Bioresource Technology 101 (2010) 4461–4471

Contents lists available at ScienceDirect

Bioresource Technology

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

Multivariate statistical analysis of X-ray data from cellulose: A new method to determine degree of crystallinity and predict hydrolysis rates

Prabuddha Bansal, Mélanie Hall, Matthew J. Realff, Jay H. Lee, Andreas S. Bommarius *

School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, 311 Ferst Drive, N.W., Atlanta, GA 30332-0100, USA

ARTICLE INFO

Article history:
Received 24 June 2009
Received in revised form 8 January 2010
Accepted 17 January 2010
Available online 20 February 2010

Keywords:
Crystallinity index
X-ray
Principal component analysis
Principal component regression
Enzymatic hydrolysis

ABSTRACT

The enzymatic hydrolysis of cellulose by cellulases is one of the major steps in the production of ethanol from lignocellulosics. However, cellulosic biomass is not particularly susceptible to enzymatic attack and crystallinity of the substrates is one of the key properties that determine the hydrolysis rates. In this work, by quantifying the respective contributions of amorphous and crystalline cellulose to the X-ray diffraction spectra of cellulose with intermediate degrees of crystallinity, a new method to obtain consistent crystallinity index values was developed. Multivariate statistical analysis was applied to spectra obtained from phosphoric acid pretreated cellulose samples of various intermediate (but undetermined) crystallinity indices to reduce their dimensionality. The crystallinity indices obtained were found to be linearly related to the enzymatic hydrolysis rates. The method was validated by predicting the degree of crystallinity of samples containing various ratios of microcrystalline cellulose and amorphous cellulose, both of known crystallinity indices. Dimensionality reduction of the spectra was also used to predict the enzymatic hydrolysis rates of various cellulose samples from X-ray data. The method developed in this work could be generalized to accurately assess the degree of crystallinity for a wide range of varieties of cellulose.

1. Introduction

Cellulose hydrolysis, which leads to the production of fermentable sugars and ultimately liquid fuels, has attracted the interest of a lot of researchers in the past few years. Resistance of cellulosic biomass to enzymatic attack, resulting from its highly crystalline nature and poor accessibility to enzymes, is one of the main challenges to overcome to make cellulosic ethanol commercial (Himmel et al., 2007). Crystallinity of cellulose has been recognized as one of the major substrate properties that determine the hydrolysis rate and has been the subject of investigation in many studies (Lynd et al., 2002; Mansfield et al., 1999; Zhang and Lynd, 2004). Accurate quantification of the crystalline content in cellulose, termed crystallinity, is thus of prime importance, as it gives an estimation of the recalcitrance of the substrate to the enzymatic attack.

Crystalline regions and lattices are formed due to hydrogen bonds between the cellulose chains and van der Waals forces between the glucose molecules. The degree of crystallinity, an average property, is the fraction of the crystalline content in the sample under consideration. The techniques used for determining the degree of crystallinity of cellulose include X-ray powder diffraction, solid-state ^{13}C NMR, density measurements (Krassig, 1993) and more recently FT Raman spectroscopy (Schenzel et al., 2005), with X-ray diffraction being most widely followed. While ^{13}C NMR is a reliable method for calculating crystallinity, it usually requires extensive acquisition time to obtain good peaks resolution and tends to be not applicable to low degrees of crystallinity, as the crystalline and amorphous peaks are hardly distinguishable. A number of methods to calculate the degree of crystallinity of cellulose from X-ray diffraction spectra have been published (Table 1). One major feature of all the methods (except for the peak-height method (Segal et al., 1959) and method 1 of Wakelin et al. (1959)) is subtraction of the amorphous spectrum as background. While doing so by scaling the acquired spectrum of an amorphous polymer (to bring it below the crystalline spectrum) may be physically meaningful for spectra with sharp peaks, for cellulose it is not as simple due to considerable peak overlaps (the different crystal planes for cellulose I are labeled in Fig. 1). With the advent of software programs such as JADE®, functional deconvolution of spectra with respect to a chosen background is simple; the issue then is the choice of the background. The easy-to-use method of Segal et al. (1959), which is still the most widely used, does not need background subtraction but the definition of a baseline, and is based on peak heights. The degree of crystallinity of cellulose I is given by comparing the minimum in intensity above baseline at $2\theta = 18^\circ$ (I_{am}), and the maximum in intensity at $2\theta = 22.5^\circ$

---

* Corresponding author. Tel.: +1 404 385 1334; fax: +1 404 894 2295.
E-mail address: Andreas.Bommarius@chbe.gatech.edu (A.S. Bommarius).

0960-8524/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.biortech.2010.01.068

Table 1
Published methods on the calculation of crystallinity index from X-ray spectra (for detailed explanations, the reader may refer to the original works).

