SUPPORTING INFORMATION: First Hyperpolarizability of Cellulose Nanocrystals: an Experimental and Theoretical Investigation

Thibaut Legat[1,2], François Mairesse[3], Ahmet R. Dok[1], Yovan de Coene[4], Wim Thielemans[2], Benoît Champagne[3], Stijn Van Cleuvenbergen[1]**

[1]Molecular Imaging and Photonics, Department of Chemistry, KU Leuven, Campus Kulak Kortrijk, Etienne Sabbelaan 53, 8500 Kortrijk, Belgium

[2]Sustainable Materials Lab, Department of Chemical Engineering, KU Leuven, Campus Kulak Kortrijk, Etienne Sabbelaan 53, 8500 Kortrijk, Belgium

[3]Theoretical Chemistry Laboratory, Unit of Theoretical and Structural Physical Chemistry, NISM (Namur Institute of Structured Matter), University of Namur (UNamur), B-5000 Namur, Belgium

[4]Molecular Imaging and Photonics, Department of Chemistry, KU Leuven, Celestijnlaan 200D, 3001 Heverlee, Belgium

*stijn.vancleuvenbergen@kuleuven.be

*benoit.champagne@unamur.be

KEYWORDS: Hyper Rayleigh Scattering, Cellulose Nanocrystals, First Hyperpolarizability, Ab inito,

ESI. 1: EXPRESSIONS FOR DEGREE OF SUBSTITUTION

Habibi *et al.* suggested an expression for the calculation of the fraction of chains exposed on the surface of the nanocrystals.[1] The total number of chains at the surface, denoted by n_S, is determined by counting the rods around the edge. This corresponds to twice the width (L_2) and twice the height (L_1) divided by their respective plane spacing distance, $d_{(110)}$ and $d_{(1\bar{1}0)}$. The total number of chains, $\sum n$, is given by the cross sectional area of the rod divided by the cross sectional area of half the unit cell ($d_{(110)}\, d_{(1\bar{1}0)}$), which contains a single chain. This calculation simplifies to eqn (S1) below:

$$R_C = \frac{n_S}{\sum n} = \frac{2\left(\frac{L_1}{d_{(110)}}\right) + 2\left(\frac{L_2}{d_{(1\bar{1}0)}}\right)}{\frac{L_1 L_2}{d_{(110)}\, d_{(1\bar{1}0)}}} = 2\left(\frac{d_{(1\bar{1}0)}}{L_2}\right) + 2\left(\frac{d_{(110)}}{L_1}\right) \qquad (S1)$$

The edge of the rod can be thought of as being a cellulose chain with tg hydroxymethyl conformation which is held by intramolecular hydrogen bonding between O3/O5 and O2/O6 hydroxyl groups. Because of the two-fold helical twisting of subsequent anhydroglucose units, the hydroxymethyl and secondary hydroxyl groups point out of the crystal on alternating AGU units. Some ^{13}C-NMR evidence that shows that there could be a certain amount of gg conformation of the hydroxyl groups at the surface however these contributions are small.[2]

To determine the degree of grafting on CNC surface, we must first express the number of hydroxyl groups at the surface before chemical modification. The number of OH moles per gram on the surface, N_{OH}, of the CNC is expressed using eqn (S2).

$$N_{OH} = \frac{n_1 + n_2}{\rho N_A L_1 L_2 c}\left(\frac{L_1 + L_2}{d_{(110)}} + \frac{L_1 + L_2}{d_{(1\bar{1}0)}}\right) + 2\left(\rho N_A L_3 d_{(110)} d_{(1\bar{1}0)}\right)^{-1} \quad (S2)$$

where n_1 is the number of primary hydroxyl groups facing (110) (or $(1\bar{1}0)$) in the unit cell and n_2 is the number of secondary hydroxyl groups facing (110) (or $(1\bar{1}0)$) in the unit cell. If we consider the chains at the edge of the rods to retain a tg conformation similar to the bulk, then n_1 = 1 and n_2 = 2. There are two chains in the unit cell, but since these are alternating this results in half the total amount of exposed hydroxyl groups at the surface. The other parameters include, N_A, Avogadro's constant, c the lattice parameter, and ρ the cellulose density. The first term of eqn (S2) corresponds to the hydroxyl groups located around the edges of the nanocrystal, while the second term corresponds to the terminating hydroxyl groups located at the ends of nanocrystal. The terminating hydroxyl groups correspond to the ones located at C1 and C4. Notably, the number of terminating hydroxyl groups per gram of cellulose is inversely proportional to the length L_3 of the crystal. However, due to the large aspect ratio L_3/L_1 of the CNC, the terminating hydroxyl groups are a very small percentage of the total amount.

The degree of surface substitution, DS_{surf}, is expressed by the number of modified moles per gram at the surface, N_{mod}, divided by the total number of available hydroxyl groups, N_{OH}. N_{mod} can be determined experimentally through conductometric titration. The total degree of substitution, DS is calculated as the product of the ratio of surface chains, R_C, multiplied by DS_{surf}. These equations are expressed in eqn (S3) and eqn (S4):

$$DS_{surf} = \frac{N_{mod}}{N_{OH}} \quad (S3)$$

$$DS = R_C\, DS_{surf} \quad (S4)$$

ESI. 2: DEGREE OF SUBSTITUTION – CONDUCTOMETRIC TITRATION

The surface charge density is determined by conductometric titration. A 15 mL sample of CNC suspension at 0.86 wt. % was passed through an H^+ ion exchange column before measurement. The suspension was held at 25 °C and a conductimetry probe was added to the suspension to track the conductivity. An aqueous NaOH solution (0.01 M) was added at a feeding rate of 0.1 mL/min with continuous stirring. The change in conductivity was recorded using a conductivity meter (Horiba Ltd., DS-51). The neutralization of free H^+ ions occurs through the addition of NaOH, forming Na^+ ions which bind to the CNC sulfate groups. This leads to a decrease in the concentration of positively charged free ions in the suspension, leading to a decrease in conductivity. Once all sulfate groups have been neutralized, additional of NaOH increases conductivity once more. The titration to 2.88 mL of 0.01 M NaOH accounts to a 0.224 mmol/g sulfate groups per gram of cellulose.

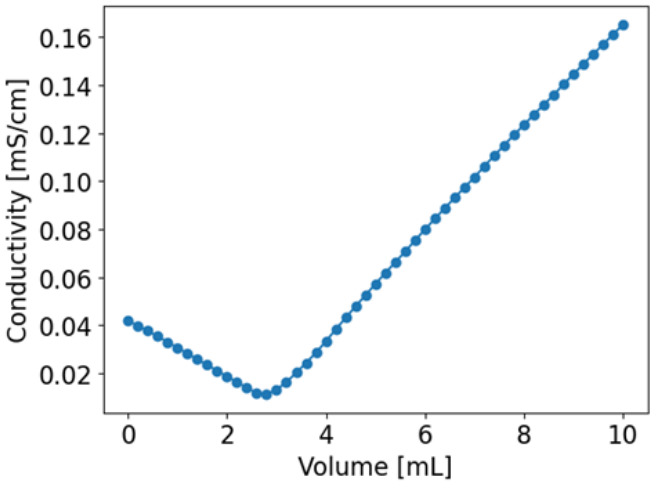

Figure S1: Titration curve. The point of inflection is used to determine the amount of sulfated grafted groups though the addition NaOH

ESI. 3: DEGREE OF SUBSTITUTION OF CNC

Using equations S1 and S2 and the AFM dimensions of the CNC, the ratio of exposed chains and number of exposed hydroxyl group on the surface of the CNC can be calculated. The average dimensions for L_1 = L_2 = 7.51 nm. The number of sulfate groups is obtained from the conductometric titration. The degree of substitution at the surface is given by eqn. S3. The calculations are developed hereunder:

R_C : Ratio of chains exposed on the surface of the CNC

$$R_C = \frac{n_s}{\sum n} = \frac{2\left(\frac{L_1}{d_{110}}\right) + 2\left(\frac{L_2}{d_{1\bar{1}0}}\right)}{\frac{L_1 L_2}{d_{(110)} \times d_{(1\bar{1}0)}}} = \frac{2\left(\frac{7.51}{0.531}\right) + 2\left(\frac{7.51}{0.596}\right)}{\frac{7.51 \times 7.51}{0.531 \times 0.596}} = 0.30$$

N_{OH} *:* Number of hydroxyl groups on the surface of the CNC (mmol/g)

$$N_{OH} = \frac{n_1 + n_2}{\rho N_A L_1 L_2}\left(\frac{L_1 + L_2}{d_{(110)}} + \frac{L_1 + L_2}{d_{(1\bar{1}0)}}\right) + 2(\rho N_A L_3 d_{110}) = 2.88 \text{ mmol/g}$$

N_{SO_4-} *:* Number of sulfate groups per gram of cellulose (mmol/g)

$$N_{SO_4-} = \ 0.22 \text{ mmol/g}$$

DS_{surf} : Degree of substitution at the surface

$$DS_{surf} = \frac{N_{mod}}{N_{OH}} = \frac{0.22}{2.88} = \ 0.076$$

DS: Degree of substitution of all chains:

$$DS = R_C \, DS_{surf} \ = \ 0.023$$

In total, 7.6 % of unit cells at the surface and 2.3 % of all unit cells have been modified. The sulfate grafted unit cells remain small with respect to the non-grafted ones. Furthermore, the sulfate groups are not aligned in the same orientation as is the case with cellulose chains. Indeed, they are grafted on the (110) and $(1\bar{1}0)$ planes and not on equivalent hydroxyl groups. For this reason, the

effective second order hyperpolarizability of the sulfate groups cannot be added cumulatively. Consequently, it is expected that the overall contributions of the grafted sulfate moieties to the second order nonlinear response remain negligible compared to the bulk response of the cellulose chains.

