Comprehensive Multiphase NMR - A Powerful Tool to Understand and Monitor Molecular Processes During Biofuel Production

Paris Ning[†], Daniel Lane[†], Ronald Soong[†], Daniel Schmidig[‡], Thomas Frei[‡], Peter De Castro[‡], Ivan Kovacevic[‡], Stephan Graf[‡], Sebastian Wegner[§], Falko Busse[§], Jochem Struppe[‖], Michael Fey[‖], Henry J. Stronks[⊥], Martine Monette[⊥], Myrna J. Simpson[†], and André J. Simpson*[†]

[†]Environmental NMR Centre, University of Toronto, Toronto, ON, Canada, M1C 1A4

[‡]Bruker BioSpin AG, Industriestrasse 26, 8117 Fällanden, Switzerland

[§]Bruker BioSpin GmbH, Silberstreifen 4, 76287 Rheinstetten, Germany

[‖]Bruker BioSpin Corp., 15 Fortune Drive, Billerica, Massachusetts, USA, 01821-3991

[⊥]Bruker Ltd., 2800 High Point Drive, Milton, ON, Canada, L9T 6P4

*Corresponding Author: André J. Simpson (andre.simpson@utoronto.ca)

Number of Pages: 12

Number of Figures: 6

Supporting Information

S1. NMR Experimental Parameters

Spectral Editing has been previously introduced and discussed in details by Courtier-Murias *et al.*.[1] As such only key parameters are provided here. ^{1}H experiments were performed with presaturation for water suppression. 90° pulses were calibrated for each sample and *garp-4* decoupling was used during acquisition to remove ^{13}C J-coupling. A standard inversion recovery experiment was also performed for each sample to measure T_1 and the recycle delay was set to 3.5 s, which is ~5 times the measured T_1 value (note the relaxation was fast due to the ^{13}C enrichment). All experiments were performed with 256 scans and 16K time domain points. Spectral width was set to 16 ppm. Spectral editing was performed with a bipolar pulse pair longitudinal encode-decode (BPPLED) sequence with encoding/decoding gradients of 1.8 ms at 50 gauss/cm and a diffusion time of 180 ms. ^{1}H spectral editing is shown in supporting information Figure S3 and more general editing details are provided in the main paper (section "*Spectral Editing*"). A line broadening of 0.3 Hz was used for the reference and IDE spectra while 5 Hz was used for other ^{1}H experiments.

All ^{13}C experiments were recorded with a spectral width of 400 ppm. 90° pulses and T_1 were measured as described for the ^{1}H experiments. 1944 (243*8) scans were recorded for all ^{13}C observe experiments. This value was chosen because the sideband suppression technique used in CP experiment (a modified version of TOSS termed TOSS.243)[2,3] requires a multiple of 243 scans. For stepped decoupling experiments, decoupling was performed by using high-power *spinal64* and low-power *waltz*16 in succession as previously described.[4] Sideband suppression for stepped decoupling experiments was done through various speed magic angle spinning (VSMAS).[4,5] For swollen fractions, spectral editing employed the same BPPLED sequence as mentioned above with the same gradient strength and diffusion time. In solid-state experiments, a Carr-Purcell-Meiboom-Gill (CPMG) relaxation filter on ^{1}H (2 × 15 μs echoes) was employed prior to a fixed cross-polarization contact time of 1ms to attenuate ^{1}H magnetisation in crystalline components. These values were found optimal for the differentiation of dynamic and crystalline solid

standards as discussed by Courtier-Murias *et al*..[1] A line broadening of 15 Hz was used for all ^{13}C observe experiments except for CP and CP_T_2 where 25 Hz was used.

S2. Quantitative Estimates of Different Phases

The distribution of biomass in each phase can provide vital insights into how much of the organic material is accessible for solvent extraction (i.e. swellable) and how much is protected/inaccessible (i.e. non-swellable). The estimation of such distribution in each phase has been introduced previously,[1,6] however, it is rather complex and thus worth revisiting here. Readers should be reminded that this is not absolute quantification; for example, it is challenging to define exactly where the solid domain stops and the gel domain starts. This stated, if applied rigorously, the approaches here provide an estimation of the amount of organic materials of each phase in a sample, and certainly allow relative comparison between samples in a series. As examples, different feedstock species and changes in extraction treatment are discussed in the paper. The following section outlines the steps needed to differentiate solution, gels, rigid gels, dynamic solids and true rigid solids by CMP-NMR. In this example, a whole, non-extracted intact *C. reinhardtii* algae sample is used.

