Supplementary information for

Quantifying cellulose accessibility during enzyme-mediated deconstruction using 2 fluorescence-tagged carbohydrate binding modules

Vera Novy[a,1], Kevin Aïssa[a,1], Fredrik Nielsen[a], Suzana K. Straus[b], Peter Ciesielski[c], Christopher G. Hunt[d], and Jack Saddler[a,2]

[a] Department of Wood Science, Faculty of Forestry, The University of British Columbia, 2424 Main Mall, Vancouver, BC, V6T 1Z4, Canada
[b] Department of Chemistry, University of British Columbia, 2036 Main Mall, Vancouver, BC, V6T 1Z1, Canada
[c] Biosciences Center, National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden CO 80401, USA
[d] Forest Products Laboratory, US Department of Agriculture, One Gifford Pinchot Drive, Madison, Wisconsin 53726, USA

[1] equally contributing first authors
[2] corresponding author:
Jack Saddler
Department of Wood Science
Faculty of Forestry
The University of British Columbia
2424 Main Mall, Vancouver BC, Canada, V6T 1Z4
+1 778 980 9152
jack.saddler@ubc.ca

This pdf includes:

Supplementary Figures S1 to S4
Supplementary Methodology

Supplementary Figure S1

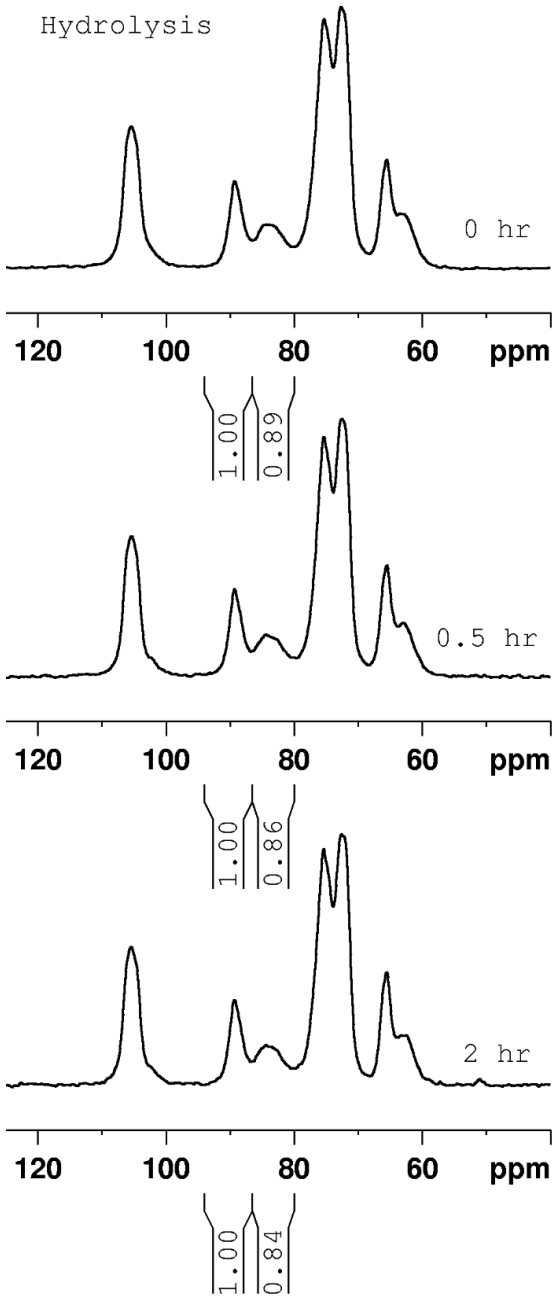

Figure S1. ^{13}C NMR analysis of NBSK fiber hydrolyzed for 0, 0.5, and 2h. After 2 h of hydrolysis, the changes occurring in the substrate drastically changes its size and supramolecular organization. The ^{13}C NMR spectra will reflect these changes, making it impossible to attribute changes in the CI to the enzymatic deconstruction mechanism.

