PAPER

Unlocking the potential of biofuels *via* reaction pathways in van Krevelen diagrams†

Diana Catalina Palacio Lozano, [ID] [a] Hugh E. Jones, [ID] [a,b] Tomas Ramirez Reina, [ID] [c] Roberto Volpe [ID] [d] and Mark P. Barrow [ID] *[a]

Production of fuels and targeted chemicals from biomass represents a current challenge. Pyrolysis of biomass generates liquid bio-oils but these are highly complex mixtures. In order to obtain the desired products, optimized reaction conditions are required and this, in turn, drives the need for a fundamental understanding of the complex reaction network. Bio-oils are a complex mixture of thousands of individual molecular compositions, with differing numbers of carbon, hydrogen, nitrogen, and oxygen atoms (c, h, n, and o, respectively). The compositional spaces of such complex mixtures with high oxygen contents are commonly plotted using van Krevelen diagrams, where the H/C *versus* O/C ratios are displayed. For a bio-oil to be effectively used in engines, further upgrading is necessary to drive the compositions towards low oxygen and high hydrogen content (thus, low O/C and high H/C values). Here, we propose reaction vectors in van Krevelen diagrams to outline the possible reaction routes that favour the production of molecules with increased energy density, using examples of bio-oils produced from citrus waste (lemon and orange peel) and olive pulp. When reactions such as the addition or loss of CO, CO_2, CH_4, and H_2O occur, a displacement of the compositions of molecules in terms of H/C and O/C coordinates is observed. The direction and magnitude of the displacement along each axis in van Krevelen diagrams depends upon the specific reaction route and the elemental content of each molecule. As a consequence of the wide diversity of compositions, different reaction routes are suggested that include multi-step upgrading processes, including hydrogenation and the elimination of oxygen in the form of CO and CO_2. The detailed molecular composition of the starting material, plotted in van Krevelen diagrams for visualization, paves the way for greater insight into potential reaction pathways for components within these highly complex mixtures. In turn, the equations proposed hold potential to inform future production strategies, increasing the energy density of bio-oils whilst also reducing the undesirable char formation.

Introduction

Pyrolysis is a highly desirable technology for the conversion of solid biomass into higher value carbonaceous solids and hydrocarbon liquids.[1] The process consists of heating the solid feedstock, in a purpose made reactor, to temperatures typically above 500 °C, in an inert atmosphere to induce cleavage of molecular bonds and feedstock decomposition without complete oxidation. Numerous reactions that occur during the pyrolysis of solid fuels are affected by feedstock and sample characteristics (physical and chemical) and process parameters;[2] such as heating rate, peak temperature, inert gas flow rate, and residence time at peak temperature.[3] Pyrolysis is a technology as old as humankind,[4] since some 30 millennia ago, human beings have used pyrolysis to make charcoal. In the past 3 to 4 decades, thousands of experimental, modelling, and numerical studies have been produced in the attempt to elucidate the fundamentals of the thermochemical breakdown of solid fuels, however, the process has not yet been fully understood.[5] The complexity of the process and the vast set of possible reactions occurring in both the homogenous and heterogeneous solids, vapour, and gas phases, make the process particularly complex to control at molecular level.[6] As a result, ensuring the quality of pyrolysis products is challenging.

Pyrolysis products can be classified into three principal types: liquids, chars (carbonaceous material), and non-condensable gases (*e.g.* CO_2, H_2, CO, and CH_4). The properties of

[a]Department of Chemistry, University of Warwick, Coventry, CV4 7AL, UK.
E-mail: M.P.Barrow@warwick.ac.uk
[b]Molecular Analytical Science Centre of Doctoral Training, University of Warwick, Coventry, CV4 7AL, UK
[c]Department of Chemical and Process Engineering, University of Surrey, GuilfordSurrey GU2 7XH, UK
[d]School of Engineering and Materials Science, Queen Mary University of London, Mile End Road, London E1 4NS, UK
†Electronic supplementary information (ESI) available. See DOI: 10.1039/d1gc01796a

each product depend upon the pyrolysis conditions, the diversity of compositions of the raw material (*i.e.* cellulose, hemicellulose, and lignin), their moisture content; among others.[7,8] The char is primarily carbon present in aromatic polycyclic structures. Therefore, chars have a high energy density that can be treated and converted into activated carbons for energy storage applications and for process heat.[9] The compositions of the non-condensable fraction may serve as a source of hydrogen.[8]

Bio-oil, in particular, the liquid product of pyrolysis, is a highly complex liquid, made up of oxygenated, aliphatic, and polyaromatic hydrocarbons, whose process of formation during reaction, and composition is largely unknown. Yet, bio-oil is a potentially high value chemical product that can be used as a precursor for virtually all the chemicals and products currently derived from crude oil, and in particular fuel for transport such as biodiesel.[10,11]

The molecular composition of the pyrolytic bio-oil can be determined by ultrahigh resolution mass spectrometry (UHRMS). A resolving power of 300 000 FWHM at m/z 300 is typically necessary to distinguish individual molecular species with very small mass differences (*e.g.* C_3/SH_4).[12] In particular, Fourier transform ion cyclotron resonance mass spectrometry (FTICR MS), can resolve species with a mass difference lower than the mass of an electron and, for special experimental set-ups, it is possible to acquire a mass spectrum with constant ultrahigh resolving power.[13,14] Whereas elemental analysis provide information of an average molecule with a certain C, N, and O content, FTICR MS analysis provides the detailed elemental compositions (*e.g.* $C_cH_hN_nO_oS_s$ where c,h,n,o and s, are the number of carbon, hydrogen, nitrogen, oxygen and sulfur atoms respectively) within a complex sample. Pyrolysis bio-oils have shown to be composed of thousands of individual molecular compositions.[15–17]

In comparison with conventional crude-oils, bio-oils are less complex in terms of number of compositions and typically contain a greater number of highly oxygenated species whose presence hampers the direct application of bio-oil as fuels, requiring further downstream upgrading processes.[15,18] As a consequence of the high oxygen content, molecular species of bio-oils are better represented in van Krevelen diagrams,[19] where the atomic H/C *vs.* O/C ratio of the molecular compositions are plotted. Ultrahigh resolution data represented in van Krevelen plots have been widely used to classify compounds in categories based on solely the H/C *vs.* O/C ratio or using a multidimensional stoichiometric compound classification.[20,21]

Van Krevelen plots were introduced in 1950 by D. W. van Krevelen as a statistical method to understand coal thermochemical processes.[19] As revisited by Kim *et al.* in 2003,[22] the principal reactions such as oxidation, dehydration, demethanation, decarboxylation, and hydrogenation are represented by straight lines in van Krevelen diagrams. Reaction processes in van Krevelen diagrams have been used to understand the mechanisms of the thermal transformation of organic matter, such as coal or cellulose. However, this approach has never been employed in conjunction with detailed information

about molecular compositions to understand the reaction processes involved in bio-oil compositions. The aim of this paper is to propose a novel use of the van Krevelen diagrams as a method to investigate the potential routes for the upgrading of bio-oils obtained from pyrolysis of different biomass sources. Starting from van Krevelen diagrams of four different pyrolysis bio-oils, we will show how an interpretation of such diagrams helps to unveil the complex chemical nature of bio-oils and provides a practical indication aimed at choosing optimal reaction routes for the upgrading of bio-oils into valuable green chemicals. We will also show how the proposed use of van Krevelen plots provides important insights for the synthesis of catalysts that are 'tailored' to a specific bio-oil produced from a specific feedstock and pyrolysis route. This information is particularly relevant for the catalysis community that currently faces significant challenges to synthesising catalysts that are suited for bio-oil upgrading. Such fundamental understanding is urgently needed in the context of a low-carbon future, in which renewable fuels and chemicals derived from biomass resources are meant to play a crucial role in the transition towards sustainable societies.

Materials and methods

Pyrolysis reaction set-up

Peel waste from lemon, "Femminello" variety (Lem-P), and blonde orange, "Tarocco" variety (Or-P), and Olive pulp from milled olives "Moresca" variety (Ol-P) were collected as a wet pulp and then oven dried for approximately 6 h at 105 °C to eliminate moisture. Dried samples were then ground to <850 μm particle size, and then again oven dried for 12 h at 105 °C to eliminate any moisture uptake. The dried feedstock was then sieved to select a particle size range of between 200 and 850 μm, before being pyrolysed.

A horizontal fixed-bed type reactor derived from the standard Gray-King (GK) assay test on coal was purposely modified to conduct pyrolysis tests. The reactor design and full details of the set-up have been reported previously.[23] Briefly, the system consists of a quartz cylindrical reactor 340 mm long and 20 mm internal diameter closed at one end. A special quartz cap, equipped with an 8 mm internal diameter quartz inner tube, allows a flow of inert gas through the feedstock to sweep volatiles away into a cold trap used to collect condensable products.

Bio-oil collection

Bio oils were condensed by means of condensation in a quartz U-shaped tube placed in a dry-ice water-ethylene glycol (Sigma-Aldrich) bath kept at a temperature of −27 °C in a Dewar flask. After the pyrolysis reaction was completed, the tar trap and fittings were washed in a 1 : 4 chloroform/methanol (both chemicals from Sigma-Aldrich) solution to recover the condensed bio-oil. Note that whilst chloroform was used in this instance, later experiments have revealed that using methanol alone is also viable; this is in line with the aim of reducing usage of sol-

vents that are harmful to the environment.[24] The organic solution was then filtered in pre-weighed Whatman no. 4 filter to remove remaining char particles. Bio-oil was then recovered by rotary evaporation at 350 mbar and 40 °C for two hours, thus, allowing the recovery of compounds with a boiling point equal or higher than benzene.[23,25] The bio-oils were collected by means of condensation at a temperature of −27 °C and were subsequently analysed by ultrahigh resolution mass spectrometry as shown below.

Pyrolysis run

Pyrolysis experiments at a peak temperature of 500 °C were carried out under a flow of 1.5 L min^{-1} of nitrogen (Air Liquide, 5.0 Alphagaz 1). Additionally, a pyrolysis run at 400 °C was carried out on Lem-P. Thus, three bio-oils obtained at 500 °C (Lem-P 500, Ol-P 500, Or-P 500) and a bio-oil from lemon obtained at 400 °C (Lem-P 400) are analysed in this paper. Approximately 10 g (dry basis – db) of sample was heated to the desired peak temperature with a 50 °C min^{-1} heating rate and held at peak value for 30 minutes. Gas residence time was calculated to be approximately 3.3 s. Char yields (Mchar/Mraw %w/w db, not reported in this study) were calculated to compare results obtained in previous studies only with the purpose of validating the runs. Runs showing char yields beyond a ±5% error range were repeated. Neither chars nor gases were analysed in the present study.