Step 1 - Estimating the Solvent Accessible/Swellable Fraction. First, the multiphase sample needs to be divided into two fractions – swellable and non-swellable. To do so, the freeze-dried sample is first placed into a 4 mm MAS rotor. The rotor contains inserts (in this case, a bottom insert made of zirconia and an extended-length Kel-F top plug), both designed to maintain the sample within the lower section of the coil. Only a relatively small amount of each sample (~10 mg) is used which only fills the 20% of the coil region (see Figure S1a). Next, a cross-polarization (CP) experiment is performed on the dry sample (Figure S1c). Cross polarization is highly efficient when strong dipoles form between ^{1}H and ^{13}C nuclei in the sample, and in the dry state, all components would undergo efficient cross polarization.[1] Next, an excess of D_2O (30 µL) is added to the same rotor to swell the sample and it is thoroughly mixed and allowed to swell for 12 hrs. Great care is taken to ensure all the sample stays within the coil region (see Figure S1b).

The extended screw plug is now replaced with a normal length screw plug to account for the volume increase during swelling. Next, the CP experiment is performed again (see Figure S1d). For the swollen components, bonds become more dynamic, and in turn, this modulates the ^{1}H-^{13}C dipoles, decreasing the efficiency of cross polarization. For example, it was shown in a multiphase sample, CP does not occur in solutions, is inefficient in swollen gels, and retains normal efficiency for true solids.[1,7] As such, the reduction of signal between the dry CP experiment (everything) and the swollen CP experiment (dominated by solid components) gives an estimation as to the percentage of organic matter that interacts with the solvent. In this particular *C. reinhardtii* sample, the signal is decreased by 36.7% when the sample becomes swollen (Figure S1e). This indicates ~37% of the sample is swellable while 63% has no interaction with the solvent (i.e. non-swellable).[6] Again, it is important to understand these values are best estimates and should not be taken as absolute values; however, these values are valid for comparison to estimate relative change in phases across a series of samples.

Step 2 - Estimating the Subfractions within the Non-swellable Fraction. The second step is to provide an estimation of the subfractions (dynamic solids and true solids) within the non-swellable fraction (Figure S2). Components with varying degrees of local mobility and dynamics have different spin-spin relaxation behavior (T_2). A very short T_2 filter (2×15 µs echoes)[1] is employed on the ^{1}H channel prior to cross polarization such that the signal with extremely broad ^{1}H profiles (i.e. true rigid solids) relaxes, leaving signals only from components that are solid but have local mobility (Figure S2b).[6] While not directly relevant to this study, to give the readers some context, rubber is an example of a dynamic solid, where the material is a solid but at the molecular level the bonds are flexible and dynamic.[7] As such, the signal loss observed between standard CP and CP with a T_2 filter (CP_T_2) is proportional to the amount of true rigid solids in the sample. The spectrum for the dynamic solids is represented by the CP_T_2 experiment (Figure S2b) while the rigid solids spectrum (Figure S2c) is created by difference (normal CP minus CP_T_2, "CP_Inv_T_2") which is essentially the rigid solids.[1,6]

Step 3 - Estimating the Subfractions within the Swellable Fraction. The last step is to differentiate the contribution of components in the swellable fraction. Because of strong ^{1}H-^{1}H dipolar interactions in solids, ^{1}H spectra give very broad lineshape for true solids which can be many KHz wide. As such, solids cannot be detected using ^{1}H NMR without the use of multiple approaches such as combined rotation and multiple-pulse NMR spectroscopy (CRAMPS).[8] Thus, when conventional low power "solution state" ^{1}H NMR experiments are performed, only the dynamic fractions with at least some interaction with the solvent (i.e. gels and solution) are observed.[1,9] This swellable fraction is further broken down into 3 subfractions or domains: swollen solids/rigid gels (Recovery Arising from Diffusion Editing, RADE), restricted diffusion such as dynamic gels and macromolecules (Diffusion Editing, DE), and soluble and mobile components (Inverse Diffusion Editing, IDE). Estimations as to the contribution of each subfraction can be performed though it is not obvious.

