Supplementary Information

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

Diana Catalina Palacio Lozano,[1] Hugh E. Jones,[1,2] Tomas Ramirez Reina,[3] Roberto Volpe,[4] and Mark P. Barrow[1*]

[1]Department of Chemistry, University of Warwick, Coventry, CV4 7AL, United Kingdom
[2]Molecular Analytical Science Centre of Doctoral Training, University of Warwick, Coventry, CV4 7AL, United Kingdom
[3]DICP-Surrey Joint Centre for Future Materials, Department of Chemical and Process Engineering, and Advanced Technology Institute, University of Surrey, Guilford, Surrey GU2 7XH, UK
[4]School of Engineering and Materials Science, Queen Mary University of London, Mile End Road, London E1 4NS, UK

1. Mass spectra

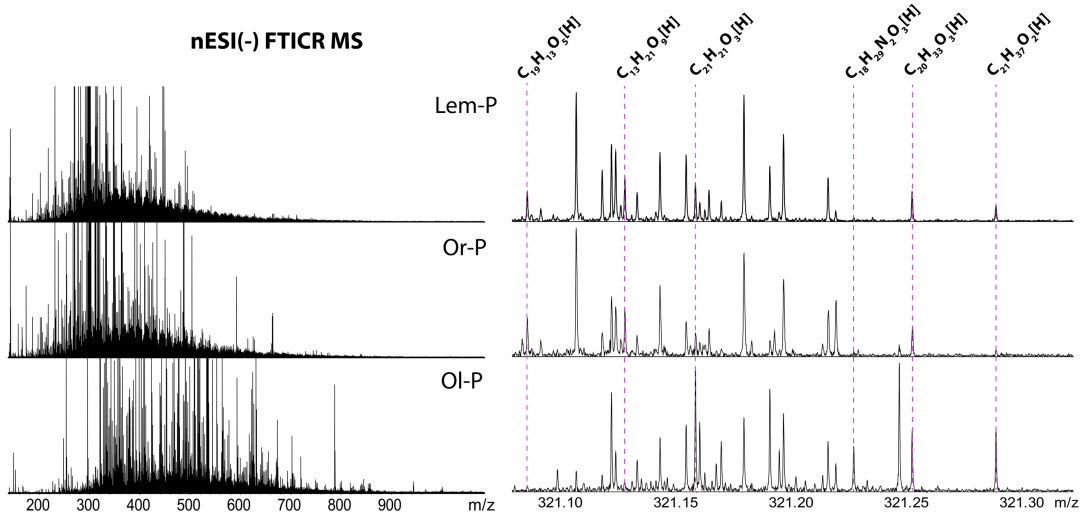

Figure S1. Left: Mass spectra of the bio-oils Lem-P, Or-P and Ol-P obtained at a peak temperature of 500 °C. The mass spectra were obtained in negative-ion mode nano-electrospray. Right: zoomed in mass spectra at 321 Da, some selected assignments are showed on top of the figure.

2. In house software VKSim: Reaction simulations

   - The reaction vector in VKSim are represented as following:

   $\pm c, \pm h, \pm o\ (n)$, where n is the number of losses/additions of the reaction and c, h, o are the carbon, hydrogen, and oxygen to be added/reduced respectively.

   - Pseudo code:

```
Molecule <C,H,O>
Reaction <c,h,o>

For each Iteration {

    Product = Molecule + Reaction //Apply reaction vector to molecule

    IF(
        any of (C, H, O) <0 //Ensure reaction is feasible
        OR
        H/C < 0.2 OR H/C > 3.1 //Ensure H/C value is within limits
suggested by Kind and Fiehn[1]
    ){
        RETURN Molecule //Molecule cannot react via current reaction,
orginal molecule is returned
    } ELSE {
        RETURN Product //Molecule can
react, and product is returned
        Molecule = Product //Product is now set as the Molecule for the
next Iteration
    }


}
```

- Format of input data: .csv file where first line has the column names as C, H, O (in capital letters), followed by the corresponding carbon, hydrogen, and oxygen content of the molecular compositions.

- Graphic interchange format (.gif) of consecutive reactions can be downloaded via the right click context menu and then selecting *save as (eg. file_name.gif).*

3. Petroleomics

Detailed composition analysis

A mass spectrometer with ultrahigh resolving power and high mass accuracy is able to resolve and identify thousands of individual and unique molecular compositions simultaneously in a single mass spectrum. As shown in Figure S2, bio-oils obtained by pyrolysis of Lem-P, Or-P, and Ol-P contain up to 8,000 individual molecular compositions classified in four main heteroatomic grouped classes e.g. O_o, N_1O_o, N_2O_o, and N_3O_o with oxygen atoms (o) of up to 11. Compositions with three nitrogen atoms and high oxygen content (e.g. N3O5-N3O8) were only detected in Lem-P 400 and Ol-P 500. As a general trend, lower oxygen-containing species were found within the bio-oil obtained from the Ol-P. Additionally, the class distribution and DBE plots (see Figure S3) show a similar chemical profile for the bio-oils of lemon obtained with a pyrolysis temperature of 400 ºC and 500 ºC.

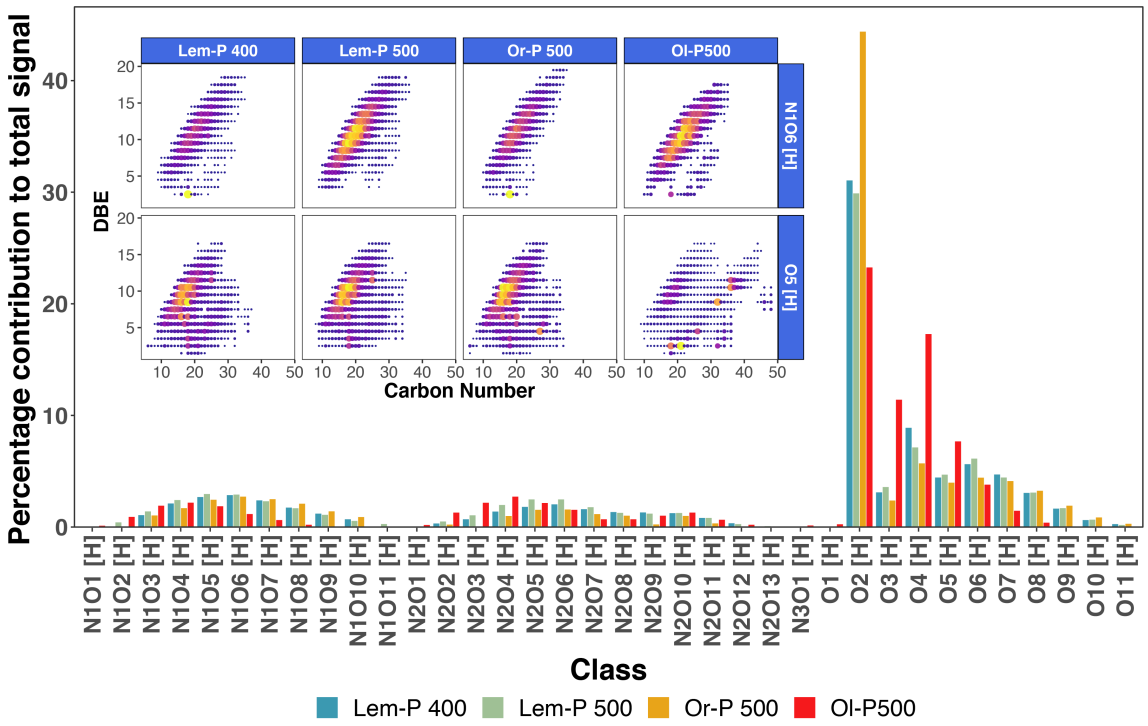

Figure S2. Heteroatomic class distribution of the bio-oils obtained from Lem-P 500, Or-P 500 and Ol-P 500 and Lem-P 400. The class distribution shows deprotonated classes with >0.5 relative abundance. The insets show DBE vs carbon number plots of representative classes and the total number for monoisotopic molecular formulas detected.

