Carbohydrate Polymers 84 (2011) 539–549

Contents lists available at ScienceDirect

Carbohydrate Polymers

journal homepage: www.elsevier.com/locate/carbpol

Chemometric analyses of the ^{1}H–^{13}C cross-polarization build-up of celluloses NMR spectra: A novel approach for characterizing the cellulose crystallites

C. Rondeau-Mouro*, H. Bizot, D. Bertrand

UR1268 Biopolymères, Interactions, Assemblages-INRA, Rue de la Géraudière, BP 71627, 44316 Nantes cedex 3, France

ARTICLE INFO

Article history:
Received 20 September 2010
Received in revised form 3 December 2010
Accepted 6 December 2010
Available online 13 December 2010

Keywords:
Cellulose
CPMAS
NMR
PCA
MCR
Chemometrics

ABSTRACT

Chemometric analyses of ^{13}C CPMAS NMR spectra (without signal curve-fitting) recorded on celluloses from various origins and by varying the contact time of the cross-polarization process, indicated a direct relation between the crystallinity of cellulose and the PCA scores. The complex exponential increase of CPMAS signals was shown to be directly dependent on the cellulose crystallinity. Estimation of a crystallinity index was possible using the proton spin diffusion time values estimated by adjustment of the experimental NMR spectra obtained while cross-polarization built-up. Calculation of the corresponding lateral dimensions of cellulose crystallites has been carried out taking into account a variable thickness of the microfibrils surface layer. Using a multilinear regression method with a constraint of positivity (MLRC), modeling of the observed spectra of celluloses from various origins has been considered as a linear combination of available NMR spectra of pure forms (Iβ, Iα and amorphous forms). The concentration of these pure forms allowed a crystallinity index to be calculated as well as estimating the proportion of the cellulose Iβ allomorph. The proposed procedures based on chemometrics are rapid, robust, even on impure samples. They need no assumption about samples and very few manipulations of the NMR signals (no curve-fitting).

1. Introduction

Cellulose is a linear β-(1–4)-linked homopolymer of anhydroglucose (Purves, 1954). The glycosidic linkages can bend, which permits intramolecular hydrogen bonds and special secondary structures (Hermans & Weidinger, 1949). From crystallographic studies (X-ray and electron diffraction), it was concluded that the secondary structures of native celluloses are ribbon-like conformations approximating twofold helical structures. The "parallel-up" chain-packing organization favored by Sarko and Mugli (1974) has been confirmed by electron microscopy studies (Koyama, Sugiyama, & Itoh, 1997). The crystal structure and hydrogen bonding system in cellulose I was elucidated by the combined use of synchrotron X-ray and neutron fiber diffraction (Nishiyama, Langan, & Chanzy, 2002). The polymorphism of cellulose was assumed early (Davis, Barry, Peterson, & King, 1943; Hess & Kissig, 1941; Petitpas and Mering, 1956; Sarko & Mugli, 1974) and apart from the native cellulose I, five other crystalline forms labeled II, III_I, III_{II}, IV_I and IV_{II} have been described (Isogai, 1994).

Studies carried out by ^{13}C solid state NMR on algal and bacterial celluloses have shown that two crystallographic phases, i.e., Iα and Iβ were present in cellulose (Atalla & VanderHart, 1984, 1999; VanderHart & Atalla, 1984). It was then demonstrated that theses phases, characterized in NMR by distinct resonance lines, were also present in many cellulosic sources. These observations were confirmed by other methods such as infrared spectroscopy and electron micro-diffraction (Sugiyama, Persson, & Chanzy, 1991; Sugiyama, Vuong, & Chanzy, 1991). The Iα form corresponds to a single-chain triclinic crystallographic symmetry, whereas Iβ is monoclinic and characterized by two parallel chains. Native cellulose is a composite of these two forms in which the ratio between Iα and Iβ depends on its source. Whereas Iα rich celluloses are present in the cell walls of some algae and in bacterial cellulose (Imai, Sugiyama, Itoh, & Horii, 1999), Iβ rich ones have been found in cotton, wood and ramie fibers (Horii, Hirai, & Kitamaru, 1987; Sugiyama, Okano, Yamamoto, & Horii, 1990; Sugiyama, Vuong, et al., 1991). Additional studies have shown that *Tunicin* from the small sea animal *Halocynthia roretzi*, consists of nearly pure (90%) Iβ phase (Belton, Tanner, Cartier, & Chanzy, 1989; Larsson, Westermark, & Iversen, 1995), whereas the cellulosic walls of *Valonia* consist mainly of Iα phase (60%) (Heux, Dinand, & Vignon, 1999; Horii, 2001; Imai & Sugiyama, 1998; Larsson, Wickholm, &

Abbreviations: CPMAS, cross polarization magic angle; MCR, multiple curve resolution; MLRC, multilinear regression with a constraint; PCA, principal component analysis; PLS, partial least square regression; RH, relative humidity; WAXS, Wide Angle X-ray Scattering.

* Corresponding author. Tel.: +33 02 40 67 50 50; fax: +33 02 40 67 50 84.
E-mail address: rondeau@nantes.inra.fr (C. Rondeau-Mouro).

0144-8617/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carbpol.2010.12.018

Iversen, 1997; Yamamoto & Horii, 1993). It was also shown that the Iα phase was metastable and could be converted into the Iβ form by a hydrothermal treatment at 260 °C (Yamamoto & Horii, 1993). Another source of variability between celluloses is associated with the occurrence of non-ordered cellulose chains, often localized at the surface of the "in-core" crystalline chains. This amorphous cellulose contains a possible contribution from the surfacial component of the crystallites in the microfibrils (Earl & VanderHart, 1981; Horii, Hirai, & Kitamaru, 1982; Newman, 1998). The additional contribution from structural defects has also been proposed based on the conformational irregularity associated with disordered hydrogen bonding of the CH_2OH groups (Masuda et al., 2003).

The super-molecular arrangement of the cellulose chains, depending on the cellulose source, modulates the lateral fibril dimensions and its degree of order. These parameters may influence the degree of crystallinity of cellulose. This degree can be evaluated via X-ray scattering using the Scherrer equation (Jakob, Fengel, Tschegg, & Fratzl, 1995; Jandura, Kokta, & Riedl, 2000) or by direct measurement on images from transmission electron microscopy (Jakob et al., 1995). Since the eighties, various authors have proposed to evaluate the degree of crystallinity by ^{13}C solid-state NMR spectroscopy (Heux et al., 1999; Newman, 1999; Newman & Hemmingson, 1990). The chemical shifts associated with ordered and disordered cellulose chains indeed result in separated lines, easily distinguishable for the C6 and C4 carbons of the glycosyl moieties. The crystallinity degree can thus be estimated by integrating signals assigned to the crystalline and the amorphous part of cellulose. However, the assignments of chemical shifts associated with the disordered regions of cellulose are imprecise and still controversial (Larsson et al., 1997). Moreover, special care has to be taken for the quantitative analysis of the CPMAS NMR measurements. In order to get reliable and quantitative solid state CPMAS ^{13}C NMR spectra of celluloses, Newman (1999) and Heux et al. (1999) have proposed an experimental procedure dedicated to the evaluation of an "apparent crystallinity index". Newman (1999) has developed a proton spin-relaxation edited NMR method (PSRE) to estimate the lateral dimensions of pure cellulose crystallites within lignocellulosic samples. The procedure used by Heux et al. (1999) was based on the NMR signals integration which consisted in mathematically modeling of the spectrum as the sum of simple curves derived from Gaussian or Lorentzian functions. Each of these curves was characterized by its chemical shift, its amplitude (calculated by integration of the intensity) and the nature of the fitted function. The quality of the adjustment was quantified by a predefined mismatch criterion. A minimization algorithm was carried out in order to vary the different parameters of the model so as to minimize this mismatch criterion. Starting with an approximate solution, the fitting procedure consists of computing and adding an increment to try to converge to a local minimum (the algorithm used is a constraint gradient protocol (Massiot et al., 2002; Press, Teukolsky, Vettering, & Flannery, 1997)). It is important to mention that any curve-fitting program such as the one used for NMR data only proposes a mathematically optimized solution and does not guarantee its physical or chemical interpretation, which remains for the user. The more the user knows about the system, the more constraints he can introduce in the algorithm, in order to limit the number of optimized parameters so as to improve the quality of the interpretation. The iterative algorithm used in such a minimization does not ensure the unicity of the obtained solution that depends on the initial conditions and is thus operator-dependant. This method of quantification is time-consuming, can be a tedious work on collections of NMR data for complex samples and requires a reference sample with known crystallinity. Moreover, such fittings are carried out independently for each relevant peak and each spectrum. The spectral data collection is thus not taken into account as a single set. Some authors have used a chemometrical approach for processing NMR spectra of cellulosic products. Chemometrics mainly consists in the application of statistical and mathematical methods for extracting relevant information from chemical and physical data. From the CPMAS NMR spectra acquired on pulp fibers of different wood species (31 samples in total), Lennholm and Iversen (1995) have classified samples using principal component analysis (PCA). They were able, based only on 31 spectra, to identify three qualitative groups related to the different lignin, cellulose and polysaccharides amount and/or structure. Wickholm, Hult, Larsson, Iversen, & Lennholm (2001) have quantified various cellulose forms (crystalline, para-crystalline cellulose and accessible/inaccessible cellulose surfaces) by ^{13}C CPMAS NMR coupled with curve-fitting. They have basically processed in two steps. The first one consisted of a non-linear curve-fitting of the C4 region. The surfaces of the different peaks identified in this way were considered as actual concentrations of the cellulose forms. In a second step, these concentrations were tentatively predicted by PLS regression from the NMR spectra. They obtained accurate models directly applicable on the unprocessed spectra. Chemometric analyses of solid-state NMR spectra from other kind of polysaccharides are rather rare. Salomonsen, Jensen, Larsen, Steuernagel, & Engelsen (2009) proposed the coupling of solid-state NMR and chemometrics in order to quantify the monomer ratio manuronic acid/guluronic acid (M/G) for different samples of alginate. They have first carried out PCA for describing their data and then attempted to calibrate and predict the composition of the monomers using partial least squares (PLS). They concluded that ^{13}C CPMAS NMR can be suitable for accurate analysis of the M/G ratio, but further demonstration without a priori knowledge have to be done.