The RADE spectrum represents the signals lost when the delays required for diffusion editing are set. These components are the fastest relaxing components in the swellable fraction and could be attenuated if diffusion editing is used alone. These components relax quite rapidly (thus have some "solid-like" character) but have enough dynamics to be narrow enough for easy ^{1}H NMR detection. This fraction is recovered by subtracting the delays-on spectrum from the reference spectrum (delays off, gradients off) as shown in Figure S3A. The contribution from this domain is easy to estimate as it simply is represented by the intensity lost when the delays are set (Figure S3A-3).

Obtaining the contribution from soluble (IDE) and restricted diffusion (DE, dynamic gel/macromolecules) domains is slightly more challenging. As seen from above, the "delays-on" spectrum equals Reference minus RADE. In turn this means the "delays-on" spectrum is the sum of the remaining domains i.e. restricted diffusion (DE, dynamic gel/macromolecules Figure S3B-2) and the soluble components (IDE) (Figure S3B-1). The components with restricted diffusion (dynamic gel/macromolecules) are easily obtained as these are the components detected during the diffusion editing experiment (Figure

S3B-2). However, in addition to suppressing the soluble molecules, the diffusion gradient also supresses the signal from large components to some extent. As such, the weighting of DE cannot be estimated directly from its signal intensity and must be given additional considerations.

The soluble components are selected via a weighted subtraction of the diffusion editing result from the "delays-on" spectrum (Figure S3A-2) but again, as the DE spectrum is used in the weighting, it is not possible to take the signal intensity directly from IDE. However, the relative contributions of IDE and DE to the mixture can be obtained by weighting their relative contribution until their sum best matches the "delays-on" spectrum. Figure S3 shows the concept and that when the weighting of DE and IDE is correct, the "weighted sum" spectrum (Figure S3B-3) equals the "delays-on" spectrum (Figure S3B-4). The correct weighting of the DE and IDE spectrum can then be used estimate is contribution to the swellable fraction. The result is that contributions of 3 swellable domains can be estimated. When combined with the non-swellable fraction, all 5 domains can be combined to show their approximate contributions to the mixture as whole (see Figure S4).

Interestingly, Ning et al.[4] introduced a ^{13}C-detect experiment based on stepped decoupling technique that allows all phases to be detected equally and quantitatively. As such, while stepped decoupling cannot be used for editing or phase differentiation, it does provide a quantitative overview of the whole sample. Combining editing steps described above, the percentage of biomass distributed in each of the five domains (true solid (Figure 1f), dynamic solid (Figure 1e), swollen solid/rigid gel (Figure 1d), gel (Figure 1c), and solution (Figure 1b)) can be estimated. To validate the estimates are representative, the ^{13}C spectra of the above domains were summed together with their appropriate weightings (Figure S5) and compared to the quantitative spectrum. The summed spectrum (Figure S5f) matches the stepped decoupling spectrum quite well (Figure S5g), confirming that the weightings estimated based on the above methods are relative accurate and there is no significant exaggeration or discrimination in any domains (solids, solution etc.).