Upset plots have been used as an alternative visualisation tool to traditional Venn diagrams when the relationship between data sets is required [2–4]. The UpSet plots were used to study the commonality within the molecular compositions of the data sets. As shown in Figure S4, each sample is located in rows at the bottom of the Figure along with their intersections (common between two or three samples, or unique, molecular compositions) in a matrix of columns. The total number of common compositions in each intersection is represented with a bar chart at the top of the figure. The UpSet plots showing the intersection between the molecular compositions of the bio-oils obtained from the pyrolysis of Lem-P at different temperatures and the bio-oils obtained at the same temperature from different biomass sources are shown in Figure 2. From a total of 9,365 different molecular formulae among Lem-P 400 and Lem-P 500, about 60% correspond to common compositions. Despite the similarities in the chemical profile of these samples, 2,390 molecular compositions were found to be unique to Lem-P 500.

Only 114 compositions were found in the intersection of the chemical profile within Or-P 500 and Ol-P 500, showing a higher dissimilarity between the compositions of these samples. In contrast, 1,694 molecular formulae were found in the intersection of Or-P 500 and Lem-P 500 compositions. A total of 2,464 unique molecular compositions found in Ol-P 500 indicates the dissimilarity between Ol-P 500 and the bio-oils obtained from pyrolysis of citrus waste (Lem-P 400/Lem-P 500 and Or-P 500).

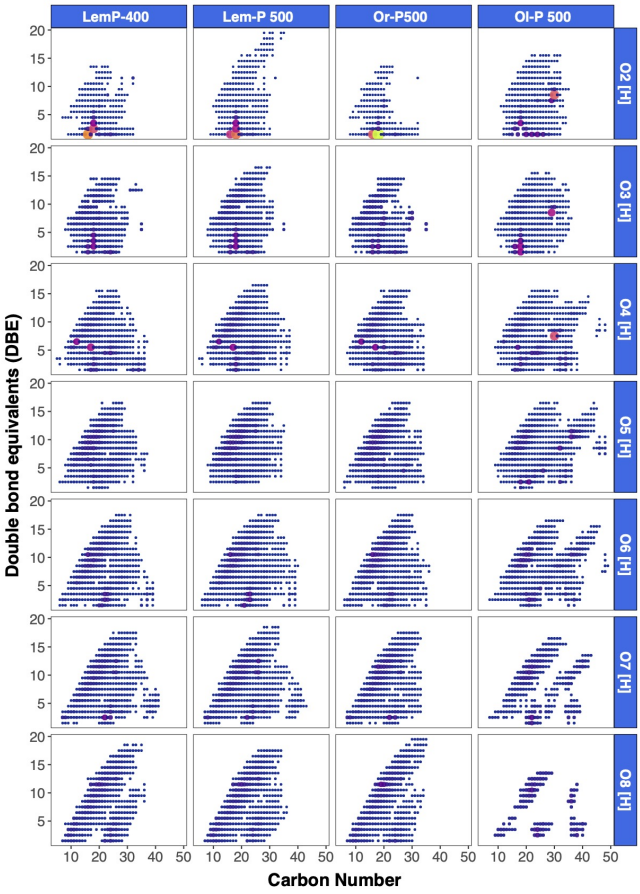

Figure S3. Double bond equivalents versus carbon number plots of oxygen-containing species detected in the bio-oils.

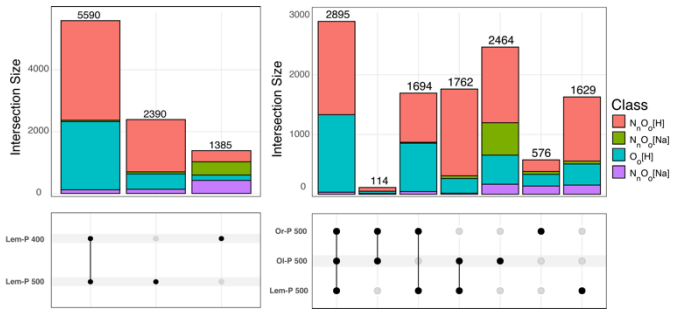

Figure S4. UpSet plot showing the intersection size (number of molecular compositions) within the bio-oils obtained from Lem-P at different temperatures (left) and the bio-oils obtained from different biomass feedstocks (right). Samples with molecular compositions present in any given intersection are coloured by a filled black circle, otherwise they are shown as a light grey circle. The bars are coloured by grouped heteroatomic classes where 1 <n < 3 and 1< o < 11.

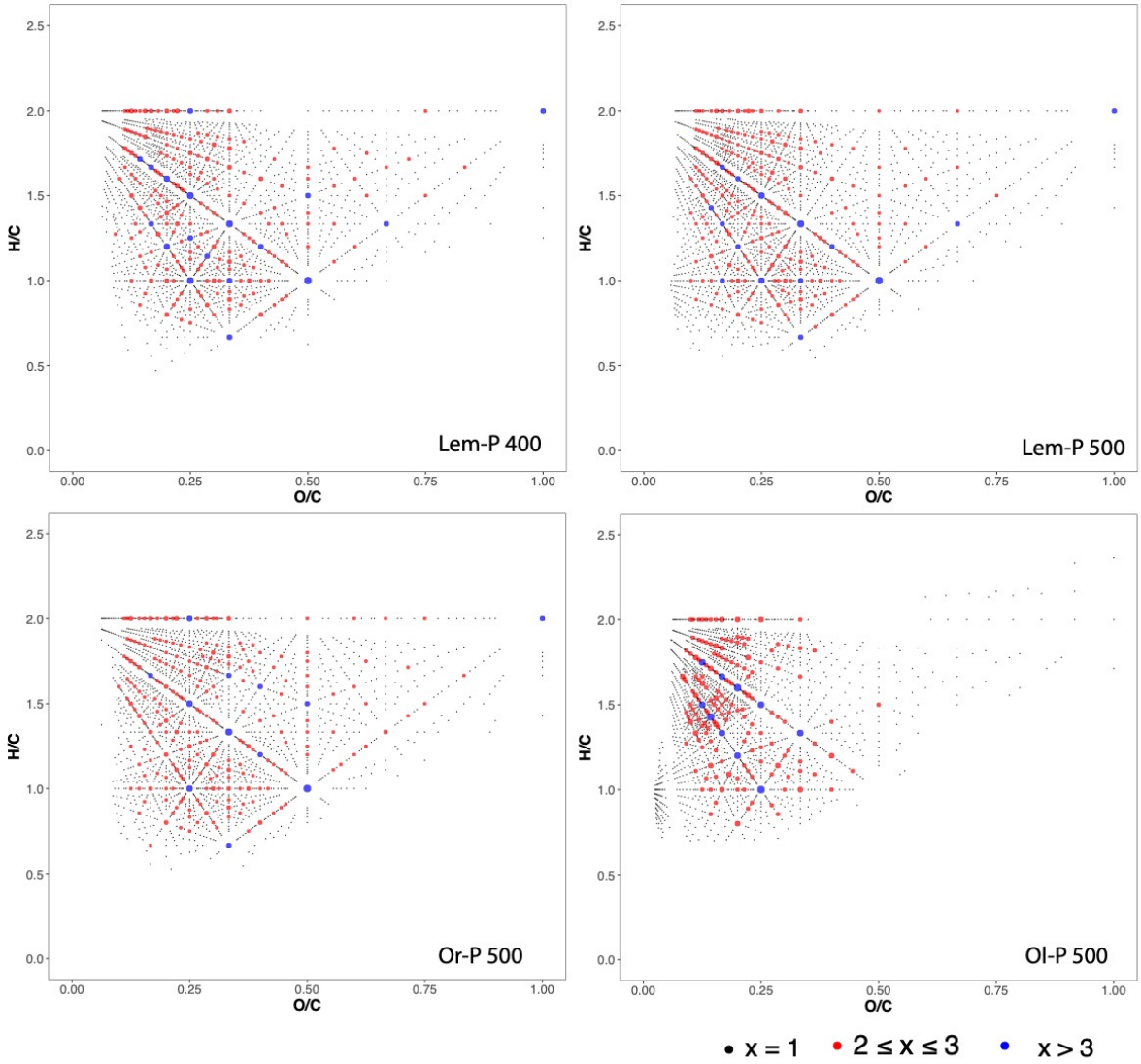

Figure S5. van Krevelen diagram of the bio-oil. The colour coding represents coordinates with unique molecular compositions (black dots), 2-3 molecular compositions (red dots) and more than 3 formulae (blue dots).