We propose to use chemometrical analyses of ^{13}C CPMAS spectra of various celluloses, without curve-fitting of the NMR signals, in order to obtain model-independent information about the cellulose crystallinity and polymorphism. The objective of this work is to propose an analytical alternative more rapid, stable making it possible to classify celluloses according to their crystallites dimensions as well as Iβ proportion.

2. Materials and methods

2.1. Materials

Avicel PH 101 (code: avicel) was produced by FMC corp (http://www.fmcbiopolymer.com) from mineral acid hydrolysis of bleached wood pulp (CAS# 9004-34-6, Fluka Biochemica). Another source of microcrystalline cellulose (code: microc) was obtained from Sigma–Aldrich (lot 055 10 KS-089, powder 20 μm, CAS# 310697). The dried citrus microfibrils (code: citrus) were given by Cargill (Beaupte). Microfibrils were prepared from lemon peel slurry which is a by-product of the industrial extraction of pectin. Dried cotton linters (code: linter) were a gift from the Southern Regional Research Center of The United States (SRRC, USDA, NewOrleans, USA) and were used with no further purification. *Valonia Macrophysa* monocellular vesicles (code: valomo) were obtained as a gift from the exotic fishes curator of Ocearium of Croisic, France. Purification followed the procedure of Sugiyama, Persson, et al. (1991), with minor modifications similar to the protocol of Heux et al. (1999). This cellulose source was oven-dried at a relative humidity of 40% at 40 °C for two days to get the dried *Valonia* cellulose (code: valoni). *Glaucocystis Nostochinearum* 0101 XII.03 strain (code: glauco) was obtained from the collection of algae of Charles University, Department of Botany Czech Republic. Bold's Basal Medium (BBM) with doubled nitrate was used for cultivation in agitated Fernbach culture flasks capped with Silicozen porous septa. The gaseous atmosphere was supplemented

Table 1
Predicted values for M_0, a, b with the lack of fit and estimated values for d_{exp}, d_{calc} and L_{calc} according to the cellulose origin and hydration state. X is the crystallinity index measured by WAXS.

Samples	Number of τ values	b	a	M_0	Lack of fit	X (WAXS)	d_{exp}	d_{calc}	n	L_{calc} (nm)
amorph00	14	**14.60**	**117.05**	**0.98**	**0.06**	**0**	**5.70**	**5.54**	**5**	**−0.16**
amorphst	20	12.05	277.66	0.99	0.16					
avicel00	28	**17.99**	**327.38**	**1.01**	**0.03**	**57**	**9.31**	**9.27**	**2**	**6.99**
avicel65	28	12.90	264.65	0.99	0.08					
avicel90	23	18.15	303.97	1.00	0.03					
avicelst	11	13.85	248.93	0.98	0.13					
citrus00	24	**20.24**	**292.51**	**1.02**	**0.08**	**30**	**10.08**	**8.76**	**4**	**4.20**
citrus65	15	46.57	652.09	1.01	0.34					
citrus90	25	16.17	146.58	0.97	0.13					
citrusst	33	16.16	306.37	0.99	0.09					
glaucost	30	16.69	181.24	0.99	0.13					
linter00	28	**14.14**	**262.52**	**1.02**	**0.16**	**70**	**6.98**	**8.30**	**1**	**7.16**
microc00	21	**13.39**	**189.10**	**0.97**	**0.15**	**50**	**7.78**	**7.04**	**2**	**4.76**
tunich65	37	92.38	1009.00	0.99	0.63					
tunich90	11	12.49	254.90	0.97	0.15					
tunics00	32	**21.35**	**435.13**	**1.00**	**0.08**	**80**	**10.80**	**10.68**	**1**	**9.54**
tunics65	29	14.18	326.67	1.00	0.06					
tunics90	35	15.70	371.51	1.00	0.06					
tunicsst	18	20.93	405.06	1.00	0.03					
valomost	33	13.02	294.64	1.00	0.07					
valoni00	22	**16.32**	**506.37**	**1.01**	**0.12**	**80**	**10.80**	**10.52**	**1**	**9.38**

with ^{13}C enriched CO_2 so as to reach 10% in a Phytotron greenhouse with diurnal enlighting (actizoo Neon cycles of 16 h/8 h at 75 μ Einstein/m^2 s). Purification was similar to *Valonia* except for disrupting cell walls which required an osmotic shock after incubation in 25% glucose syrup; no hydrochloric treatment was applied and bleaching was replicated 4 times. Two sources of *Tunicin* were available. A first one (code: tunich) was obtained from *Ascidiella Mentula* mantle provided by Station Biologique de Roscoff (France) and extracted following the same procedure as for *Valonia* except that the higher dry matter content in suspensions and the tougher material required three successive alkali and bleach treatment to remove proteoglycans (Vandaele, Revol, Gaill, & Goffinet, 1992) and more intense grinding (Van den Berg, Capadona, & Weder, 2007). This source of *Tunicin* was then desorbed over saturated salt solutions (see later). A second source of *Tunicin* (code: tunics) came from extracts of the bacteria Halocynthia *roretzi* (Japan). This material was a gift from CERMAV (Grenoble, France). It was hydrolyzed by 2.5 N hydrochloric acid hydrolysis in reflux for 10 h and then oven-dried at a relative humidity of 40% at 40 °C for two days. Amorphous cellulose (code: amorph) was obtained by regeneration of bacterial cellulose (DP around 2000). Bacterial cellulose was produced from a non-acidifying *Gluconacetobater xylinus* strain ATCC53524 at pH 5 (Johnson & Neogi, 1989), in Hestrin and Schram browth (Hestrin & Schramm, 1954). The bacterial cellulose was first dissolved in a non-derivatizing solvent (lithium chloride/N,N-dimethylacetamide) followed by a solvent exchange procedure as described by Dupont (2003).

In order to control their relative humidity (RH), samples have been stored at 25 °C over saturated salt solutions (prepared with D_2O) of $CuCl_2$ and $BaCl_2$ for at least one weak under reduced pressure samples (Multon, Bizot, & Martin, 1991; Nyqvist, 1983). Such conditions correspond to 65 and 91% relative humidity (RH) and increase the cellulose water content to 7 and 13%, respectively (defined on wt./wt. basis, i.e., as grams water per 100 g of solid materials) (Multon et al., 1991). "Saturated" cellulose (referred as "st" in the followings) was obtained by first soaking materials in heavy water. The use of such a solvent has been proposed by Tang, Wang, & Belton (2000) in order to limit the number of proton reservoir to non-exchangeable protons. These protons are then the only implied in magnetization transfer to carbons during the NMR cross-polarization experiment. Such a precaution was also kept in order to limit microorganism degradations.

The excess of capillary water was drained by blotting and hand pressing between wipers (Kimberly Clark 33370 CREW II POLY WHT wipers) before rotor filling for NMR studies.

In the followings, the samples are coded by the name of the kind of cellulose, followed by an indication on the hydration level (00, 65, 91, st, respectively for oven-dried samples, samples conditioned at RH = 65%, those conditioned at RH = 91% and those saturated with D_2O). For the sake of clarity, see sample codes mentioned in Table 1 and 2 (first column).

2.2. Wide Angle X-ray Scattering (WAXS)

Measurements were performed using a spectrometer equipped with a XRG 3000 generator (Inel Orléans, France) working at 40 kV and 30 mA, a quartz monochromator (λ = 0.15405 nm) and a curve position sensitive detector (Inel CPS 120). The procedure to estimate the crystallinity has already been described in Rondeau-Mouro et al. (2003) and Wakelin, Virgin, & Crystal (1959).

2.3. NMR

NMR experiments were performed on a Bruker DMX-400 spectrometer operating at a ^{13}C frequency of 100.62 MHz and equipped with a double resonance H/X CP-MAS 4 mm probe. The MAS rate was fixed at 7000 Hz and each experiment was recorded at ambient temperature (294 ± 1 K). The cross polarization pulse sequence used a 3.4 μs 90 proton pulse, a variable contact time at 74 kHz and a 6 s recycle delay for an acquisition time of 34 ms during which a dipolar decoupling of 74 kHz was applied (Bennett, Rienstra, Auger, Lakshmi, & Griffin, 1995). A typical number of 3072 scans were acquired for each spectrum. Transfer of polarization was achieved by varying the contact time (between 11 and 37 points varying between 5 μs and 5 ms) (Wu, Zhang, & Wu, 1988). Chemical shifts were calibrated by using the peak at 176.03 ppm of glycine as an external standard.

To enhance the signals from rare nuclei such as ^{13}C, the cross polarization technique was used. The transfer of polarization is based on the principle that magnetization flows from highly polarized nuclei (protons) to nuclei with lower polarizations (carbons for instance) when the two are brought into contact. As spin flips are not energy-conserving at high magnetic fields, two continuous radiofrequency (RF) fields are required, one at the resonance fre-

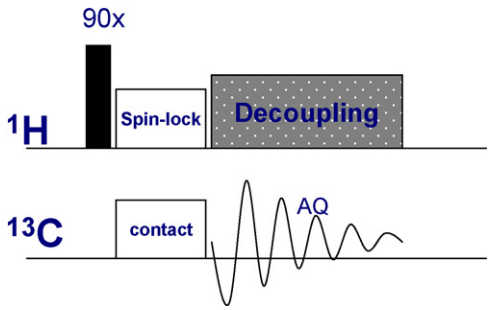

Fig. 1. ^{13}C CPMAS (cross polarization magic angle) sequence.

quency of the protons and the other at the resonance frequency of carbons. The Hartmann–Hahn condition, essential to achieve the polarization transfer, requires the equalization of the nutation frequencies of protons and carbons in order to induce an energy conserving dipolar contact between the two spins reservoir. The experimental implementation of this principle is shown in Fig. 1.