ESI. 4: CONCENTRATION MEASUREMENTS VIA THERMAL ANALYLSIS

A Netzsch F3 Tarsus thermogravimetric analyser was used to determine the wt. % concentration of the CNC suspension. The method was carried out in nitrogen atmosphere and involved heating of the sample during two segments: the first segment included heating to 85°C at a rate of 10°C/minute, followed by an isothermal period of 30 minutes. During the second segment heating was continued to 150°C at a rate of 10°C /minute. The water content was determined in the linear region of the resulting TGA curve, as the mass loss at 85°C.

ESI. 5: ABSORPTION SPECTRA – UV-VIS

HRS raw data has not been corrected with the absorption spectra which would reduce the amount of measurable signal. These have nonetheless been carried out using UV-Vis and shown in Fig. S2. The absorption spectra is small for low weight concentrations and furthermore, the laser beam was aimed at the edge of the quartz cuvette at less than 10 % of the total optical path of the cuvette. For the highest concentrations, we observe less than 0.2 absorption which accounts for less than 0.02 absorption when we consider the reduced optical path length of the HRS measurement.

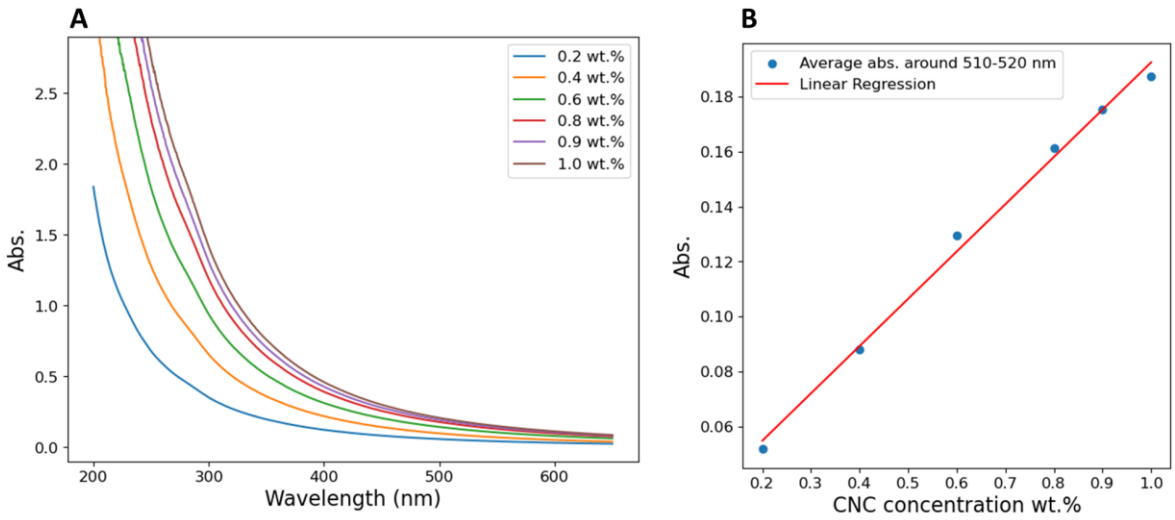

Figure S2: Absorption spectra (A) UV-Vis spectra of CNC concentration series (B) Linear plot of average abs. around 510-520 nm as a function of concentration. The HRS signal would experiences minimal signal loss since the laser is focused on the edge of the cuvette.

ESI. 6: QUADRATIC REGIME – ABSENCE OF HIGHER ORDER EFFECTS

To ensure that we were within the quadratic response of our CNC and to ensure that there was no occurrence of third order effects in our signal. A power sweep was performed to check for the quadratic regime of the response. Additionally, the power sweeps for each concentration was fitted to a quadratic function of the form $f(x) = B2\,x^2 + cst$. The fitted B2 coefficients are then plotted against CNC concentration. Both these plots are shown in Fig. S3 which clearly demonstrate a quadratic power dependence. This quadratic regime demonstrates the absence of higher order effects (for instance Kerr effects) which would deviate the response.

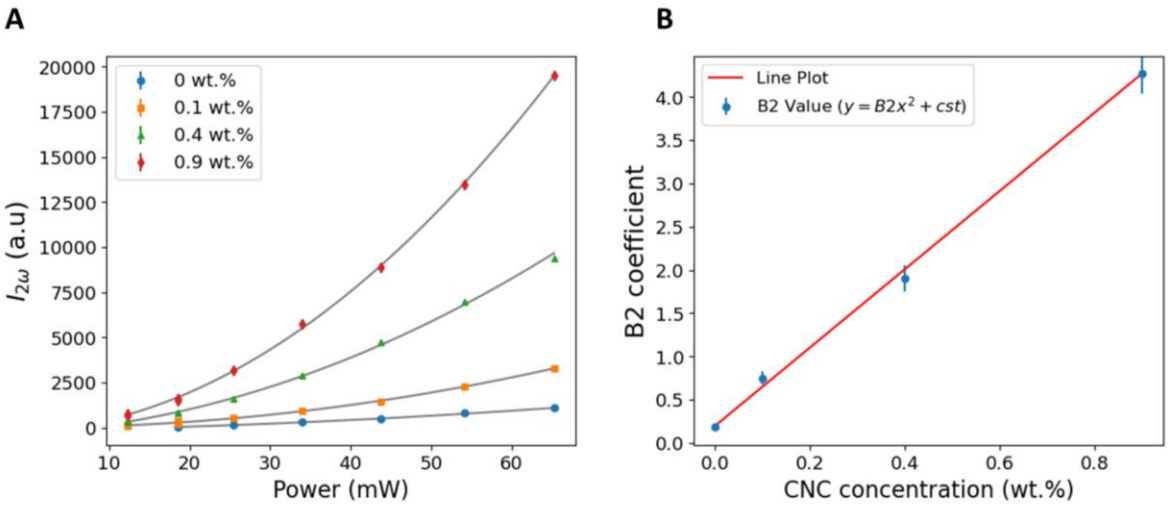

Figure S3: Power sweep. Confirming that our measurement are carried out in the quadratic regime, without the occurrence of third order effects. (A) Second harmonic response with respect to varying powers for four different samples (0, 0.1 wt.%, 0.4 wt.% and 0.9 wt.%). (B) The quadratic coefficient (B2) of the power sweep with respect to the CNC concentration.

ESI. 7: HYPER-RAYLEIGH SCATTERING THEORETHICAL BACKGROUND

Hyper-Rayleigh scattering (HRS) involves the scattering of second-harmonic light, which is governed by the molecular first hyperpolarizability, β. In an HRS experiment, the concentration of the solute is kept low to ensure that each scatterer is independent and randomly oriented. In other terms, the hyper-Rayleigh signal originates from individual nonlinear scatters that do not exhibit specific phase relations with respect to one another in the solution. For uncorrelated scatterers, the hyper-Rayleigh intensity is given by eqn (S5) (Clays et al. 1994):

$$I_{2\omega} = \frac{16\pi^5}{c\lambda^4 r^2} N\, f_{\omega}^4\, f_{2\omega}^2\, \beta_{\mathrm{HRS}}^2\, I_{\omega}^2 \qquad (S5)$$

where N is the concentration of chromophores, and f_{ω} and $f_{2\omega}$ are local field factors. I_{ω} the intensity at the fundamental wavelength, $I_{2\omega}$ the intensity at the second harmonic, r the distance to the scattering molecule, and c the speed of light in vacuum.

The scattering geometry for HRS is defined using the surface of the optical table as the horizontal reference plane (XY). The fundamental light beam is propagating in the X-direction and polarized in the Z-direction. The scattered light is collected in the Y-direction (i.e at 90° of the fundamental light beam) and through the use of an analyzer, SH signal polarized in the X direction and the Z direction can be isolated. These measurable quantities are defined as β_{ZZZ}^2 and β_{XZZ}^2. Here the first subscript (Z or X) refers to the polarization state of the frequency double light in the laboratory coordinate system. The second and third subscript correspond to the two photons polarized in Z of the incoming beam. The scattering geometry showing the macroscopic and molecular frames are provided in Fig. S4.