4. Processing reactions in van Krevelen plots

General equations:

$$C_c H_h N_n O_o S_s$$

$$C_c H_{2c+Z} O_o$$

$$Z = -2(DBE) + n + 2 = -2(DBE) + 2, \qquad if\ nitrogen = 0$$

$$Z = 2, \qquad Z \leq 0, \qquad even$$

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

Hydration

$$(x, y) = \left(\frac{O}{C}, \frac{H}{C}\right)$$

$$(x_1, y_1) = \left(\frac{O+1}{C}, \frac{H+2}{C}\right)$$

The slope a is defined as:

$$a = \frac{y_1 - y}{x_1 - x} = \frac{\frac{H+2}{C} - \frac{H}{C}}{\frac{O+1}{C} - \frac{O}{C}} = \frac{H+2-H}{O+1-O} = 2$$

$$\frac{H}{C} = 2\frac{O}{C} + b$$

$$\frac{H}{C} = 2\frac{O}{C} + \left(\frac{H}{C}\right)_q$$

Where $b = (H/C)_q$ is the intercept when $O/C = 0$

General equation:

$$C_c H_{2c+z} O_{\frac{2c+z}{2} - \frac{cb}{2}}$$

$$\frac{2c+Z}{2} - \frac{cb}{2} \geq 0$$

If the oxygen content is reduced by dehydration (O=0):

$$\frac{2c+z}{2} - \frac{cb}{2} = 0$$

The product will be a composition with:

$$\frac{H}{C} = \left(\frac{H}{C}\right)_q$$

For instance:
If $b = 0, \left(\frac{H}{C}\right)_q = 0$:

$$\frac{H}{C} = 2\frac{O}{C}$$

$$O = \frac{H}{2}$$

$$C_c H_{2c+Z} O_{\frac{2c+z}{2}}$$

A partial dehydration will reduce the H content by 2 per oxygen atom removed. A total dehydration will produce C_c

If $\left(\frac{H}{C}\right)_q$ =1/3

$$\frac{H}{C} = 2\frac{O}{C} + \frac{1}{3}$$

$$2\frac{O}{C} = \frac{H}{C} - \frac{1}{3}$$

$$O = \frac{1}{2}H - \frac{C}{6} = \frac{2c+z}{2} - \frac{c}{6}$$

$$C_c H_{2c+z} O_{\frac{2c+z}{2} - \frac{c}{6}}$$

A total dehydration will produce a product where H/C= 1/3, $C = 3H$: a composition with three H atoms per carbon atom

If $\left(\frac{H}{C}\right)_q$ =1/2

$$C_c H_{2c+z} O_{\frac{2c+z}{2} - \frac{c}{4}}$$

If total dehydration: $H/C = 1/2$, compositions with $C = 2H$

If $\left(\frac{H}{C}\right)_q = 1$

$$C_cH_{2c+z}O_{\frac{2c+z}{2}-\frac{c}{2}}$$

If total dehydration: $H/C = 1$, compositions with $C = H$

Hydrogenation

$$(x, y) = \left(\frac{O}{C}, \frac{H}{C}\right)$$

$$(x_1, y_1) = \left(\frac{O}{C}, \frac{H+2}{C}\right)$$

$$\frac{O}{C} = \left(\frac{O}{C}\right)_y$$

Where $(O/C)_y$ is the intercept in O/C-axis when $H/C = 0$

If O/C = 1/4 then O=C/4

$$C_cH_{2c+z}O_{\frac{c}{4}}$$

Then, carbon number must be a multiple of 4

If $\left(\frac{O}{C}\right)_y = 0{,}5$ then $O = C/2$

$$C_cH_{2c+z}O_{\frac{c}{2}}$$

Then, carbon number must be a multiple of 2

CO_2 elimination (Decarboxylation)

$$(x, y) = \left(0\ , \left(\frac{H}{C}\right)_q\right)$$

$$(x_1, y_1) = (2, 0)$$

$$a = \frac{y_1 - y}{x_1 - x} = \frac{0 - \left(\frac{H}{C}\right)_q}{2 - 0} = -\frac{1}{2}\left(\frac{H}{C}\right)_q$$

Decarboxylation lines intercept at O/C=2 when H/C=0, therefore:

$$0 = a\frac{2}{1} + b$$

$$b = -2a = \left(\frac{H}{C}\right)_q$$

$$\frac{H}{C} = -\frac{1}{2}\left(\frac{H}{C}\right)_q \frac{O}{C} + \left(\frac{H}{C}\right)_q$$

$\frac{H_h}{C_c}$ is the final product after a total CO_2 elimination

If $\left(\frac{H}{C}\right)_q = 2$

$$\frac{H}{C} = -\frac{O}{C} + 2$$

$$O = H - 2C$$
$$C_cH_{2c+z}O_{2c-(2c+z)} = C_cH_{2c+z}O_{-z}$$

Z is even for Nitrogen=0, therefore the oxygen must to be even.

If $\left(\frac{H}{C}\right)_q = 1$

$$\frac{H}{C} = -\left(\frac{1}{2}\right)\frac{O}{C} + 1$$

$$O = 2(C - H)$$

$$C_cH_{2c+z}O_{2(c-2c-Z)} = C_cH_{2c+z}O_{-2(c+Z)}$$

CO elimination

$$(x, y) = \left(0\ , \left(\frac{H}{C}\right)_q\right)$$

$$(x_1, y_1) = (1, 0)$$

$$a = \frac{y_1 - y}{x_1 - x} = \frac{0 - \left(\frac{H}{C}\right)_q}{1 - 0} = -\left(\frac{H}{C}\right)_q$$

$\frac{H_h}{C_c}$ is the final product after a total CO elimination

Decarboxylation lines intercept at O/C=1 when H/C=0, therefore:

$$0 = a1 + b$$

$$b = -a = \left(\frac{H}{C}\right)_q$$

$$\frac{H}{C} = -\left(\frac{H}{C}\right)_q \frac{O}{C} + \left(\frac{H}{C}\right)_q$$

If $\left(\frac{H}{C}\right)_q = 2$

$$\frac{H}{C} = -2\frac{O}{C} + 2$$

$$O = -\frac{H - 2C}{2} = -\frac{2c + z - 2c}{2} = -\frac{z}{2}$$

$$C_cH_{2c+z}O_{-Z/2}$$

If $\left(\frac{H}{C}\right)_q = 1$

$$\frac{H}{C} = -\frac{O}{C} + 1$$

$$O = -(n + Z)$$

$$C_cH_{2c+z}O_{-(c+Z)}$$

$$(n + Z) < 0$$

Demethanation

$$(x, y) = (0\,, 4)$$

$$(x_1, y_1) = \left(\left(\frac{O}{C}\right)_y, 0\right)$$

$$a = \frac{y_1 - y}{x_1 - x} = \frac{0 - 4}{\frac{O_o}{C_c} - 0} = -4\left(\frac{O}{C}\right)_y^{-1}$$

$$\frac{H}{C} = -4\left(\frac{O}{C}\right)_y^{-1}\frac{O}{C} + b$$

H/C= 4 if O/C=0 therefore b=4

$$\frac{H}{C} = -4\left(\frac{O}{C}\right)_y^{-1}\frac{O}{C} + 4$$

If $\left(\frac{O}{C}\right)_y = 2$

$$\frac{H}{C} = -2\frac{O}{C} + 4$$

$$-H + 4C = 2O$$

$$O = -\frac{H}{2} + 2C$$

$$C_cH_{2c+z}O_{2c-\frac{2c+z}{2}} = C_cH_{2c+z}O_{\frac{4c-2-z}{2}} = C_cH_{2c+z}O_{\frac{2c-z}{2}}$$