The proton magnetization is first excited by a $\pi/2$ pulse at the proton resonance frequency, followed by a spin-locking field consisting in applying simultaneously a RF fields to protons and carbons for a period τ. During that time, exchange of magnetization occurs between protons and carbons in their respective rotating frame. Finally the carbons are detected while dipolar carbons–protons interactions are suppressed by a high-power proton decoupling. In addition to the sensitivity enhancement, the CP experiments show the advantage to depend on the relaxation time of abundant nuclei. As the protons in solid-state relax more rapidly than most of other spin 1/2 nuclei, many more scans can be acquired than for direct detection of the rare ^{13}C nuclei.

The increase of the ^{13}C magnetization during the CP mixing period τ depends on the strength of the ^{1}H–^{13}C dipolar coupling. The polarization buildup for unprotonated carbons requires a longer contact time than for carbons which are quickly polarized by their bound protons (Alemany, Grant, Pugmire, Alger, & Zilm, 1983a, 1983b). The CP magnetization transfer can then be interpreted using the two proton reservoir model proposed by Müller, Kumar, Baumann, & Ernst (1974) to explain transient oscillations in $^{1}H/^{13}C$ CP experiments on ferrocene crystals. In these crystallized systems, a first reservoir assigned to protons directly bound to carbons while the other protons (hydroxyls and hydrogen bonded water molecules over surfaces) belong to a second reservoir (called R2) whose interactions with carbons are neglected. However, the interactions between the carbon-bound protons and the protons from R2 can be described by an isotropic spin-diffusion process with the rate $1/a$. For short contact times τ and semi-crystalline systems, rotating-frame spin–lattice relaxations are neglected and the observed magnetization can be expressed as (Jarvis, Fenwick, & Apperley, 1996; Müller et al., 1974; Wu et al., 1988):

$$M(\tau) = M_0(1 - s.e^{-\tau/a} - (1-s).e^{-3\tau/2a}e^{-\tau^2/2b^2}) \quad (1)$$

In which a is the proton spin-diffusion time, $b = 1/2(\hbar\gamma_H\gamma_C/r_{CH}^3)(3\cdot\cos^2\theta - 1)$ is the dipolar coupling between ^{13}C and its directly bound proton, r_{CH} the distance between them. γ_H and γ_C are the gyromagnetic ratio of ^{1}H and ^{13}C, respectively, $\hbar$ is the Plank's constant divided by 2π.

The parameters allow the number of protons bound to carbon (0.5 for CH group as it is the case for cellulose).

2.4. Chemometrics

The sample collection eventually consisted of eleven kinds of cellulosic materials. Some of them were conditioned at various hydration levels (Table 1) ("00", "65", "91", "st"). The NMR spectra of the conditioned samples were acquired with varying contact times. The spectral data collection eventually contained 528 spectra corresponding to 21 kinds of sample differing by the origin of the cellulose, the hydration and the contact time. Due to the magnet drift imposing a chemical-shift calibration relative to an external reference (see Section 2.3), the NMR spectrometer did not record the spectra at exactly the same intervals in ppm. In order to apply linear methods, the data points were first linearly interpolated in order to have spectra ranging from 115 to 50 ppm at intervals of 0.1 ppm (651 data points). The interpolated spectra were eventually gathered in a data matrix $\boldsymbol{X}$ dimensioned $n \times p$ with n and p equal to 528 and 651, respectively.

2.4.1. Bilinear models: principal component analysis and linear regression

Principal component analysis (Jollife, 2002) was performed on the unscaled and centered data collection. PCA makes it possible to emphasize the similarities and differences between a large number of observations (here spectra) including many variables. It can be seen as a rotation in a multidimensional space. PCA calculates new variables, called "principals components" which are linear combinations of the old ones. The scores $\boldsymbol{c}_i$ are linear combinations of the original variables such as

$$\boldsymbol{c}_i = \alpha_{i1}\boldsymbol{x}_1 + \alpha_{i2}\boldsymbol{x}_2 + \ldots + \alpha_{ip}\boldsymbol{x}_p \quad (2)$$

The factors α_{ij}, with $j = 1, 2 \ldots p$ are elements of the PC- loading associated with dimension i representing the importance of a variable for the corresponding component. The loadings have the same number of elements as a spectrum, and can be usefully plotted as curves. In such curves, the local maxima and minima can be interpreted in terms of spectroscopy.

The elements α_{ij} can be gathered in a matrix $\boldsymbol{A}$ with p rows and k columns,with k, the number of dimensions retained in the PCA model.

The general model of PCA is a bilinear model, which can be expressed as: $\boldsymbol{C} = \boldsymbol{XA}$, in which $\boldsymbol{C}$ is the matrix of components, $\boldsymbol{X}$ the centered data matrix and $\boldsymbol{A}$ the loadings (in rows).

When knowing $\boldsymbol{C}$ and $\boldsymbol{A}$, it is possible to "reconstruct" the spectra, according to:

$$\hat{\boldsymbol{X}} = \boldsymbol{CA}^T + \boldsymbol{E} \quad (3)$$

in which $\hat{\boldsymbol{X}}$ is the spectra approximately reconstructed and $\boldsymbol{E}$ a matrix of error-terms. The accuracy of the reconstruction of $\boldsymbol{X}$ depends on the number of components and corresponding loadings which have been retained.

It must be noticed that (3) has formally the same shape that the one involved in the modeling of spectral mixture $\boldsymbol{X}$ when knowing the "pure spectra" $\boldsymbol{S}^T$ and their proportion in the mixture $\boldsymbol{P}$:

$$\hat{\boldsymbol{X}} = \boldsymbol{PS}^T + \boldsymbol{F} \quad (4)$$

in which $\boldsymbol{F}$ is a matrix of error expressing the lack of fit of the model.

The main difference between Eqs. (3) and (4) is that the PCA model does not constraint the elements of matrices $\boldsymbol{C}$ and $\boldsymbol{A}$ to have only positive or null elements. On the contrary, when working with an additive spectral model, it is natural to impose in (4) that the pure spectra and the proportions take only positive or null values. In most situations, $\boldsymbol{S}^T$ and $\boldsymbol{P}$ are not known and must be derived from $\boldsymbol{X}$, the experimental spectral matrix. In these cases, methods such as multiple curve resolution (MCR) (De Juan & Tauler, 2003) can be used. In the present work, the model associated with (4) is simplified because the spectra of pure cellulosic forms are experimentally available. We have indeed considered that products tunicsst, glaucost and amorphst were representative of pure forms of cellulose, respectively, Iβ, Iα and amorphous forms (see Section

3.1). The corresponding spectra were gathered in order to form the matrix $\boldsymbol{S}$ in (4). It was then possible to estimate the proportion $\boldsymbol{P}$ of the cellulosic forms by multilinear regression with a constraint of positivity (MLRC) for the elements of $\boldsymbol{P}$. For this purpose, the function *lsqnonneg* available in the Matlab® environment has been used. This function uses the algorithm described by Lawson and Hanson (1974). The algorithm starts with a set of possible basis vectors and computes the associated dual vector λ. It then selects the basis vector corresponding to the maximum value in λ in order to swap it out of the basis in exchange for another possible candidate. This continues until $\lambda \leq 0$. Prior to the computation, the non-centered spectra were normalized in order to have the same surface. The lack of fit of the model was expressed from (4) by $\left\|\boldsymbol{F}\right\|^2 / \left\|\boldsymbol{X}\right\|^2$ in which "|| ||" designates the Frobenius norm.

2.5. Fitting theoretical models

We attempt to independently fit the physical model as described in (1) on all of the 21 available samples. For this purpose, we first apply MRCP on each of the 21 independent series. In this case, the matrix $\boldsymbol{S}$ in Eq. (4) was reduced to a single "reference spectrum", which was chosen as the spectrum in the series under study having the highest NMR intensity (highest contact time). Every spectrum of the studied series is supposed to be proportional to this reference spectrum. The matrix of computed proportions $\boldsymbol{P}$ in Eq. (4) was thus reduced to a single column-vector, which is now referred as p. An element of p corresponds to a given time contact τ. If the physical model were applicable, $p(\tau)$ is proportional to $M(\tau)$ as given in Eq. (1). The physical model to be fitted is thus depending on 5 parameters, M_0, s, a, b, τ.

$p(\tau)$ and τ are known, whereas M_0, s, a and b must be determined by an optimization algorithm. The criterion to be optimized was here the sum of square of the residual between the experimental proportions in p and the calculated proportions derived from the computed values of M_0, R and b (least-square minimization). The optimization was carried out using the Matlab® function *fminsearch*. This function is based on the simplex method as described in Lagarias, Reeds, Wright, & Wright (1998). The lack of fit of the model was expressed as $\|\boldsymbol{r}\|^2 / \|\boldsymbol{p}\|^2 \times 100$ in which $\boldsymbol{r}$ is the residual, i.e., the difference between the actual values of the proportion in $\boldsymbol{p}$ and the computed values derived from the adjusted model.

