Conference paper

H. N. Cheng*

NMR analysis of compositional heterogeneity in polysaccharides

DOI 10.1515/pac-2016-1020

Abstract: Many copolysaccharides are compositionally heterogeneous, and the composition determined by the usual analytical or spectroscopic methods provides only an average value. For some polysaccharides, the NMR data contain copolymer sequence information, such as diad, triad, and tetrad sequence intensities. In such cases, it is possible to estimate the extent of compositional heterogeneity through NMR. Two general types of methodologies can be used: perturbed Markovian and discrete component approaches. The theoretical bases for these approaches are reviewed in this work, and three examples are shown of the application of these NMR methodologies to copolysaccharides.

Keywords: alginate; chitosan; composition distribution; heterogeneity; ICS-28; multicomponent model; NMR; pectin; perturbed model; polysaccharides; sequence distribution.

Introduction

Many copolysaccharides are compositionally heterogeneous, containing variations in monosaccharide composition and sequence distributions [1–4]. This heterogeneity may originate from different sources: 1) It can be caused by variations in biosynthesis due to different stages of plant maturation, seasonal variations, or differences in temperature, drought, and nutrient profile during plant growth. 2) Different plant varieties or cultivars may produce copolymers with slightly different composition. 3) The process of extraction of the polysaccharides from a plant or downstream processing may vary, leading to minor differences in composition. 4) If the polysaccharide has been subjected to post-harvest modification (e.g. alkaline hydrolysis, enzymatic reaction, or chemical derivatization), the variations in reaction conditions or process fluctuations may cause changes in composition and sequence distribution. 5) Sometimes a commercial material may be a blend of different batches of materials. In view of the possibility of heterogeneity, NMR analysis of these polysaccharides should be done with care. Frequently the NMR sequence information (e.g., diad and triad sequence intensities) is fitted to simple Bernoullian (**B**), first-order Markovian (**M1**) or second-order Markovian (**M2**) models. For compositionally heterogeneous polymers, the use of these conventional models can be misleading.

A number of NMR methodologies have been developed to address compositional heterogeneity [5–17]. For simplicity, these treatments can be grouped into two general types: continuous function and discrete-component approaches.

Article note: A collection of invited papers based on presentations at the XXVIII International Carbohydrate Symposium (ICS-28), New Orleans, USA, July 17–21, 2016

***Corresponding author: H. N. Cheng,** USDA Agricultural Research Service, Southern Regional Research Center, 1100 Robert E. Lee Blvd., New Orleans, LA 70124, USA, e-mail: hn.cheng@ars.usda.gov

In the continuous function treatment, one can use the perturbed Markovian models, such as symmetric [11] and non-symmetric functions [15] or the general (function-independent) formalism [15]. For example, the exponentially modified Gaussian (EMG) function [15] can be utilized, where the Bernoullian probability is represented not by one value but by a distribution of values, f(z):

$$f(z) = \frac{N}{\tau\sigma\sqrt{2\pi}} \int_0^\infty \exp\left[-\frac{(z - P_1 - z')^2}{2\sigma^2} - \frac{z'}{\tau} \right] dz'$$

where z is the Bernoullian probability, N the area under the Gaussian, σ the standard deviation, τ the skew factor, z′ the dummy variable of integration, and P_1 the average value of Bernoullian probability in the absence of exponential modification. The EMG model for the first-order Markovian probabilities can be similarly expressed [15]. For these models, the theoretical expressions for polymer composition, diad, triad, and tetrad sequences have been previously derived [15]. The observed NMR sequence intensities can be fitted to the theoretical expressions to obtain P_1, σ, and τ.

In the discrete-component approaches [7–9], the polymer is regarded as the mixture of two or more discrete components. The components may be separate chains or joined together as a block copolymer. In this approach, the composition and the sequence intensities are the weighted averages of the corresponding compositions and sequence intensities of all the components.

$$I_\beta = \sum_\kappa w_\kappa \cdot I_{\beta\kappa}$$

where I_β is the total intensity for sequence β, w_κ is the weight fraction for component κ, and $I_{\beta\kappa}$ is the intensity for sequence β and component κ. In simple cases, the observed and the calculated NMR sequence intensities can be fitted to the two-component **B**/**B** or **M1**/**M1** model; if NMR sequence data on polymer fractions are available, it may be possible to combine the NMR data of the fractions and fit them to three- or four-component models.

The methodologies described above were initially developed to treat compositional heterogeneity observed in synthetic polymers [7–16]. The same approaches have also been shown to be applicable to copolysaccharides [18–22]. In this work a review is made of the NMR analysis of three polysaccharides that exhibit compositional heterogeneity.

Chitosan

Chitin is a homopolymer of 2-acetamido-2-deoxy-β-D-glucopyranose (GlcNAc). Chitosan is partially deacetylated chitin and may be considered a copolymer of GlcNAc residue and 2-amino-2-deoxy-β-D-glucopyranose (GlcN) residue [3].

The ^{1}H and ^{13}C NMR spectra of chitosan have been previously assigned by Varum et al. [23]. They have also reported triad sequence intensities from both heterogeneous and homogeneous deacetylation reactions [3, 23, 24]. Their data can be fitted to both conventional (unperturbed) and perturbed Markovian models [22]. The NMR triad data from two chitosan samples, one made via homogeneous reaction, and the other one via heterogeneous reaction, together with analysis, are given in Table 1, where A = acetylated unit (GlcNAc residue) and D = deacetylated unit (GlcN residue).