If $\left(\frac{O}{C}\right)_y = 1$

$$\frac{H}{C} = -4\frac{O}{C} + 4$$

$$H - 4C = -4O$$

$$O = C - \frac{H}{4}$$

$$C_cH_{2c+z}O_{c-\frac{2c+z}{4}} = C_cH_{2c+z}O_{\frac{4c-2c-z}{4}} = C_cH_{2c+z}O_{\frac{2c-z}{4}}$$

If $\left(\frac{O}{C}\right)_y = 0.5$

$$\frac{H}{C} = -8\frac{O}{C} + 4$$

$$H - 4C = -8O$$

$$O = \frac{C}{2} - \frac{H}{8}$$

$$C_cH_{2c+z}O_{\frac{c}{2}-\frac{2c+z}{8}} = C_cH_{2c+z}O_{\frac{4c-2c-z}{8}} = C_cH_{2c+z}O_{\frac{2c-z}{8}}$$

General equation		General equations for the reaction lines		Related examples	Z
Hydration $\frac{H}{C} = 2\frac{O}{C} + \left(\frac{H}{C}\right)_q$	1	$\frac{H}{C} = 2\frac{O}{C} + 1$	$C_cH_{2c+Z}O_{(2c+Z)/2-(c/2)}$	$C_{16}H_{20}O_2$[H]	-12
				$C_{24}H_{30}O_3$[H]	-18
				$C_{12}H_{18}O_3$[H]	-6
				$C_{16}H_{24}O_4$[H]	-8
	2	$\frac{H}{C} = 2\frac{O}{C}$	$C_cH_{2c+Z}O_{(2c+Z)/2}$	$C_{10}H_8O_4$[H]	-12
				$C_{12}H_{12}O_6$[H]	-12
				$C_{18}H_{18}O_9$[H]	-18
				$C_9H_6O_3$[H]	-12
				$C_{10}H_{12}O_6$[H]	-8
Decarboxylation $\frac{H}{C} = -\frac{1}{2}\left(\frac{H}{C}\right)_q\frac{O}{C} + \left(\frac{H}{C}\right)_q$	3	$\frac{H}{C} = -\frac{O}{C} + 2$	$C_cH_{2c+Z}O_{-Z}$	$C_8H_{12}O_4$[H]	-4
				$C_{16}H_{24}O_8$[H]	-8
				$C_{12}H_{18}O_6$[H]	-6
				$C_{10}H_{16}O_4$[H]	-4
				$C_{15}H_{24}O_6$[H]	-6
	4	$\frac{H}{C} = -\left(\frac{1}{2}\right)\frac{O}{C} + 1$	$C_cH_{2c+Z}O_{-2(c+Z)}$	$C_{10}H_8O_4$[H]	-12
				$C_{15}H_{12}O_6$[H]	-18
				$C_{14}H_{12}O_4$[H]	-16
				$C_{28}H_{24}O_8$[H]	-32
				$C_{25}H_{24}O_2$[H]	-26
Decarbonylation $\frac{H}{C} = -\left(\frac{H}{C}\right)_q\frac{O}{C} + \left(\frac{H}{C}\right)_q$	*5*	$\frac{H}{C} = -2\frac{O}{C} + 2$	$C_cH_{2c+Z}O_{-Z/2}$	$C_8H_8O_4$[H]	-8
				$C_{10}H_{10}O_5$[H]	-10
				$C_{10}H_{12}O_4$[H]	-8
				$C_{15}H_{18}O_6$[H]	-12
				$C_{36}H_{64}O_4$[H]	-8
	6	$\frac{H}{C} = -\frac{O}{C} + 1$	$C_cH_{2c+Z}O_{-(c+Z)}$	$C_{15}H_{12}O_3$[H]	-18
				$C_{20}H_{16}O_4$[H]	-24
				$C_{14}H_{12}O_2$[H]	-16
				$C_{16}H_{12}O_4$[H]	-20
				$C_9H_6O_3$[H]	-12
Demethanation $\frac{H}{C} = -4\left(\frac{O}{C}\right)_y^{-1}\frac{O}{C} + 4$	*7*	$\frac{H}{C} = -2\frac{O}{C} + 4$	$C_cH_{2c+Z}O_{2c-(2c+Z)/2}$	$C_6H_{12}O_6$[H]	0
				$C_7H_{14}O_7$[H]	0
				$C_9H_{18}O_9$[H]	0
				$C_6H_{12}O_6$[H]	0
	8	$\frac{H}{C} = -4\frac{O}{C} + 4$	$C_cH_{2c+Z}O_{c-(2c+Z)/4}$	$C_{12}H_{24}O_6$[H]	0
				$C_{16}H_{32}O_8$[H]	0
				$C_{16}H_{28}O_9$[H]	-4
				$C_6H_8O_4$[H]	-4
Hydrogenation $\frac{O}{C} = \left(\frac{O}{C}\right)_y$	*9*	$\frac{O}{C} = 0.5$	$C_cH_{2c+Z}O_{c/2}$	$C_8H_8O_4$[H]	-8
				$C_{10}H_{10}O_5$[H]	-10
				$C_{12}H_{12}O_6$[H]	-12
				$C_{16}H_{20}O_8$[H]	-12
				$C_{16}H_{26}O_8$[H]	-6
	10	$\frac{O}{C} = 1$	$C_cH_{2c+Z}O_c$	$C_6H_{12}O_6$[H]	0
				$C_7H_{14}O_7$[H]	0
				$C_9H_{16}O_9$[H]	-2
				$C_6H_{10}O_6$[H]	-2
				$C_7H_{10}O_7$[H]	-4

Table S1. Summary of general processing reaction lines with related examples as shown in Figure 4. The molecular formulae shown correspond to molecules detected in the bio-oils for each particular line. $(H/C)_q$ is the intercept when $O/C = 0$ and $(O/C)_y$ is the intercept when $H/C = 0$.

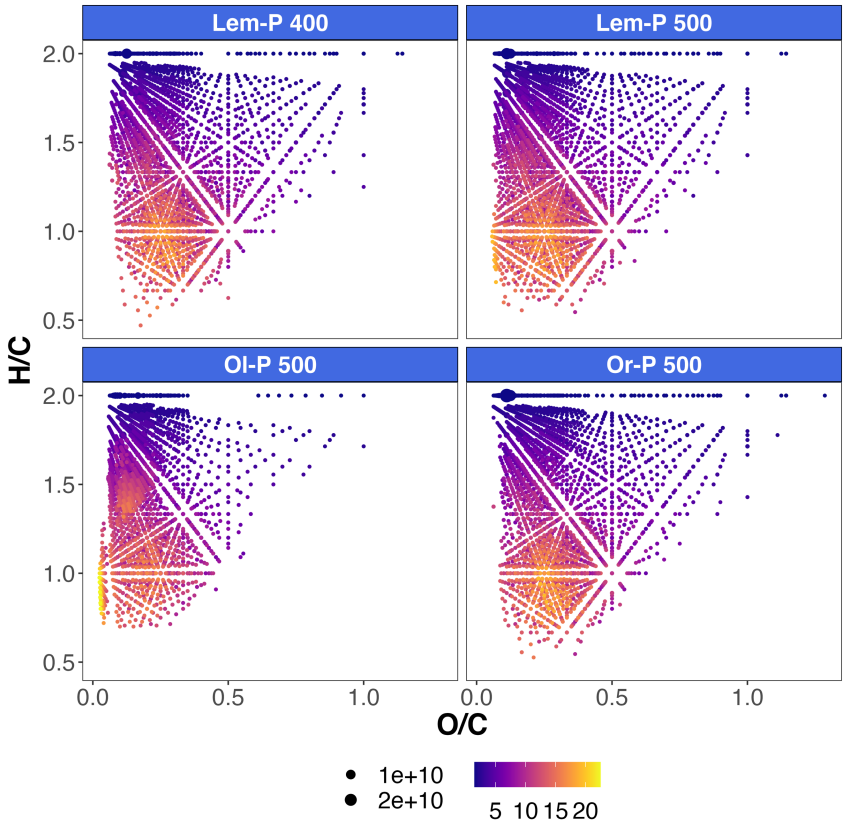

Figure S6. van Krevelen diagrams colour coded by double bond equivalents (DBE).