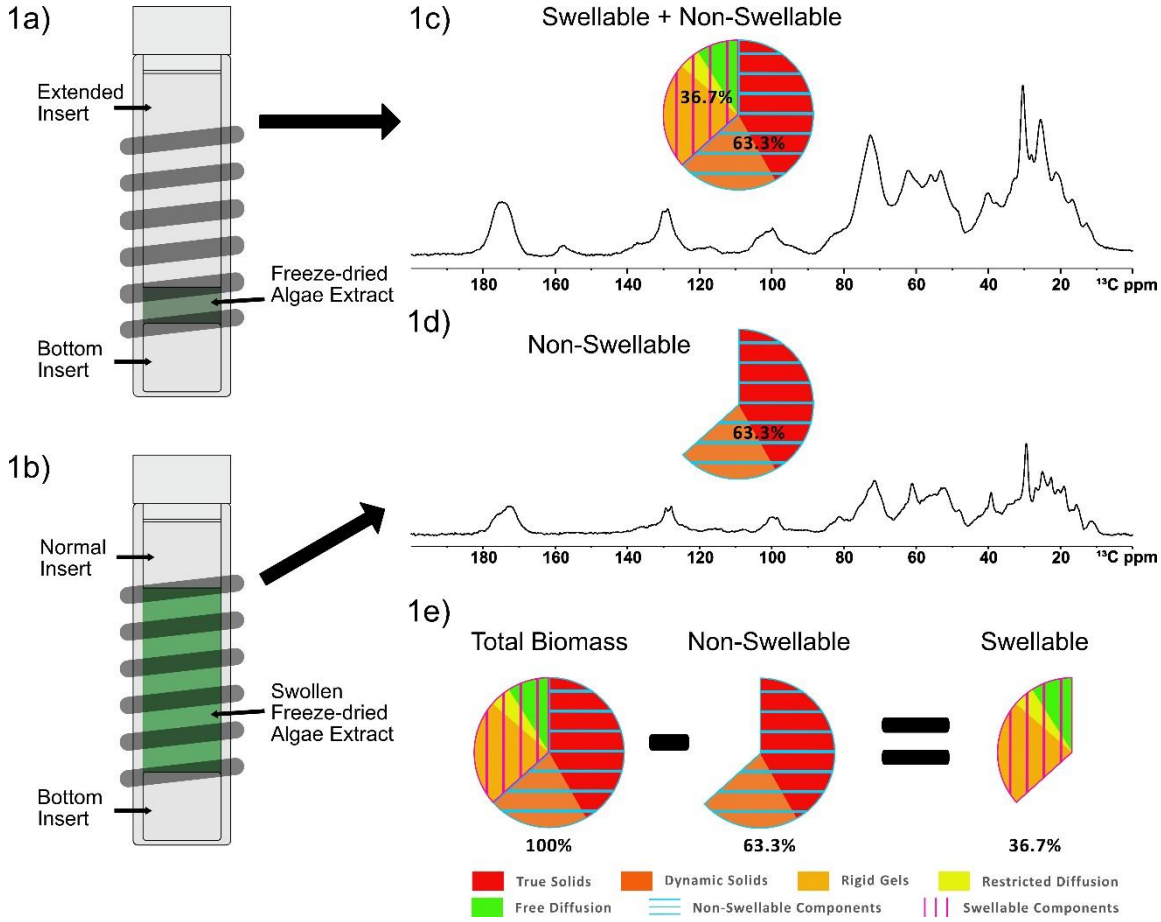

Figure S1 Differentiation of swellable and non-swellable fractions by CP experiments on intact C. reinhardtii. Spectrum (c) is obtained from freeze-dried algae sample (a) and the spectrum represents all carbon in the sample. Spectrum (d) is obtained when sample (a) is swollen with excess amount of water (b). Because CP experiments detect based on strong dipolar interaction (rigid, solid structures), spectrum (d) represents the remaining solid fraction after swelling (non-swellable fraction). The difference between the percentage represented by spectra (c) and (d) is the amount of carbon whose signal is lost after swelling (i.e., swellable fraction (e)).

138

139

140

141

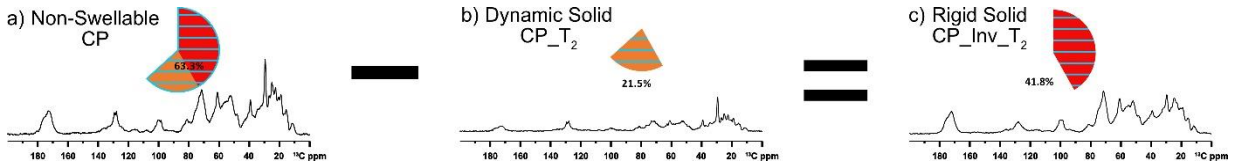

Figure S2 Differentiation of domains within the non-swellable fraction. By deploying a T_2 relaxation filter prior to CP (CP_T_2), signals from fast-relaxing components (rigid solids) are filtered out, leaving only slower-relaxing, dynamic solids (middle). Conversely, subtracting this spectrum from the CP spectrum of the non-swellable domain (left) yields the signals from the rigid solid domain (right spectrum, CP inverse T_2 filter).

