA and feruloyl-CoA to *p*-coumaraldehyde and coniferaldehyde, respectively [31]. Since down-regulation of the *CCR* gene in annual model plants resulted in significantly reduced lignin content [32–38], down-regulating *CCR* in woody plants was seen as a promising strategy to improve wood quality for pulping [39] and bioethanol production [40]. Wood from 6-month-old greenhouse-grown *CCR* down-regulated line (FS3) had a decrease of about 18% in the Klason lignin content as compared to wood from the wild-type line while the thioacidolysis-derived syringyl/guaiacyl (S/G) ratio was not statistically different from that of the wild-type [39]. Similar to the greenhouse-grown poplars, wood derived from two 20-month-old field-grown *CCR* down-regulated lines had a 9% (FS3) and 24% (FAS13) reduction in Klason lignin content. Furthermore, the FAS13 line presented about a 7% decrease in thioacidolysis-derived syringyl/guaiacyl (S/G) ratio whereas (S/G)ratio of the FS3 line was not statistically different from that of the wild-type [38]. Data generated from greenhouse-grown trees cannot be directly extrapolated to field-grown trees without any experimentation on a field since greenhouse-grown trees do not experience the annual cycles of growth and dormancy, and hence their xylem does not mature as in the case of field-grown trees. Therefore, field trials play an important role in the translation of fundamental knowledge produced under laboratory conditions to real life conditions [40].

Here, field-grown *CCR* down-regulated poplars and wild-type poplars have been pyrolyzed to investigate the influence of lignin composition of the lignocellulosic biomass on the composition of fast pyrolysis bio-oils. A multi-way technique, i.e. parallel factor analysis 2 (PARAFAC2) was used for data decomposition and extraction of the relevant information from complex gas chromatography-mass spectrometry (GC–MS) data. Then, principal component analysis (PCA) and partial least squares-discriminant analysis (PLS-DA) were performed for classification and extraction of the systematic variation in the bio-oils produced from the genetically engineered and wild-type poplars. This study provides guidance for plant genetic engineers that allow to design and optimize lignocellulosic biomass for the thermochemical production of chemicals.

2. Materials and methods

2.1. Plant material and sample preparation

The genetically engineered, *CCR* down-regulated and the corresponding wild-type poplars (*Populus tremula x alba* cv. 717-1B4) were sampled from a field trial at Orléans (France). Down-regulation of *CCR* was obtained by transforming poplar with sense (FS3 lines) and antisense constructs (FAS13 lines) [39–41]. Prior to planting, they were simultaneously micropropagated in vitro and 120 ramets of each line were grown in the greenhouse. They were planted in the field in July 2008, after obtaining suitable authorization (application #B/FR/07/06/01, authorization #07/015 from the “Direction Générale de l’Alimentation” from the French “Ministère de l’Agriculture et de la Pêche” (on September 21, 2007 for a 5-year period) after a positive ruling from the French “Commission du Génie Biomoléculaire”). FS3 and FAS13 correspond to the names WT52-3 and WT62-13 in the #B/FR/07/06/01 document. The 120 plants of each line were dispatched in 5 randomized blocks (with 24 trees per block planted in two double rows, Fig. 2).

The plant density was chosen according to short-rotation coppice practice: the space between trees of one double row was 0.55 m whereas the interspace between the two double rows was 1.5 m, and the planting distance within a row was 1 m. To prevent edge effects, the experimental plantation was bordered with one row of wild-type (WT) trees. During the growing season, the poplars were drip irrigated. In March 2010, the stems were coppiced to initiate regrowth as short-rotation coppice. Plants were harvested after a first full rotation of 2 years, in February 2012 [40,42].

At harvest, the stems were debarked and air-dried. The samples were derived from a 40 cm long stem fragment at the basis of the main stem from 3 out of the 6 biggest trees from each block. The 3 samples