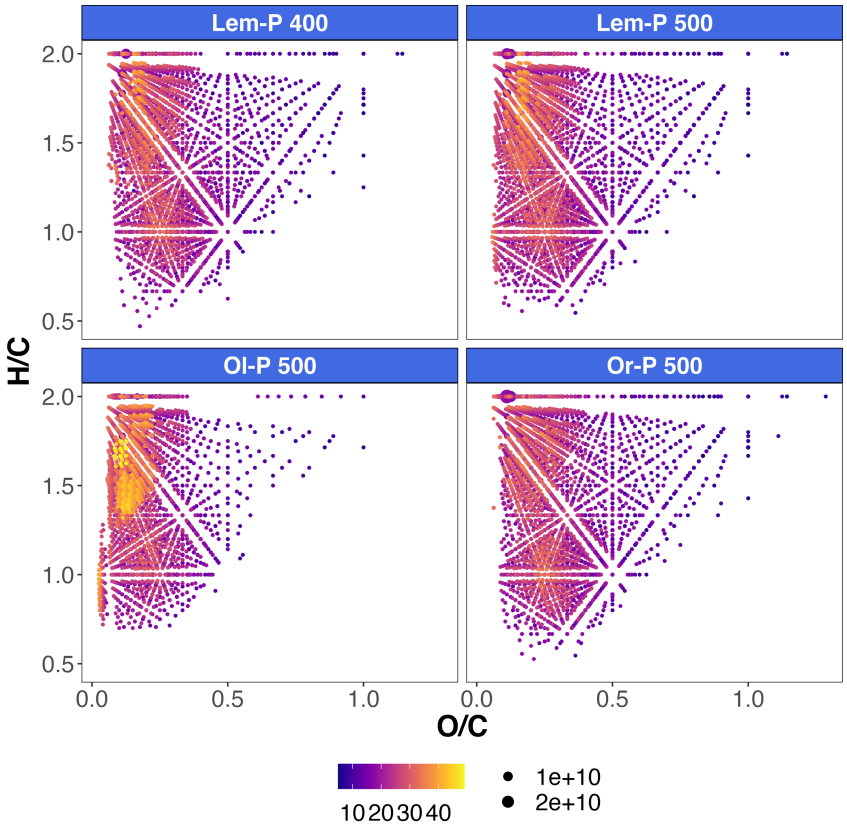

Figure S7. van Krevelen diagram colour coded by carbon number.

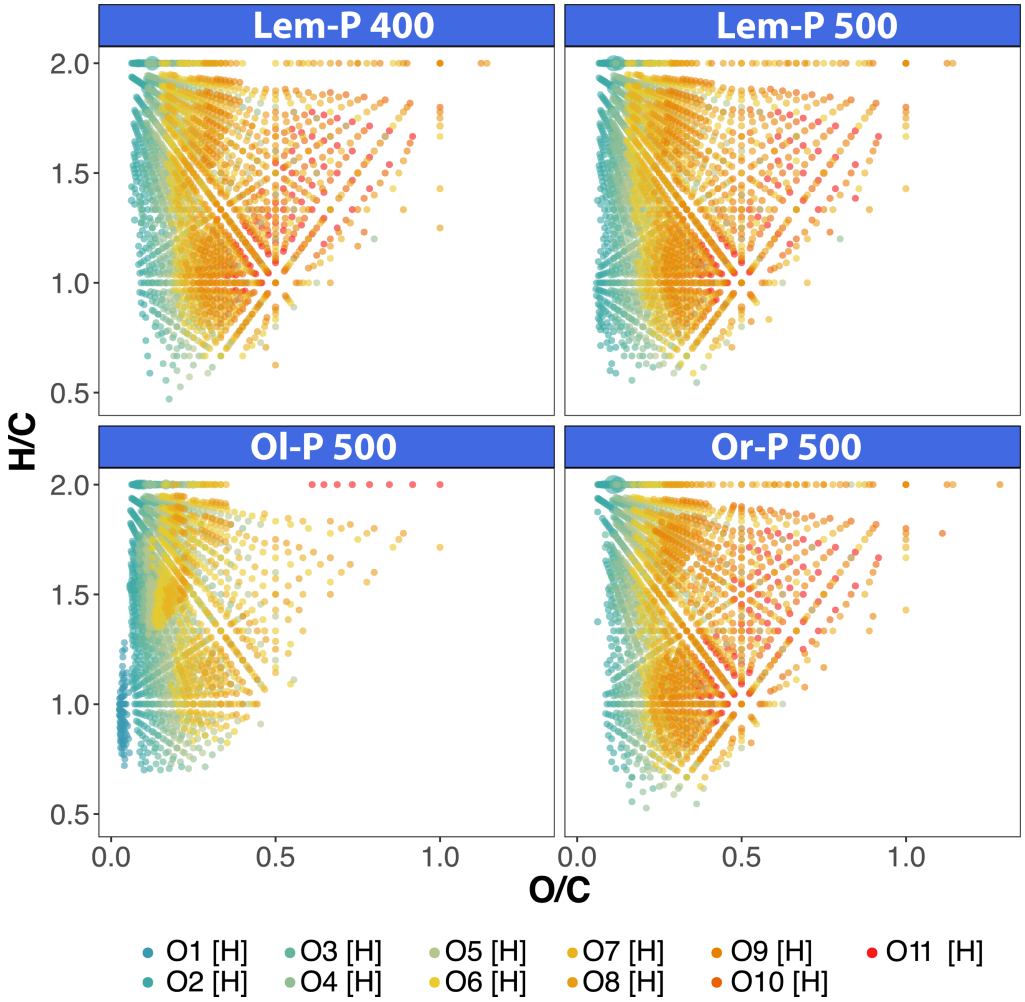

Figure S8. van Krevelen distribution colour coded by the oxygen content of the molecule.