Ultrahigh resolution mass spectrometry

Molecular compositions of the bio-oil were obtained by a 12 T solariX FTICR mass spectrometer (Bruker Daltonik GmbH, Bremen, Germany) equipped with a custom nano-electrospray source (nESI) operating in negative-ion mode for the detection of the highly polar and acidic compositions of the samples. The samples were diluted in methanol (HPLC grade 99.9%, Honeywell, Bracknell, UK) at 0.05 mg mL^{-1}. Diluted samples were directly infused, and ions were accumulated for 0.120–0.200 s in the collision cell prior to being transferred to the ICR cell. The mass spectra were acquired by co-adding 150 data sets at 8 MW. A resolving power of 1 M at m/z 200 was achieved. With these acquisition parameters, a mass spectrum can then be acquired in about 20 minutes. It worth to note that the new solariX 2XR can offer a similar performance with half the acquisition time.[26] The mass spectra can be found in Fig. S1.†

The mass spectra were externally calibrated using internal recalibration with abundant homologous series corresponding to compositions with O_4 and O_6. Composer version 1.5.6 (Sierra Analytics, CA, USA) was used to assign the molecular compositions of bio-oils with molecular formulae $C_{c \leq 200}H_{h \leq 1000}N_{n \leq 3}O_{o \leq 20}S_{s \leq 1}$, a mass accuracy <1 ppm, and up to 40 double bond equivalents (DBE). The double bond equivalents are calculated according to eqn (1)

$$DBE = c - h/2 + n/2 + 1 \qquad (1)$$

where c, h and n are the number of carbon, hydrogen, and nitrogen atoms respectively. Hydrocarbon compositions can be expressed in terms of the hydrogen deficiency, Z, as C_cH_{2c+Z} where,

$$Z = -2(DBE) + n + 2 \qquad (2)$$

According to eqn (2), saturated hydrocarbons (DBE = 0) have a Z-value of 2, species with one double bond or one ring have a Z-value of 0, and species with more than one rings plus double bond have negative, even Z-values. Thus, molecules with low hydrogen deficiency are enriched with hydrogen atoms and therefore, have an increased energy density (higher H/C-values).

In typical petroleomics analysis, each elemental composition is classified by its heteroatomic class (*e.g.* $N_nO_oS_s$ with a label "[H]" or "[Na]" to indicate deprotonated and sodium adducts respectively), carbon number, DBE, H/C and O/C ratio. For instance, a composition $[C_8H_8O_4–H]^-$ is classified as belonging to the O_4[H] compound class, with carbon number of 8 and a DBE of 5.5 as the deprotonated species (or 5 when considering the neutral species). The detailed compositional analysis of the samples, *e.g.* class distribution, DBE plots and the commonality within the molecular composition of the samples can be found in the ESI.† A summary of the total number of oxygenated molecular compositions, their carbon and oxygen atom range, and their DBE and hydrogen deficiency can be found in Table 1.

Data visualization and simulation

The data mining and visualization was performed using an in-house software developed at the University of Warwick, named KairosMS.[27] The software also offers the possibility for density plot visualizations (Kernel density plots) that implements the stat_density_2d function from the ggplot2 package and Upset plots[28,29] for the inspection of the intersection of the molecular formulae between the samples.

Additional software was developed, named VKSim, to simulate the effect of different upgrading reactions upon the bio-oil compositions. Pseudo code for the key algorithm can be found in the ESI.† In brief, the software allows the visualisation of the van Krevelen diagram for a complex mixture after the addition or removal of small molecules including, but not restricted to: CO, CO_2, H_2O, CH_4, and H_2. As the first attempt to develop a graphical-mathematical method to simulate the reactions based on reaction vectors and van Krevelen diagrams, the molecules obtained after a reaction vector (products) are currently only restricted to the following golden

Table 1 Carbon number and oxygen content range detected for the oxygenated hydrocarbons of the bio-oils ($C_cH_hO_o$). The DBE, hydrogen deficiency range, and the percentage contribution to the total signal are also presented

Sample	Molecular species	%	C_c	O_o	DBE	Z
Lem-P 400	3970	63	C_{5-41}	O_{2-11}	1–18	0–(−34)
Lem-P 500	4067	61	C_{4-45}	O_{2-11}	1–19	0–(−36)
Or-P 500	3798	65	C_{6-38}	O_{2-11}	1–19	0–(−36)
Ol-P 500	3583	71	C_{7-49}	O_{1-8}	0–22	2–(−34)

rules for heuristic filtering of molecular formulas:[30] (1) restriction for the number of elements (*e.g.* negative values of C, H, and O are not allowed for products of a reaction), (2) an elemental ratio of $0.2 \leq H/C \leq 3.1$, and (3) an elemental ratio of $0 \leq O/C \leq 1.2$. Additionally, the minimum DBE value of a molecule is restricted to DBE = 0 (saturated hydrocarbon). The current simulation does not incorporate secondary reactions such as C–C, C–O bond cleavage, polymerization, or isomerization. Additionally, the simulations are performed considering the neutral forms of the monoisotopic oxygenated compositions. The number of molecules used for simulation per sample are as follows; Lem-P 400: 2375, Lem-P 500: 2573, Or-P 500: 2340, and Ol-P 500: 2062.

Different visualization tools are available and the results are downloadable *via* the VKSim interface. Examples include: density plots to visualise the distribution of molecules according to the H/C and O/C ratios in van Krevelen plots, bar charts containing the information of the reactive and non-reactive molecules after performing a particular reaction, the molecular class distribution after each reaction iteration, and violin plots where the distribution of the molecules in terms of H/C ratio of selected heteroatomic classes can be plotted (*e.g.* hydrocarbons as HC class, compositions with one oxygen atom as O_1, *etc.*). Additionally, consecutive reactions can be visualised as animations using the graphic interchange format (.gif).

In the current version, a file containing the information of the neutral molecular compositions (C, H and O) can be uploaded for simulation (see details in the ESI†). Therefore, the software can be used to simulate the reaction steps across multiple molecular compositions within a starting mixture, as obtained from any analytical technique that can provide elemental formulae. A link to the software can be found in the supplementary data.

Discussion

Petroleum-derived liquid hydrocarbon possesses high energy density and optimum combustion characteristics, ideal for the current transportation system. In contrast, the high oxygen and the consequent low hydrogen content of bio-oils negatively affects their energy density.[31] In order to achieve a sustainable production of biomass-derived fuels and chemicals, deoxygenation accompanied by an increased H/C ratio are necessary. The number of potential routes for transforming bio-based chemicals into fuel and useful chemicals is colossal, however.[32] For instance, more than 80 chemicals have been reported from the transformation of furfural[32] (hereafter FUR, $C_5H_4O_2$, furan-2-carbaldehyde).

Reaction pathways to the production of different chemicals can be illustrated in van Krevelen diagrams. For instance, different pathways for the production of high-valuable chemicals from FUR are shown in Fig. 1-left (adapted from Mariscal *et al.*,[32] Sun *et al.*,[33] Runnebaum *et al.*,[34] and Sitthisa *et al.*,[35] and Bond *et al.*[36]). Pathway 1 is explained as follows: furfuryl alcohol (FOL), can be synthesised *via* catalytic hydrogenation of FUR, which then can be transformed into levulinic acid (LA) *via* acid-catalysed ring opening in H_2O (+H_2O). LA is then converted to γ-valerocaltone (GVL) by hydrogenation over a RuSn/
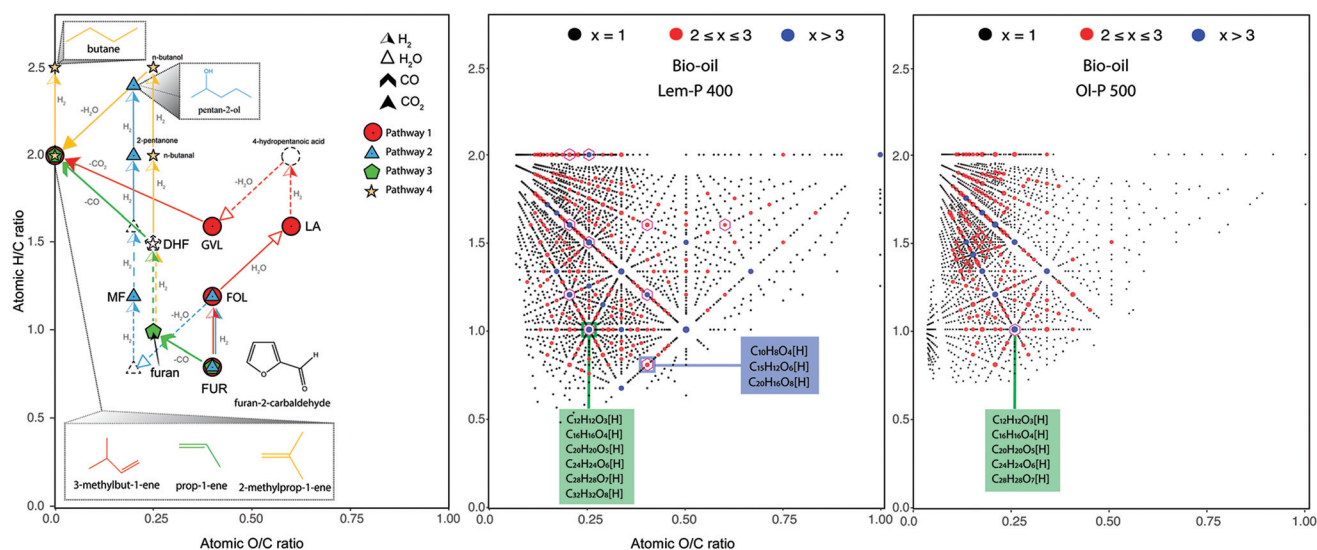

Fig. 1 Left: Reaction pathways for the production of chemicals from furan-2-carbaldehyde (FUR, $C_5H_4O_2$). Pathway 1 shows the production of 3-methyl-1-ene (C_5H_{10}), pathway 2 shows the production of pentan-2-ol ($C_5H_{12}O$), pathway 3 the formation of prop-1-ene (C_3H_6) and pathway 4 the formation of 2-methylprop-1-ene (C_4H_8) (adapted from Mariscal *et al.*,[32] Sun *et al.*,[33] Runnebaum *et al.*,[34] Sitthisa *et al.*,[35] and Bond *et al.*[36]). Right: van Krevelen diagrams of the oxygenated bio-oil compositions detected in the bio-oil Lem-P 400 and Ol-P 500 samples (2375 and 2062 molecular compositions, respectively). In the figure, *x* represents the number of compositions present at each point, coloured as such: in black only one molecular composition, red points containing 2 or 3 formulae and in blue points with more than 3 compositions. A hexagon denotes the location within the van Krevelen diagram of various precursor chemicals present in the reaction pathways shown in the left-hand side figure.