Supplementary Figure S2

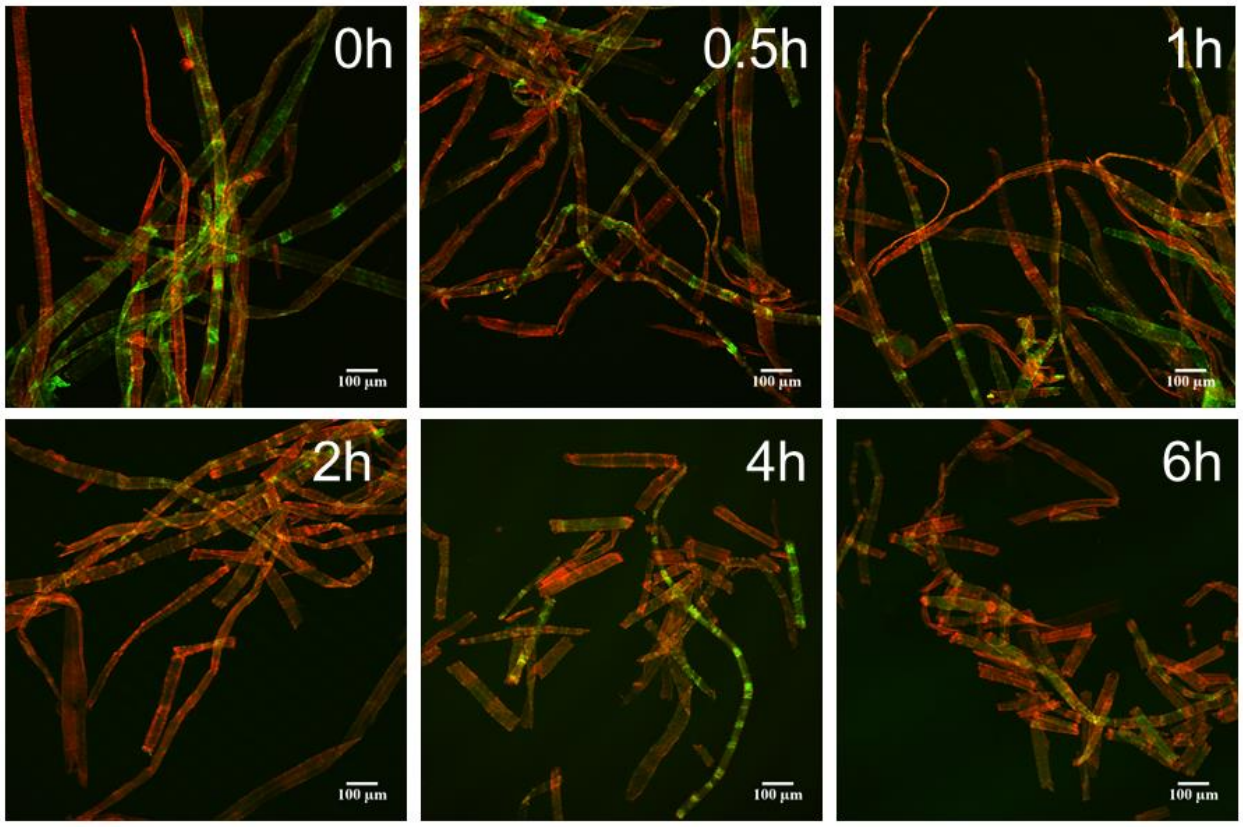

Figure S2. CLSM micrographs from the hydrolysis reactions using the never-dried NBSK. Incubation times as indicated. Scale bar is 100 µm.