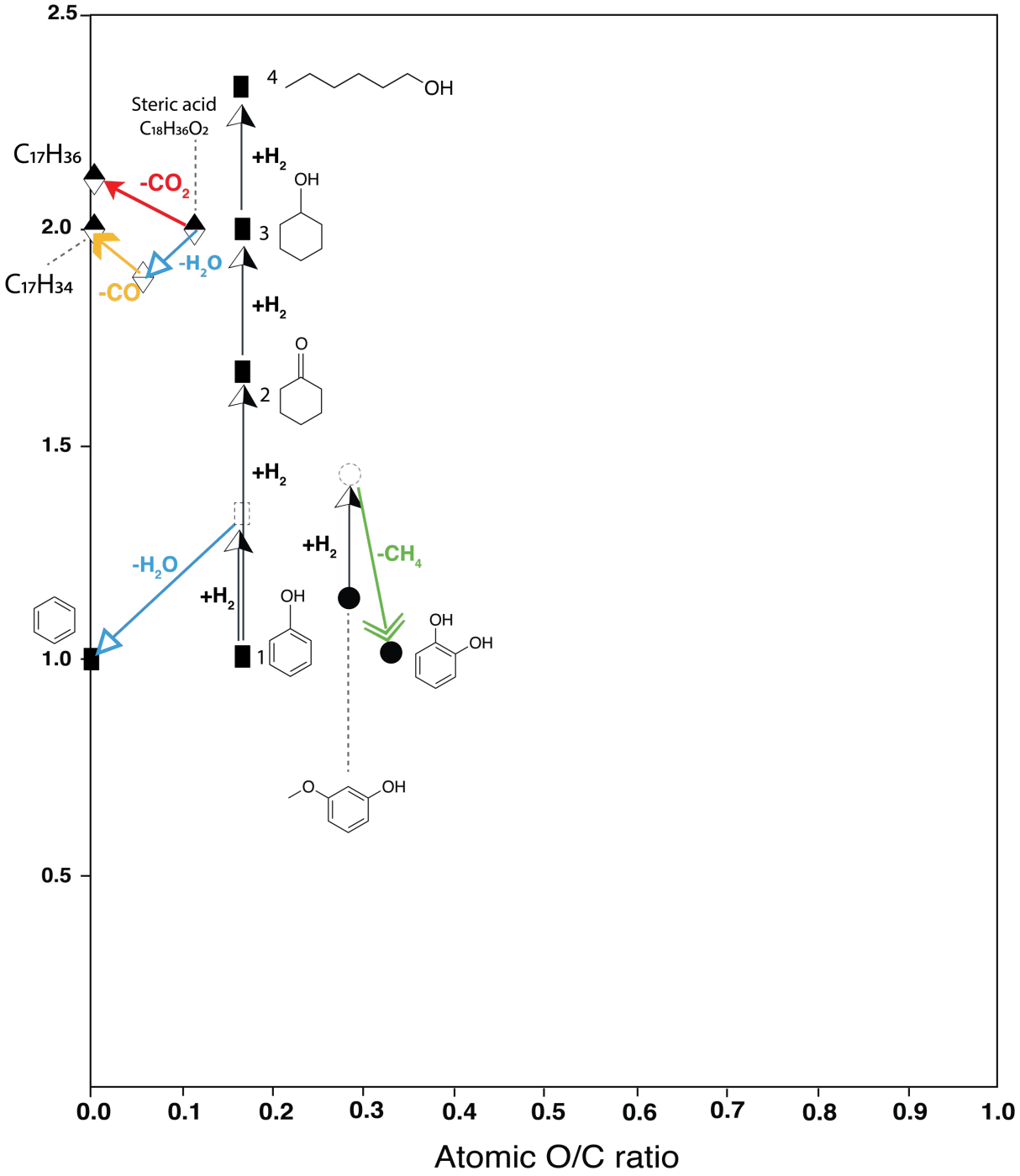

Figure S9. van Krevelen diagram of proposed reaction of standard compositions [5–7].

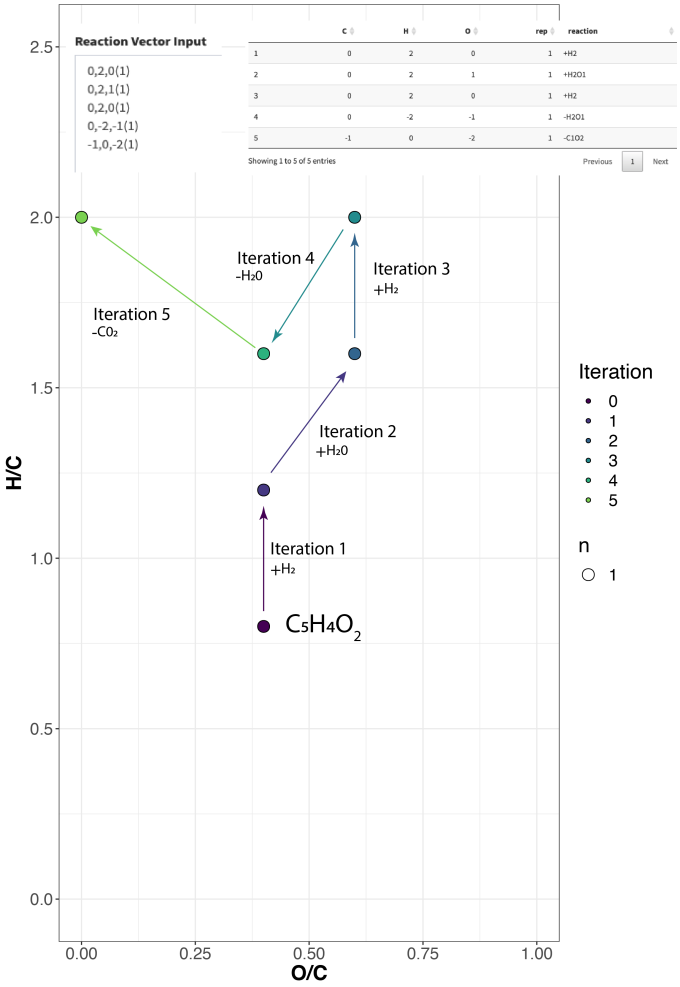

Figure S10. Simulation of the reaction pathway for the production of 3-methyl-1-ene (C_5H_{10}) (pathway 1): consecutive reactions $+H_2 \rightarrow +H_2O \rightarrow +H_2 \rightarrow -H_2O \rightarrow -CO_2$. The simulation can be performed in VKSim, the insets show the reaction vector input in the software, and the list of reaction routes.

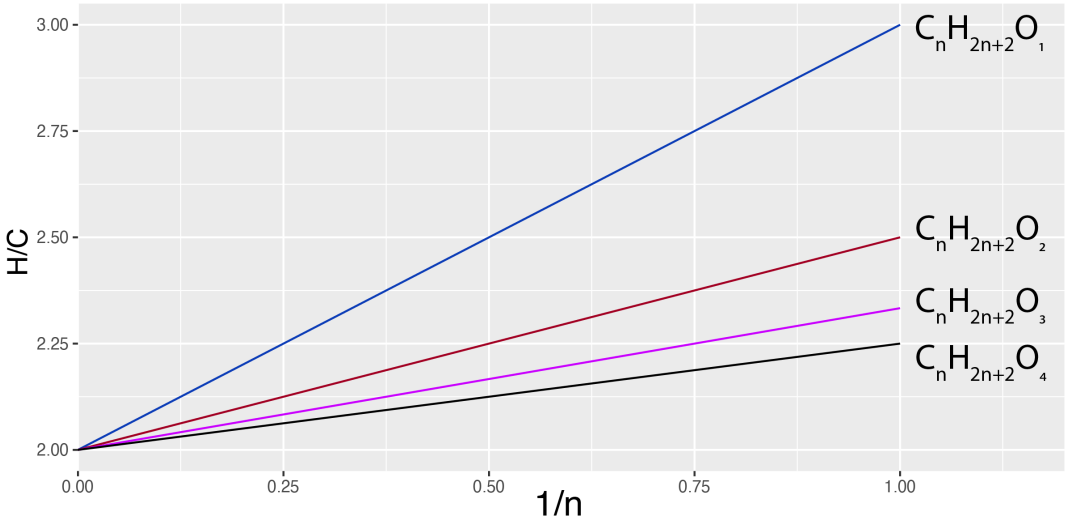

Figure S11. Distribution of saturated acyclic ethers and alcohols in van Krevelen diagrams.