Reference	Mathematical methodology
Hermans and Weidinger (1948)[a]	Total crystalline and amorphous intensity is calculated from the diffraction pattern (area under the curve) by marking the crystalline and amorphous portions in the spectra. Crystallinity is then expressed as $X_c = \frac{I_c}{I_c + K \times I_a}$ (1) where X_c – crystallinity, I_c – crystalline portion intensity, I_a – amorphous portion intensity, K – empirical constant
Segal et al. (1959)[a]	Ratio of intensities at $2\theta = 22.5°$ to that at $2\theta = 18°$ gives the ratio of crystalline to amorphous fractions (cellulose I) or $2\theta = 19.5°$ to that at $2\theta = 16°$ (cellulose II). 2θ – diffraction angle: $\mathrm{Cr}I = 100 \times (I_{200} - I_{am})/I_{200}$ (2) I_{am} – minimum in intensity above baseline at $2\theta = 18°$, I_{200} – maximum in intensity above baseline at $2\theta = 22.5°$ (I_{200})
Wakelin et al. (1959)[a]	Method 1: correlation of the difference in the intensities of sample and amorphous with difference in intensities of crystalline standard and amorphous. Method 2: area between the sample spectrum and the amorphous spectrum. Relative crystallinity is given by the ratio of this area to that calculated with the crystalline standard
Ruland (1961), Vonk (1973)	Separation of amorphous and crystalline spectra. Amorphous spectrum scaled to match the spectrum of partially crystalline sample at regions where peaks are absent. Crystallinity is given by: $X_c = \frac{\int_{s_0}^{s_1} s^2 I_c ds}{\int_{s_0}^{s_1} s^2 I ds} \times \frac{\int_{s_0}^{s_1} s^2 \bar{f}^2 ds}{\int_{s_0}^{s_1} s^2 \bar{f}^2 D^2 ds}$ (3) where X_c – crystallinity, I_c – intensity of crystalline portion, I – total intensity, $s = 2\sin\theta/\lambda$, 2θ – diffraction angle, $\bar{f}^2$ – mean square of scattering, D – disorder function
Chung and Scott (1973)	Amorphous spectrum expressed as a Gaussian like function and is subtracted as background from the sample spectrum. Crystallinity and a constant k determined by use of following equations: $I_a = k_a x_a$ (4) $I_c = k_c x_c$ (5) $x_c + x_a = 1$ (6) I_a – amorphous portion intensity (area under the diffraction curve), I_c – crystalline portion intensity, x_a – fraction of amorphous component, x_c – fraction of crystalline component, k_a, k_c – constants
Soltys et al. (1984)[a]	Crystalline diffraction pattern was obtained after removing the linear background and scaling the amorphous sample spectrum. Crystallinity calculated as the ratio of area under the crystalline diffraction peaks to the total area
Polizzi et al. (1990)	The background is expressed as a function of the amorphous spectra, and has crystallinity and disorder factor as parameters. These two parameters along with the parameters of the fitting function for the sample spectrum are optimized for the best fit. The background is given by: $I_B(s) = (1 - X_c)I_{am}(s) + X_c\langle f(s^2)\rangle\{1 - \exp(-ks^2)\}$ (7) $I_B(s)$ – background scattering, X_c – degree of crystallinity, $I_{am}(s)$ – experimental intensity of amorphous sample, $\langle f(s^2)\rangle$ – mean square atomic scattering factor, k – disorder factor, $s = 2\sin\theta/\lambda$, 2θ – diffraction angle
Majdanac et al. (1991)[a]	Instrument background subtracted from spectrum, amorphous scattering expressed by a Gaussian function, the peaks in the spectra expressed as Gaussian or Lorentzian functions. Crystallinity is given by the area under the curves (other than the amorphous Gaussian curve) divided by the total area

[a] Developed for cellulose.

(I_{200}), accounting for the amorphous part and the crystalline part (major diffraction from the 200 plane), respectively ($\mathrm{Cr}I = 100 \times (I_{200}–I_{am})/I_{200}$). However, it is clear from Fig. 1 that the trough at 18°, which is assumed to account for the amorphous portion, is shifted to lower angles compared to the actual reflection from a pure amorphous sample (maximum intensity at $2\theta = 19.5°$). Nevertheless, the method is useful for relative comparison and results should be carefully interpreted when used for absolute crystallinity index determination. Thygesen et al. (2005) applied four different methods (peak-height method (Segal et al., 1959), Ruland–Vonk (Ruland, 1961; Vonk, 1973), profile refinement method (Rietveld, 1969) and Debye calculations (Debye, 1915)) to calculate the crystallinity index of four different substrates and showed that the results can vary within a range of up to 20% depending on the method. For Avicel, a pure microcrystalline cellulose I sample and one of the substrates used in this work, calculated crystallinities from X-ray spectra in the literature vary over a large range (54–92.97%) (Table 2). Although differences in the ways to handle the samples (drying method and temperature, amount of sample analyzed) are likely to introduce variations in the results, the work of Thygesen et al. (2005) shows that different crystallinity values can be extracted using different analytical methods on the same spectrum. Given the importance of cellulose crystallinity in the enzymatic hydrolysis and its role in evaluating efficient pretreatment methods (or how to render a cellulose sample more amorphous), we have revisited the topic of crystallinity calculation from X-ray data. We use X-ray powder diffraction for measuring the intensities of beams (averaged over the sample used in the setup) at various diffraction angles to calculate the degree of crystallinity (weight fraction of the crystalline content). While it may be possible to calculate microscopic properties such as the crystallite dimensions corresponding to different phases (Garvey et al., 2005) and structural determination in terms of atomic coordinates (Nishiyama et al., 2002) using the X-ray diffraction data, we do not try to quantify any microscopic property of cellulose other than its degree of crystallinity.

In this work, a data-driven method for calculating the degree of crystallinity of cellulose was developed. Cellulose samples of intermediate degree of crystallinity were prepared by treating Avicel and fibrous cellulose (FC) with varying (and controlled) concentrations of phosphoric acid. Purely amorphous samples were obtained for both types of cellulose. To calculate crystallinity indices, normalized X-ray diffraction spectra were expressed as a linear combination of the normalized untreated cellulose (Avicel or FC) and amorphous cellulose spectra. Principal component analysis (PCA) was also applied to the spectroscopic data (separately to Avicel and FC spectra sets) and the principal component scores were related to calculated crystallinities. This revealed the dimensionality