Supplementary Figure S3

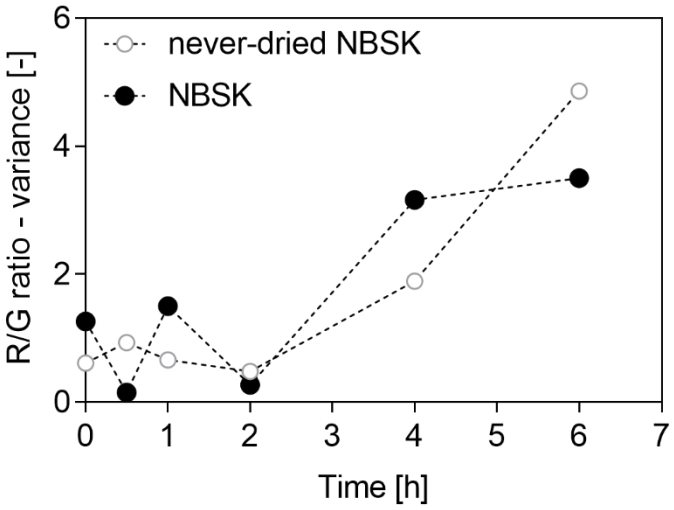

Figure S3: The variance of the R/G ratios in samples hydrolyzed with increasing time.

Supplementary Figure S4

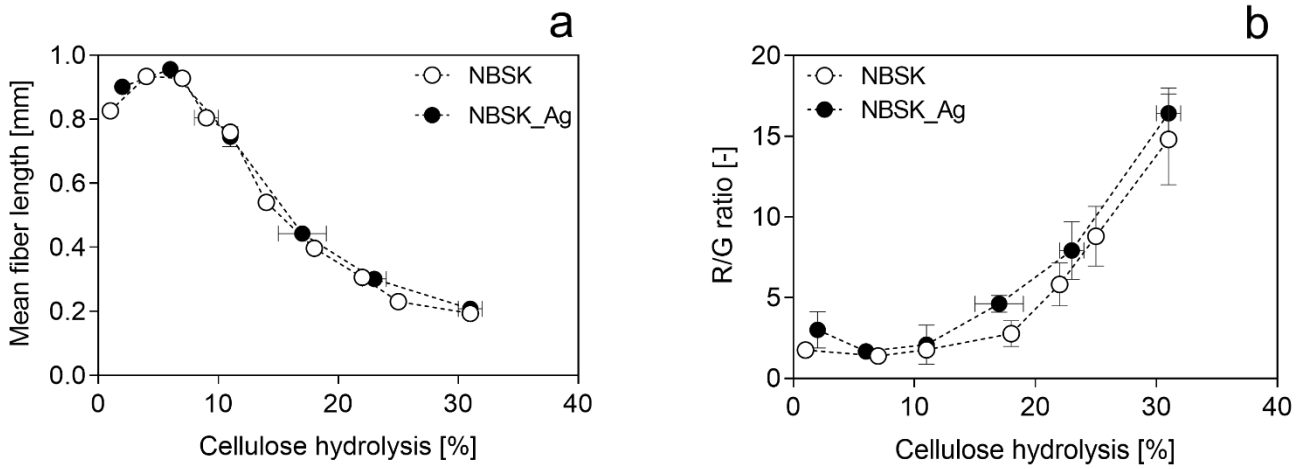

Figure S4. The effect of agitation on the change in mean fiber length (a) and R/G ratio (b) with increasing cellulose hydrolysis. NBSK_Ag was incubated in shaken flasks under continuous agitation in an orbital shaker (150 rpm), as presented in the main manuscript. NBSK was incubated in 2 mL tubes in a heating block, with regular periodical vortexing.

Supplementary methodology: Quantitative image analysis

1. Data acquisition

Confocal laser scanning microscopy (CSLM) imaging was performed with an Olympus FV1000 (Olympus, Japan), using a 10x (NA 0.3-0.4) air objective, and a Zeiss 710 LSM (Carl Zeiss AG, Germany), using a 10x (NA 0.3-0.4) air objective and a 40x (NA 1.2) water objective. The respective instruments capture micrographs of 1024×1024 and 512×512 pixels (picture elements), respectively. Following the specifications of the dyes, CBM17-FITC was excited at 570 nm and emissions detected at 590-650 nm. The excitation and emission of CBM2a-RRedX was at 490 nm and 510-550 nm, respectively. Images were acquired in 5-6 µm (10x) and 0.4-0.5 µm (40x) thick optical sections. Bleed-through of fluorescence emission was tested and can be neglected for the presented image quantification method.