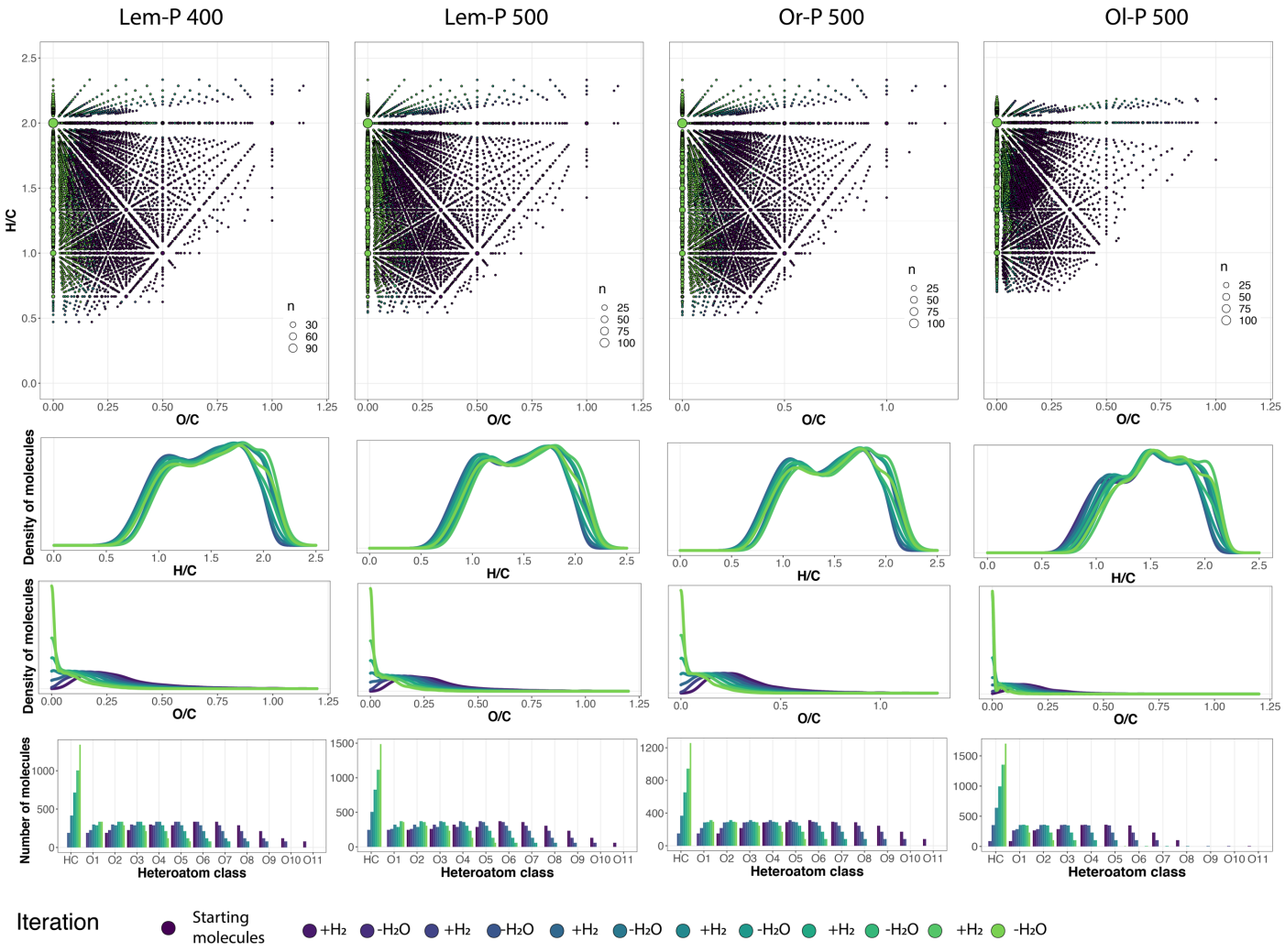

Figure S12. Simulation of the molecular distribution of the bio-oils by consecutive addition/losses of H_2 and H_2O. The reduction of up to six oxygen atoms was simulated using VKSim. Reaction vector (0,2,0)→(0,-2,-1) →(0,2,0)→(0,-2,-1) →(0,2,0)→(0,-2,-1) →(0,2,0)→(0,-2,-1) →(0,2,0)→(0,-2,-1) →(0,2,0)→(0,-2,-1) →

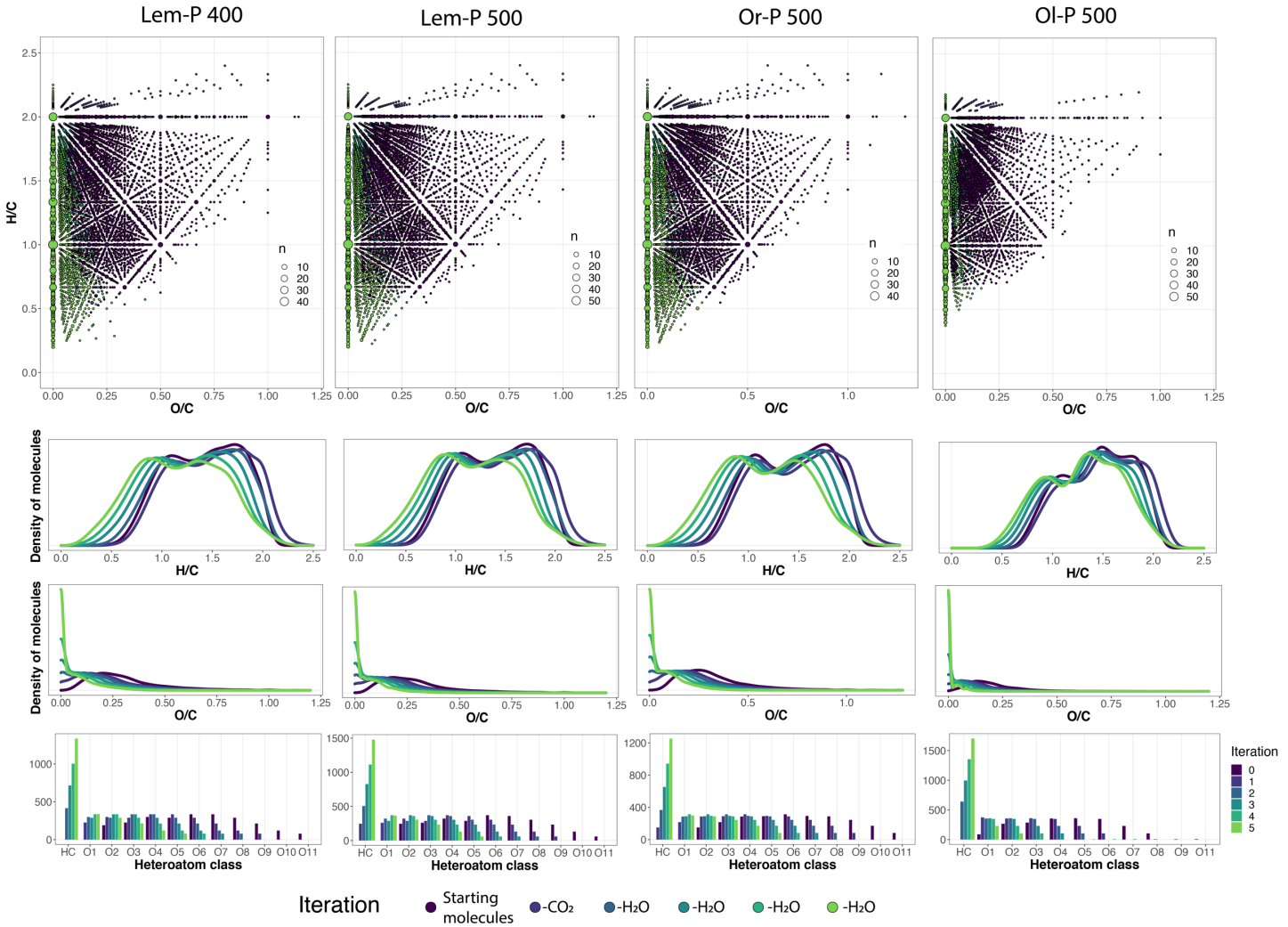

Figure S13. Simulation of a reaction where 6 oxygen atoms are removed by a removal of a CO_2 molecule followed by the removal of 4 oxygen atoms by dehydration. Reaction vector (-1,0,-2) →(0,-2,-1) →(0,-2,-1) →(0,-2,-1) →(0,-2,-1).

