Supplementary Information

How lignin sticks to cellulose—insights from atomic force microscopy enhanced by machine-learning analysis and molecular dynamics simulations†

Diego M. Nascimento[a], Felippe M. Colombari[b], Bruno Focassio[a,c], Gabriel R. Schleder [a,c], Carlos A. R. Costa[a], Cleyton A. Biffe[a], Liu Y. Ling [b], Rubia F. Gouveia [a,c], Mathias Strauss[a], George J. M. Rocha[b], Edson Leite[a,d], Adalberto Fazzio [a,c], Rodrigo B. Capaz[a,e], Carlos Driemeier *[b] and Juliana S. Bernardes *[a,c]

[a]Brazilian Nanotechnology National Laboratory (LNNano), Brazilian Center for Research in Energy and Materials (CNPEM), CEP 13083-970 Campinas, São Paulo, Brazil. E-mail: juliana.bernardes@lnnano.cnpem.br.

[b]Brazilian Biorenewables National Laboratory (LNBr), Brazilian Center for Research in Energy and Materials (CNPEM), CEP 13083-970 Campinas, São Paulo, Brazil. E-mail: carlos.driemeier@lnbr.cnpem.br.

[c]Center for Natural and Human Sciences, Federal University of ABC (UFABC), CEP 09606-070 Santo André, São Paulo, Brazil.

[d]Department of Chemistry, Federal University of São Carlos (UFSCAR), CEP 13565905 São Carlos, São Paulo, Brazil.

[e]Instituto de Física, Universidade Federal do Rio de Janeiro (UFRJ), CEP 21941-972 Rio de Janeiro, Rio de Janeiro, Brazil.