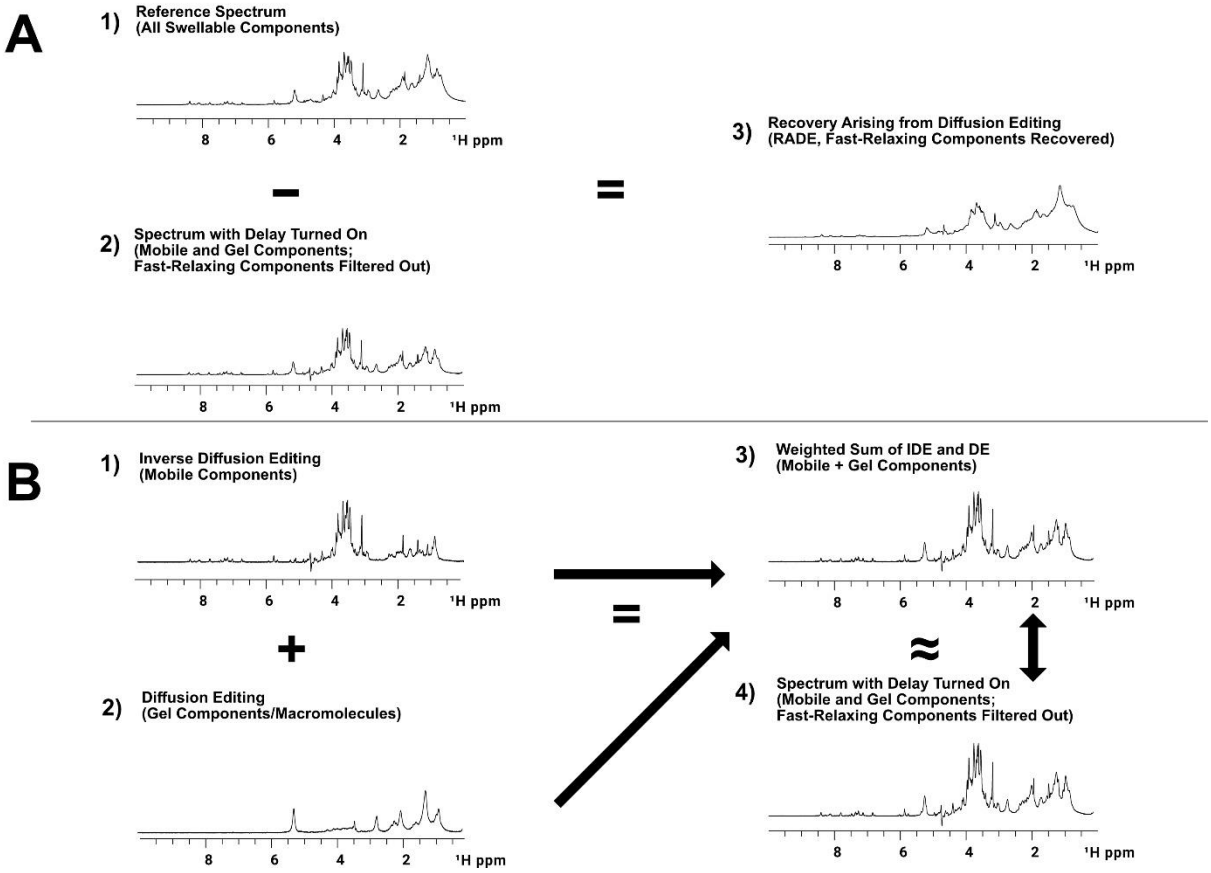

Figure S3 Differentiation of domains within the swellable fraction using ^1H diffusion editing. Delays associated with this technique would filter out fast-relaxing components (rigid gels). Subtracting the delays-on spectrum (A-2) from the reference spectrum (A-1) gives the signals from the fast-relaxing components (spectrum A-3, recovery arising from diffusion editing (RADE)). With appropriate weightings of the mobile components (B-1) and gel components (B-2), the sum of these two spectra shows an identical spectrum as the delays-on spectrum that contains signals from gel and mobile components.