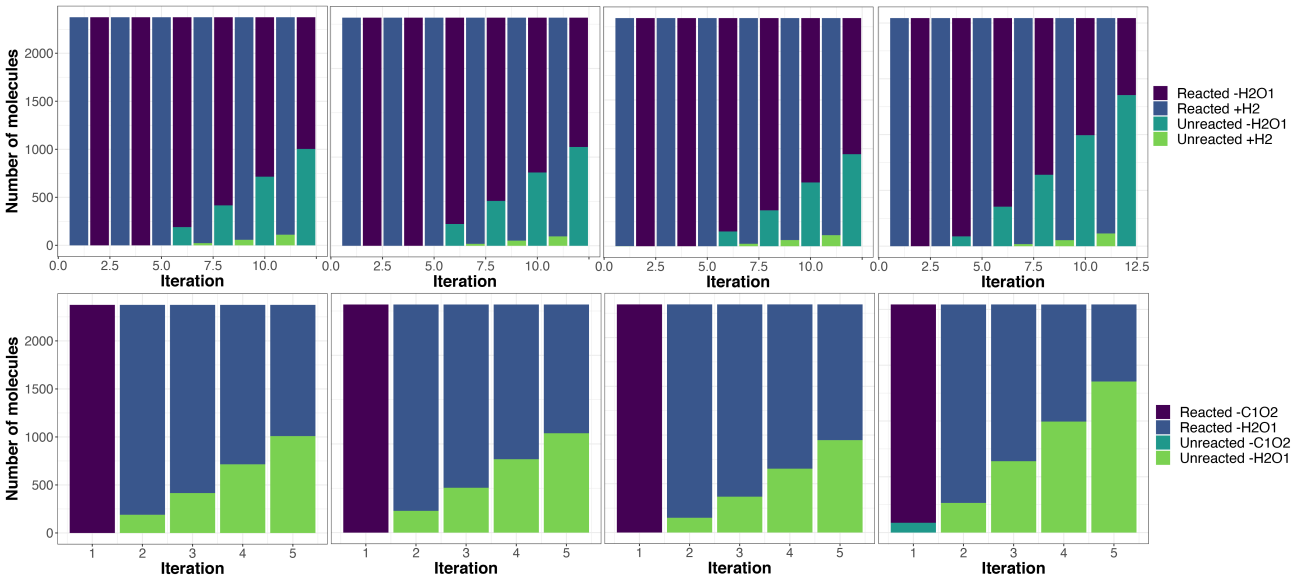

Figure S14. Number of molecules completing a reaction after each iteration.

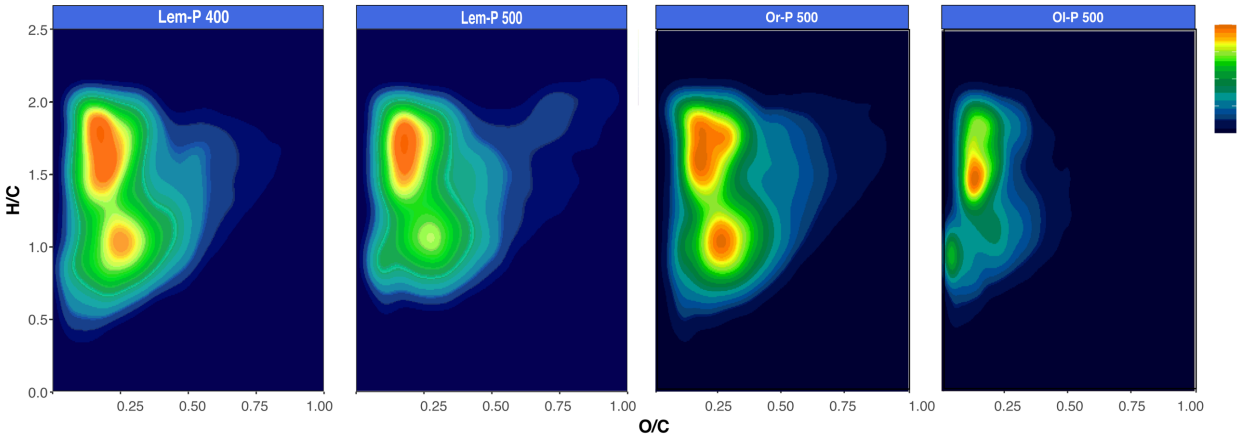

Figure S15. van Krevelen diagrams of the oxygenated species of the bio-oils plotted as Kernel density plots.

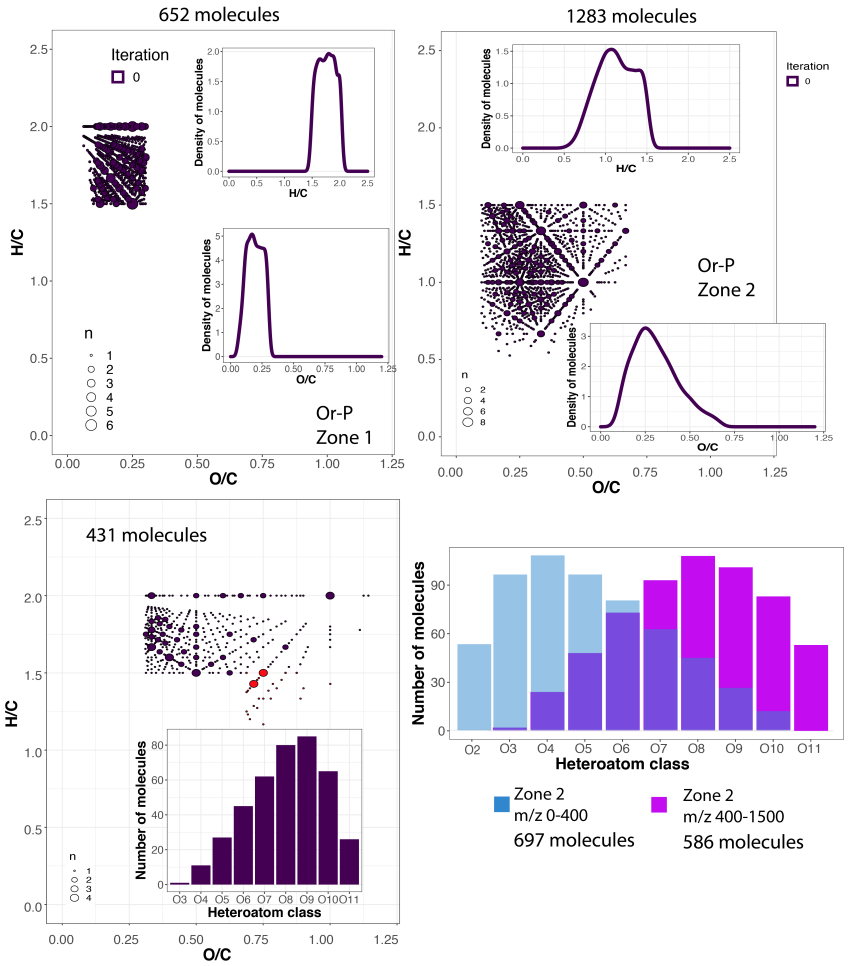

Figure S16. Molecular distribution of the bio-oil Or-P 500 in different zones of the van Krevelen diagram. 652 molecules corresponded to compositions located at 1.5<H/C<2, 0<O/C<0.3 (zone 1), 1283 are located in zone 2 (Z2) 0.5<H/C<1.5, 0<O/C<0.6). And 431 molecules are distributed in zone 3. The distribution of the heteroatomic compositions of different areas is shown.