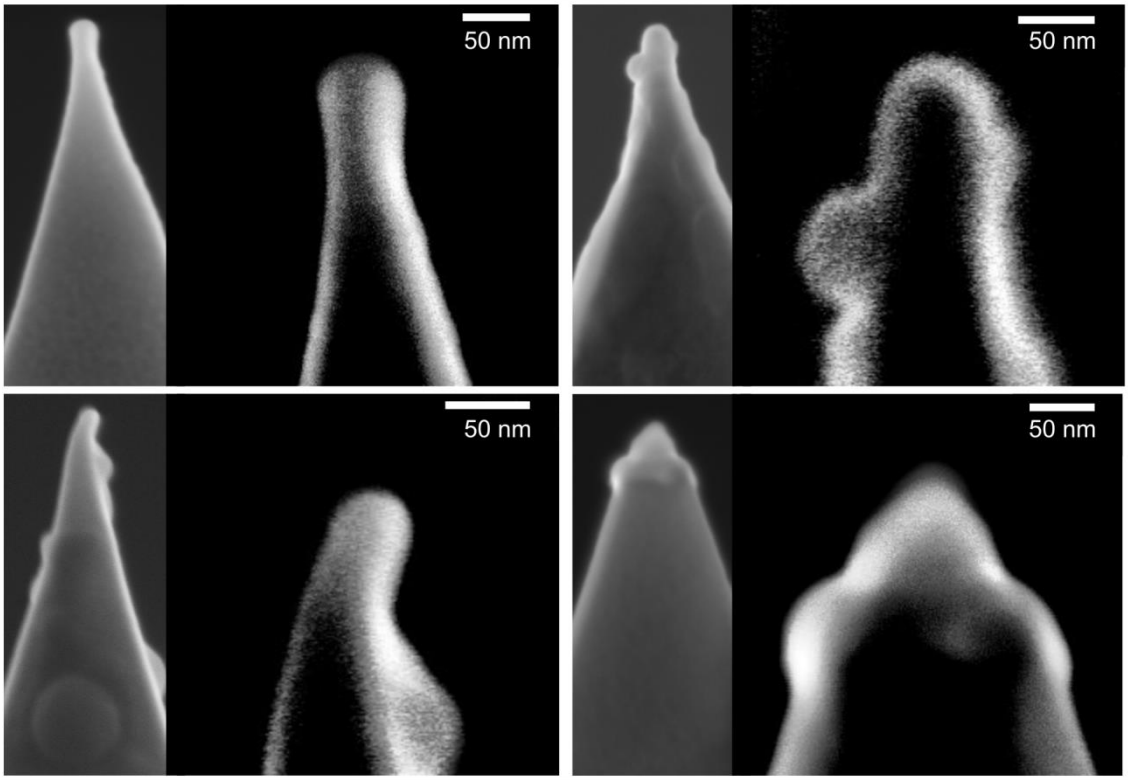

Fig. S1. SEM images of lignin-coated tips. All images were collected at acceleration voltage of 5.0 kV under high vacuum.

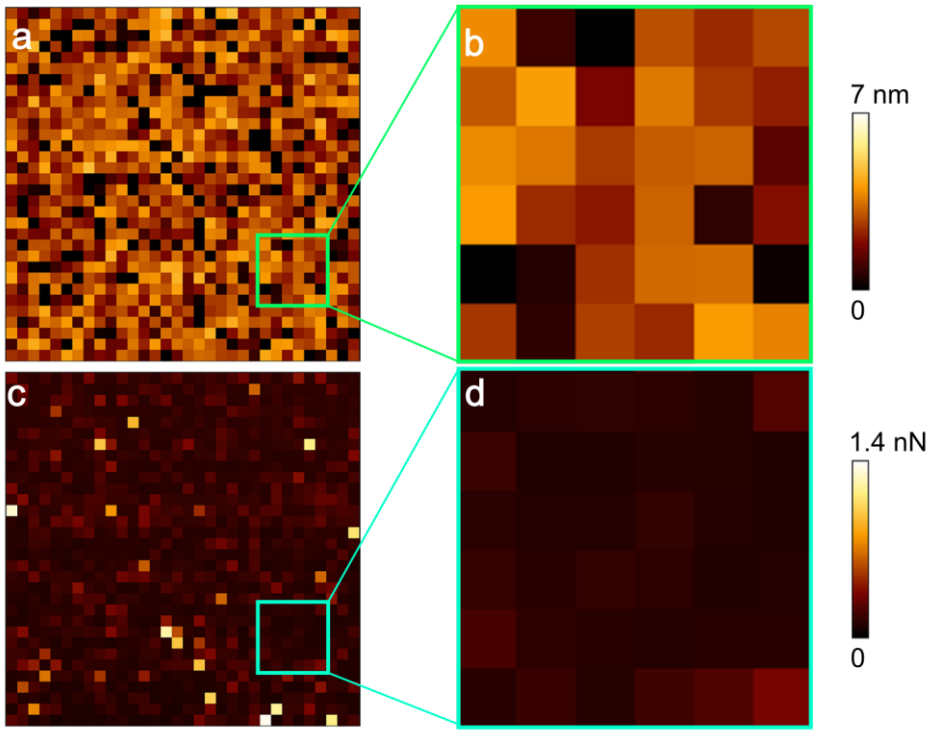

Fig. S2. AFM Imaging of the MFC using lignin-coated tip in liquid water. a-b Topography. c-d Adhesion maps. The green and blue squares in a-c indicate the zoom-in of the topography and adhesion maps, respectively.