3. Results and discussion

3.1. Spectral collection

Fig. 2 displays the CPMAS NMR spectra of cellulose microfibrils of different species in a dried state: amorphous (amorph00), citrus (citrus00), commercial microcrystalline celluloses (avicel00 and microc00), cotton linters (linter00), *Tunicin* (tunics00) and *Valonia* celluloses (valoni00). The upfield region from 60 to 70 ppm is assigned to C6 carbon. The more intense signals from 70 to 80 ppm are assigned to C2, C3 and C5 carbons and the region from 80 to 90 ppm corresponds to the C4 carbon. Eventually, the downfield signal at 105 ppm is assigned to the C1 carbon. The large broadening of signals due to the anisotropic nature of the magnetic interactions observed by NMR attests the solid state of samples but also a severe signals overlapping. It has indeed been shown that the existence of two crystallographic phases for native cellulose, i.e., Iα and Iβ, induces distinct signals at separate chemical shifts: 105.5 ppm and 103.8 ppm for the C1 resonances of the Iβ form and one resonance around 104.9 ppm for the C1 of the Iα form (Atalla & VanderHart, 1999; Larsson et al., 1995). These specific chemical signals are clearly visible on NMR spectra from *Valonia* and *Tunicin* (Fig. 2F and G, respectively).

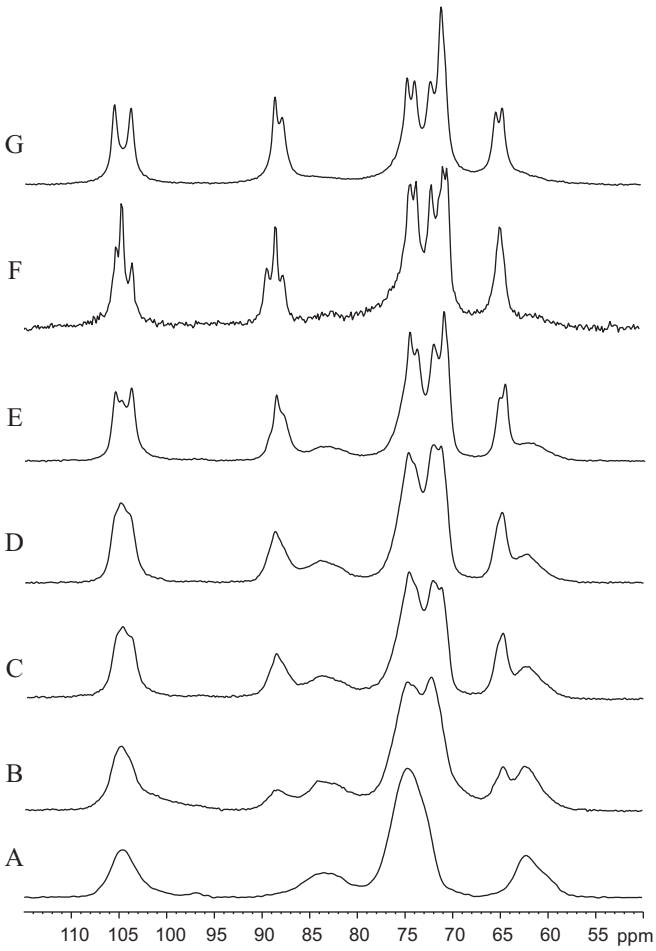

Fig. 2. ^{13}C CPMAS NMR spectra of A, amorph00; B, citrus00; C, microc00; D, avicel00; E, linter00; F, valoni00; G, tunics00.

Another variant among the different cellulose spectra is the proportion of the signal intensity between 86–92 ppm and 78–86 ppm assigned respectively to crystalline and surface chains (Heux et al., 1999; Larsson et al., 1997; Newman, 1999; Newman & Hemmingson, 1990) (called the 'amorphous phase') of C4 carbons. As expected, the NMR spectrum of "amorphous" cellulose displays no signal in the C4 neither in the C6 resonance region assigned to the 'in-core' crystalline material.

In order to distinguish each cellulose type relative to various fibril order levels, the contact time during the CP experiment has been increased. This induced an increase of NMR signal for short contact times due to the magnetization transfer from protons to carbons modulated by the thermal mixing time T_{CH} (noted b in Eq. (1)) and the proton spin diffusion T_{HH} (noted a in Eq. (1)). For longer contact times signal decay has been observed, due to the rotating-frame relaxation times of protons and carbons $T_{1\rho}$H and $T_{1\rho}$C, respectively. The present work focuses on the cross-polarization process for short contact times, permitting to reveal different order levels associated with various proton spin-diffusion times T_{HH} and/or ^{1}H–^{13}C dipolar couplings T_{CH}. In the case of cellulose, these times must depend on the microfibrils cristallinity, their hydration level and/or on the allomorph type (Iα/Iβ).

3.2. Principal component analyses

PCA has been applied on the unscaled spectra. The first plan (Fig. 3) represented 72.6% of the total sum of square. On this graph, a series of spectra of a given product only varying by the contact

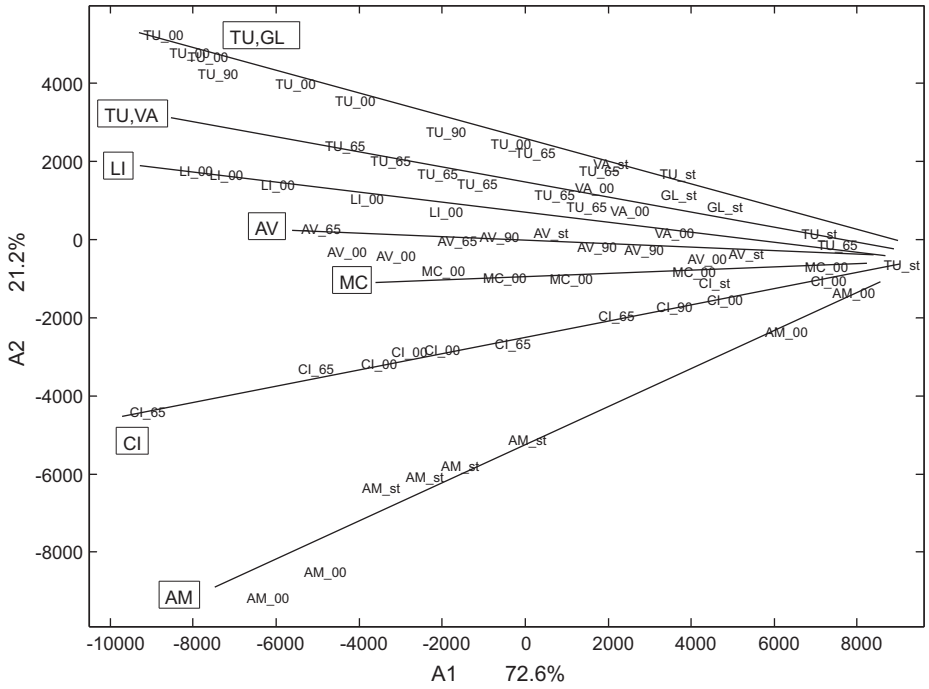

Fig. 3. Principal component analysis of NMR spectra of cellulosic materials (whole spectral collection). Scatter plot of scores #1–#2. MC, microc; AM, amorph; AV, avicel; CI, citrus; LI, linter; TU, tunics and tunich; 00, dried powder; 65, 65% relative humidity; 91, 91% relative humidity; st, saturation. The straight lines indicate the direction from low contact-times (right) to high contact-times (left) for a given material. For the sake of clarity, only one third of the data points are actually plotted. Such a reduction does not change the visual interpretation of the figure.

time were clearly positioned on straight lines. In these series, the spectra were indeed almost proportional to each other, and there projection in the PCA-space logically built up these characteristic straight lines. These straight lines have a single intersection point which corresponds to a short contact time (5 μs). At this time, the signal-to-noise ratio is low, and thus makes a unique point in the PCA space. For a given contact time greater than 5 μs, the spectra are sorted on the second axis according to the degree of ordering of the products, with the more crystalline celluloses at the top of the graph. *Tunicin* (encoded TU), *Valonia* (encoded VA) and *Glaucosystis* (encoded GL) microfibrils display higher degree of order than hydrolysed microfibrils (LI, AV and MC) the plant cell wall microfibrils issue from lemon pulps (encoded CI) either the amorphous cellulose (encoded AM). These results are in good agreement with those obtained from the estimation of the crystallinity by WAXS (Table 2), as well as other published NMR data based on curve fitting of CPMAS spectra (Heux et al., 1999; Lennholm & Iversen, 1994; Newman, 1999; Rondeau-Mouro et al., 2003; Wickholm et al., 2001).

The first loading (Fig. 4) is mainly representative of the "time effect", independently of the studied material. It is very close to the average spectrum of the whole spectral collection. The second loading (Fig. 5) is in accordance with this classification. The local maxima and minima of the second loading emphasize the peaks presenting some intensity variability and are comparable to PC subspectra obtained by Lennholm and Iversen (1995). It opposes negative values representative of the amorphous cellulosic forms to positive ones associated with crystalline forms. The position of the positive signals in Fig. 5 is in accordance with the signal chemical shifts found for highly crystalline cellulose as *Tunicin* for instance (see Fig. 2G) with C1–C6 chemical shifts at 105.7, 103.9, 88.8, 71.2 and 65.4 ppm. The negative peaks of Fig. 5 correspond to signals assigned to amorphous or fibrils surface chains as shown by several authors (Heux et al., 1999; Lennholm & Iversen, 1994; Newman, 1999; Rondeau-Mouro et al., 2003; Wickholm et al., 2001) and localized for C1 and C4 at 104.6 and 83.9 ppm, for C2, C3 and C5 between 73 and 79 ppm and at 62.2 ppm for C6.