C catalyst which then, undergoes decarboxylation (–CO_2) over SiO_2/Al_2O_3 to form butene isomers and CO_2. A different pathway can be used for the production of 2-methyl furan (MF) *via* hydrogenolysis of the C–OH bond (pathway 2). Further hydrogenation of MF can be performed for the production of 2-pentanone and pentan-2-ol. Supercritical CO_2 in the absence of H_2 has also been used as a reaction medium for the decarbonylation (–CO) of FUR to furan (pathway 3). Under the high H_2 pressures achieved in this process, the furan ring is hydrogenated to yield dihydrofuran (DHF, intermediate). The formation of propane and propene can be explained by a sequence of hydrogenations, hydrogenolysis and decarbonylation reactions. Finally, the conversion of FUR in H_2 over mono metallic Ni catalysts (pathway 4) have shown the production of FOL, furan and C4 products which include butanal, butanol and butane from the ring opening reaction *via* C–O hydrogenolysis of the furan ring.

As shown in Fig. 1, many products can be obtained from the same chemical depending on the reaction conditions and the nature of the catalyst used. The main drawback to the production of chemicals under catalytic reactions include the inevitable production of reaction products *via* undesired reactions.[32] The reactions as shown in Fig. 1-left, present the complexity of the multiple reactions that can undergo from a single chemical (*e.g.* FUR). Upgrading of bio-oils to produce for instance, branched alkanes for inclusion in aviation fuel blends involves the simultaneous reactivity of thousands of compositions. In Fig. 1-right it is shown the van Krevelen diagram (H/C *versus* O/C atomic ratio) of the 2375 and 2062 molecular compositions detected within the oxygenated species of the bio-oil from lemon peel and olive pulp, respectively. The following discussion aims to illustrate how van Krevelen diagrams can be used and interpreted to predict possible reaction routes for the production of various chemicals from bio-oil upgrading.

Processing reactions in van Krevelen diagrams

In a van Krevelen diagram each molecular formula is represented as a single point with the coordinate (O/C, H/C), therefore thousands of molecular compositions can be plotted in a single van Krevelen diagram. To simplify the discussions and analysis of processing reactions, we consider only processing reactions performed on oxygen-containing classes ($O_o[H]$) since oxygen concentration is vital to assess the bio-oil quality and its potential for catalytic upgrading to obtain added value fuels and chemicals.

It is well known, that although each molecular species is plotted as a single coordinate in a van Krevelen diagram, each point in a van Krevelen diagram can contain multiple data points corresponding to different molecular compositions with the same H/C and O/C ratio. An example of this is shown in Fig. 1-right and Fig. S5 in the ESI.† The point sizes and colour coding in Fig. 1-right indicates H/C *vs.* O/C coordinates that contain one or multiple molecular composition.

As discussed by Kim *et al.*,[22] the data is distributed in the form of a pattern as a consequence of restraints in the C, H,

and O atoms in the molecular assignments. Consider for instance the elemental composition expressed in terms of the well-known "hydrogen deficiency", Z:

$$C_c H_{2c+Z} O_o \tag{3}$$

where c and o are the carbon and oxygen atoms within the molecule. The hydrogen deficiency of molecules without nitrogen atoms is defined as:

$$Z = -2(DBE) + 2 \tag{4}$$

With Z values of 2, 0 or even-values, it is clear that the number of hydrogen atoms in the molecules are only even numbers. It is interesting to note that the density of the points in the van Krevelen diagrams differ within the different samples (see Fig. 1 and Fig. S5†), indicating particular chemical profiles within the samples.

Loss or gain of elements in specific molecular ratios such as H_2O, H_2, CO_2, CO, and CH_4 shift the molecular compositions in specific directions in van Krevelen diagrams.[19,22] The reaction lines as defined by van Krevelen in 1950,[19] can be demonstrated by mathematical calculations as revisited in section 3 in the ESI.† Briefly, a line drawn between any two of points in a van Krevelen diagram can be described by the following general formula:

$$\frac{H}{C} = a \frac{O}{C} + b \tag{5}$$

where *a* is the slope of the linear equation and *b* is the intercept with the coordinate. With thousands of points in van Krevelen diagrams, multiple series of reaction lines can be plotted.[37] Some related examples are shown in Fig. 2 and a summary of the general equation of the processing reaction lines can be found in Table S1.† As shown in Fig. 2, the molecules in each line exhibit a characteristic molecular formula that constrains the oxygen content in terms of the c atom number or the hydrogen deficiency of the molecule. Consider for instance the general formulae of a molecule in the decarboxylation line 4 in Fig. 2 (more details can be found in section 3 in the ESI†):

$$C_c H_{2c+Z} O_{-2(c+Z)} \tag{6}$$

According to eqn (6), the oxygen-content of the molecules detected in this line is exactly −2(c + Z) (see related examples in Table S1 in the ESI†). It is important to note that some molecules can be found along multiple reaction lines. As shown in Fig. 2, a molecular formula such as $C_{10}H_8O_4[H]$ is located in an intersection between line 4 and 2, thus, the oxygen-content of this molecule is in addition constrained by (2c + Z)/2 and, therefore Z = −6c/5 which implies that molecules in this coordinate must have a carbon number multiple of five (*e.g.* $C_{15}H_{12}O_6[H]$, $C_{20}H_{16}O_8[H]$). In this particular point in the van Krevelen diagram, the elimination of oxygen in form of CO_2 would drive the molecular formula towards H/C = 1. In contrast, the elimination of oxygen in form of water would shift the composition towards H/C = 0 (C_c, carbon-
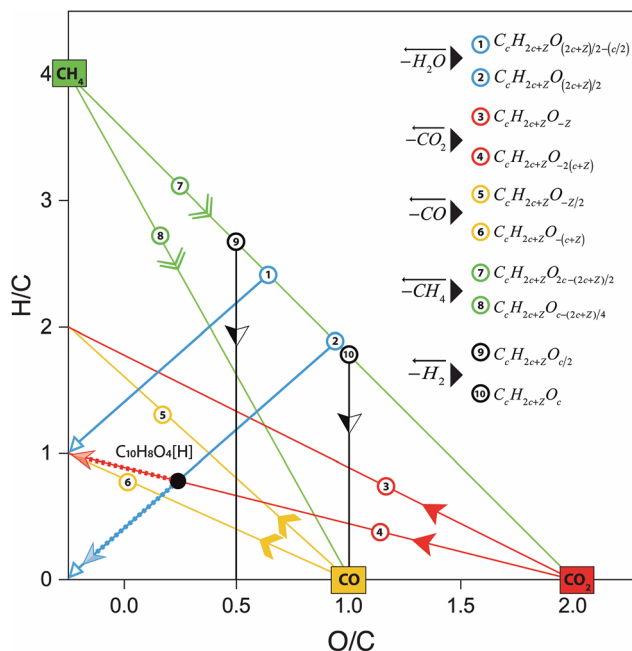
Fig. 2 Example of reaction processing lines in van Krevelen diagrams corresponding to losses of molecules of water, carbon dioxide, carbon monoxide, methane, and hydrogen. The characteristic molecular formulae of molecules in each line are shown on the right corner and the CO, CO_2, and CH_4 poles are shown along the axis, $2c + Z$ correspond to the total hydrogen atoms within the molecule.

aceous material). Thus, a selective reaction process can displace the molecules towards specific products. This information is extremely valuable to design the bio-oil downstream upgrading strategy. For instance, oxygen removal may happen *via* decarboxylation or hydrodeoxygenation (HDO) reactions as reported elsewhere.[38] Given the importance of maintaining a high H/C in the upgraded bio-fuel, removal of oxygen as water, in other words, *via* HDO routes is the preferred option and hence the catalytic processes should be selectively oriented towards this pathway. The particular trendlines observed in van Krevelen diagrams are then a consequence of the constraints of the molecular formulae.[22] Molecular constraints as the ones shown in Fig. 2 and Table S1† can potentially be used in molecular assignment algorithms.

It is important to notice from eqn (7) that the H/C ratio of each alkylation series is distributed along the van Krevelen diagram according to the carbon number distribution, thus for $Z \leq 0$, H/C $\rightarrow$ 2 when the carbon number increases. As expected, molecular species with higher hydrogen deficiency are located at lower H/C ratios (see Fig. S6 and S7†).

From the mathematical point of view, a molecule $C_cH_{2c+Z}O_o$ can potentially lose oxygen in form of H_2O, CO, or CO_2. It is also possible to increase the H/C content by the addition of H_2 or CH_4 molecules. As shown in Fig. 2, those addition or losses of non-condensable gases are characterised by a direction in the van Krevelen diagrams or "poles". Thus, the addition of CO, CO_2, and CH_4 shift the compositions towards the CO (O/C = 1), CO_2 (O/C = 2) and CH_4 (H/C = 4) pole respectively (see

Fig. 2). As we demonstrate below, reaction processes can be additionally characterised by a magnitude that depends on the carbon content.

The initial (i) elemental $(H/C)_i$ and $(O/C)_i$ ratio of a molecule described by the eqn (3) can be written as:

$$\left(\frac{H}{C}\right)_i = \frac{2c + Z}{c} = 2 + \frac{Z}{c} \tag{7}$$

$$\left(\frac{O}{C}\right)_i = \frac{o}{c} \tag{8}$$

When the molecule loses a molecule of water for instance, the final ratios, $(H/C)_f$ and $(O/C)_f$, are given by eqn (9) and (10), respectively.

$$\left(\frac{H}{C}\right)_f = \frac{2c + Z - 2}{c} = \left(2 + \frac{Z}{c}\right) - \frac{2}{c} \tag{9}$$

$$\left(\frac{O}{C}\right)_f = \frac{o - 1}{c} = \frac{o}{c} - \frac{1}{c} \tag{10}$$

Thus, the displacement of the H/C and O/C coordinates in a van Krevelen diagram of a molecule that loses oxygen in form of water are given by:

$$\Delta\frac{H}{C} = \left(\frac{H}{C}\right)_f - \left(\frac{H}{C}\right)_i = -\frac{2}{c} \tag{11}$$

$$\Delta\frac{O}{C} = \left(\frac{O}{C}\right)_f - \left(\frac{O}{C}\right)_i = -\frac{1}{c} \tag{12}$$

In a similar way, it can be calculated the magnitude along the coordinates H/C and O/C of the different reaction processes. A summary of those equations is shown in Table 2.