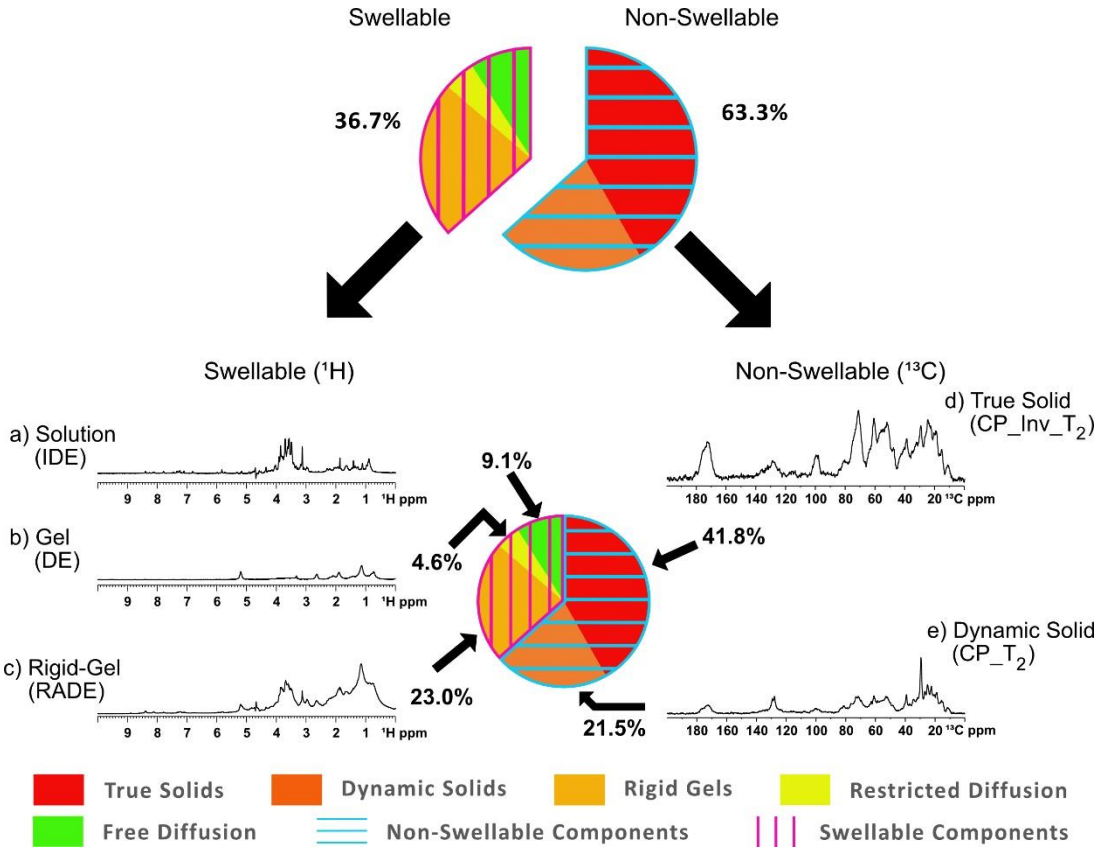

Figure S4 Differentiating components in all phases. Combining the swellable and non-swellable fractions, information on the distribution of different chemical components in the multiphase sample can be extracted with the help of CMP-NMR.