The effect of the water content of celluloses was observable on the PCA scatter plot in Fig. 3. The cellulosic materials showed slight differences associated with their water content. These differences were particularly marked for the highly crystalline *Tunicin*, which is taken as an example. Fig. 6 shows the PCA of the *Tunicin* spectra. As mentioned in Section 2, the set of *Tunicin* samples included one oven-dried sample (tunics) and a hydrated sample (tunich) prepared at four hydration levels (00, 65, 91 and st – see Table 1). The Fig. 6 indicates that two groups can be discriminated (two straight lines noted h and s) in relation to the hydration state of samples, dried before rehydration (tunics) or dried by slow desorption of water (tunich). Hydration has few effects on signal chemical-shifts

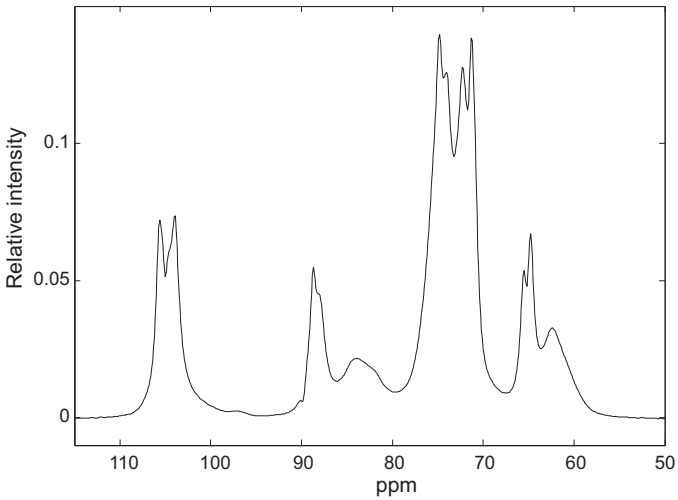

Fig. 4. Principal component analysis of NMR spectra of cellulosic materials. First PC-loading.

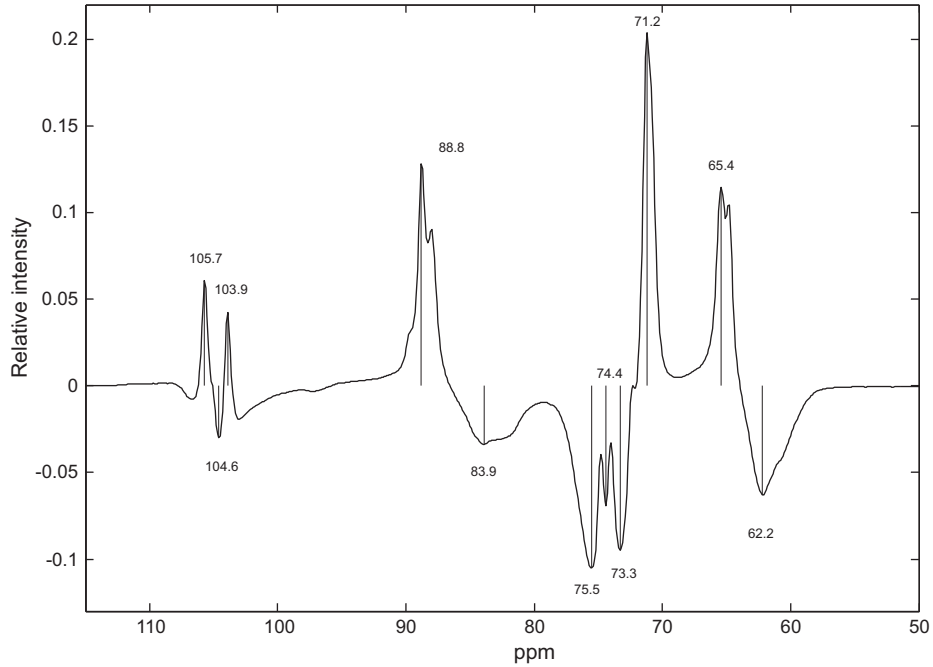

Fig. 5. Principal component analysis of RMN spectra of cellulosic material. Second PC-loading (local maxima and minima are indicated).

while it induces net resolution improvement as demonstrated for starch (Paris, Bizot, Emery, Buzaré, & Buléon, 1999). Therefore discrimination observed in Fig. 6 demonstrates that the kinetics of the magnetization transfer between protons and carbons in cellulose microfibrils is different according to the hydration or drying process. This result is in agreement with different works related to the hornification of cellulose fibres (Diniz, Gil, & Castro, 2004; Kato & Cameron, 2002; Newman, 2004). This phenomenon refers to polymeric structural changes upon drying or water removal (Jayme & Hunger, 1957). Drying of microfibrils improves the hydrogen bonding formation between them, making difficult or at least different the rewetting and improving the cocrystallization crystal domains (Newman, 2004). This phenomenon will be also discussed later (see Section 3.4 and Table 2).

3.3. Data fitting

The parameters M_0, a, b and s from Eq. (1) were tentatively estimated by a simplex procedure. A first series of adjustment (not reported here) showed that the parameter s had values close to 0.5, which is in agreement with previously reported work carried out on Citrus cell walls (Jarvis et al., 1996) and starchy substrates

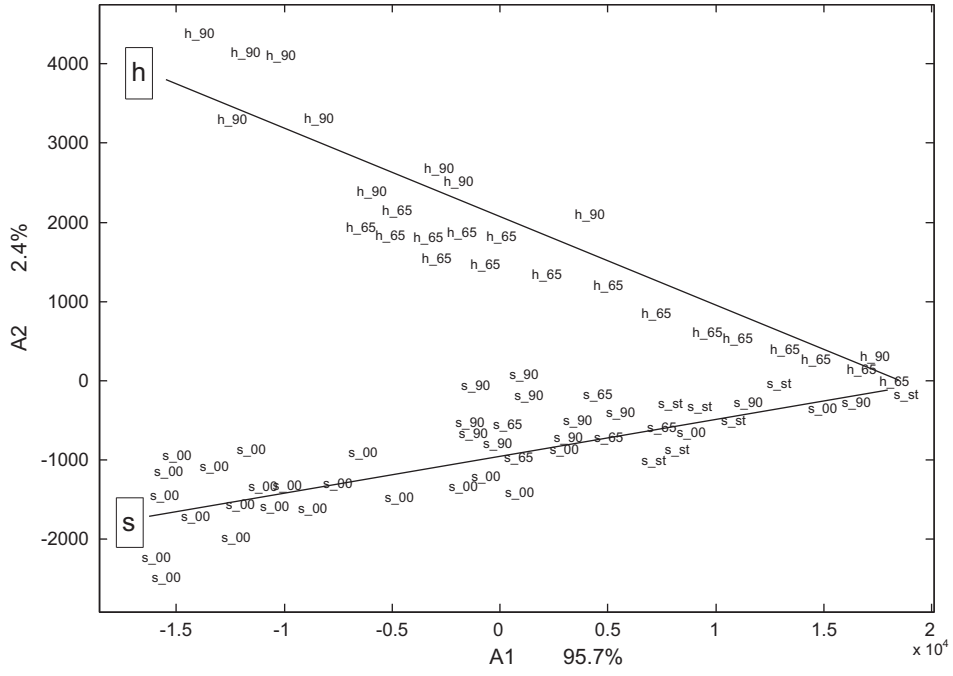

Fig. 6. Principal component analysis of NMR spectra of tunicin. Scatter plot of scores #1–#2. h, tunich; s, tunics. 00, dried powder; 65, 65% relative humidity; 91, 91% relative humidity; st, saturation. The straight lines indicate the direction from low contact-times (right) to high contact-times (left) for a given material.

(Paris, Bizot, Emery, & Buléon, 2001). This value indicates that glycosyl carbons are bound to a single proton. The data fittings were then carried out, with fixing s to 0.5.

Table 1 gathers the predicted values for M_0, a, b according to the cellulose origin and hydration state. As expected for normalized data, M_0 is close to 1. The lacks of fit of the 21 cellulose samples adjusted independently varied from 0.03% to 0.63% with two abnormal values for citrus65 and tunich65 (0.13 and 0.63%, respectively). Except these two samples, values for the thermal mixing time b (T_{CH}) varied from 12 to 21 μs while the spin diffusion time, a (T_{HH}), was found to differ from 117 to 506 μs. These values agree with reported data for Citrus cell walls (Jarvis et al., 1996) and starchy substrates (Paris et al., 2001). The impact of the samples drying or moistening on the a and b values is difficult to interpret. While few changes are detectable on the thermal mixing time b values, a correlation can be drawn between the spin diffusion time a values determined on dried samples (00) and the crystallinity index measured by the WAXS method (Table 2). The more crystalline is the cellulose, the higher is the spin diffusion time T_{HH} (a value). This result is surprising regarding the proton spin diffusion process supposed to arise between protons of the surrounding shell, non-bonded to the glycosyl carbons. According to several authors (Müller et al., 1974; Wu et al., 1988) the slow mobility of nuclei and molecules in a crystallized system may induce a faster proton spin diffusion process, then lower T_{HH} values. However, the relation between the nuclei mobility and the spin diffusion phenomenon has to be considered by taking into account the mean interproton distances. Spin diffusion corresponds to a spatial transfer of magnetization between dipolar-coupled spins and depends on the spins dipolar couplings as well as their average distances (Bloembergen, 1949). The spin diffusion coefficient is then the ratio between the covered surface by the magnetization and the time taken to cover this surface. A proton spin diffusion of $D = 9.7 \times 10^{-16}$ m^2 s^{-1} was proposed by Cheung and Gerstein (1981) in Whatman cellulose based on proton spin–spin relaxation time measurements ($T_2 = 8.9$ μs) and using the relation $D = 0.13 \cdot d^2/T_2$ in which d was chosen equal to 0.26 nm. This relation supposed that spin diffusion is the only process occurring during the T_2 measurements and within very short distances (0.26 nm). However, spin diffusion can be theoretically measurable by NMR through a surface between 1 and 200 nm of diameter (Bloembergen, 1949). Our measurements of the proton spin diffusion occurring within the carbons surrounding shell of non-bonded protons permitted to define the relation between the (non-bonded proton) spin diffusion coefficient D_L and the lateral size of cellulose microfibrils, assuming a certain thickness of the surface chain layer which varies with the origin of cellulose. Fig. 7 shows the good correlation (root mean square error = 0.3) between the microfibrils lateral size d_{exp} estimated using the whole NMR dataset (50–115 ppm) and two different formula, both proposed by Heux et al. (1999) and Newman (1999):

$$d_{exp} = \frac{2 \cdot n \cdot h}{1 - \sqrt{X}} \quad (5)$$

where n is the number of surface chains layers, assuming that the thickness of one layer is $h = 0.57$ nm, and X is the measured (WAXS) crystallinity index and:

$$d_{calc} = \sqrt{\frac{D_L \cdot T_{HH}}{K}} \quad (6)$$

where D_L is the spin diffusion coefficient (m^2 s^{-1}), T_{HH} the spin diffusion time and K is a constant.