Thus, the $\Delta(H/C)$, under losses of CO or CO_2, is higher for compositions with low carbon number and high H/C-value ($2 + Z/c$). In contrast, the $\Delta(H/C)$ under dehydration/dehydrogenation depends uniquely in the carbon content. A displacement towards lower O/C can be also achieved by losses of H_2O, CO, and CO_2, as well as by CH_4 losses. Decarbonylation, de-

Table 2 General equations of the magnitude of the displacement along the axis H/C and O/C corresponding to losses of molecules of H_2O, H_2, CH_4, CO, and CO_2 of a molecule with a general chemical formula $C_cH_{2n+Z}O_o$

Principal reactions	Reaction vector (c,h,o)	Δ(O/C)	Δ(H/C)
Dehydration ($-H_2O$)	$(0,-2,-1)$	$-\frac{1}{c}$	$-\frac{2}{c}$
Dehydrogenation ($-H_2$)	$(0,-2,0)$	0	$-\frac{2}{c}$
Decarbonylation ($-CO$)	$(-1,0,-1)$	$\frac{1}{c-1}\left(\frac{o}{c}-1\right)$	$\frac{1}{c-1}\left(2+\frac{Z}{c}\right)$
Decarboxylation ($-CO_2$)	$(-1,0,-2)$	$\frac{1}{c-1}\left(\frac{o}{c}-2\right)$	$\frac{1}{c-1}\left(2+\frac{Z}{c}\right)$
Demethanation ($-CH_4$)	$(-1,-4,0)$	$\frac{1}{c-1}\left(\frac{o}{c}\right)$	$\frac{1}{c-1}\left[\left(2+\frac{Z}{c}\right)-4\right]$

carboxylation, and demethanation processes reduce the O/C value as a consequence of losses of either carbon or oxygen atoms in the molecule with an inverse dependency with the carbon number.

The displacement of the molecules in van Krevelen diagrams can indicate the type of hydrocarbon that can be potentially produced upon upgrading. According to van Krevelen,[19] there exist a well-defined relationship between the type of hydrocarbon (*e.g.* paraffins, cycloparaffins, and aromatics), the size of the molecule measured by the total carbon number, and the H/C ratio (see Fig. 3). Condensed aromatic structures as defined by Koch *et al.*[39,40] have an aromaticity index (AI) greater or equal to 0.67 which correspond to the H/C value of molecules with a general formula $C_{6c}H_{4c+2}$ when c is infinite (H/C = 0.67 + 2/c). This can be seen in the atomic H/C ratio *vs.* 1/c plot shown in Fig. 3. Similarly, a molecule containing one ring (naphthene/cycloparaffin, see Fig. 3(a)) with a molecular formula $C_{6c}H_{10c+2}$ has an H/C = 1.67. Thus, appropriate reaction processes can potentially drive the compositions towards a specific compositional space in the van Krevelen diagram.

Different reaction mechanisms have been suggested for model compounds present in bio-oil and has been revisited in different reviews.[17,41,42] Phenol hydrogenation has been shown as a reaction route for the formation of cyclohexanol and further hydrocracking of this composition can produce small paraffinic hydrocarbons. Hydrogenation of a phenol followed by a loss of H_2O can be used to produce benzene. Also, catalytic deoxygenation of fatty acids has been proposed *via* decarboxylation/decarbonylation reactions.[38] These reactions can be plotted in a van Krevelen diagram as shown in Fig. S8.† It is important to note that the mechanisms proposed in literature for the addition/losses of H_2, H_2O, CO, CO_2, and CH_4 for these standard molecules follow the reaction processing vectors shown in this paper (see Fig. 1 and Fig. S8†). The later highlights the power of this advanced characterisation study which allows a deep understanding of the molecular structure of fairly complex bio-oil serving a toolkit to design, predict and guide further conversion/upgrading reactions.

Results

Bio-oils must be treated extensively before they can be used as final fuel products.[43] The production of a hydrocarbon-like fuel from a bio-oil involves the removal of oxygen, the reduction of the average molecular weight, and increasing the ratio of atomic hydrogen to carbon.[8] Two main reaction processes have been extensively studied for the upgrading of bio-oils: hydrodeoxygenation (HDO) and catalytical cracking.[8,44,45] HDO eliminates oxygen primarily *via* loss of water through sequential hydrogenation, and dehydrogenation reactions. On the other hand, catalytic cracking accomplishes deoxygenation through simultaneous losses of carbon oxides (primary CO_2) and water.[46] Although the deoxygenation of bio-oils have been achieved by catalytical cracking and HDO, the resulting hydrocarbon products still significantly differ from petroleum-like fuels.[47]

As shown in Fig. 1 and Fig. S5,† pyrolyzed bio-oils are a mixture of thousands of highly oxygenated compounds with unique chemical profiles. The upgrading process then involves the simultaneous reactions of thousands of elemental compositions. As presented in the previous section, a detailed compositional map can be used to understand reactions when applied to complex mixtures.

VKSim was developed as a visualization tool to simulate reactions involving losses or additions of small molecules
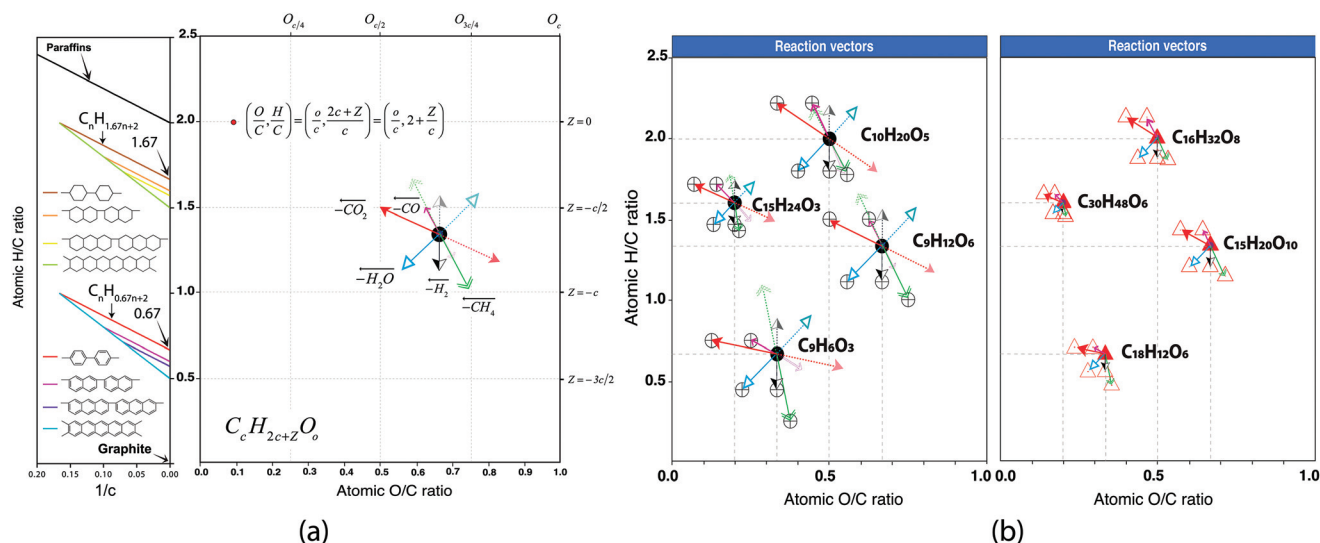

Fig. 3 (a): Right: processing reaction vectors represented in van Krevelen diagrams for a molecule with a general formula $C_cH_{2c+Z}O_o$ where c and o are the carbon and oxygen atoms, respectively and Z is the hydrogen deficiency. (a) Left: atomic H/C ratio of hydrocarbons *versus* 1/c (adapted from the original van Krevelen paper from 1950 19). (b) Reaction vectors for example molecular compositions detected in Or-P 500. The species on b-left and b-right, represent molecules with the same H/C and O/C value but with differing molecular formulas (higher carbon number on the right).

given a starting material, such as from the detailed molecular composition of the bio-oils. In the following discussion, VKSim was used to simulate two different reactions pathways:

-Pathway 1: six sequential sets of a H_2 addition and a H_2O loss to allow the reduction of up to six oxygen atoms. This simulates an HDO process.

-Pathway 2: a loss of six oxygen atoms was simulated by a loss of two oxygen atoms *via* decarboxylation ($-CO_2$) followed by 4 sequential water losses ($-4H_2O$). Water is the dominant mechanism in catalytic cracking; therefore, this reaction aims to simulate a catalytic reaction process.

The proposed pathways correspond to representative reactions and are helpful to understand the effect of hydrogenation, dehydration, and decarboxylation reactions on complex mixtures. The following considerations were taken for the evaluation of the hydrocarbons and oxygen-containing species produced after the simulated reactions:

-The H/C ratio can be used to indicate the type of hydrocarbons produced after each reaction (see Fig. 3). In order to do this, the hydrocarbons produced under deoxygenation reactions were categorised in four main groups that indicates the main structural character of the composition: condensed aromatics (H/C ratio ≤0.67), cycloalkane-like character (0.67 < H/C ≤ 1.67), olefins-type character 1.67 < H/C ≤ 2 and paraffins (H/C > 2). It is important to note that some uncertainty is involved in classifying hydrocarbon structures according to the H/C-value alone. A qualitative structural analysis by means of spectroscopic techniques, MS/MS, or GC × GC would be recommended for more detailed insights into the structures.

-Without the use of chromatography or fragmentation experiments, a traditional mass spectrum cannot provide information regarding the oxygen functionality, and therefore, it is not possible to draw definitive conclusions from the H/C–O/C − 1/n diagram.[19] It is however, possible to determine saturated acyclic ethers and alcohols in van Krevelen diagrams (H/C > 2 and O/C > 0, see Fig. S11†).

Simulation of reaction routes

Van Krevelen diagrams can be complex and difficult to understand when thousands of compositions are plotted simultaneously. Therefore, a simplified van Krevelen diagram is presented in Fig. 4 to illustrate the effects of reaction pathways upon a limited number of compositions (a .csv file with the information of the molecular formula of the compositions in Fig. 4 is included in the ESI†). In this Figure, 6 of the 8 the molecules shown in Fig. 3 are selected for reactions *via* Pathways 1 and 2. The molecules can be divided in 3 pairs of compositions, where each pair is characterised by the same H/C and O/C-values but each molecule has a unique molecular formula. One molecule of each pair contains a higher number of carbon and oxygen atoms. The findings can be summarised as follows:

Pathway 1

-The displacements along the H/C axis upon hydrogenation is inversely proportional to 1/C (see Table 2). This implies that molecules with twice the carbon number require twice the
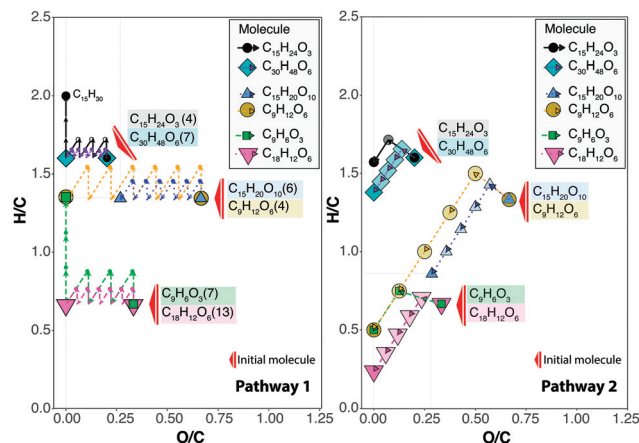

Fig. 4 van Krevelen plots showing the reaction Pathways 1 and 2, applied to three pairs of molecules. Each pair has the same H/C and O/C value but each molecule has a unique molecular formula. The index of hydrogen deficiency is shown in brackets. It can be seen that there are different end points, despite shared H/C and O/C ratios, depending upon each molecular formula.

number of H_2 in order to reach the same H/C-value in a hydrogenation reaction. For instance, the displacement of the molecules $C_{15}H_{24}O_3$ and $C_{30}H_{48}O_6$ from H/C = 1.6 to an H/C = 2 requires $3H_2$ and $6H_2$ respectively.