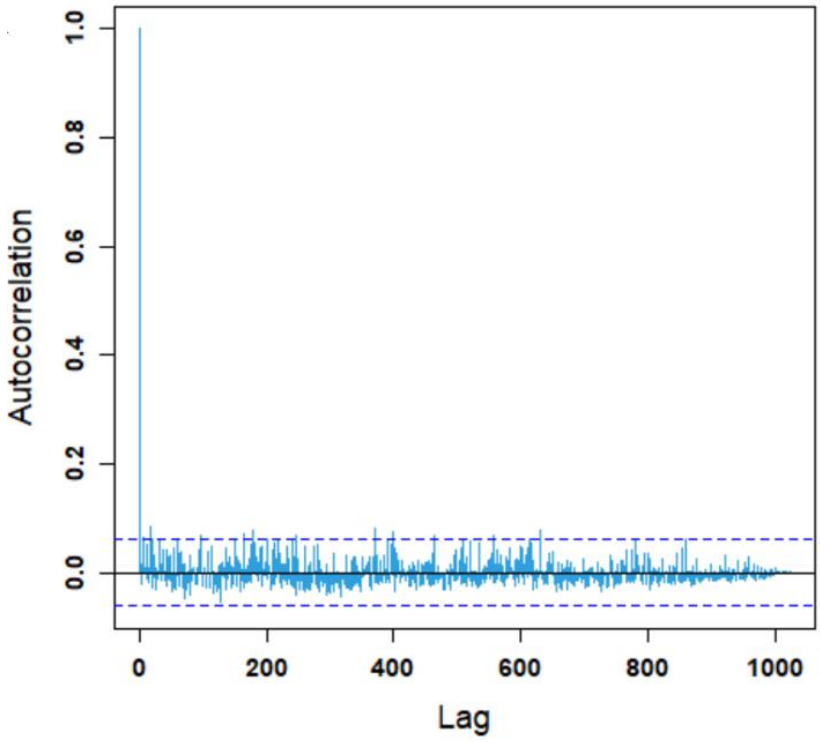

Fig. S3. Representative autocorrelation function of the adhesion force for MFC.

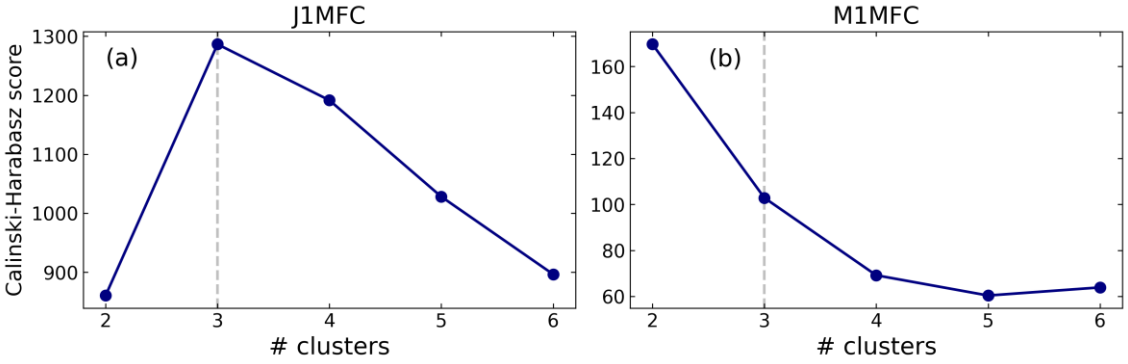

Fig. S4. Calinski-Harabasz (CH) score (variance ratio criterion) for two representative force curve sets (a) J1MFC and (b) M1MFC, each set obtained with a different lignin-coated tip. The best clustering choice is suggested by high CH scores, reflecting dense and separated clusters. Results suggest 3 clusters for J1MFC and 2 clusters for M1MFC. Therefore, 3 clusters should be chosen for consistency across the full set of lignin-coated tips. Inspection of the M1MFC dataset (Fig. 3a of the main text) suggests the lacking third cluster in the CH plot (b) is due to the low frequency (therefore statistical low weight) of the high-adhesion force curves. Conversely, relatively higher frequency of the high-adhesion curves in the J1MFC dataset results in the prominent peak that suggests 3 clusters in the CH plot (a).