For the sake of clarity, Fig. 7 displays only correlations for dried samples (encoded 00).

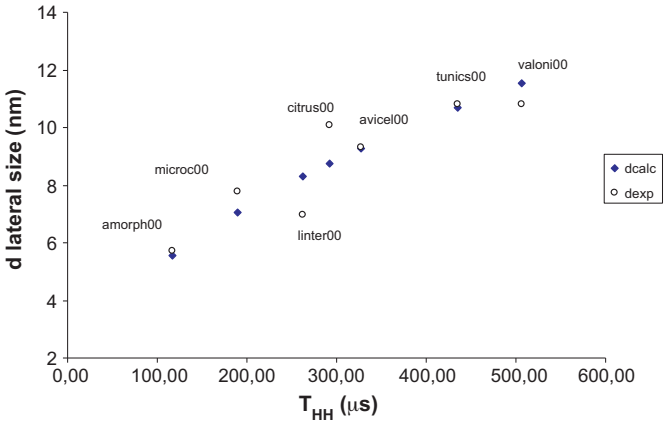

Fig. 7. Correlations (for dried powder encoded 00) between the predicted NMR spin-diffusion time values T_{HH} (a in Eq. (1)) and the calculated microfibrils lateral sizes according to Eqs. (5) and (6) (see text).

The optimization of the d_{calc} values (Table 1) compared to d_{exp} with varying n values, allowed the evaluation of the proton spin diffusion coefficient $D_L = 4.28 \times 10^{-14}$ m^2 s^{-1} through the microfibrils for a constant $K = 0.163$. The lateral size of crystals could be calculated according the following equation:

$$L_{calc} = d_{calc} - 2 \cdot n \cdot h \quad (7)$$

The negative value −0.16 nm, close to 0, for the amorphous cellulose illustrates the good agreement between the spin diffusion time measured and the hypothesis that amorphous cellulose is formed by 5 layers of non-ordered chains. This result is consistent with the description of amorphous cellulose as disordered non-crystalline chains of β-(1–4)-linked anhydroglucose, with a possible contribution from the surfacial component of the crystallites in the microfibrils (Earl & VanderHart, 1981; Horii et al., 1982; Newman, 1998).

The values obtained for citrus and avicel agreed with those reported by Newman (1999) and Rondeau-Mouro et al. (2003). The low values obtained for high crystalline celluloses (*Valonia* and *Tunicin*) compared to previous estimated lateral sizes close to 20 nm (Sugiyama, Harada, Fujiyoshi, & Uyeda, 1985), can be explained by the short mixing time used in these studies (CPMAS contact times shorter than 5 ms). In this short time condition and by considering cellulose microfibrils as biphasic systems (crystalline/amorphous) with a cylinder dimensionality (Clauss, Schmidt-Rohr, & Spiess, 1993), the spin diffusion proceeds over shorter surfaces (<10 nm).

The presence of amorphous cellulose chains as well as free water molecules, may result in a higher local mobility (less ordered), reducing the spin diffusion process. As a consequence, the data predicted for D_2O saturated samples (st) are averaged, cancelling any correlation with the d_{exp} estimated by WAXS measurements.

3.4. Regression with constraints of positivity

In the present study, a given spectrum can be seen as a mixture of pure spectra of amorphous and crystalline Iα and Iβ forms. Such a model (Eq. (4)) is in principle applicable at each time of contact, providing that the spectra were sufficiently intense. According to the literature (Belton et al., 1989; Imai et al., 1999; Larsson et al., 1995), the *Tunicin* is mainly constituted of the Iβ form with a crystallinity next to 90% while *Glaucosystis* is mainly composed of the Iα form with the same crystallinity level. *Glaucosystis* and *Tunicin* have fibrils of approximately equal lateral size (20 nm).

The amorphous cellulose (amorph) is supposed to present the non-ordered chains of cellulose at the surface of the 'in-core' crystals. These considerations justify the choice of tunicsst, glaucost and

Table 2
Predicted values for the crystallinity index *X* and Iβ level compared to crystallinity percentage estimated by WAXS and average values of Iβ from published data (see references in the text).

Samples	Predicted crystallinity (%)			Predicted Iβ (%)			Measured	
	X^a	X^b	X^c	$I\beta^a$	$I\beta^b$	$I\beta^c$	X (WAXS)[d]	Iβ (average)[e]
amorph00	**0.0**	**7.4**	**0.3**	**0.0**	**100.0**	**0.0**	**0**	**0.0**
amorphst	0.0	0.0	0.0	0.0	0.0	0.0		
avicel00	**55.0**	**40.1**	**51.1**	**69.0**	**69.2**	**60.7**	**57**	**68.0**
avicel65	57.6	40.6	52.7	69.6	68.7	61.4		
avicel90	58.5	47.4	56.0	69.3	65.7	57.7		
avicelst	62.0	35.2	55.8	70.8	70.6	65.4		
citrus00	**28.9**	**20.5**	**18.2**	**73.0**	**73.5**	**55.3**	**30**	**49.0**
citrus65	29.3	22.9	19.0	71.5	69.0	53.2		
citrus90	31.0	10.6	18.3	74.4	73.4	67.4		
citrusst	37.7	23.1	20.7	73.6	62.3	70.6		
glaucost	100.0	100.0	100.0	0.0	0.0	0.0		
linter00	**67.3**	**55.2**	**67.2**	**81.1**	**86.4**	**73.8**	**70**	**62.7 ± 2.0**
microc00	**49.4**	**34.6**	**46.0**	**71.9**	**73.1**	**65.1**	**50**	–
tunich65	77.7	64.4	76.5	86.6	94.3	81.3		
tunich90	82.8	71.9	84.5	88.6	93.1	77.5		
tunics00	**88.7**	**71.5**	**84.0**	**89.9**	**100.0**	**91.7**	**80**	**80.7 ± 5.9**
tunics65	90.2	74.5	87.9	91.8	100.0	92.4		
tunics90	88.6	73.3	85.5	91.9	100.0	90.7		
tunicsst	100.0	100.0	100.0	100.0	100.0	100.0		
valomost	100.0	100.0	100.0	40.7	43.7	11.8		
valoni00	**90.5**	**56.9**	**91.9**	**42.4**	**34.6**	**57.5**	**80**	**39. 5 ± 6.7**

[a] Values estimated by chemometrics of the whole NMR spectra between 115 and 50 ppm.
[b] Values estimated by chemometrics of the C1 region of the NMR spectra between 112 and 98 ppm.
[c] Values estimated by chemometrics of the C4 region of the NMR spectra between 92 and 78 ppm.
[d] Crystallinity pourcentage estimated by WAXS in this study.
[e] Average values from published data, see references in text.

amorphst has representative of pure forms of cellulose, respectively Iβ, Iα and amorphous forms. The pure spectra involved in Eq. (4) corresponded to the spectra of these materials acquired with the highest available contact time. Observed spectra have been considered as a linear combination of the spectra of these pure forms. The concentrations were mathematically constrained to be null or positive. Crystallinity of each sample have been calculated by addition of the concentration of Iα and Iβ forms divided by the total concentrations (Iα + Iβ + amorphous). The estimation of the Iβ levels was equal to the percentage of the *Tunicin* concentration relative to the total crystalline concentration (Iα + Iβ). Table 2 gathers the various crystallinity indexes and Iβ levels determined for various spectral regions: whole spectra between 50 and 115 ppm; C1 spectral region between 98 and 112 ppm and C4 region between 78 and 92 ppm.

The accuracy of the estimated crystallinity is reduced when using only the C1 spectral region. The results using the whole spectrum are comparable with the ones obtained from the C4 region, and agree with already published data (Agarwal, Reiner, & Ralph, 2010; Bansal, Hall, Realff, Lee, & Bommarius, 2010; Heux et al., 1999; Lennholm & Iversen, 1994; Newman, 1999; Park, Johnson, Ishizawa, Parilla, & Davis, 2009; Rondeau-Mouro et al., 2003; Thygesen, Oddershede, Lilholt, Thomsen, & Stahl, 2005; Wickholm et al., 2001). Fig. 8 shows correlations between the average concentrations of Iβ, calculated from published data and the estimations obtained in the present work (for the sake of clarity, Fig. 8 displays only correlations for dried samples (encoded 00)). The crystallinity indices measured by WAXS for the dried samples are highly correlated (slope = 0.95, R^2 = 0.975) with the crystallinity predicted from the whole spectra (50–115 ppm). While the dataset obtained for the C4 spectral region is not better for the crystallinity prediction, estimation of the Iβ percentage in this reduced region is rather good except for *Valonia*. The whole spectrum and the C1 regions give close values of Iβ which are in agreement with already published results (Belton et al., 1989; Heux et al., 1999; Horii, 2001; Horii et al., 1987; Imai & Sugiyama, 1998; Imai et al., 1999; Larsson et al., 1995, 1997; Lennholm & Iversen, 1994; Rondeau-Mouro et al., 2003; Sugiyama, Persson, et al., 1991; Yamamoto & Horii, 1993). The results for the whole NMR dataset (50–115 ppm) are rather satisfying. Fig. 8 displays a straight line with a slope of 0.84, and a R^2 equal to 0.920.