-Partial deoxygenation (final O/C > 0) was observed for the molecules with an oxygen content higher than 6 (*e.g.* $C_{15}H_{20}O_{10}$), which is directly related with the reduction of up to 6 oxygen atoms in the form of water *via* this pathway. Therefore, additional consecutive reactions involving H_2/H_2O need to be considered for a complete deoxygenation of species with increased oxygen content.

-A total dehydroxygenation followed by hydrogenation is achieved for molecules with low oxygen content. Therefore, for species of low oxygen content, additional dehydroxygenation is not required, leading to potential saving for H_2 usage.

-The hydrogen deficiency (defined in eqn (4)) plays an important role when trying to account the number of molecules of H_2 required to obtain the corresponding saturated molecular composition. This number is traditionally known as the index of hydrogen deficiency (IHD)[48] and is defined as follows:

$$IHD = (2c + 2 - h)/2 \qquad (13)$$

As can be seen in Fig. 4, the molecule $C_{15}H_{24}O_3$ (IHD 4), requires $3H_2/-H_2O$ reactions for a complete deoxygenation followed by the addition of $4H_2$ to obtain a saturated hydrocarbon ($C_{15}H_{32}$). Therefore, the molecule $C_{15}H_{24}O_3$ following the pathway 1, requires an addition of $1H_2$ to reach its saturation limit. Considering that molecules with DBE < 0 (*Z* > 2) are not allowed for a molecular formula, the addition of two or more H_2 will no longer affect the H/C-value of this molecule. In contrast, the molecules $C_{18}H_{12}O_6$, with an IHD of 13, requires a total of $19H_2$ molecules for the production of a saturated hydrocarbon. This highlights that there is an upper limit

of H_2 uptake that depends on the hydrogen deficiency of the molecule, after this point, the molecule no longer reacts to hydrogenation reactions. The movie included in the ESI† helps to visualise the upper limit of H_2 uptake of molecules with different IHD-values and oxygen content. Given the high pressures used in hydrogenation, it is possible that secondary reactions such as cracking and isomerisation start to play a more important role after the saturation limit is reached.

Pathway 2

-The reduction of water in absence of an H-donor, clearly reduces the H/C-value of the molecules. Therefore, a greater aromatic character of the molecules can be observed following this reaction.

-Given that on dehydrogenation the $\Delta(H/C) = 2/c$, the molecules with the lowest carbon number are displaced sharply to lower H/C-values with the dehydration reaction. Therefore, the molecules with the lowest carbon number are prone to produce char upon dehydrogenation reactions.

-The decarboxylation vector shifted the compositions with lower carbon number to a higher H/C-value in comparison with a molecule located in the same initial point but with higher carbon number.

-Decarboxylation requires the presence of at least 2 oxygen atoms (such as a carboxylic acid group). Therefore, compositions with, for instance, only 3 oxygen atoms can only lose one CO_2 molecule.

-Molecules with lower oxygen content (*e.g.* $C_{15}H_{24}O_3$ and $C_9H_6O_3$) can only lose one CO_2 molecule and one H_2O molecule. A further reduction of water would drive the compositions to O < 0, which is not a valid molecular formula. In this case, the product reported by VKSim correspond to the hydrocarbon after the loss of the total oxygen content.

The effect of reaction routes on bio-oil compositions

A simulation of the reaction Pathways 1 and 2 was performed with the full set of bio-oil compositions. As an example, the application of Pathway 2 to the Or-P 500 sample is shown in Fig. 5. The compositional space of the bio-oils after each iteration (reaction vector) can be visualised using van Krevelen diagrams (see animation in ESI†). Additionally, the density of molecules along the H/C and O/C axes allows a quick visual inspection of the hydrogen and oxygen content. In general, a higher H/C along with a low O/C are desired. Finally, the heteroatomic class distribution allows the visualisation of the number of hydrocarbon molecules (class HC) and oxygen-containing species after each reaction vector is applied. The reactions Pathways 1 and 2 for all bio-oil samples can be seen in Fig. S12 and S13.†

As shown in Fig. 4, 5, S12 and S13,† both the class distribution and the distribution of molecules along the O/C axis indicates a partial deoxygenation of the bio-oils and a concentration of hydrocarbons widely dispersed along the H/C axis after each iteration. In both reaction pathways, a major production of hydrocarbons is observed for the bio-oil produced from olive pulp (~1700 compositions) followed by the bio-oil obtained from lemon peel at 500 °C (~1490 compositions). In contrast, the bio-
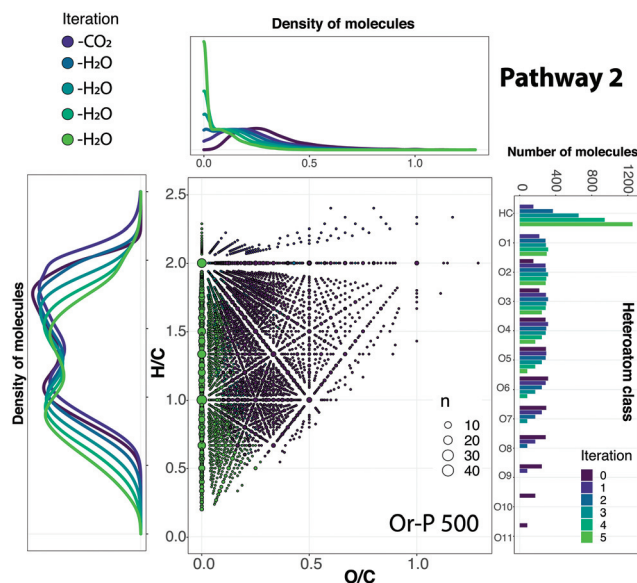

Fig. 5 van Krevelen diagram of Or-P 500 showing the steps comprising Pathway 2, involving one loss of CO_2 and four sequential losses of water molecules, illustrating the changes in composition of a complex mixture during upgrading. The density of molecules distributed along the H/C and O/C ratio after each iteration and the class distribution are also shown. Reaction vector $(-1,0,-2) \rightarrow (0,-2,-1) \rightarrow (0,-2,-1) \rightarrow (0,-2,-1) \rightarrow (0,-2,-1)$. An animation of the simulation of the molecules in a van Krevelen diagram after each iteration can be found in the ESI.†

oil obtained from the pyrolysis of lemon peel at 400 °C and Or-P 500 seems to have the lowest potential to produce hydrocarbons (~1300 and ~1200 compositions, respectively).

Given the combination of the reduced number of oxygenated species and the molecular constrains applied for a molecule to perform a reaction (see Materials and Methods section), our mathematical model shows that the degree of deoxygenation decreases with the number of iterations (see Fig. S14†). It is interesting to note that only 34% of the molecules in the bio-oil Ol-P 500 reacted in the last iteration, which can be an indication of the reduced need of hydrogenation/dehydrogenation reactions in this oil for a similar deoxygenation degree. Future studies, currently outside the scope of this paper, can explore the potential relation between the degree of deoxygenation and the observation of secondary reactions such as cracking, and to evaluate optimum removal of oxygen.

A comparison of the hydrocarbon types produced from Pathways 1 and 2 is shown in Fig. 6. In general, the violin plots show that a reaction pathway simulating the consecutive addition of H_2, and reduction of H_2O clearly yields a hydrocarbon with higher H/C-value in comparison with Pathway 2. In contrast, a reaction following Pathway 2 produces a higher content of condensed aromatics. In both reactions, naphthenes and olefin-type of molecules are produced with the highest yield. These simulations then indicate that in general, HDO is a preferred route for the production of a higher quality of hydrocarbons, which is in good agreement with previous literature.[8,44]
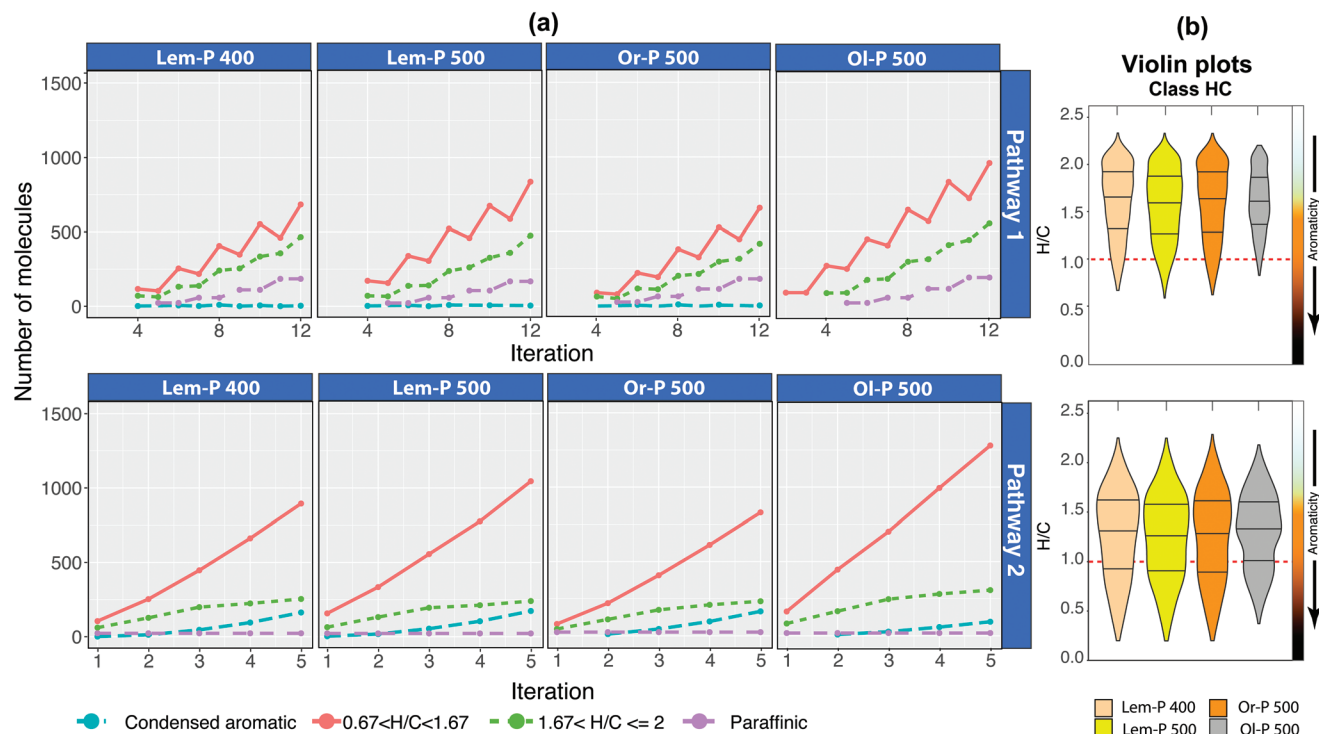
Fig. 6 A comparison of the hydrocarbons-type obtained per sample when two different reaction pathways are applied. (a) The hydrocarbons-type are presented per iteration. Pathway 1 corresponds to the elimination of 6 oxygens by consecutive $H_2/-H_2O$ reactions and pathway 2 corresponds to the loss of 6 oxygen atoms by losses of a molecules of CO_2 and 4 losses of water. (b) Violin plots showing the H/C-value distribution of the hydrocarbons produced in the last iteration. The line at H/C = 1 highlights the point where a benzene ring can be found.