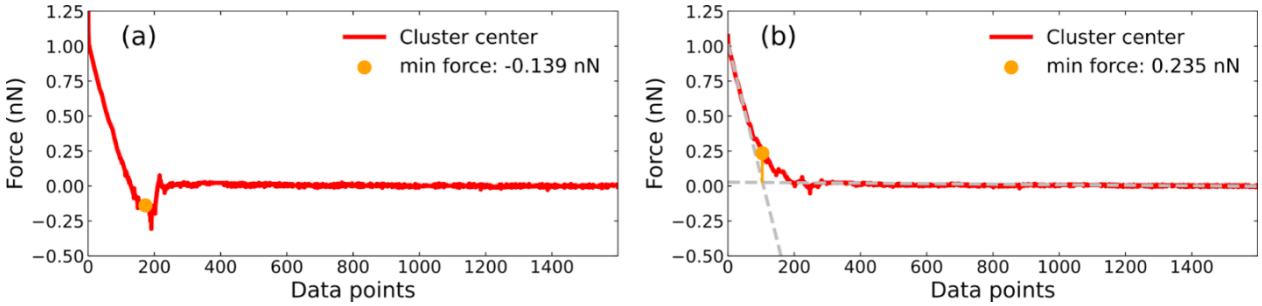

Fig. S5. Minimum force criteria used for extracting the adhesion force. (a) Force curve where the minimum force is negative. The minimum point (orange circle) is determined after application of the Savitzky-Golay filter to reduce noise. (b) Force curve with purely repulsive forces between tip and surface. The contact point is defined by the crossing point between the two linear portions of the curve, represented by dashed grey lines. The repulsive force at the contact point is then given by the orange circle.

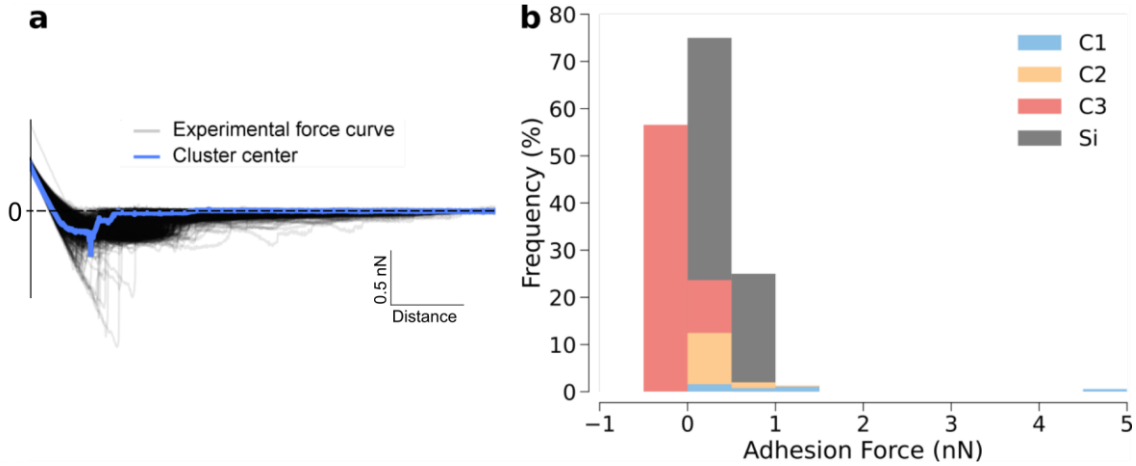

Fig. S6. AFM pull-off force measurements. F-d curves during retraction (a) for SiO_2/Si and representative frequency histograms of adhesion force (b) for MFC (C1, C2, and C3) and SiO_2/Si probed with lignin-coated tip.

Intermolecular energy components. Average intermolecular energy components were obtained for each Umbrella Sampling window to understand further the driving force behind the cellulose-lignin binding free energies and attractive forces (Fig. 5a-b). MD simulations allow to decompose the intermolecular interaction energy contributions into electrostatic (Coulombic) and van der Waals (Lennard-Jones) for various components (lignin, cellulose, and water) of the system. For each contribution, the energy variation along the path was obtained from the difference between bounded and unbounded states. Thus, for each separation distance trajectory of each cellulose surface, interaction energy components of cellulose-lignin, cellulose-water, and water-water pairs were analyzed, as shown in Supplementary Fig. S7.