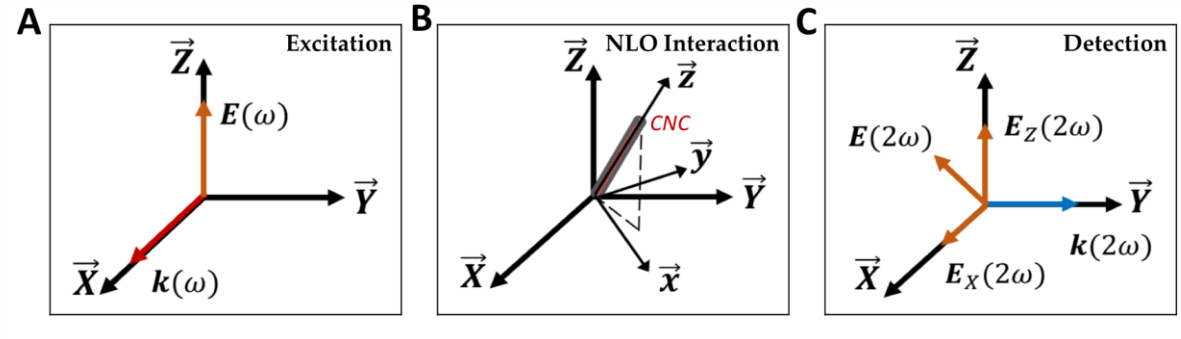

Figure S4: Scattering geometry of HRS setup. ***(A)*** *Incoming beam with frequency ω, propagating in the X-direction and polarized in the Z-direction of the macroscopic frame.* ***(B)*** *A CNC with arbitrary orientation and it's molecular frame at an angle to the macroscopic frame. This illustration depicts a single CNC however a large number of CNC over all orientations are found within interaction volume.* ***(C)*** *The detected SH signal at frequency 2ω is used to determine* β^2_{ZZZ} *and* β^2_{XZZ} *by the use of an analyzer in the Z-direction for* $E_Z(2\omega)$ *and the X-direction for* $E_X(2\omega)$ *respectively.*

The macroscopic measurable quantities are related by the molecular first hyperpolarizability components β_{ijk} by the eqn (S6) and eqn (S7) which have been demonstrated by the works of Cyvin et al.[3] These are as follows:

$$\beta^2_{ZZZ} = \frac{1}{7}\sum_{i}\beta^2_{iii} + \frac{6}{35}\sum_{i\neq j}\beta_{iii}\beta_{ijj} + \frac{9}{35}\sum_{i\neq j}\beta^2_{iij} + \frac{6}{35}\sum_{ijk,cyclic}\left(\beta_{iij}\beta_{jkk} + \frac{12}{35}\beta^2_{ijk}\right) \qquad (S6)$$

$$\beta^2_{XZZ} = \frac{1}{35}\sum_{i}\beta^2_{iii} - \frac{2}{105}\sum_{i\neq j}\beta_{iii}\beta_{ijj} + \frac{11}{105}\sum_{i\neq j}\beta^2_{iij} - \frac{2}{105}\sum_{ijk,cyclic}\left(\beta_{iij}\beta_{jkk} + \frac{8}{35}\beta^2_{ijk}\right) \qquad (S7)$$

Whereby ijk refers to the molecular coordinate system. The sum over "ijk, cyclic" corresponds to ijk being equal to either xyz, zxy, or yzx. Since both polarizations are detected with equal sensitivity, the total orientational average β^2_{HRS} is the sum of both these macroscopic responses.

$$\beta_{HRS}^2 = \beta_{ZZZ}^2 + \beta_{XZZ}^2 \quad (S8)$$

ESI. 8: INTERNAL REFERENCE METHOD

In the internal reference method, the first hyperpolarizability, $\beta_{HRS,x}$ of the sample is determined directly from the first hyperpolarizability of the solvent, $\beta_{HRS,S}$. For very dilute solutions such as those used in HRS experiments, the refractive index of the solvent is assumed not to change. In a two-component system consisting of the solvent and randomly oriented molecules, the total measured HRS signal will be the sum of both uncorrelated scatterers. The HRS signal is the cumulative response of the solvent (*S*) and the randomly oriented molecule (*x*) to give:

$$I_{2\omega,x} = g\left[N_S\beta_{HRS,S}^2 + N_x\beta_{HRS,x}^2\right] I_\omega^2 \quad (S9)$$

This equation can be rearranged and expressed as a linear expression as a function of solute concentration, N_x whereby

$$\begin{aligned} I_{2\omega,x} &= (g\,\beta_{HRS,x}^2\, I_\omega^2).N_x + N_S\,(g\,\beta_{HRS,S}^2\, I_\omega^2) \\ &= a\,N_x + b \end{aligned} \quad (S10)$$

With a = $(g\,\beta_{HRS,x}^2\, I_\omega^2)$ being the slope and b = $N_S(g\,\beta_{HRS,S}^2\, I_\omega^2)$ being the y-intercept. These values can be experimentally determined through the collection of a concentration series of the solute. From the known slope and the y-intercept of the linear fit and rearranging the expression, we obtain for $\beta_{HRS,x}$:

$$\beta_{HRS,x} = \sqrt{\frac{slope}{y-intercept}N_S}\beta_{HRS,S} \quad (S11)$$

For the first hyperpolarizability of the solvent, the values used are the ones by Campo et al.[4] The static first hyperpolarizability, $\beta_{zzz,0}$ are provided, and by using the two-state model we can obtain the β_{zzz} at a specific wavelength

$$\beta_{zzz} = \frac{\beta_{zzz,0}}{\left[1 - \left(\frac{\lambda_{eg}}{\lambda}\right)^2\right]\left[1 - \left(\frac{\lambda_{eg}}{\lambda/2}\right)^2\right]} \quad (S12)$$

Where λ_{eg} is the transition wavelength and λ is the fundamental wavelength. For uniaxial dipolar molecule, the first hyperpolarizability for HRS of the solvents can be calculated by the following:

$$\beta_{\mathrm{HRS,S}} = \sqrt{\frac{6}{35}}\beta_{zzz,S} \quad (S13)$$

Using the linear plot of the concentration series to determine the slope and the y-intercept, and the known first hyperpolarizability of the solvent, one can determine the first hyperpolarizability of the sample.

ESI. 9: ERROR PROPAGATION OF THE INTERNAL REFERNCE METHOD WHEN DERIVING β_{HRS}

The reported error for the HRS intensity is the standard deviation from five successive acquisitions for each data point. Fig S5. provides an example of raw HRS data that was obtained for the 0.6 wt. % CNC suspension.

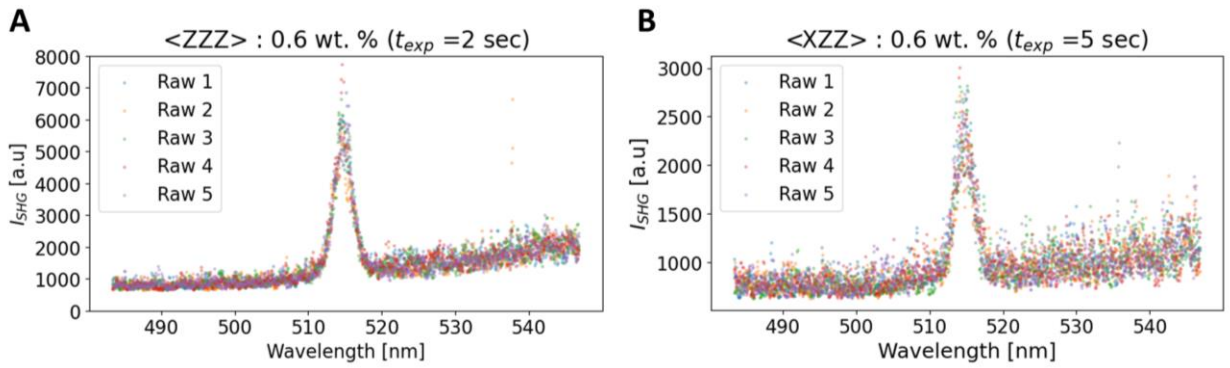

Figure S5: Raw data acquisition: *Each data point was acquired five times, with each spectrum individually fitted to determine the area under the curve. The error for each data point was calculated as the standard deviation of the five acquisitions. Exposure times were adjusted based on relative concentrations to prevent detector saturation with all other acquisition parameters remaining constant. An example of five raw spectra detections for 0.6 wt.% CNC is shown in (**A**) <ZZZ> with a 2-second exposure time and in (**B**) <XZZ> with a 5-second exposure time.*

A linear regression with weighted sum of least squares on each data point gave an overall fit error for the linear regression for the slope ($a \pm \Delta a$) and y-intercept ($b \pm \Delta b$). The equation for β_{HRS} is then provided by the following:

$$\beta_{HRS} \pm \Delta\beta_{HRS} = \sqrt{\frac{a \pm \Delta a}{b \pm \Delta b} N_S \beta_{HRS,S}} \qquad (S14)$$

The propagation of the error, $\Delta\beta_{HRS}$, is then calculated from the sum of the partial derivatives and their own errors. It expands to the following:

$$\Delta\beta_{HRS} = \left|\frac{\partial\beta(a,b)}{\partial a}\right| \Delta a + \left|\frac{\partial\beta(a,b)}{\partial b}\right| \Delta b \qquad (S15)$$

where we have for both partial derivatives:

$$\frac{\partial\beta(a,b)}{\partial a} = \frac{Cst}{2b}.\frac{1}{\sqrt{a/b}} \quad (S16)$$

$$\frac{\partial\beta(a,b)}{\partial b} = -\frac{Cst}{2b}.\frac{a}{\sqrt{ab^3}} \quad (S17)$$

With $Cst = \sqrt{N_S}\,\beta_{\mathrm{HRS,S}}$

From eqns S15 – S17, the error for $\Delta\beta_{\mathrm{HRS}}$ can be determined.