Some of the reaction products correspond to saturated acyclic ethers and alcohols (H/C > 2 and O/C > 0). A higher yield of this type of molecule was observed under $H_2/-H_2O$ reactions (see Fig. S11 and S12†). The production of alcohols has also been reported in literature,[45] and correspond to the products of hydrogenation of carboxylic acids and aldehydes.

The molecular profile of the bio-oils obtained from the citrus wastes are very similar (Lem-P and Or-P 500), while a more distinctive chemical profile is observed in the bio-oil obtained from olive pulp (Fig. S5 and S14†). As can be seen in Fig. 6, a more distinctive hydrocarbon profile is observed for Ol-P 500 sample. For instance, a higher number of molecules categorised as cycloalkane and olefin-type was observed in both reaction pathways. Therefore, the hydrocarbon molecules produced from the olive pulp can potentially produce hydrocarbons with higher H/C ratio with a lower production of condensed aromatics. Additionally, a higher deoxygenation of the bio-oil from olive pulp was observed (see class distribution in Fig. S12 and S13†).

The reaction simulated in Pathway 2, is considered a reaction where oxygen atoms are reduced primarily by the losses of water molecules. It is interesting to note that different ratios of $-CO_2/-H_2O$ have a noticeable effect on the hydrocarbon type. Consider for instance, a reaction pathway 2a, 2b and 2c as described in Table 3. As can be seen in Table 3, the removal of 6 oxygen atoms purely by decarboxylation (reaction 2c) produces a higher number of saturated hydrocarbons along with a

Table 3 Evaluation of the hydrocarbon types produced using three reaction routes with differing loss ratios of CO_2/H_2O. Up to six oxygen atoms can be removed in each pathway. The simulation is performed in Or-P 500

Hydrocarbon type	Number of molecules		
Reaction pathway	2a $-CO_2/-4H_2O$	2b $-2CO_2/-2H_2O$	2c $-3CO_2$
Condensed aromatic	165	48	3
Cycloalkane	830	746	376
Olefinic	233	380	225
Paraffinic	25	52	80
Total HC	**1253**	**1226**	**684**
Total $O_1/O_{0>1}$	299/788	290/824	754/902

minimal production of condensed aromatics, but it is, however, the reaction route that produces the lowest number of hydrocarbons (~700 molecules). As discussed previously, decarboxylation requires a minimum of two oxygen atoms within the molecule, therefore, the reduction of the polyoxygenated species to mainly HC and O_1, as observed in a reaction pathway (c), limits the number of molecules that can react to produce hydrocarbons (see Table 3). Additionally, as can be seen in Table 3, a deoxygenation dominated by losses of water (2a) can potentially produce a higher number of condensed aromatic species. This is a consequence of the reduction of the number of hydrogen atoms in the molecule. The carbon

efficiency for converting carbon in the feedstock to carbon in the bio-oil is lower however, in a reaction following the pathway 2.[8]

Effect of the molecular diversity in bio-oil upgrading

An interesting bimodal distribution of the molecules along the H/C-value was observed after the simulation of the reaction pathways 1 and 2 (see Fig. 5, S12 and S13†). The detailed compositional map of the molecular species in the van Krevelen diagrams can explain this particular trend.

Fig. 7 demonstrates how upgrading strategies may be further optimised if specific groups of molecules can be targeted for different reaction routes. As shown in Fig. 7(a) most of the molecular compositions of the bio-oil obtained from orange peel are distributed in three main zones. The molecules in zone 1 (652 molecules) are characterised by a higher H/C-value (lower DBE values, see Fig. S6†), a higher carbon number (see Fig. S7†), and oxygen content shifted to lower

oxygen values (lower O/C, see Fig. S8†). In contrast, a higher number of molecules were detected in zone 2 (1283) and are characterised by their lower H/C-value (high hydrogen deficiency). An additional 402 molecules were detected at higher O/C and high H/C-value (zone 3). The molecules in zone 3 are then characterised by their high oxygen content. A summary of the main molecular differences between these zones is shown in Fig. S16.†

A further separation of the molecules in zone 2 by mass value indicates that molecules with a mass <400 Da contain lower oxygen content than the molecules with a mass >400 Da. The different characteristics of the compositions in these zones is believed to be the primary reason for the bimodal distribution of the molecules produced by the reactions.

Considering the reaction process vectors and the previous discussion, it is important to gain deeper insight into the effect of the reactions upon the different zones of the bio-oil, which are in turn, unique for each sample (see Fig S15†). In
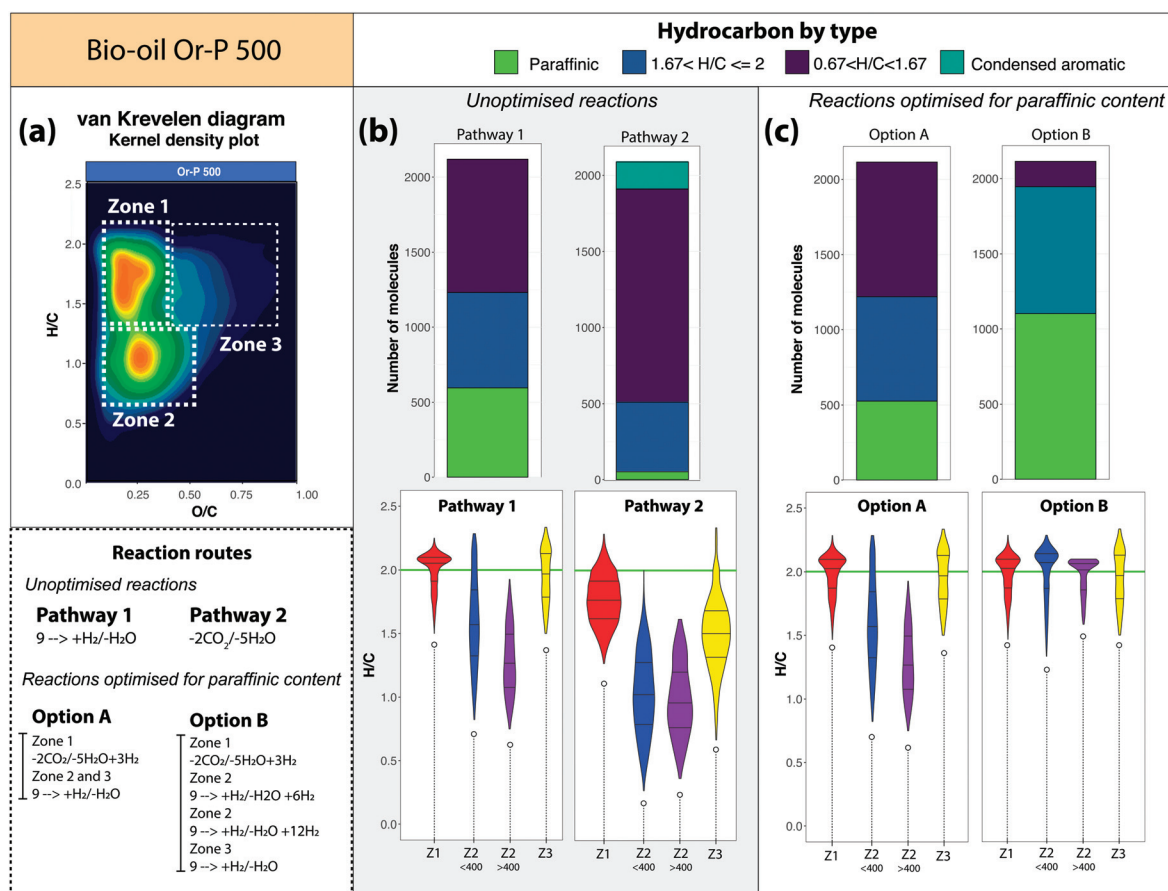

Fig. 7 Comparison of the starting material and the hydrocarbon distribution by type obtained after the simulation of different reaction pathways. (a) van Krevelen diagrams of the oxygenated species of the bio-oil obtained from orange pulp plotted as Kernel density plot. (b) and (c) – Distributions of the hydrocarbons obtained by the simulation of unoptimised reactions and reactions optimised for paraffinic content, respectively. Violin plots are used to illustrate the distribution of the hydrocarbons in the H/C-value by zone. The reactions were applied to four different fractions of the bio-oil Or-P 500. Zone 1 (Z1) $1.5 \leq$ H/C < 2, 0 < O/C $\leq$ 0.3 (652 molecules), zone 2 (Z2) 0.5 < H/C < 1.5, 0 < O/C < 0.67 (1283 molecules), and zone 3 (Z3) 1.5 < H/C < 2, O/C > 0.3 (402), the fraction of molecules in zone 2 was also divided by mass: <400 Da and >400 Da (697 and 586 molecules respectively). All the reactions were performed to reduce 9 oxygen atoms from the molecular formulae. About 2100 hydrocarbon molecules were obtained after the deoxidation reactions were applied.

order to do so, we first applied the reaction Pathways 1 and 2 in the different zones of the bio-oil Or-P 500, which produced the lowest number of hydrocarbons out of the bio-oils used. Given that the molecular compositions of the bio-oil obtained from the olive pulp can contain up to 11 oxygen atoms, the reaction Pathways 1 and 2 aimed to eliminate up to 9 oxygen atoms. The results are presented in Fig. 7b.

The hydrocarbons produced after the removal of oxygen atoms following Pathways 1 and 2 are very similar to the ones observed for the full data set. In summary, more aromatic-type of compositions are observed after a reaction that removes $2CO_2$ molecules followed by 5 consecutive dehydrogenation reactions.