For the lignin-cellulose interaction, the energy changes become more favorable when the lignin globule moves closer to the cellulose surfaces. Additionally, the sum of these energetic contributions (solid green line in Fig. S7) does not show a notable difference between the cellulose surfaces, regardless of the lignin being close to or far away from the surface. Nevertheless, when the lignin reaches closer to the (100) surface, the van der Waals interaction (solid purple line in Fig. S7) is significantly more favored than the electrostatic interaction. Our results are consistent with the previous MD simulations about cellulose-lignin interactions.[3,4,5]

As regards cellulose-water interactions, this energy change is expected to become more positive as lignin approaches the cellulose surface since fewer solvent molecules will be in the solvation shell. Further, the magnitude of such changes can be attributed to the different interactions of cellulose surfaces with water. Interestingly, in Fig. S7, more significant energy changes were observed for (010) and $(1\bar{1}0)$, whereas lower were found for (100) and (110). This result indicates that escaping water molecules of these latter surfaces is more favorable and can lead to more favorable cellulose-lignin binding.

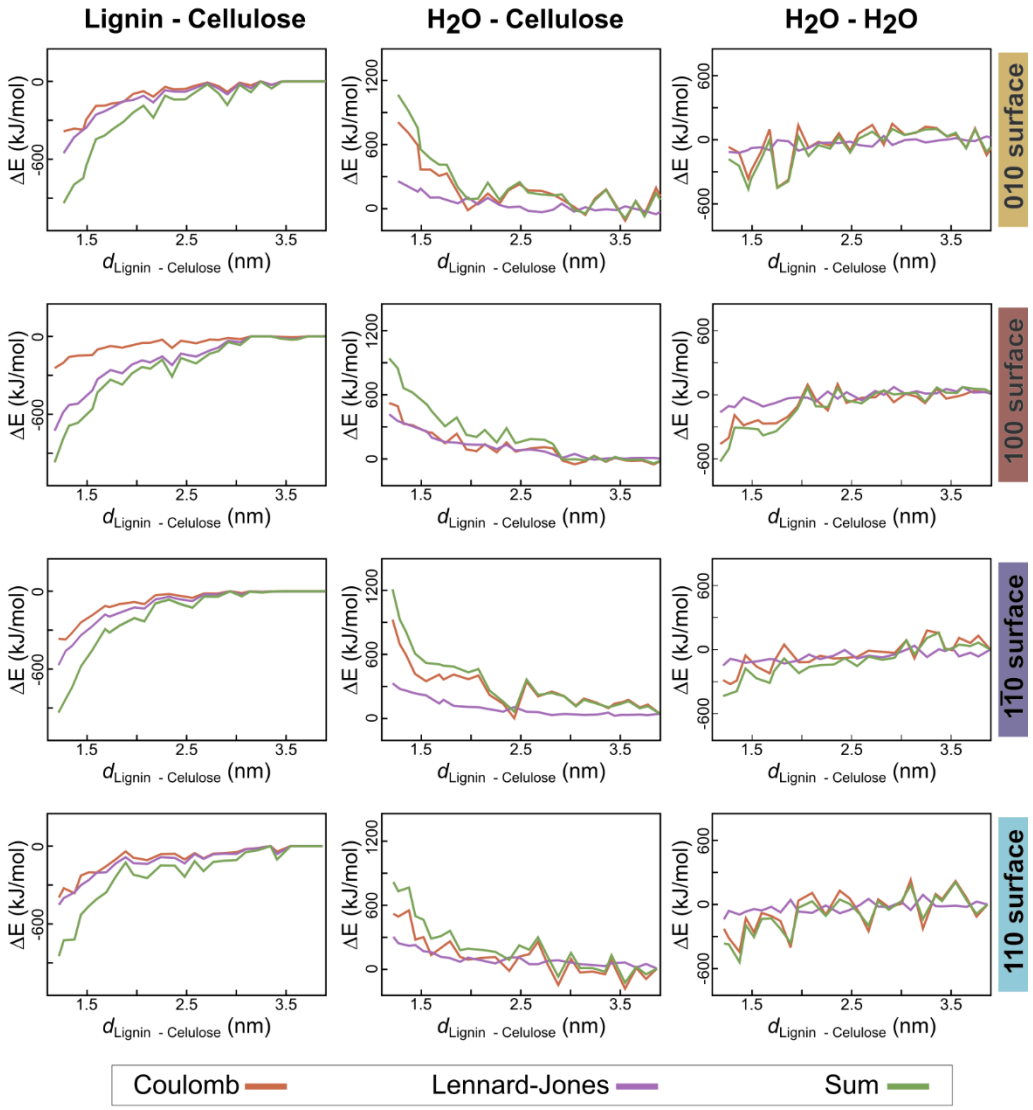