2. Software

Quantitative image analysis of the CLSM images were performed in a Python 2.7 environment (build 2.7.10, Python Software Foundation). Image data acquisition, processing, and visualization were performed using the open source toolboxes Bio-Formats (build 5.7.1, Open Microscopy Environment (OME), https://www.openmicroscopy.org/)[1]with a python wrapper (python-bioformats 1.4.0, OME) [2], the open source computer vision library for python (opencv-python 3.4.3.18, https://opencv.org/) [3], the python imaging library (PIL) fork Pillow (build 5.2.0, https://python-pillow.org/) [4], Matplotlib (build 2.2.3, https://matplotlib.org/) [5], NumPy (build 1.15.1, http://www.numpy.org/) [6], and the SciPy library (build 1.2.0, https://www.scipy.org/) [7]. Supporting micrograph metadata was extracted using ImageJ (build 1.52a, https://imagej.net) [8].

3. Image processing

Quantitative image analysis of the acquired CLSM images was performed to assess changes in the relative proportion of the adsorbed fluorophores and the distribution of features in the specimens with increasing cellulose conversion. High throughput could be achieved with a high degree of automation of the image analysis. The micrographs were batch processed. Batch data and user supplied inputs were retrieved from a spreadsheet and the acquired micrographs was analyzed in a 3-step process: preprocessing, segmentation, and analysis. The workflow of the image analysis is illustrated in Figure S5 and is described more in detail below.

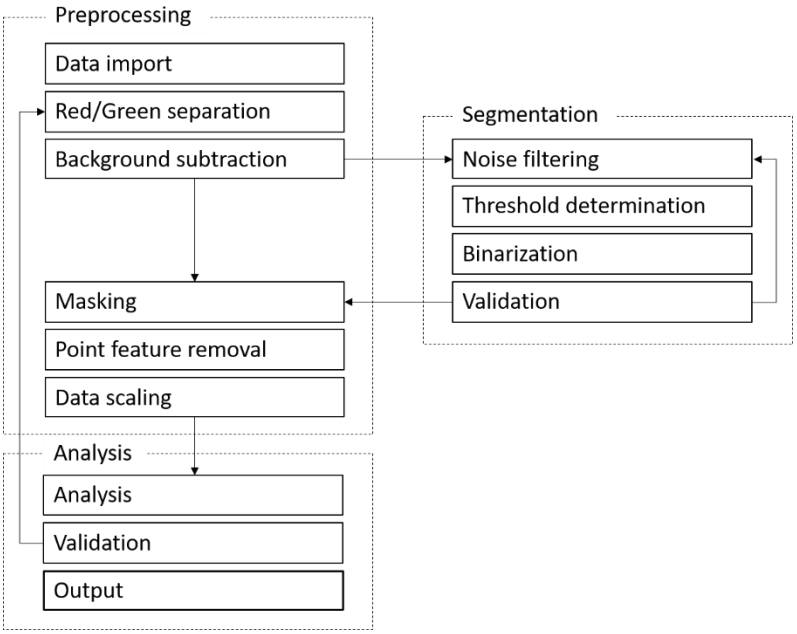

Figure S5. Overview of the quantitative image analysis workflow.

3.1 Preprocessing.

The image preprocessing stage encompassed import of user specified micrographs and metadata to the processing environment, extraction of intensity data, and formatting of data for fiber identification (i.e. segmentation) and image analysis. The data preprocessing module was built on code by Dr Carl J. Houtman (USDA Forest product laboratory, Madison, WI) [9].