ESI. 10: NEGLIGIBLE COHERENCE CONTRIBUTION

For an ensemble of molecules with fixed relative positions, phase relations result in the appearance of coherent effects. If the dimensions of this ensemble remain small compared to the scattering wavelength, then we can still consider them as point sources.[5] Additionally, as mentioned earlier, since the NLO response of the bulk of the response depends on the square of the volume of the aligned harmonophores (V^2), the contribution of the bulk will quickly dominate over surface effects.

We can demonstrate that due to the relatively small size of the CNC combined with the wavelengths of our experiment, that the angular dependence can be neglected. We considered the NIST SANS Data reduction for a parallelepiped of dimensions A,B and C.[6] In this context L_1=A, L_2=B and L_3 = C. Although we have in our main discussion L_1=L_2, this equality cannot be used in the model therefore we consider $L_2 > L_1$ and increase the size of one of the edges by a small amount. The aim here is not to get an accurate determination of the angle dependence but rather to demonstrate that it remains small for the 90° HRS setup to remain valid i.e isotropic scattering even for larger CNC crystals. The response of the parallelepiped model for dimensions L_1 = 7 Å, L_2 = 7.5 Å and L_3 = 170 Å is provided in Fig. S6. The SANS equations are expressed through the reduced lengths, normalized to the middle ones whereby

$$a = \frac{A}{B}; \quad b = \frac{B}{B} = 1; \quad c = \frac{C}{B} \tag{S18}$$

The form factor is calculated as a double integral by the following:

$$P(q) = \frac{scale}{V_P} \int_0^1 \phi_Q \left(\mu\sqrt{1-\sigma^2}, a\right) [S(uc\sigma/2)]^2 d\sigma \tag{S19}$$

With

$$\sigma_Q(\mu, a) = \int_0^1 \left(S\left[\mu/2cos\left(\frac{\pi}{2}u\right)\right].S\left[\mu a/2sin\left(\frac{\pi}{2}u\right)\right]\right)^2 du \tag{S20}$$

Where the function $S(x)$ is defined as

$$S(x) = \frac{sin(x)}{x} \tag{S21}$$

And

$$\mu = qB \tag{S22}$$

Whereby q, is the scattering vector and expressed by:

$$q = \frac{4\pi}{\lambda} sin\left(\frac{\theta}{2}\right) \tag{S23}$$

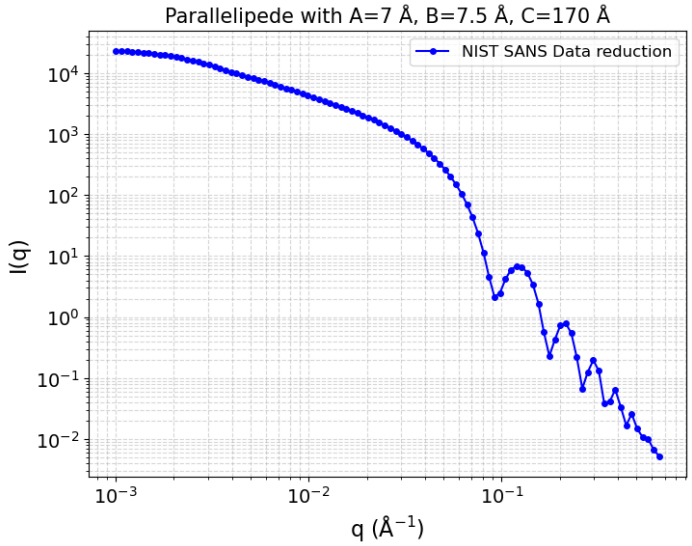

Figure S6: Parallelepiped scattering curve: *Parallelepiped with dimensions of 7 Å, 7.5 Å, and 170 Å, plotted on a logarithmic scale.*

The numerical aperture (NA) of our setup is 0.165, corresponding to a total collection angle (2θ) of approximately 15°. Fig. S7 displays the responses of parallelepipeds of increasing sizes, using the scattered wavelength intensity of 515 nm for angles from 0 - 180°. The angular dependence becomes more pronounced as the size of the parallelepiped increases, in this case the length c (or L_3). However, even for the longest rods at $L_1 = L_2 = 10$ nm and $L_3 = 400$ nm minimal angle dependence around 90° is observed. Indeed, even at the upper limit of these CNCs, the angular dependence remains small (less than 10%). This is an extreme case as the proportion of CNCs measuring 400 nm or longer is very low. According to our probability density function for L_3 of our AFM measurements (eqn. 15) fewer than 0.52% of rods exceed 400 nm in length.

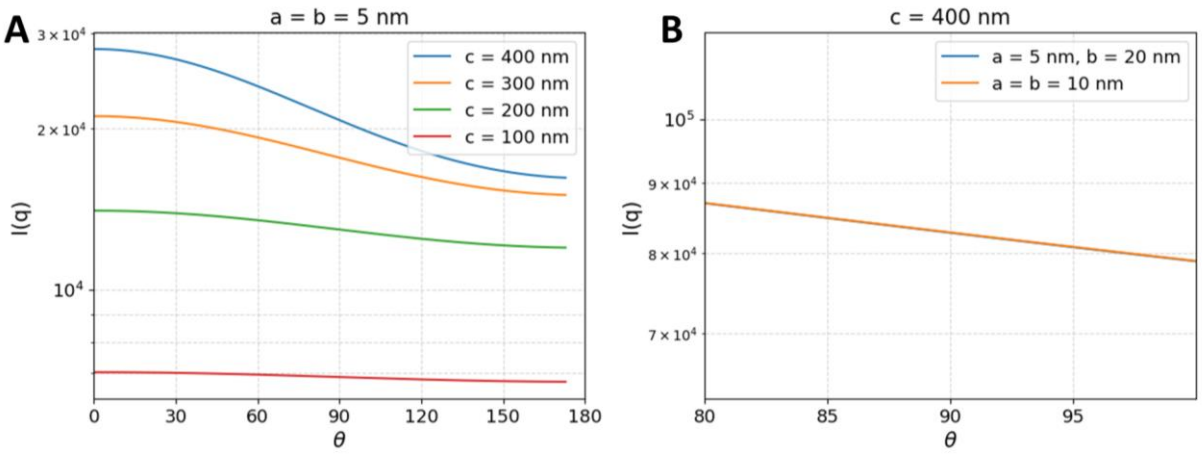

Figure S7: Angle dependence for different CNC dimensions*. (**A**) Parallelepiped scattering curves for different dimensions of a, b and c. (**B**) Scattering curve for c=400 nm and a=10 and b = 10 nm which shows that in the upper limit of long rods, angular dependance remains small (<10%) around values of 90° ± 15°.*

ESI. 11: GEOMETRY OPTIMIZATION OF THE CRYSTAL UNIT CELL AT THE DFT LEVEL WITH PERIODIC BOUNDARY CONDITIONS – THEORETICAL FRAMEWORK

The following calculation parameters were employed: TOLINTEG 9 9 9 30 70, SHRINK 7 7. The latter is related to the density of the grid sampling the first Brillouin zone in view of the k summations/integrations. The TOLINTEG keyword controls the truncations of the lattice summations involved in computing k-dependent matrices and the energy per unit cell. The choice of the XCF, the PBC-related parameters, as well as the basis set were made considering the results of Mairesse et al.[7] Further details about the theoretical framework are available in Dovesi et al.[8]

ESI. 12: UNIT CELL FIRST HYPERPOLARIZABILITY AT THE DFT LEVEL WITH PERIODIC BOUNDARY CONDITIONS – THEORETICAL FRAMEWORK

Within Periodic Boundary Conditions, Bloch functions are combined linearly to give crystalline orbitals. A Bloch function φ is defined as:

$$\varphi_q = \sum_{\boldsymbol{g}} \chi_q^{\boldsymbol{g}} e^{i\boldsymbol{k}\cdot\boldsymbol{g}} \tag{S24}$$

where $\boldsymbol{g}$ is a vector pointing from the origin and running to every crystalline cell, $\chi_q^{\boldsymbol{g}}$ is the qth atomic orbital in cell $\boldsymbol{g}$ and $\boldsymbol{k}$ is a vector in the first Brillouin zone.