Despite the high degree of deoxygenation observed in the reaction Pathway 1, our simulations show that after 9 consecutive addition/losses of H_2/H_2O, 52% of the molecules in zone 1 cannot further react *via* the addition of H_2. This implies that at this point, 52% of the molecules have reached their saturation limit.

Due to the reduced oxygen content of molecules in zone 2 mass <400 Da, Pathway 1 allowed the full saturation of some hydrocarbons in this fraction (81 paraffinic compositions) along with a minor production of species containing few oxygen atoms (5.8% O_1–O_2). In contrast, paraffinic species were not produced from reaction Pathway 1 of the molecules in zone 2 with mass >400 Da and 20% of the molecules still contain O_1–O_2 oxygen atoms. This indicates that the molecules located in zone 2 and mass >400 Da need a higher number of hydrodeoxygenation reactions, and an even higher number of hydrogen molecules for the production of paraffinic-olefinic hydrocarbons.

The hydrocarbons obtained from Pathway 2 (see Fig. 7), are more aromatic. It is noticeable for instance, that the reaction pathway in zone 2 (both low and high mass) produced condensed aromatic compositions. Molecules in this zone can then easily form char/aromatics if molecules of water are lost, therefore losses of H_2O without the presence of a hydrogen donor (H_2) is not recommended.

Finally, compositions with a higher oxygen content are located at higher O/C ratio (zone 3). This zone has been described in the literature as the immature zone of oil formation[19,49–51] and corresponds to the diagenesis stage of kerogen evolution. The compositions in this area have probably the lowest oil potential and might be challenging to upgrade. Or-P 500 contain 17% of compositions in this zone in comparison with only 9.5% in the bio-oil obtained from olive pulp, this may indicate a lower potential to produce hydrocarbons from this bio-oil. It is however noticeable that this area has the lowest number of compositions overall which is evidence that the pyrolysis parameters implemented to produce the bio-oils have led to molecules with an increased heating value, thus as compositions with higher H/C and lower O/C-values.[52] A hydrodeoxygenation route applied to the compositions in this zone (zone 3), seems to produce more paraffinic/olefinic type hydrocarbons.

In summary, if a reaction pathway is applied to different fractions of the bio-oil (in this example zone 1, 2 <400 Da, 2 >400, and 3), hydrocarbons with different characteristics can be obtained. The molecules in zone 1 and 3 have the highest potential to produce paraffinic-olefinic hydrocarbons whereas zone 2 is prone to form aromatic-type of compositions.

The application of different reaction routes in different fractions of the bio-oil compositions can then lead to hydrocarbons with higher H/C-value. The compositions classified as zone 2 lies in a particular area of the van Krevelen diagram that is traditionally associated with lignins. Lignins are defined as having 0.7 < H/C < 1.5 and 0.1 < O/C < 0.67.[53] Lignins are known for their high aromatic character and for leading to recalcitrant condensed aromatics upon upgrading.[54] The high aromatic character of the compositions in zone 2 as shown in our simulations is then in-line with previous literature. Interesting approaches such as lignin-first biorefinery has been already suggested in literature as an attractive alternative method for the solvent extraction of lignin from lignocellulosic material in a presence catalysis as a H-donor (typically RANEY® Ni).[55] Given the low H/C-value of this material, the addition of H_2 is crucial for successfully upgrading this material.

The solvent fractionation of bio-oils have been also reported in literature for the separation of fatty acids and resin acids (*n*-hexane soluble, potentially molecules in zone 1), low-molecular mass lignin (water insoluble but dichloromethane soluble) and high-molecular mass lignin (water insoluble but dichloromethane insoluble).[56] However, few studies have reported the detailed molecular composition resulting from the fractionation of bio-oils. Miettinen *et al.*,[57] is perhaps one of the most relevant studies reporting the elemental composition of the oily and aqueous phase of bio-oils obtained from pine wood. As reported by the van Krevelen diagrams in Miettinen *et al.* work, species with different oxygen content and containing two types of lignin material can be separated in the oily and the aqueous phase. Therefore, this fractionation method can potentially separate the compositions located in zone 2. Other fractionation methods reported in literature can be used in order to reduce the diversity of compositions in bio-oils.[58] Future studies for the evaluation of the detailed molecular composition of fractions of bio-oils obtained with different separation strategies are then needed as they can inform further advance upgrading strategies.

Proposed reactions to optimise for paraffinic content

With the potential to separate compositions found within the different zones and subsequently target them for different reactions,[56,58] it becomes possible to further optimise upgrading strategies, based upon knowledge of the composition of individual bio-oils. Two examples of such targeted strategies, named here as Option A and Option B, are described for the bio-oil derived from orange pulp.

Option A: As can be seen in Fig. 7(b), the reaction pathway 2, applied to the molecules in zone 1, produced molecules distributed within a narrow H/C-value (more than 50% of the molecules are located between H/C = 1.5–2). As shown in Fig. 4, this reaction pathway presents a smaller displacement

of the compositions along the H/C-axis upon dehydration. Therefore, we suggest Option A, involving the losses of CO_2 and H_2O, followed by the addition of $3H_2$ (hydrogenation). As shown in the bar chart in Fig. 7(c), this pathway can potentially shift the hydrocarbons towards a higher H/C-value, with a final product similar to the one obtained by Pathway 1 but with a lower requirement of gas-phase molecular hydrogen. The addition of molecules of H_2 is however required if the formation of paraffinic structures is desired. In this case, a hydrogenation catalyst (*i.e.* Ni, Pd, Pt based systems) could help to achieve the upgrading to paraffinic molecules.

Option B: This option considers the case where a highly paraffinic-olefinic type of hydrocarbons are desired. In order to achieve this, it is possible to perform Option A for the molecules in zone 1. On the other hand, hydrodeoxygenation reactions appear to be the best alternative for the species in zone 2 (in particular zone 2 < 400) and zone 3. It is important to consider however, that as shown in previous discussion, species in zone 2 and >400 Da require a higher number of H_2 molecules to produce similar paraffinic/olefinic hydrocarbons. Therefore, in Option B, $6H_2$ and $12H_2$ molecules were added in zone 2 with mass <400 Da and zone 2 with mass >400 Da, respectively. As can be seen in Fig. 7(c), the highest percentage of produced hydrocarbons by reaction option B correspond to paraffins, followed by olefinic and cycloalkane hydrocarbons.

It is known that HDO reactions offer better carbon efficiency but at expense of the addition of H_2 either by the addition of gaseous molecular hydrogen or by using a solvent capable of undergoing a hydrogen transfer reaction.[59] The higher H_2 pressures and temperatures traditionally required for bio-oil upgrading translate to higher processing costs, however. Compositions with a more aromatic-naphthenic character can be achieved from the species in zone 2 and mass >400 Da with lower H_2 additions. Therefore, these species can be a better source of aromatic chemicals.

According to the latest discussions, bio-oil upgrading is challenging as a consequence of the complexity and diversity of compositions that can lead to the formation of distinct products and residues. The data displayed in the van Krevelen diagrams in this study correspond to the molecular compositions as obtained by the direct infusion FTICR MS analysis of the bio-oils. Thus, it is not possible to separate compositions with the same molecular formula but containing different functional groups (structural isomers) which can influence the products upon upgrading. However, several studies conducted on model compounds with different functional groups (*i.e.* alcohols, aldehydes, ketones, acids, and esters) have shown as a general trend that catalytic upgrading of compositions with low effective hydrogen index (defined as $(H/C)_{eff} = (h - 2o)/c$) promote coke formation and the formation of aromatic hydrocarbons.[8] Dehydration has been proposed as dominant mechanism to explain this phenomenon which is in good agreement with the displacement of molecular compositions in van Krevelen diagrams by losses of water molecules. In agreement with our discussions, Vispute *et al.*, has proposed reactions that include hydro-processing followed by subsequent catalytic

upgrading over zeolites to reduce oxygen in form of water of compositions with low hydrogen content.[60]

The combination of the detailed molecular composition of a bio-oil and the reaction pathways simulated and visualised in van Krevelen diagrams as discussed in this paper can give an indication of reaction routes potentially more suitable than others. The reaction processes performed in the samples can displace the compositions towards either paraffinic or more aromatic moieties and therefore, the more efficient reaction route depends upon the chemical profile of the sample and the desired chemical product (*i.e.* chars, bio-oils, gasses). Overall, models based on reaction processes in van Krevelen diagrams can be used to infer upgrading routes that improve the economic potential of biofuels production. In order to reduce the production of undesired products such as carbon formation it is necessary to investigate in greater detail if upgrading yields can be improved; by the separation of compositions with very different reaction routes (such as compositions in zones 1 and 2) prior to upgrading, multistep upgrading methodologies, and pyrolysis parameters that produce more homogeneous chemical profiles of the bio-oils when plotted in van Krevelen diagrams.

Conclusions

A detailed analysis of bio-oils obtained by pyrolysis was performed with ultra-high resolution mass spectrometry. The chemical profile among the bio-oils obtained from citrus wastes was shown to be similar whereas the compositions of the bio-oil obtained from olive pulp contain species with lower oxygen-content. In this paper, we also revisited the reaction processing lines introduced by van Krevelen in 1950 and we proposed mathematical equations to calculate the magnitude of the displacement of the molecules along the axis H/C and O/C upon reaction processes. Considering that reactions such as addition or losses of CO, CO_2, H_2O, and CH_4, have both a direction and a magnitude, we defined the reaction processes as vectors. In general, the displacements along each axis is inversely dependent on the carbon number, thus, the higher the carbon number, the smaller the displacement. Reaction vectors including a reduction of the carbon atoms such as CO, CO_2, and CH_4 also depend on the H/C-value of the molecule, while hydrogenation/hydration reactions depend uniquely to the inverse number of carbons. Finally, reaction processes that can increase the oil potential of the bio-oils upon upgrading were suggested. Software developed in this study, named VKSim, allows the simulation of diverse set of reaction routes as applied to a full bio-oil compositional set or a fraction of its compositions. The software can be used to simulate reactions of the multiple molecular compositions, as determined from any analytical technique that can provide elemental formulae. This approach can then be used by researchers in the field of bio-oil upgrading.

From the results shown in this paper, it is clear that challenges in upgrading and bio-oil production are a consequence

of the wide dispersion of the molecular species that produce a diverse number of products under losses of non-condensable gasses. The reaction vectors proposed here aim to guide researchers in the understanding on the potentially more suitable reaction routes, based on knowledge of individual bio-oil starting materials. Molecular information regarding the volatiles and semi-volatiles, isomeric composition (functional groups), and the molecular composition of the heavier bio-oil material can be used in conjunction with the model. Through more advanced understanding of the detailed molecular compositions of individual bio-oils and of the available reaction pathways, it becomes possible to tailor reaction routes and, in turn, design optimised upgrading strategies for production of desirable compounds.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 P. Basu, *Biomass Gasification and Pyrolysis Handbook*, Burlington, 2010.