Fig. S7. Interaction energy for each Umbrella Sampling. Interaction energy between cellulose (010), (100), (1$\bar{1}$0), and (110) surfaces and lignin, cellulose, and water, and between water molecules as a function of the lignin – cellulose approximation coordinate. Energy values were separated into electrostatic (Coulomb potential) and Van der Waals (Lennard-Jones potential) contributions.

Umbrella Sampling details. For each simulation window (a total of 32 for each cellulose surface), the lignin structure samples a different portion of the separation coordinate. This sampling strategy can be visualized by plotting the probability histograms as a function of the lignin-cellulose separation distance along the coordinate (Figure S8). It is important to highlight that each one of the 32 simulations start with a discrete position but allows back-and-forth fluctuation of the lignin position through a harmonic restraining potential. As a result, the overlap of these individual histograms results in the full coverage of the separation coordinate.

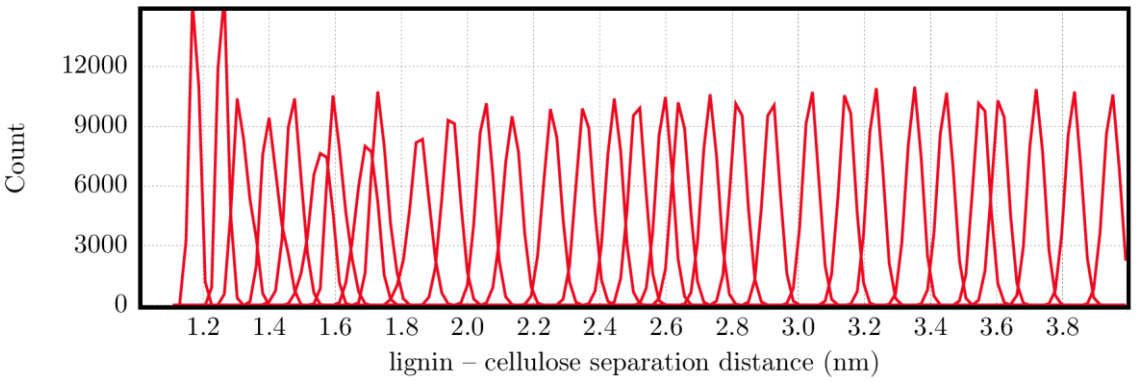

Fig. S8. Probability histograms of the lignin position along the separation coordinate. Each histogram corresponds to an Umbrella Sampling simulation of 100 ns and covers a r ± dr distance.

References

1 T. Caliński and J. Harabasz, *Commun. Stat.-Theory Methods*, 1974, **3**, 1–27.
2 P. J. Rousseeuw, *J. Comput. Appl. Math.*, 1987, **20**, 53–65.
3 S. Besombes and K. Mazeau, *Plant Physiol. Biochem.*, 2005, **43**, 277–286.
4 J. V. Vermaas, M. F. Crowley and G. T. Beckham, *ACS Sustain. Chem. Eng.*, 2019, **7**, 19570–19583.
5 S. Youssefian and N. Rahbar, *Sci. Rep.*, 2015, **5**, 1–13.