The polarizability and first hyperpolarizability of a crystalline unit cell can be computed from the LCAO coefficient matrix, its derivatives (with respect to the electric field), some integral matrices, and the eigenvalues of the corresponding crystalline orbital:

$$\alpha_{\zeta\eta}(-\omega;\omega) = \frac{-2}{N_{\boldsymbol{k}}} \Re \sum_{\boldsymbol{k}} \wp_{\zeta/-\omega,\eta/\omega} \left\{ \sum_p \sum_q \sum_i \sum_a C^{(0)*}_{pi,\boldsymbol{k}} M^{\zeta}_{pq,\boldsymbol{k}} C^{(0)}_{qa,\boldsymbol{k}} U^{\eta}_{ai,\boldsymbol{k}}(\omega) \right\} \tag{S25}$$

$\beta_{\zeta\eta\tau}(-2\omega;\omega,\omega)$

$$= \frac{-1}{N_{\boldsymbol{k}}} \Re \sum_{\boldsymbol{k}} \sum_{i} \sum_{a} \wp_{\zeta/-2\omega,\eta/\omega,\tau/\omega} \left\{ U^{\zeta *}_{ai,\boldsymbol{k}}(-2\omega) \left[\sum_{b} G^{\eta}_{ab,\boldsymbol{k}}(\omega) U^{\tau}_{bi,\boldsymbol{k}}(\omega) \right.\right.$$

$$\left.\left. - \sum_{j} U^{\tau}_{aj,\boldsymbol{k}}(\omega) G^{\eta}_{ji,\boldsymbol{k}}(\omega) + \iota \frac{\partial U^{\tau}_{ai,\boldsymbol{k}}(\omega)}{\partial k_{\eta}} \right] \right\} \qquad (S26)$$

where ζ, η, τ are cartesian directions, p and q are Bloch functions indices, $\wp$ is the permutation operator, $\Re$ takes the real part, ι is the imaginary number, $N_{\boldsymbol{k}}$ is the number of vectors sampling the first Brillouin zone, and $C^{(0)}_{qm,\boldsymbol{k}}$ is the LCAO coefficient of the q$^{\text{th}}$ basis function in the m$^{\text{th}}$ crystalline orbital at the $\boldsymbol{k}$ vector (the (0) superscript refers to the unperturbed solution). The i, j and a, b indices refer to occupied and unoccupied crystalline orbitals, respectively. When ζ, η, τ are in superscript, it means that the element is differentiated with respect to the module of the applied electric field in that direction, around zero amplitude. The M and G matrices are defined as

$$M^{\zeta}_{pq,\boldsymbol{k}} = \sum_{\boldsymbol{g}} \left\langle \chi^{0}_{p} \middle| (\boldsymbol{r} + \iota \boldsymbol{\nabla}_{\boldsymbol{k}}) e^{\iota \boldsymbol{k}\cdot\boldsymbol{g}} \middle| \chi^{\boldsymbol{g}}_{q} \right\rangle_{\zeta} = \sum_{\boldsymbol{g}} e^{\iota \boldsymbol{k}\cdot\boldsymbol{g}} \left\langle \chi^{0}_{p} \middle| (\boldsymbol{r} - \boldsymbol{g})_{\zeta} \middle| \chi^{\boldsymbol{g}}_{q} \right\rangle \qquad (S27)$$

$$G^{\zeta}_{\boldsymbol{k}}(\omega) = C^{(0)\dagger}_{\boldsymbol{k}} \left(F^{\zeta}_{\boldsymbol{k}}(\omega) C^{(0)}_{\boldsymbol{k}} + M^{\zeta}_{\boldsymbol{k}} C^{(0)}_{\boldsymbol{k}} + \iota S_{\boldsymbol{k}} \frac{dC^{(0)}_{\boldsymbol{k}}}{dk_{\zeta}} \right) \qquad (S28)$$

where the electric dipole moment operator is replaced by $\boldsymbol{r} + \iota\boldsymbol{\nabla}_{\boldsymbol{k}}$ to prevent breaking the periodicity of the potential, $S_{\boldsymbol{k}}$ is the overlap matrix of the Bloch functions, $F_{\boldsymbol{k}}$ is the Fock matrix defined according to

$$F_{pq,\boldsymbol{k}} = \sum_{\boldsymbol{g}} \left\langle \chi^{0}_{p} \middle| \hat{F} e^{i\boldsymbol{k}\cdot\boldsymbol{g}} \middle| \chi^{g}_{q} \right\rangle \qquad (S29)$$

where $\hat{F}$ is the Fock operator, defined and discussed in Dovesi et al.[8] Finally, the $U_{\boldsymbol{k}}$ matrix, whose off-diagonal elements are computed with

$$U^{\zeta}_{ia,\boldsymbol{k}}(\pm\omega) = \frac{G^{\zeta}_{ia,\boldsymbol{k}}(\pm\omega)}{\varepsilon^{(0)}_{a,\boldsymbol{k}} - \varepsilon^{(0)}_{i,\boldsymbol{k}} \mp \omega} \tag{S30}$$

where ω is the frequency of the incident electric field and $\varepsilon^{(0)}_{\boldsymbol{k}}$ are the crystalline orbital eigenvalues, allows to get $C^{\zeta}_{\boldsymbol{k}}$:

$$C^{\zeta}_{\boldsymbol{k}}(\omega) = C^{(0)}_{\boldsymbol{k}} U^{\zeta}_{\boldsymbol{k}}(\omega) \tag{S31}$$

Equations S27 – S31 are solved self-consistently through the iterative CPKS scheme (called TDDFT in the dynamic case).[9–11]

The unit cell first hyperpolarizability tensors of both the neutron diffraction and optimized structure of I_β cellulose were computed using CRYSTAL17. The selected levels of approximation are i) the 6-311G** basis set, with ii) the CAM-B3LYP [12], ωB97X, and LC-BLYP [13] XCFs. The calculation parameters were: TOLINTEG 9 9 9 30 70, SHRINK 6 6. These choices were made in line with the results of Mairesse et al.[14] In particular, range-separated hybrid XCFs were employed because the NLO responses are intrinsically non-local properties and the XCFs should contain a substantial amount of HF exchange because it displays the correct asymptotic behavior. This condition is fulfilled using global hybrids, where the amount of HF exchange does not depend on the interelectronic distance, provided this amount is large[15] or with range-separated hybrids, where i) the 1/r operator is split into short- and long-range parts by using a smooth function and ii) the short-range part is associated with local/semilocal exchange and the long-range part with HF exchange. Several studies demonstrated that RSHs perform better than local/semilocal XCFs[16] These conclusions were further substantiated in Mariesse *et al,* which deals with organic and inorganic crystals.[14]

ESI. 13: UNIT CELL FIRST HYPERPOLARIZABILITY AT THE DFT LEVEL WITH PERIODIC BOUNDARY CONDITIONS – RESULTS

Table S1. *$\chi^{(1)}$ non-zero components of I_β cellulose, calculated with various XCFs considering the experimental and the DFT-optimized structure, for a* ***static*** *incident field*

	$\chi_{XX}^{(1)}$	$\chi_{XY}^{(1)}$	$\chi_{YY}^{(1)}$	$\chi_{ZZ}^{(1)}$	Band gap (eV)
CAM-B3LYP	0.94	-0.01	1.26	1.41	10.38
ωB97X	0.96	-0.01	1.26	1.40	12.07
LC-BLYP40	1.13	-0.01	1.42	1.58	7.44
LC-BLYP30	1.19	-0.01	1.48	1.63	6.53
LC-BLYP20	1.25	-0.01	1.54	1.70	5.63
LC-BLYP20-optimized	1.56	0.00	1.69	1.84	6.04

Table S2. *$\chi^{(1)}$ non-zero components of I_β cellulose, calculated with various XCFs considering the experimental and the DFT-optimized structure, for a* ***1064 nm*** *wavelength incident field*

	$\chi_{XX}^{(1)}$	$\chi_{XY}^{(1)}$	$\chi_{YY}^{(1)}$	$\chi_{ZZ}^{(1)}$	Band gap (eV)
CAM-B3LYP	0.95	-0.01	1.26	1.42	10.38
ωB97X	0.97	-0.01	1.26	1.41	12.07
LC-BLYP40	1.14	-0.01	1.44	1.59	7.44
LC-BLYP30	1.20	-0.01	1.49	1.65	6.53
LC-BLYP20	1.26	-0.01	1.56	1.71	5.63
LC-BLYP20-optimized	1.58	0.00	1.71	1.86	6.04