Data import. The acquired micrographs were imported to the image processing framework as 3D NumPy arrays using OME Bio-Formats, and micrograph metadata, embedded and user supplied (read from spreadsheets), for the image analysis were appended. The acquired 12-bit micrograph, read as 16-bit image scaled from 0 to 1 by the data import environment, were reshaped to 12-bit data with a dynamic range from 0 to 4095.

Red/green image separation. The spatial FITC and RRedX fluorescence intensities were extracted from the acquired 3D NumPy array as separate grayscale images, i.e. 2D NumPy arrays containing intensity information from respective fluorescent label. FITC and RRedX fluorescence data are henceforth referred to as “green” and “red” layer.

Background subtraction. The modes of the background intensities of the “red” and “green” layers, acquired from blank slide micrographs analyzed with ImageJ, were subtracted from the “red” and “green” layers of the analyzed micrograph. Duplicate layers were created for the image segmentation process.

Masking. The pixels in the layers that represent fiber area, the region of interest (ROI), were isolated from the background. The corrected layers were masked with the binary mask created by the image segmentation process (section 3.2. Segmentation).

Point feature removal. Saturated pixels and negative values in the masked “red” and “green” layers were removed.

Data scaling. The intensities in the “red” and “green” layers were scaled with the respective laser transmission factors used to control the excitation of the fluorescent labels (set to maximize the signal-to-noise ratio within the dynamic range of the detector).

3.2 Segmentation

The stage encompassed unsupervised identification of fiber from the background, definition of a ROI, and the creation of a binary mask for isolation of the ROI sparse matrix in the image analysis. Duplicate “red” and “green” layers were created and manipulated for the segmentation operations.

Noise filtering. A median filter with a 7×7 kernel was applied to reduce spatial noise in the copies of the “red” and “green” layers. An OpenCV subroutine was used to replace each pixel with the center value of the neighbors in the kernel, efficiently reducing “salt-and-pepper” noise.

Threshold determination. The filtered layers were segmented based on contrasting intensities between fibers (foreground) and background by binary thresholding, using OpenCV subroutines. Nonparametric and unsupervised methods were used to automate the selection of threshold value, *T*. The threshold value for the strongly bimodal histogram of the “red” layer was selected using Otsu’s method [10], which find the optimal threshold value by minimizing the intra-class variances of the foreground and background. The lower signal-to-noise ratio (S/N-ratio) of the “green” layer, and the weaker bimodality, made segmentation more complex. The low S/N-ratio makes the segmentation sensitive to various disturbances, e.g. illumination, resulting in multi-modal background in the “green” layer histogram. Robust selection of threshold value for the “green” layer was achieved by setting the mode of the background class as the left boundary of the domain over which the threshold is computed, using contrast stretching, and selecting the threshold value with the triangle method [11].

Binarization. The duplicated “red” and “green” layers were segmented by applying the determined threshold values on the respective layers, creating binary representations of the layers. The ROI was defined as the union of the binary ROI subsets identified in the “red” and “green” layers. The identified ROI was used to mask the corrected fluorescence layers.

Verification. In order to verify that the previous steps lead to correct identification of the ROI, the contours of the identified ROI were superimposed on the original micrograph and the “red” and “green” layers for manual comparison. Correction of irregular features in the micrograph that does not coincide with fibers, such as the boundaries of air bubbles and nonspecifically labelled debris in the specimen, were manually removed from the duplicated “red” and “green” layers before being reprocessed.

3.3 Analysis and output

In the analysis step, the relative proportion of “red” and “green” fluorescence were estimated for every pixel, descriptive statistics calculated, the numerical appropriateness of the methodology assessed, and the outputs exported.