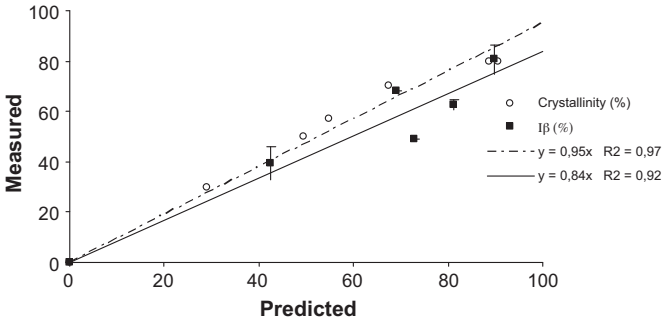

Fig. 8. Correlations (for dried powder encoded 00 – see Table 2) between the data predicted (crystal index and the Iβ proportions) by chemometrics of NMR data (in the whole spectrum 115–50 ppm) and the measured values (crystallinity index by WAXS and averaged Iβ levels from published data).

The different crystallinity levels estimated for the various samples of *Tunicin* (tunics, dried before rehydration and tunich, dried by slow desorption of water) are interesting. The regression procedure used here confirms that samples tunics are more crystalline than tunich. This result is in agreement with the PCA and refers to hornification of cellulose fibres (Diniz et al., 2004; Kato & Cameron, 2002; Newman, 2004) which is supposed to improve the cocrystallization of crystal domains (Newman, 2004).

4. Conclusion

We demonstrated that chemometric analyses of solid-state NMR spectra of celluloses are a more rapid and reliable alternative method to characterize the polymorphism and crystallinity of celluloses from various origins. This tool can be used without knowledge about samples, without curve-fitting of the NMR signals which is a time-consuming and operator-dependant method.

Two bi-linear models, namely PCA and MLRC, have been successfully used in the present work. These methods share in common the fact that they allow the fitting of the data as a matrix product of "weight" by "spectral signals" (Eqs. (4) and (5)). In PCA, the weight corresponds to PC scores and the spectral signals are the PC loadings. In MLRC, the weights are expected to be representative of the intensity of the spectral signals. However, the spectra of the pure components must be known before applying it. These methods may have particular applications in function of the needs of the user. The computation of PCA is extremely fast and gives very quickly first pieces of information on the data set. In the present case, it made it possible to immediately assess the linear dependence of the intensity of the spectra as a function of the contact time. Moreover, it gave a first unsupervised classification of the cellulosic materials which appeared to be relevantly related to their degree of crystallinity. The PC loadings can be successfully interpreted in terms of spectroscopy. However, the model of PCA does not impose that both the elements of the scores or those of the loadings were positive. This can be a problem when one tries to build a mathematical model to be related to the theoretical basis of NMR.

MCR (not used here) and MLRC give models respecting these constraints of positivity. In the present work, as the "pure spectra" were actually available, MLRC appeared as a relevant method. Results obtained using this new method were in agreement with crystallinity indices measured by WAXS and comparable with others published data concerning the cellulose Iβ proportion. The NMR data were also successfully fitted using a simplex procedure on the whole collection of data. This method permitted, for each type of cellulose, a proton spin diffusion time to be determined that have been shown to be directly related to the crystallite dimensions (lateral size).

These chemometric methods have been proved efficient for the interpretation of a collection of NMR spectra of lignocellulosic extracts of maize (publication in progress), so is applicable to impure samples.

Acknowledgements

Access to the NMR facilities of the BiBS platform of INRA Angers-Nantes is acknowledged. We gratefully acknowledge Anne-Marie Brossard and Bruno Pontoire for their involvement in the NMR and WAXS analyses, respectively. We thank Dr. A. French (SRRC, USDA, NewOrleans, USA), Dr. H. Chanzy (CERMAV, Grenoble, France) and Dr. S. Nováková (Charles University, Department of Botany, Czech Republic) for their kind gift of cellulose from cotton linters, *Glaucosystis* and hydrolysed *Tunicin*, respectively. This work was in part supported by French National Research Agency (ANR), in the project MAGIC (BLAN08-1_310638).

References

Agarwal, U. P., Reiner, R. S., & Ralph, S. A. (2010). Cellulose I crystallinity determination using FT-Raman spectroscopy: Univariate and multivariate methods. *Cellulose*, *17*, 721–733.

Alemany, L. B., Grant, D. M., Pugmire, R. J., Alger, T. D., & Zilm, K. W. (1983a). Cross polarization and magic angle sample spinning NMR-spectra of model organic-compounds. 1 Highly protonated molecules. *Journal of the American Chemical Society*, *105*, 2133–2141.

Alemany, L. B., Grant, D. M., Pugmire, R. J., Alger, T. D., & Zilm, K. W. (1983b). Cross polarization and magic angle sample spinning NMR-spectra of model organic-compounds. 2 Molecules of low or remote protonation. *Journal of the American Chemical Society*, *105*, 2142–2147.

Atalla, R. H., & VanderHart, D. L. (1984). Native cellulose – a composite of two distinct crystalline forms. *Science*, *223*, 283–287.

Atalla, R. H., & VanderHart, D. L. (1999). The role of solid state C-13 NMR spectroscopy in studies of the nature of native celluloses. *Solid State NMR*, *15*, 1–19.

Bansal, P., Hall, M., Realff, M. J., Lee, J. H., & Bommarius, A. S. (2010). Multivariate statistical analysis of X-ray data from cellulose: A new method to determine degree of crystallinity and predict hydrolysis rates. *Bioresource Technology*, *101*, 4461–4471.

Belton, P. S., Tanner, S. F., Cartier, N., & Chanzy, H. (1989). High-resolution solid-state ^{13}C NMR spectroscopy of Tunicin, an animal cellulose. *Macromolecules*, *22*, 1615–1617.

Bennett, A. E., Rienstra, C. M., Auger, M., Lakshmi, K. V., & Griffin, R. G. (1995). Heteronuclear decoupling in rotating solids. *Journal of Chemical Physics*, *103*, 6951–6958.

Bloembergen, N. (1949). On the interaction of nuclear spins in a crystalline lattice. *Physica*, *15*, 386–426.

Cheung, T. T. P., & Gerstein, B. C. (1981). H-1 nuclear magnetic resonance studies of domain-structures in polymers. *Journal of Applied Physics*, *52*, 5517–5528.

Clauss, J., Schmidt-Rohr, K., & Spiess, H. W. (1993). Determination of domain sizes in heterogeneous polymers by solid-state NMR. *Acta Polymerica*, *44*, 1–17.

Davis, W. E., Barry, A. J., Peterson, F. C., & King, A. J. (1943). X-ray studies of reactions of cellulose in non-aqueous systems II. Interaction of cellulose and primary amines. *Journal of the American Chemical Society*, *65*, 1294–1300.

De Juan, A., & Tauler, R. (2003). Chemometrics applied to unravel multicomponent processes and mixtures – Revisiting latest trends in multivariate resolution. *Analytica Chimica Acta*, *500*, 195–210.

Diniz, J. M. B. F., Gil, M. H., & Castro, J. A. A. M. (2004). Hornification – its origin and interpretation in wood pulps. *Wood Science and Technology*, *37*, 489–494.

Dupont, A. L. (2003). Cellulose in lithium chloride/N,N-dimethylacetamide, optimisation of a dissolution method using paper substrates and stability of the solutions. *Polymer*, *44*, 4117–4126.

Earl, W. L., & VanderHart, D. L. (1981). Observations by high-resolution carbon-13 nuclear magnetic resonance of cellulose I related to morphology and crystal structure. *Macromolecules*, *14*, 570–574.

Hermans, P. H., & Weidinger, A. (1949). X-ray studies on the crystallinity of cellulose. *Journal of Polymer Science*, *4*, 135–144.

Hess, K., & Kissig, H. (1941). Zur Kenntnis der Hochtemperatur-Modifikation der Cellulose (Cellulose IV). *Zietschrift Physikalische Chemie B*, *49*, 235–239.

Hestrin, S., & Schramm, M. (1954). Synthesis of cellulose by *Acetobacter xylinum* II. Preparation of freeze-dried cells capable of polymerizing glucose to cellulose. *Biochemical Journal*, *58*, 345–352.

Heux, L., Dinand, E., & Vignon, M. R. (1999). Structural aspects in ultrathin cellulose microfibrils followed by ^{13}C CP-MAS NMR. *Carbohydrate Polymers*, *40*, 115–124.

Horii, F., Hirai, A., & Kitamaru, R. (1982). Solid-state high-resolution 13C NMR studies of regenerated cellulose samples with different crystallinities. *Polymer Bulletin*, *8*, 163–170.

Horii, F., Hirai, A., & Kitamaru, R. (1987). CPMAS C-13 NMR spectra of the crystalline components of native cellulose. *Macromolecules*, *20*, 2117–2120.

Horii, F. (2001). Structure of cellulose: Recent developments in its characterization. In D. N.-S. Hon, & N. Shiraishi (Eds.), *Wood and cellulose chemistry* (2nd ed., pp. 83–107). New York: Marcel Dekker.

Imai, T., & Sugiyama, J. (1998). Nanodomains of Iβ and Iα cellulose in algal microfibrils. *Macromolecules*, *31*, 6275–6279.

Imai, T., Sugiyama, J., Itoh, T., & Horii, H. (1999). Almost pure I-alpha cellulose in the cell wall of glaucocystis. *Journal of Structural Biology*, *127*, 248–257.

Isogai, A. (1994). Allomorphs of cellulose and other polysaccharides. In R. D. Gilbert (Ed.), *Cellulosic polymers blend and composites* (pp. 1–24). Munich: Hanser Publisher.