Table S3. *$\chi^{(1)}$ non-zero components of I_β cellulose, calculated with various XCFs considering the experimental and the DFT-optimized structure, for a* ***532 nm*** *wavelength incident field*

	$\chi_{XX}^{(1)}$	$\chi_{XY}^{(1)}$	$\chi_{YY}^{(1)}$	$\chi_{ZZ}^{(1)}$	Band gap (eV)
CAM-B3LYP	0.97	-0.01	1.29	1.45	10.38
ωB97X	0.99	-0.01	1.29	1.44	12.07
LC-BLYP40	1.18	-0.01	1.48	1.63	7.44
LC-BLYP30	1.24	-0.01	1.54	1.70	6.53
LC-BLYP20	1.32	-0.01	1.61	1.77	5.63
LC-BLYP20-optimized	1.65	0.00	1.76	1.92	6.04

In Table S4, we have gathered the first hyperpolarizability tensor components of the I_β cellulose unit cell, and the (HOMO-LUMO) electronic band gap, for each of the employed XCFs. The quantities β_{ZZZ}^2, β_{XZZ}^2, β_{HRS}^2, $\beta_{HRS} = \sqrt{\beta_{HRS}^2}$ and DR are also reported. These were calculated from eqn (S6) – (S8). Experimental values are provided as a comparison.

Table S4. *β components and related quantities of I_β cellulose bulk per unit cell computed with several exchange correlation functionals (in 10^{-30} esu, B convention) considering the Nishiyama neutron diffraction geometry, except for LC-BLYP20-optimized where the structure has been optimized with ωB97X, and where the number next to LC-BLYP is the tuned maximum percentage of Hartree-Fock exchange at infinite distance.*

β	LC-BLYP20-optimized	LC-BLYP20	LC-BLYP30	LC-BLYP40	ωB97X	CAM-B3LYP	Experimental
XXX	0.000	0.000	0.000	0.000	0.000	0.000	
XXY	0.000	0.000	0.000	0.000	0.000	0.000	
XXZ	0.448	0.376	0.278	0.209	0.080	0.063	
XYX	0.000	0.000	0.000	0.000	0.000	0.000	
XYY	0.000	0.000	0.000	0.000	0.000	0.000	
XYZ	-0.446	-0.078	-0.005	0.033	0.056	0.067	

XZX	0.448	0.376	0.278	0.209	0.080	0.063	
XZY	-0.446	-0.078	-0.005	0.033	0.056	0.067	
XZZ	0.000	0.000	0.000	0.000	0.000	0.000	
YXX	0.000	0.000	0.000	0.000	0.000	0.000	
YXY	0.000	0.000	0.000	0.000	0.000	0.000	
YXZ	-0.426	-0.043	0.015	0.045	0.058	0.070	
YYX	0.000	0.000	0.000	0.000	0.000	0.000	
YYY	0.000	0.000	0.000	0.000	0.000	0.000	
YYZ	-0.994	-0.928	-0.807	-0.727	-0.584	-0.566	
YZX	-0.426	-0.043	0.015	0.045	0.058	0.070	
YZY	-0.994	-0.928	-0.807	-0.727	-0.584	-0.566	
YZZ	0.000	0.000	0.000	0.000	0.000	0.000	
ZXX	0.398	0.374	0.276	0.207	0.079	0.063	
ZXY	-0.397	-0.013	0.037	0.061	0.065	0.075	
ZXZ	0.000	0.000	0.000	0.000	0.000	0.000	
ZYX	-0.397	-0.013	0.037	0.061	0.065	0.075	
ZYY	-0.982	-0.968	-0.839	-0.752	-0.593	-0.575	
ZYZ	0.000	0.000	0.000	0.000	0.000	0.000	
ZZX	0.000	0.000	0.000	0.000	0.000	0.000	
ZZY	0.000	0.000	0.000	0.000	0.000	0.000	
ZZZ	1.560	1.570	1.340	1.160	0.757	0.896	
β^2_{ZZZ}	0.459	0.391	0.277	0.206	0.097	0.114	
β^2_{XZZ}	0.240	0.211	0.153	0.117	0.061	0.066	
β^2_{HRS}	0.699	0.603	0.429	0.322	0.159	0.180	
β_{HRS}	0.836	0.776	0.655	0.568	0.398	0.425	0.193 ± 0.011
DR	1.916	1.851	1.816	1.767	1.587	1.729	3.111 ± 0.668
Band Gap (eV)	6.04	5.63	6.53	7.44	12.07	10.38	

ESI.14: REDUCING THE TENSOR COMPLEXITY

Table S5. *New first hyperpolarizability components of cellulose nanocrystals (10^{-30} esu, B convention) computed from the depolarization ratio and β_{HRS} obtained from DFT-PBC calculations at 1064 nm, assuming β_{zzz} and β_{zyy} are the only nonzero independent components*

	Initial quantities		Discarded solution			Selected solution		
	β_{HRS}	DR	β_{zzz}	β_{zyy}	$R_{C_{2v}} = \frac{\beta_{zyy}}{\beta_{zzz}}$	β_{zzz}	β_{zyy}	$R_{C_{2v}} = \frac{\beta_{zyy}}{\beta_{zzz}}$
LC-BLYP20-opt	0.836	1.916	0.455	-1.453	-3.197	1.953	-0.954	-0.489
LC-BLYP20	0.776	1.851	0.491	-1.353	-2.755	1.784	-0.922	-0.517
LC-BLYP30	0.655	1.816	0.450	-1.143	-2.543	1.490	-0.797	-0.534
LC-BLYP40	0.568	1.767	0.434	-0.991	-2.285	1.270	-0.713	-0.561
ωB97X	0.398	1.587	0.455	-0.687	-1.509	0.803	-0.571	-0.712
CAM-B3LYP	0.425	1.729	0.352	-0.741	-2.106	0.936	-0.547	-0.584

ESI. 15: ELECTRIC FIELD INSIDE A DIELECTRIC IN A CONTINUOUS MEDIUM – FULL CALCULATIONS

The ratio between the electric field in a cavity and the Maxwell field in a surrounding solvent (here, water with a ε_r = 1.7583 at 1064 nm) is, considering that the shape of the cavity is approximated by an ellipsoid of principal axis a, b, and c, computed with the following formula, see equation 2.79 of Böttcher et al.[17]

$$\frac{E_{\alpha,cellulose}}{E_{\alpha,Maxwell}} = \text{ratio}_\alpha = \frac{\varepsilon_{water}}{\varepsilon_{water} + (\varepsilon_{\alpha,cellulose} - \varepsilon_{water})A_\alpha} \quad \text{(S32)}$$

where α is a cartesian direction along one of the principal axis of the ellipsoid, and A_α is a geometrical parameter defined as

$$A_\alpha = \frac{L_1 L_2 L_3}{2} \int_0^\infty \frac{1}{(s + L_\alpha^2)\sqrt{(s + L_1^2)(s + L_2^2)(s + L_3^2)}} ds \qquad \text{(S33)}$$

where L_α is the length along the α direction and s the integration variable. .

The distribution of the ratios between the nanocellulose crystal L_3/L_1 axis, where L_3 is identified with the z axis of the linear and nonlinear optical quantities (that is, along the cellulose polymer chains), is shown in the following graph.

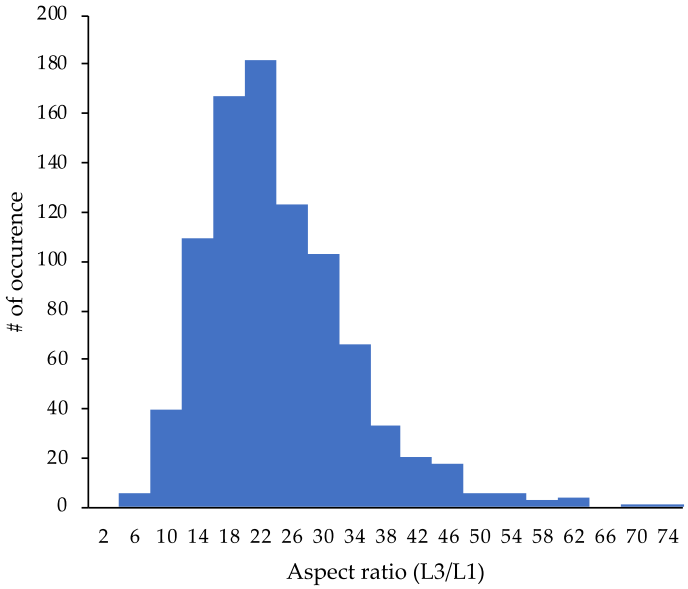

Figure S8: L3/L1 ratio distribution. *The number of occurrences are divided into classes that were obtained from the distribution analysis from the AFM images*