Analysis. A fluorescence ratio imaging approach was used to derive a consistent quantitative measure which can be used to assess the temporal change of the surface structure of the fibers during enzymatic hydrolysis. The basic premise of the method is that the fluorescence ratio of “red” to “green” (R/G-ratio) represents a measure of the relative proportion of the adsorbed fluorophores, reflecting the relative abundance of specific surface structures. Because of differences in optical properties of fluorophores (e.g. extinction coefficients and quantum yield), the R/G-ratio has an arbitrary magnitude which is proportional to the ratio of the absolute concentrations of the two fluorophores.

The mean R/G-ratio was estimated by inversion from the sparse matrices using the estimator below (eq 1), where R and G were the masked "red" and "green" layers.

$$r_{R/G} = \overline{\left(R(1/G)\right)} \qquad \text{(eq 1)}$$

The mean R/G-ratio estimate was complemented with statistical properties of estimator (variance and sample size) and descriptive statistics of the underlying data sets (min, max, mean, and variance).

Validation. The validation of the appropriateness of the segmentation and analysis results was made by visual inspection of the segmented micrographs and evaluation of the distribution of the $(1/G)$ part of the estimator, which could bias the mean. As $G_{i,j}$ approach 0, $(1/G)_{i,j}$ approach infinity, which also would be reflected in the variance of the estimator. Inflated values of $(1/G)_{i,j}$ was primarily associated with image elements along the boundary of the ROI with low S/N-ratios, unduly included in the ROI under relaxed segmentation criteria.

Output. The quantitative analysis results were output to a user specified spreadsheet. The data output is assembled in summary statistics and figures in external software.

4. References

[1] Linkert M, Rueden CT, Allan C, Burel J-M, Moore W, Patterson A, Loranger B, Moore J, Neves C, MacDonald D, Tarkowska A, Sticco C, Hill E, Rossner M, Eliceiri KW, & Swedlow JR (2010) Metadata matters: Access to image data in the real world. *J Cell Biol* 189(5):777-782. DOI: 10.1083/jcb.201004104

[2] Lamprecht MR, Sabatini DM, & Carpenter AE (2007) CellProfiler™: Free, versatile software for automated biological image analysis. *Biotechniques* 42(1):71-75. DOI: 10.2144/000112257

[3] Bradsky V, Pisarevsky WG, Itzees, et al. Open Source Computer Vision Library (version 3.4.3) [Open Source library]. OpenCV.org. Retrieved from https://opencv.org/, Released 20 September 2018.

[4] Clark A, et al. Pillow, the friendly PIL fork (version 5.2.0) [Open Source library]. Python-pillow.org. Retrieved from https://github.com/python-pillow/Pillow, Released 1 July 2018. DOI: 10.5281/zenodo.596518.

[5] Hunter JD (2007) Matplotlib: A 2D graphics environment. *Comput Sci Eng* 9(3):90-95. DOI: 10.1109/MCSE.2007.55

[6] van der Walt S, Colbert C, & Varoquaux G (2011) The NumPy Array: A structure for efficient numerical computation, computing in science & engineering, 13, 22-30. DOI: 10.1109/MCSE.2011.37

[7] Jones E, Oliphant E, Peterson P, et al. (2001-) SciPy: Open Source Scientific Tools for Python. http://www.scipy.org/ [Online; accessed 2018-12-20]

[8] Schneider CA, Rasband WS, & Eliceiri KW (2012), NIH Image to ImageJ: 25 years of image analysis, *Nat Methods* 9(7): 671-675. DOI: 10.1038/nmeth.2089

[9] Houtman CJ, Kitin P, Houtman, JCD, Hammel KE, & Hunt CG (2016). Acridine orange indicates early oxidation of wood cell walls by fungi. *PloS one*, 11(7). DOI: 10.1371/journal.pone.0159715

[10] Otsu N (1979) A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern* 9(1):62-66. DOI: 10.1109/TSMC.1979.4310076

[11] Zack G, Rogers W, & Latt S (1977) Automatic measurement of sister chromatid exchange frequency. *J Histochem Cytochem* 25(7):741-753. DOI: 10.1177/25.7.70454