147

148

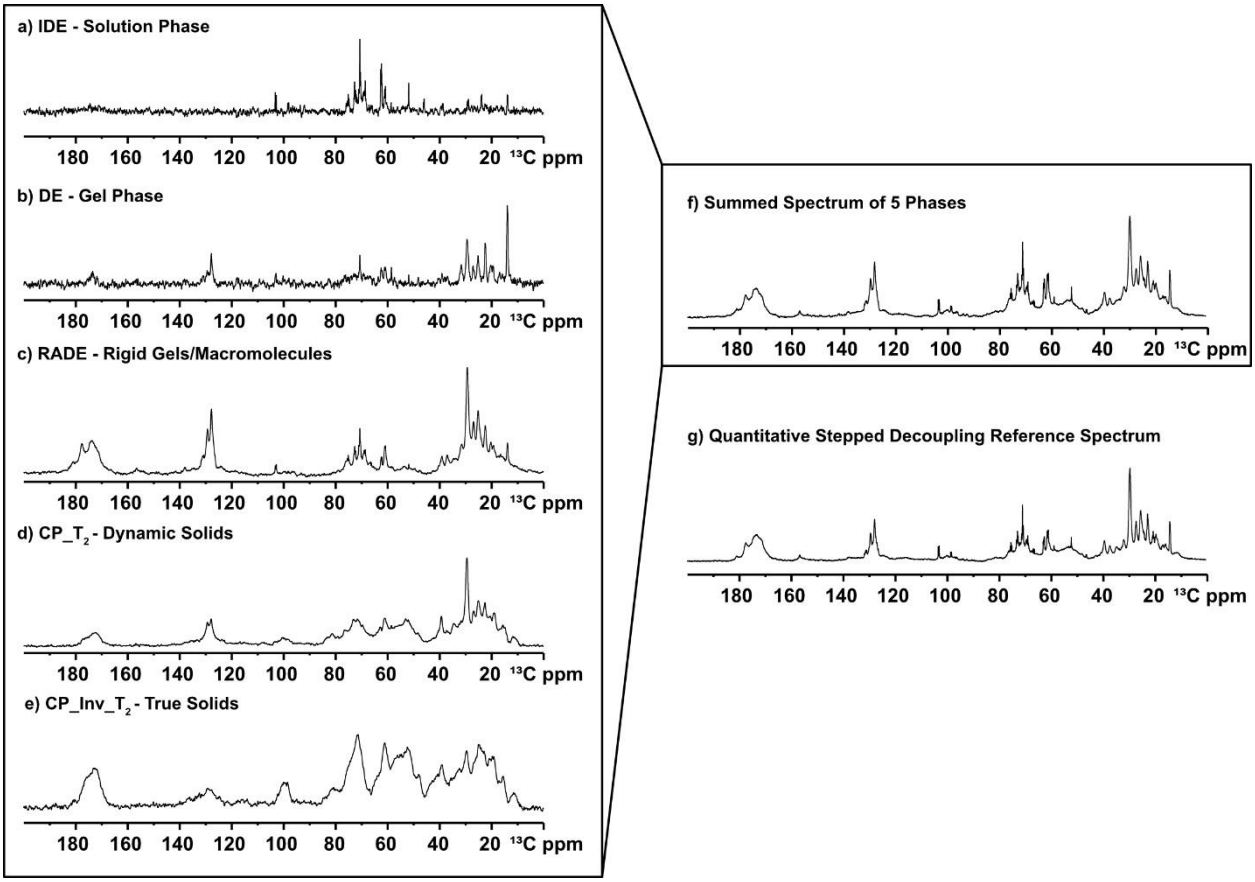

Figure S5 Comparison between the sum of signals from all domains (a – e) and a reference quantitative stepped decoupling spectrum. With appropriate weighting for spectra a – e, the sum resembles the quantitative spectrum, indicating an accurate weighting of different phases.

149

150

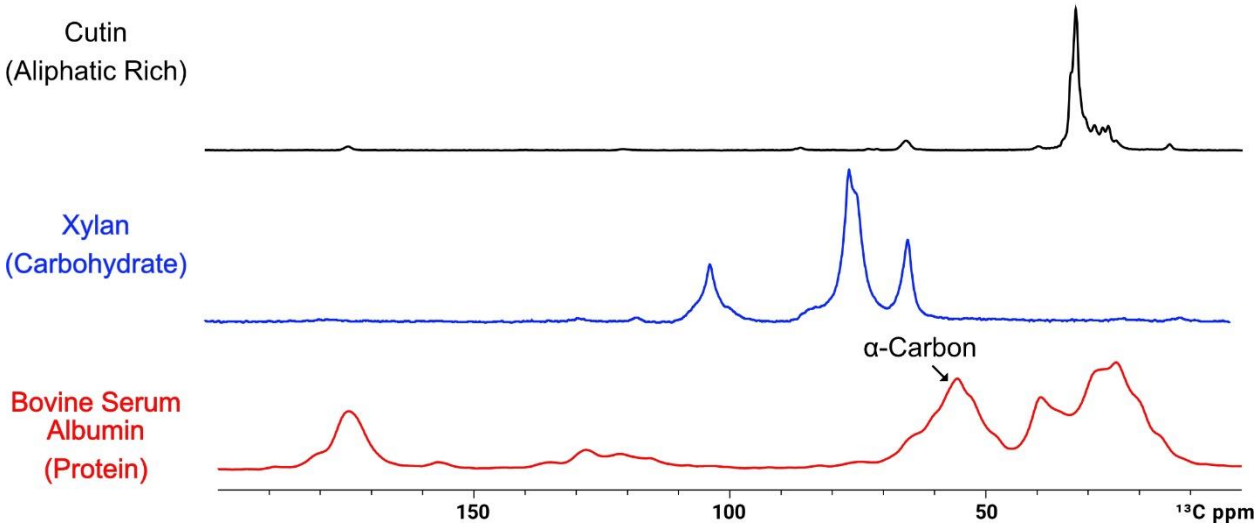

Figure S6 CP spectra of model biopolymers to represent major groups within the biomass – lipids (aliphatic rich), carbohydrates, and protein. The model compounds, cutin, xylan, and bovine serum albumin, showcase the key spectral features, such as the alpha carbon, and can help identifying corresponding groups within the algae extract. Note these spectra were collected as dry solids and thus the lineshape is broader than most algal materials in this paper, the majority of which were swollen to some extent during the experiment.