Experimentally, the nanocrystals have been divided into classes of L_3/L_1 ratios between 0 and 4, 4 and 8, 8 and 12, … up to 74, and the middle of the classes range have been taken as the representative ratio of the classes (that is: 2, 6, 10, …). This is illustrated in Fig. S8. Assigning those class representative ratios to the ratios of principal axes lengths of prolate ellipsoids, we

compute the A_α parameters. Since the integral has been carried numerically, the quality of the geometrical parameters is monitored by looking at their sum, which is in principle exactly equal to one. [18]

Table S6. *Distribution of the L3/L1 ratios, divided in classes, as well as the class weights and the corresponding ellipsoid geometrical parameters*

Class	# of occurrence	Weight in distribution	$A_x = A_y$	A_z	Sum(A_α's)
2	0	0.0000	0.41322	0.17356	1.00000
6	6	0.0068	0.47839	0.04323	1.00000
10	40	0.0450	0.48986	0.02029	1.00000
14	109	0.1227	0.49400	0.01200	1.00000
18	167	0.1881	0.49599	0.00801	1.00000
22	182	0.2050	0.49711	0.00577	1.00000
26	123	0.1385	0.49781	0.00438	1.00000
30	103	0.1160	0.49828	0.00344	1.00000
34	66	0.0743	0.49861	0.00279	1.00000
38	33	0.0372	0.49885	0.00231	1.00000
42	20	0.0225	0.49903	0.00195	1.00000
46	18	0.0203	0.49917	0.00167	1.00000
50	6	0.0068	0.49928	0.00144	1.00000
54	6	0.0068	0.49937	0.00126	1.00000
58	3	0.0034	0.49944	0.00112	1.00000
62	4	0.0045	0.49950	0.00099	1.00000
66	0	0.0000	0.49955	0.00089	1.00000
70	1	0.0011	0.49960	0.00080	1.00000

74	1	0.0011	0.49963	0.00073	1.00000
	Average		0.49661±0.01985	0.00677±0.03971	

We also give the bulk cellulose permittivity's computed with the various XCFs, as well as their first hyperpolarizability components obtained assuming C_{2v} symmetry and $\beta_{zxx} = \beta_{xzx} = \beta_{xxz} = 0$. The x and y directions are considered as being equivalent in the model, so the relative permittivity taken into account to obtain the $\frac{E_{\alpha,cellulose}}{E_{\alpha,Maxwell}}$ ratio is the same for both of them, and equal to the mean value of the computed relative permittivity's.

Table S7. *Cellulose crystal permittivity's computed at the DFT-PBC level with various XCFs for a 1064 nm incident field*

	LC-BLYP20 (optimized)	LC-BLYP20	LC-BLYP30	LC-BLYP40	ωB97X	CAM-B3LYP
ε_z	2.860	2.714	2.647	2.590	2.413	2.419
$\frac{\varepsilon_x + \varepsilon_y}{2}$	2.643	2.410	2.345	2.290	2.114	2.106

Considering the weights of the different classes of nanocrystals as well as the $\frac{E_{\alpha,cellulose}}{E_{\alpha,Maxwell}}$ ratios in the x, y, and z directions, we get that the field intensity is different in the z and in the x or y direction, so that the first hyperpolarizability tensor elements are not measured with respect to the same field intensity reference. Therefore, the ratio between β_{zyy} and β_{zzz} is affected. To take this effect into account, we compute a new $R_{C_{2v}}$ ratios by multiplying those tensor components by the square of the corresponding $\frac{E_{\alpha,cellulose}}{E_{\alpha,Maxwell}}$ ratios, which are shown in Tables S8 and S9. In these Tables,

we see that, due to the similar geometrical parameters once the L3/L1 ratio is over 10, the electric field ratios quickly converge. Also, due to the weights of the various L3+L1 ratio classes, the mean value is almost identical to the converged value, with a small standard deviation.

$$\beta_{zyy} \rightarrow \beta_{zyy}\left(\frac{E_{y,cellulose}}{E_{y,Maxwell}}\right)^2 \quad (S34)$$

$$\beta_{zzz} \rightarrow \beta_{zzz}\left(\frac{E_{z,cellulose}}{E_{z,Maxwell}}\right)^2 \quad (S35)$$

The new tensor components are shown in Tables S10 and S11, and we naturally observe the same convergence as for the electric field ratios. From these, scaled β_{HRS} and depolarization ratios are calculated, see Table S12.

Table S8. *Ratio between the electric field inside the cellulose cavity and the Maxwell electric field (1064 nm) in the solvent surrounding it, for* ***the y (or x) component****, considering the classes in the distribution of nanocrystals*

Class	LC-BLYP20-opti	LC-BLYP20	LC-BLYP30	LC-BLYP40	ωB97X	CAM-B3LYP
2	0.828	0.867	0.879	0.889	0.923	0.925
6	0.806	0.849	0.862	0.874	0.912	0.914
10	0.802	0.846	0.859	0.871	0.910	0.912
14	0.801	0.845	0.858	0.870	0.909	0.911
18	0.800	0.845	0.858	0.870	0.909	0.911
22	0.800	0.844	0.858	0.869	0.909	0.911
26	0.800	0.844	0.857	0.869	0.908	0.910
30	0.800	0.844	0.857	0.869	0.908	0.910
34	0.799	0.844	0.857	0.869	0.908	0.910
38	0.799	0.844	0.857	0.869	0.908	0.910

42	0.799	0.844	0.857	0.869	0.908	0.910
46	0.799	0.844	0.857	0.869	0.908	0.910
50	0.799	0.844	0.857	0.869	0.908	0.910
54	0.799	0.844	0.857	0.869	0.908	0.910
58	0.799	0.844	0.857	0.869	0.908	0.910
62	0.799	0.844	0.857	0.869	0.908	0.910
66	0.799	0.844	0.857	0.869	0.908	0.910
70	0.799	0.844	0.857	0.869	0.908	0.910
74	0.799	0.844	0.857	0.869	0.908	0.910
Average	0.800±0.054	0.845±0.057	0.858±0.058	0.869±0.059	0.909±0.062	0.911±0.062

Table S9. *Ratio between the electric field inside the cellulose cavity and the Maxwell electric field (1064 nm) in the solvent surrounding it, for* ***the z component****, considering the classes in the distribution of nanocrystals*

Class	LC-BLYP20-opti	LC-BLYP20	LC-BLYP30	LC-BLYP40	ωB97X	CAM-B3LYP
2	0.902	0.914	0.919	0.924	0.939	0.939
6	0.974	0.977	0.979	0.980	0.984	0.984
10	0.987	0.989	0.990	0.991	0.993	0.992
14	0.993	0.994	0.994	0.994	0.996	0.996
18	0.995	0.996	0.996	0.996	0.997	0.997
22	0.996	0.997	0.997	0.997	0.998	0.998
26	0.997	0.998	0.998	0.998	0.998	0.998
30	0.998	0.998	0.998	0.998	0.999	0.999
34	0.998	0.998	0.999	0.999	0.999	0.999
38	0.999	0.999	0.999	0.999	0.999	0.999
42	0.999	0.999	0.999	0.999	0.999	0.999
46	0.999	0.999	0.999	0.999	0.999	0.999

50	0.999	0.999	0.999	0.999	0.999	0.999
54	0.999	0.999	0.999	0.999	1.000	1.000
58	0.999	0.999	0.999	0.999	1.000	1.000
62	0.999	0.999	0.999	1.000	1.000	1.000
66	0.999	1.000	1.000	1.000	1.000	1.000
70	0.999	1.000	1.000	1.000	1.000	1.000
74	1.000	1.000	1.000	1.000	1.000	1.000
Average	0.996±0.067	0.996±0.067	0.997±0.067	0.997±0.067	0.997±0.068	0.997±0.068

Table S10. Scaled β_{zyy} *(10^{-30} esu, B convention) obtained after multiplication by the ratio of the electric field inside the cavity and the Maxwell electric field (1064 nm) in the solvent surrounding it, for the y component, considering the classes in the distribution of nanocrystals*