2 M. R. Barr, R. Volpe and R. Kandiyoti, *ACS Sustainable Chem. Eng.*, 2019, **7**, 13734–13745.

3 R. Kandiyoti, A. Herod, K. Bartle and T. Morgan, in *Solid Fuels and Heavy Hydrocarbon Liquids (Second Edition)*, ed. R. Kandiyoti, A. Herod, K. Bartle and T. Morgan, Elsevier, 2nd edn, 2017, pp. 25–123.

4 M. J. Antal and M. Grønli, *Ind. Eng. Chem. Res.*, 2003, **42**, 1619–1640.

5 M. S. Mettler, D. G. Vlachos and P. J. Dauenhauer, *Energy Environ. Sci.*, 2012, **5**, 7797–7809.

6 T. J. Morgan and R. Kandiyoti, *Chem. Rev.*, 2014, **114**, 1547–1607.

7 F. X. Collard and J. Blin, *Renew. Sustain. Energy Rev.*, 2014, **38**, 594–608.

8 M. S. Talmadge, R. M. Baldwin, M. J. Biddy, R. L. McCormick, G. T. Beckham, G. A. Ferguson, S. Czernik, K. A. Magrini-Bair, T. D. Foust, P. D. Metelski, C. Hetrick and M. R. Nimlos, *Green Chem.*, 2014, **16**, 407–453.

9 B. Khiari, M. Jeguirim, L. Limousy and S. Bennici, *Renew. Sustain. Energy Rev.*, 2019, **108**, 253–273.

10 S. P. Singh and D. Singh, *Renew. Sustain. Energy Rev.*, 2010, **14**, 200–216.

11 A. E. Atabani, A. S. Silitonga, H. C. Ong, T. M. I. Mahlia, H. H. Masjuki, I. A. Badruddin and H. Fayaz, *Renew. Sustain. Energy Rev.*, 2013, **18**, 211–245.

12 Y. Cho, Y. Qi, P. B. O'Connor, M. P. Barrow and S. Kim, *J. Am. Soc. Mass Spectrom.*, 2014, **25**, 154–157.

13 D. C. Palacio Lozano, R. Gavard, J. P. Arenas-Diaz, M. J. Thomas, D. D. Stranz, E. Mejía-Ospino, A. Guzman, S. E. F. Spencer, D. Rossell and M. P. Barrow, *Chem. Sci.*, 2019, **10**, 6966–6978.

14 R. Gavard, D. C. Palacio Lozano, A. Guzman, D. Rossell, S. E. F. Spencer and M. P. Barrow, *Anal. Chem.*, 2019, **91**, 15130–15137.

15 D. C. Palacio Lozano, C. X. Ramírez, A. J. Chaparro Sarmiento, M. J. Thomas, R. Gavard, H. E. Jones, R. Cabanzo, E. Mejia-ospino and M. P. Barrow, *Fuel*, 2020, **259**, 116085.

16 M. Staš, J. Chudoba, D. Kubicka, J. Blazek and M. Pospíšil, *Energy Fuels*, 2017, **31**, 10283–10299.

17 W. Jin, L. Pastor-Pérez, D. K. Shen, A. Sepúlveda-Escribano, S. Gu and T. Ramirez Reina, *ChemCatChem*, 2019, **11**, 924–960.

18 D. C. Palacio Lozano, M. J. Thomas, H. E. Jones and M. P. Barrow, *Annu. Rev. Anal. Chem.*, 2020, **13**, 405–430.

19 D. W. Van Krevelen, *Fuel*, 1950, **29**, 269–284.

20 A. Stubbins, R. G. M. Spencer, H. Chen, P. G. Hatcher, K. Mopper, P. J. Hernes, V. L. Mwamba, A. M. Mangangu, J. N. Wabakanghanzi and J. Six, *Limnol. Oceanogr.*, 2010, **55**, 1467–1477.

21 S. A. Brockman, E. V. Roden and A. D. Hegeman, *Metabolomics*, 2018, **14**, 1–5.

22 S. Kim, R. W. Kramer and P. G. Hatcher, *Anal. Chem.*, 2003, **75**, 5336–5344.

23 M. Volpe, C. D'Anna, S. Messineo, R. Volpe and A. Messineo, *Sustainability*, 2014, **6**, 7866–7882.

24 D. Prat, A. Wells, J. Hayler, H. Sneddon, C. R. McElroy, S. Abou-Shehada and P. J. Dunn, *Green Chem.*, 2016, **18**, 288–296.

25 R. Volpe, A. Messineo, M. Millan, M. Volpe and R. Kandiyoti, *Energy*, 2015, **82**, 119–127.

26 M. J. Thomas, E. Collinge, M. Witt, D. C. Palacio Lozano, C. H. Vane, V. Moss-Hayes and M. P. Barrow, *Sci. Total Environ.*, 2019, **662**, 852–862.

27 R. Gavard, H. E. Jones, D. C. Palacio Lozano, M. J. Thomas, D. Rossell, S. E. F. Spencer and M. P. Barrow, *Anal. Chem.*, 2020, **92**, 3775–3786.

28 J. R. Conway, A. Lex and N. Gehlenborg, *Bioinformatics*, 2017, **33**, 2938–2940.

29 A. Lex, N. Gehlenborg, H. Strobelt, R. Vuillemot and H. Pfister, *IEEE Trans. Vis. Comput. Graph.*, 2014, **20**, 1983–1992.

30 T. Kind and O. Fiehn, *BMC Bioinf.*, 2007, **8**, 105.

31 J. C. Serrano-Ruiz and J. A. Dumesic, *Energy Environ. Sci.*, 2011, **4**, 83–99.

32 R. Mariscal, P. Maireles-Torres, M. Ojeda, I. Sádaba and M. López Granados, *Energy Environ. Sci.*, 2016, **9**, 1144–1189.

33 C. Sun, Z. Cao, J. Wang, L. Lin and X. Xie, *New J. Chem.*, 2019, **43**, 2567–2574.

34 R. C. Runnebaum, T. Nimmanwudipong, J. Doan, D. E. Block and B. C. Gates, *Catal. Lett.*, 2012, **142**, 664–666.

35 S. Sitthisa, W. An and D. E. Resasco, *J. Catal.*, 2011, **284**, 90–101.

36 J. Q. Bond, A. A. Upadhye, H. Olcay, G. A. Tompsett, J. Jae, R. Xing, D. M. Alonso, D. Wang, T. Zhang, R. Kumar, A. Foster, S. M. Sen, C. T. Maravelias, R. Malina, S. R. H. Barrett, R. Lobo, C. E. Wyman, J. A. Dumesic and G. W. Huber, *Energy Environ. Sci.*, 2014, **7**, 1500–1523.

37 E. Terrell and M. Garcia-perez, DOI: 10.1021/acs.energyfuels.0c01687.

38 W. Jin, L. Pastor-Pérez, J. J. Villora-Pico, M. M. Pastor-Blas, A. Sepúlveda-Escribano, S. Gu, N. D. Charisiou, K. Papageridis, M. A. Goula and T. R. Reina, *Energies*, 2019, **13**, 1–15.

39 B. P. Koch and T. Dittmar, *Rapid Commun. Mass Spectrom.*, 2006, **20**, 926–932.

40 B. P. Koch and T. Dittmar, *Rapid Commun. Mass Spectrom.*, 2016, **30**, 250.

41 T. N. Pham, T. Sooknoi, S. P. Crossley and D. E. Resasco, *ACS Catal.*, 2013, **3**, 2456–2473.

42 E. Santillan-Jimenez and M. Crocker, *J. Chem. Technol. Biotechnol.*, 2012, **87**, 1041–1050.

43 D. M. Alonso, J. Q. Bond and J. A. Dumesic, *Green Chem.*, 2010, **12**, 1493–1513.

44 P. M. Mortensen, J. D. Grunwaldt, P. A. Jensen, K. G. Knudsen and A. D. Jensen, *Appl. Catal., A*, 2011, **407**, 1–19.

45 D. A. Ruddy, J. A. Schaidle, J. R. Ferrell, J. Wang, L. Moens and J. E. Hensley, *Recent advances in heterogeneous catalysts for bio-oil upgrading via 'ex situ catalytic fast pyrolysis': Catalyst development through the study of model compounds*, 2014, vol. 16.

46 L. Ciddor, J. A. Bennett, J. A. Hunns, K. Wilson and A. F. Lee, *J. Chem. Technol. Biotechnol.*, 2015, **90**, 780–795.

47 A. H. Zacher, M. V. Olarte, D. M. Santosa, D. C. Elliott and S. B. Jones, *Green Chem.*, 2014, **16**, 491–515.

48 V. Pellegrin, *J. Chem. Educ.*, 1983, **60**, 626–633.

49 D. W. Van Krevelen, *Org. Geochem.*, 1984, **6**, 1–10.

50 L. Patrick, M. Monthioux and J. D. Meunier, *Org. Geochem.*, 1984, **7**, 249–260.

51 F. Behar and M. Vandenbroucke, *Org. Geochem.*, 1987, **11**, 15–24.

52 M. Ahmad and H. Subawi, *Apex J.*, 2013, **2**, 295–301.

53 W. C. Hockaday, J. M. Purcell, A. G. Marshall, J. A. Baldock and P. G. Hatcher, *Limnol. Oceanogr.: Methods*, 2009, **7**, 81–95.

54 A. G. Gayubo, B. Valle, A. T. Aguayo, M. Olazar and J. Bilbao, *J. Chem. Technol. Biotechnol.*, 2010, **85**, 132–144.

55 R. Rinaldi, R. T. Woodward, P. Ferrini and H. J. E. Rivera, *J. Braz. Chem. Soc.*, 2019, **30**, 479–491.

56 M. Staš, D. Kubička, J. Chudoba and M. Pospíšil, *Energy Fuels*, 2014, **28**, 385–402.

57 I. Miettinen, M. Mäkinen, T. Vilppo and J. Jänis, *Energy Fuels*, 2015, **29**, 1758–1765.

58 Y. H. Chan, S. K. Loh, B. L. F. Chin, C. L. Yiin, B. S. How, K. W. Cheah, M. K. Wong, A. C. M. Loy, Y. L. Gwee, S. L. Y. Lo, S. Yusup and S. S. Lam, *Chem. Eng. J.*, 2020, **397**, 125406.

59 T. Renders, S. Van Den Bosch, S. F. Koelewijn, W. Schutyser and B. F. Sels, *Energy Environ. Sci.*, 2017, **10**, 1551–1557.

60 T. P. Vispute, H. Zhang, A. Sanna, R. Xiao and G. W. Huber, *Science*, 2010, **330**, 1222–1227.