Jakob, H. F., Fengel, D., Tschegg, S. E., & Fratzl, P. (1995). The elementary cellulose fibril in *Picea abies*: Comparison of transmission electron microscopy, small-angle X-ray scattering, and wide-angle X-ray scattering results. *Macromolecules*, *28*, 8782–8787.

Jayme, G., & Hunger, G. (1957). The rearrangement of microfibrils in dried cellulose and the implication of this structure alteration on pulp properties. In F. Bolam (Ed.), *Fundamentals of papermaking fibres* (pp. 263–270). Cambridge: BPBMA.

Jandura, P., Kokta, B. V., & Riedl, B. (2000). Fibrous long-chain organic acid cellulose esters and their characterization by diffuse reflectance FTIR spectroscopy, solid-state CP/MAS C-13 NMR, and X-ray diffraction. *Journal of Applied Polymer Science*, *78*, 1354–1365.

Johnson, D. C., & Neogi, A. N. (1989). Sheeted products formed from reticulated microbial cellulose. US Patent 4863565.

Jarvis, M. C., Fenwick, K. M., & Apperley, D. C. (1996). Cross-polarisation kinetics and proton NMR relaxation in polymers of Citrus cell walls. *Carbohydrate Research*, *288*, 1–14.

Jollife, L. T. (2002). *Principal component analysis* (2nd ed.). New York: Springer-Verlag.

Kato, K. L., & Cameron, R. E. (2002). Structural aspects of the thermally accelerated ageing of cellulose: Effect of cellulose source and ageing conditions. *Polymer International*, *51*, 707–714.

Koyama, M., Sugiyama, J., & Itoh, T. (1997). Systematic survey on crystalline features of algal celluloses. *Cellulose*, *4*, 147–160.

Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM Journal of Optimization*, *9*, 112–147.

Larsson, P. T., Westermark, U., & Iversen, T. (1995). Determination of the cellulose Iα allomorph content in a tunicate cellulose by CP-MAS ^{13}C NMR spectroscopy. *Carbohydrate Research*, *278*, 339–343.

Larsson, P. T., Wickholm, K., & Iversen, T. (1997). A CP/MAS ^{13}C NMR investigation of molecular ordering in celluloses. *Carbohydrate Research*, *302*, 19–25.

Lawson, C. L., & Hanson, R. J. (1974). *Solving least-squares problems*. Englewood Cliffs, New Jersey: Prentice-Hall Inc.

Lennholm, H., & Iversen, T. (1994). Determination of cellulose I-alpha and I-beta in lignocellulosic materials. *Carbohydrate Research*, *261*, 119–131.

Lennholm, H., & Iversen, T. (1995). Classification of pulp fibres from different wood species by multivariate data analysis of ^{13}C CPMAS NMR spectra. *Holzforschung*, *49*, 462–464.

Massiot, D., Fayon, F., Capron, M., King, I., Le Calvé, S., Alonso, B., et al. (2002). Modelling one- and two-dimensional solid state NMR spectra. *Magnetic Resonance in Chemistry*, *40*, 70–76.

Masuda, K., Adachi, M., Hirai, A., Yamamoto, H., Kaji, H., & Horii, F. (2003). Solid-state ^{13}C and ^{1}H spin diffusion NMR analyses of the microfibril structure for bacterial cellulose. *Solid State Nuclear Magnetic Resonance*, *23*, 198–212.

Müller, L., Kumar, A., Baumann, T., & Ernst, R. R. (1974). Transient oscillations in NMR cross-polarization experiments in solids. *Physical Review Letters*, *32*, 1402–1406.

Multon, J. L., Bizot, H., & Martin, G. (1991). In J. L. Multon (Ed.), *Mesure de l'eau adsorbée dans les aliments* (pp. 1–64). Paris: Lavoisier, Tec et Doc.

Newman, R. H. (1998). NMR of cellulose crystallite surfaces. *Holzforschang*, *52*, 157–159.

Newman, R. H. (1999). Estimation of the lateral dimensions of cellulose crystallites using ^{13}C NMR signal strengths. *Solid State NMR*, *15*, 21–29.

Newman, R. H. (2004). Carbon-13 NMR evidence for cocrystallization of cellulose as a mechanism for hornification of bleached kraft pulp. *Cellulose*, *11*, 45–52.

Newman, R. H., & Hemmingson, J. A. (1990). Determination of the degree of cellulose crystallinity in wood by carbon-13 nuclear magnetic resonance spectroscopy. *Holzforschung*, *44*, 351–355.

Nishiyama, Y., Langan, P., & Chanzy, H. (2002). Crystal structure and hydrogen-bonding system in cellulose 1 beta from synchrotron X-ray and neutron fiber diffraction. *Journal of the American Chemical Society*, *124*, 9074–9082.

Nyqvist, H. (1983). Saturated salt solutions for maintaining specified relative humidities. *International Journal of Pharmaceutical Technology and Product Manufacture*, *4*, 47–48.

Paris, M., Bizot, H., Emery, J., Buzaré, J. Y., & Buléon, A. (1999). Crystallinity and structuring role of water in native and recrystallized starches by 13C CP-MAS NMR spectroscopy 1: Spectral decomposition. *Carbohydrate Polymers*, *39*, 327–339.

Paris, M., Bizot, H., Emery, J., & Buléon, A. (2001). NMR local range investigations in amorphous starchy substrates: II. Dynamical heterogeneity probed by H-1/C-13 magnetization transfer and 2D WISE solid state NMR. *International Journal of Biological Macromolecules*, *29*, 137–143.

Park, S., Johnson, D. K., Ishizawa, C. I., Parilla, P. A., & Davis, M. F. (2009). Measuring the crystallinity index of cellulose by solid state C-13 nuclear magnetic resonance. *Cellulose*, *16*, 641–647.

Petitpas, T., & Mering, J. (1956). Molecular structure of cellulose II. *Comptes Rendus*, *243*, 47–50.

Press, W. H., Teukolsky, S. A., Vettering, W. T., & Flannery, B. P. (1997). *Numerical recipes in C*. Cambridge: Cambridge University Press.

Purves, C. B. (1954). Chemical nature of cellulose and its derivatives. In H. M. Spurlin, & M. W. Grafflin (Eds.), *Cellulose and cellulose derivatives, part 1* (pp. 29–98). New York: Interscience.

Rondeau-Mouro, C., Bouchet, B., Pontoire, B., Robert, P., Mazoyer, J., & Buléon, A. (2003). Structural features and potential texturizing properties of lemon and maize cellulose microfibrils. *Carbohydrate Polymers*, *53*, 241–252.

Salomonsen, T., Jensen, H. M., Larsen, F. H., Steuernagel, S., & Engelsen, S. B. (2009). Alginate monomer composition studied by solution- and solid-state NMR – A comparative chemometric study. *Food Hydrocolloids*, *23*, 1579–1586.

Sarko, A., & Mugli, J. (1974). Packing analysis of carbohydrates and polysaccharides III. Valonia cellulose and cellulose II. *Macromolecules*, *7*, 486–494.

Sugiyama, J., Harada, H., Fujiyoshi, Y., & Uyeda, N. (1985). Lattice images from ultrathin sections of cellulose microfibrils in the cell wall of Valonia macrophysa. *Kütz, Planta*, *166*, 161–168.

Sugiyama, J., Okano, T., Yamamoto, H., & Horii, F. (1990). Transformation of Valonia cellulose crystals by an alkaline hydrothermal treatment. *Macromolecules*, *23*, 3196–3198.

Sugiyama, J., Persson, J., & Chanzy, H. (1991). Combined IR and electron diffraction study of the polymorphism of native cellulose. *Macromolecules*, *24*, 2461–2466.

Sugiyama, J., Vuong, R., & Chanzy, H. (1991). Electron diffraction study on the two crystalline phases occurring in native cellulose from an algal cell wall. *Macromolecules*, *24*, 4168–4175.

Tang, H. R., Wang, Y. L., & Belton, P. S. (2000). C-13 CPMAS studies of plant cell wall materials and model systems using proton relaxation-induced spectral editing techniques. *Solid State Nuclear Magnetic Resonance*, *15*(4), 239–248.

Thygesen, A., Oddershede, J., Lilholt, H., Thomsen, A. B., & Stahl, K. (2005). On the determination of crystallinity and cellulose content in plant fibres. *Cellulose*, *12*, 563–576.

Vandaele, Y., Revol, J. F., Gaill, F., & Goffinet, G. (1992). Characterization and supramolecular architecture of the cellulose–protein fibrils in the tunic of the sea peach (Halocynthia-Papillosa, Ascidiacea, Urochordata). *Biology of the Cell*, *76*, 87–96.

Van den Berg, O., Capadona, J. R., & Weder, C. (2007). Preparation of homogeneous dispersions of tunicate cellulose whiskers in organic solvents. *Biomacromolecules*, *8*, 1353–1357.

VanderHart, D. L., & Atalla, R. H. (1984). Studies of microstructure in native celluloses using solid state C-13 NMR. *Macromolecules*, *17*, 1465–1472.

Wakelin, J. H., Virgin, H. S., & Crystal, E. (1959). Development and comparison of two X-ray methods for determining the crystallinity of cotton cellulose. *Journal of Applied Physic*, *30*, 1654–1662.

Wickholm, K., Hult, E. L., Larsson, P. T., Iversen, T., & Lennholm, H. (2001). Quantification of cellulose forms in complex cellulose materials: A chemometric model. *Cellulose*, *8*, 139–148.

Wu, X., Zhang, S., & Wu, X. (1988). Two-stage feature of Hartmann–Hahn cross relaxation in magic-angle sample spinning. *Physical Review B*, *37*, 9827–9829.

Yamamoto, H., & Horii, F. (1993). CPMAS ^{13}C NMR analysis of the crystal transformation induced for valonia cellulose by annealing at high temperature. *Macromolecules*, *26*, 1313–1317.