Class	LC-BLYP20-opti	LC-BLYP20	LC-BLYP30	LC-BLYP40	ωB97X	CAM-B3LYP
2	-0.654	-0.694	-0.615	-0.563	-0.486	-0.467
6	-0.620	-0.665	-0.592	-0.544	-0.475	-0.456
10	-0.614	-0.661	-0.588	-0.541	-0.473	-0.454
14	-0.612	-0.659	-0.587	-0.539	-0.472	-0.454
18	-0.611	-0.658	-0.586	-0.539	-0.472	-0.453
22	-0.611	-0.658	-0.586	-0.539	-0.471	-0.453
26	-0.610	-0.657	-0.586	-0.538	-0.471	-0.453
30	-0.610	-0.657	-0.585	-0.538	-0.471	-0.453
34	-0.610	-0.657	-0.585	-0.538	-0.471	-0.453
38	-0.610	-0.657	-0.585	-0.538	-0.471	-0.453
42	-0.610	-0.657	-0.585	-0.538	-0.471	-0.453
46	-0.610	-0.657	-0.585	-0.538	-0.471	-0.453
50	-0.609	-0.657	-0.585	-0.538	-0.471	-0.453
54	-0.609	-0.657	-0.585	-0.538	-0.471	-0.453
58	-0.609	-0.657	-0.585	-0.538	-0.471	-0.453
62	-0.609	-0.657	-0.585	-0.538	-0.471	-0.453
66	-0.609	-0.657	-0.585	-0.538	-0.471	-0.453
70	-0.609	-0.657	-0.585	-0.538	-0.471	-0.453
74	-0.609	-0.657	-0.585	-0.538	-0.471	-0.453
Average	-0.611±0.041	-0.658±0.045	-0.586±0.040	-0.539±0.036	-0.471±0.032	-0.453±0.031

Table S11. Scaled β_{zzz} *(10^{-30} esu, B convention) obtained after multiplication by the ratio of the electric field inside the cavity and the Maxwell electric field (1064 nm) in the solvent surrounding it, for the z component, considering the classes in the distribution of nanocrystals*

Class	LC-BLYP20-opti	LC-BLYP20	LC-BLYP30	LC-BLYP40	ωB97X	CAM-B3LYP
2	1.589	1.490	1.260	1.085	0.708	0.825
6	1.851	1.703	1.427	1.220	0.777	0.906
10	1.904	1.745	1.460	1.246	0.791	0.922
14	1.924	1.761	1.472	1.256	0.795	0.928
18	1.933	1.769	1.478	1.261	0.798	0.930
22	1.939	1.773	1.482	1.263	0.799	0.932
26	1.942	1.776	1.484	1.265	0.800	0.933
30	1.944	1.778	1.485	1.266	0.800	0.934
34	1.946	1.779	1.486	1.267	0.801	0.934
38	1.947	1.780	1.487	1.267	0.801	0.934
42	1.948	1.780	1.487	1.268	0.801	0.935
46	1.949	1.781	1.488	1.268	0.802	0.935
50	1.949	1.781	1.488	1.268	0.802	0.935
54	1.950	1.782	1.488	1.269	0.802	0.935
58	1.950	1.782	1.489	1.269	0.802	0.935
62	1.950	1.782	1.489	1.269	0.802	0.935
66	1.951	1.782	1.489	1.269	0.802	0.935
70	1.951	1.783	1.489	1.269	0.802	0.935
74	1.951	1.783	1.489	1.269	0.802	0.935
Average	1.936±0.131	1.771±0.120	1.480±0.100	1.262±0.085	0.799±0.054	0.931±0.063

Table S12. *Averaged nonlinear optical quantities computed from the scaled β_{zyy} and β_{zzz} for various XCFs, in 10^{-30} esu (B convention)*

	LC-BLYP20 (optimized)	LC-BLYP20	LC-BLYP30	LC-BLYP40	ωB97X	CAM-B3LYP
R_{scaled}	-0.315±0.021	-0.371±0.025	-0.396±0.027	-0.427±0.029	-0.590±0.040	-0.487±0.033
DR_{scaled}	2.541±0.172	2.289±0.155	2.195±0.149	2.089±0.141	1.720±0.116	1.920±0.130
$\beta_{HRS,scaled}$	0.773±0.052	0.719±0.049	0.606±0.041	0.524±0.035	0.364±0.025	0.398±0.027

ESI.16 : CORRESPONDING SECOND ORDER SUSCEPTIBILITY

If we consider the $\beta_{HRS,CNC,0} = 0.149 \times 10^{-30}$ esu and we have a unit cell of $V_{uc} = 6.58 \times 10^{-22}$ cm^3 with the relationship between the "cgs/esu"and the SI unit system to yield the following conversion factor $\chi_{SI}^{(2)} = 4.19 \times 10^{-3} \chi_{esu}^{(2)}$. We then get a value for the susceptibility in the SI unit system to $\chi_{CNC}^{(2)} = 0.949 \approx 1$ pm/V.

ESI.17 : GENERAL FORM OF FIRST HYPERPOLARIZABILITY TENSOR

The first hyperpolarizability tensor consists of 27 independent components; however, by considering the experimental conditions and the point group of the CNC, simplifications are possible. Since the experiments are being performed in non-resonant conditions, we assume Kleinmann symmetry to hold (i.e $\beta_{kji} \approx \beta_{ijk}$). Additionally, in the special case of second harmonic scattering, which is a degenerated case of sum-frequency scattering, the two excitation fields are indiscernible, this leads to $\beta_{ijk} = \beta_{ikj}$. Considering these approximations, the number of independent tensor components can be reduced from 27 to 18. The hyperpolarizability tensor can then be simplified to the contracted 6×3 matrix.

Furthermore, the I_β allomorph of cellulose corresponds to a monoclinic $P2_1$ space group which is associated to a C_2 point group.[19] The only non-vanishing tensor components for C_2 are β_{zzz}, $\beta_{zxx}(=\beta_{xxz}=\beta_{xzx})$, $\beta_{zyy}(=\beta_{yzy}=\beta_{yyz})$ and $\beta_{xyz}\,(=\beta_{yzx}=\beta_{zxy}=\beta_{xzy}=\beta_{yxz}=\beta_{zyx})$. The general non-vanishing tensor components for C_2 with the dominant axis along z is then expressed as follows:

$$\beta_{ijk}=\begin{pmatrix} 0 & 0 & 0 & xyz & zxx & 0 \\ 0 & 0 & 0 & zyy & xyz & 0 \\ zxx & zyy & zzz & 0 & 0 & 0 \end{pmatrix} \qquad (S36)$$

BIBLIOGRAPHY

1 Y. Habibi, L. A. Lucia and O. J. Rojas, *Chem. Rev.*, 2010, **110**, 3479–3500.

2 S. P. Rowland and P. S. Howley, *J. Polym. Sci. Part A Polym. Chem.*, 1988, **26**, 1769–1778.

3 S. J. Cyvin, J. E. Rauch and J. C. Decius, *J. Chem. Phys.*, 1965, **43**, 4083–4095.

4 J. Campo, F. Desmet, W. Wenseleers and E. Goovaerts, *Opt. Express*, 2009, **17**, 4587.

5 A. R. Dok, T. Legat, Y. de Coene, M. A. van der Veen, T. Verbiest and S. Van Cleuvenbergen, *J. Mater. Chem. C*, 2021, **9**, 11553–11568.

6 O. Kratky, *Prog. Biophys. Mol. Biol.*, , DOI:10.1016/s0079-6107(63)80015-2.

7 F. Mairesse and B. Champagne, *Int. J. Quantum Chem.*, 2024, **124**, 1–17.

8 R. Dovesi, R. Orlando, A. Erba, C. M. Zicovich-Wilson, B. Civalleri, S. Casassa, L. Maschio, M. Ferrabone, M. De La Pierre, P. D'Arco, Y. Noël, M. Causà, M. Rérat and B. Kirtman, *Int. J. Quantum Chem.*, 2014, **114**, 1287–1317.

9 M. Ferrero, M. Rérat, B. Kirtman and R. Dovesi, *J. Chem. Phys.*, , DOI:10.1063/1.3043366.

10 M. Ferrero, M. Rérat, R. Orlando and R. Dovesi, *J Comput Chem*, 2008, 1450–1459.

11 M. Ferrero, M. Rérat, R. Orlando and R. Dovesi, *J. Chem. Phys.*, , DOI:10.1063/1.2817596.

12 T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2004, **393**, 51–57.

13 Y. Tawada, T. Tsuneda, S. Yanagisawa, T. Yanai and K. Hirao, *J. Chem. Phys.*, 2004, **120**, 8425–8433.

14 F. Mairesse, L. Maschio and B. Champagne, *J. Chem. Phys.*, , DOI:10.1063/5.0137274.

15 L. Lescos, S. P. Sitkiewicz, P. Beaujean, M. Blanchard-Desce, B. Champagne, E. Matito and F. Castet, *Phys. Chem. Chem. Phys.*, 2020, **22**, 16579–16594.

16 K. Garrett, X. Sosa Vazquez, S. B. Egri, J. Wilmer, L. E. Johnson, B. H. Robinson and C. M. Isborn, *J. Chem. Theory Comput.*, 2014, **10**, 3821–3831.

17 C. J. F. Böttcher, *Theory of Electric Polarization*, Elsevier Scientific Publishing Company, New York, 1973.

18 D. Srivastava, J. Ahopelto and A. J. Karttunen, *Molecules*, 2022, **27**, 6240.

19 T. Verbiest, K. Clays and V. Rodriguez, *Second-order Nonlinear Optical Characterization Techniques: An Introduction*, 2009.