Reference

(1) Courtier-Murias, D.; Farooq, H.; Masoom, H.; Botana, A.; Soong, R.; Longstaffe, J. G.; Simpson, M. J.; Maas, W. E.; Fey, M.; Andrew, B.; Struppe, J.; Hutchins, H.; Krishnamurthy, S.; Kumar, R.; Monette, M.; Stronks, H. J.; Hume, A.; Simpson, A. J. Comprehensive Multiphase NMR Spectroscopy: Basic Experimental Approaches to Differentiate Phases in Heterogeneous Samples. *J. Magn. Reson.* **2012**, *217*, 61–76. https://doi.org/10.1016/j.jmr.2012.02.009.

(2) Mobarhan, Y. L.; Struppe, J.; Fortier-McGill, B.; Simpson, A. J. Effective Combined Water and Sideband Suppression for Low-Speed Tissue and in Vivo MAS NMR. *Anal. Bioanal. Chem.* **2017**, *409* (21), 5043–5055. https://doi.org/10.1007/s00216-017-0450-3.

(3) Antzutkin, O. N.; Levitt, M. H. Centerband Phase Shift in the TOSS Spectra of a Magic-Angle-Spinning Single Crystal. *J. Magn. Reson. - Ser. A* **1996**, *118* (2), 295–298. https://doi.org/10.1006/jmra.1996.0041.

(4) Ning, P.; Soong, R.; Bermel, W.; Lane, D.; Simpson, M. J.; Simpson, A. J. 13C Quantification in Heterogeneous Multiphase Natural Samples by CMP-NMR Using Stepped Decoupling. *Anal. Bioanal. Chem.* **2018**, *410* (27), 7055–7065. https://doi.org/10.1007/s00216-018-1306-1.

(5) Madhu, P. K.; Pratima, R.; Kumar, A. Suppression of Sidebands by Variable Speed Magic Angle Sample Spinning in Solid State NMR. *Chem. Phys. Lett.* **1996**, *256* (1–2), 87–89. https://doi.org/10.1016/0009-2614(96)00428-9.

(6) Masoom, H.; Courtier-Murias, D.; Farooq, H.; Soong, R.; Kelleher, B. P.; Zhang, C.; Maas, W. E.; Fey, M.; Kumar, R.; Monette, M.; Stronks, H. J.; Simpson, M. J.; Simpson, A. J. Soil Organic Matter in Its Native State: Unravelling the Most Complex Biomaterial on Earth. *Environ. Sci. Technol.* **2016**, *50* (4), acs.est.5b03410. https://doi.org/10.1021/acs.est.5b03410.

(7) Silva, L. M. A.; Filho, E. G. A.; Simpson, A. J.; Monteiro, M. R.; Venâncio, T. Comprehensive Multiphase NMR Spectroscopy: A New Analytical Method to Study the Effect of Biodiesel Blends on the Structure of Commercial Rubbers. *Fuel* **2016**, *166*, 436–445. https://doi.org/10.1016/j.fuel.2015.10.105.

(8) Jackson, P.; Harris, R. K. A Practical Guide to Combined Rotation and Multiple-pulse NMR Spectroscopy of Solids. *Magn. Reson. Chem.* **1988**, *26* (11), 1003–1011. https://doi.org/10.1002/mrc.1260261112.

(9) Simpson, A. J.; Kingery, W. L.; Shaw, D. R.; Spraul, M.; Humpfer, E.; Dvortsak, P. The Application of 1 H HR-MAS NMR Spectroscopy for the Study of Structures and Associations of Organic Components at the Solid−Aqueous Interface of a Whole Soil. *Environ. Sci. Technol.* **2001**, *35* (16), 3321–3325. https://doi.org/10.1021/es010607v.