With the unperturbed **M1** model, the mean deviations are relatively large (1.4–3.0 %). When the perturbed **M1-EMG** model is used, the mean deviations decrease substantially, showing a much better fit. Thus, the NMR analysis of these chitosan samples indicates that both are heterogeneous in composition. This finding is consistent with the results of fractionation and other analyses [3], which also showed chitosan to be compositionally heterogeneous.

Table 1: Observed NMR data of two samples of chitosan[a] and analysis[b].

Triad	Homogeneous reaction product			Heterogeneous reaction product		
	I_{obsd}	I_{calc}(M1)	I_{calc}(M1-EMG)	I_{obsd}	I_{calc}(M1)	I_{calc}(M1-EMG)
AAA	15	15.0	15.0	11	11.8	11.1
AAD	28	26.8	28.0	26	25.8	25.7
DAD	10	12.0	9.0	14	14.1	13.6
ADA	14	14.0	15.0	15	15.2	16.3
DDA	16	22.8	16.0	20	23.8	20.2
DDD	17	9.3	17.0	13	9.3	13.1
MD[c]		3.0	0.3		1.4	0.4
P_{AD}		0.472	0.466		0.523	0.544
P_{DA}		0.551	0.577		0.560	0.587
σ		–	0.116		–	0.111
τ		–	– 0.035		–	– 0.017

[a]The letter I denotes NMR intensities (in %); data taken from Ref. [24]. [b]Analysis reported in Ref. [22]. [c]Mean deviation between observed and calculated triad intensities (in %).

Alginates

Alginates are anionic copolysaccharides isolated from brown algae or produced through bacterial fermentation [4]. They are copolymers of β-D-mannuronate (M) and α-L-guluronate (G) residues. The monomer residues may be arranged in homopolymer, alternating, or random sequences [4], e.g.

--GGGGGG-- --MGMGMGMG-- --GMGGMMG--

NMR studies on alginates have been well documented in the literature [25–27]. The assignments for both ^{1}H and ^{13}C NMR spectra are known; diad and triad sequence distributions can be readily derived from the NMR spectra [25, 26]. It was shown earlier that previously published NMR data on alginate [25, 26] could be fitted to perturbed Markovian and/or discrete component models [19]. Furthermore, three commercial samples of sodium alginates were fractionated by size exclusion chromatography (SEC), and several SEC fractions were analyzed by NMR using the same methodology as delineated here [21]. The coupled SEC-NMR data showed that at least two separate components (and perhaps even three) could be identified. The possible linkage of these two components to the enzymatic action in the M → G epimerization reaction and their molecular weight dependence were noted in the earlier report [21].

For example, the NMR data for two (unfractionated) alginate samples [21] are given in Table 2, where large mean deviations are observed for the conventional **M1** model (around 4 %) but mean deviations decrease to < 1 % with the two-component **M1**/**M1** model.

Pectin

Commercially available pectin is obtained by extraction from citrus peel, apple pomace, and sugar beet. Pectin has a complex structure, of which about 60 % consists of homogalacturonan, with a backbone of α-1,4-linked GalA, which is partly methylated at C-6 [2]. One of the major determinants in pectin properties is the degree of esterification (DE). Sucrose is usually used to form gels with high methoxy pectin (DE > 50 %), whereas Ca^{++} is often added for the gelation of low methoxy pectin (DE < 50 %). Other than being a gelling agent, pectin can also serve as a thickener and stabilizer in food applications [2].

Table 2: NMR analysis of commercial alginate samples.[a]

Sequence	Sample 1			Sample 2		
	I_{obsd}	I_{calc} (M1)	I_{calc} (M1/M1)	I_{obsd}	I_{calc} (M1)	I_{calc} (M1/M1)
MM	19.1	19.1	19.1	32.4	32.4	32.4
MG	24.0	24.0	24.0	49.5	49.5	49.5
GG	56.9	56.9	56.9	18.1	18.1	18.1
MGM	5.9	2.1	8.8	19.9	14.3	19.1
GGM	6.4	19.8	6.4	11.3	20.9	11.3
GGG	56.4	47.0	53.7	11.7	7.6	12.4
One-component 1st order Markov (**M1**) model[b]						
		P_{GM} 0.174			P_{GM} 0.578	
		P_{MG} 0.386			P_{MG} 0.433	
		MD 4.4			MD 3.2	
Two-component 1st order Markov (**M1**/**M1**) model[b]						
Component 1			P_{GM} 0.004			P_{GM} 0.006
			P_{MG} 0.991			P_{GM} 0.998
			w_1 0.533			w_1 0.110
Component 2			P_{GM} 0.745			P_{GM} 0.774
			P_{MG} 0.382			P_{GM} 0.432
			w_2 0.467			w_2 0.890
			MD 0.9			MD 0.3

[a]Data taken from Ref. [21]. The letter I denotes NMR intensities (in %). [b]MD = mean deviation between observed and calculated intensities (in %).

A typical ^{13}C NMR spectrum of a citrus pectin with a DE of 70 % is given in Fig. 1. The assignments [28–31] have been labeled on the spectrum, where G and E refer to galacturonic acid and ester, respectively. The ester carbonyl appears on the spectrum as four peaks corresponding to the E-centered triad sequences (Fig. 1 inset). In addition, C-1, C-4, and C-5 (at 99.6, 78.6 and 71 ppm, respectively) all show splitting due to their sensitivity towards the ester formation and can be used to determine the DE of the sample. Other features of the spectrum include the methyl ester peak at 54.3 ppm, and numerous small peaks from neutral sugars, primarily rhamnose, galactose, arabinose, xylose, and fucose. In addition, the ^{1}H NMR spectrum (not shown)

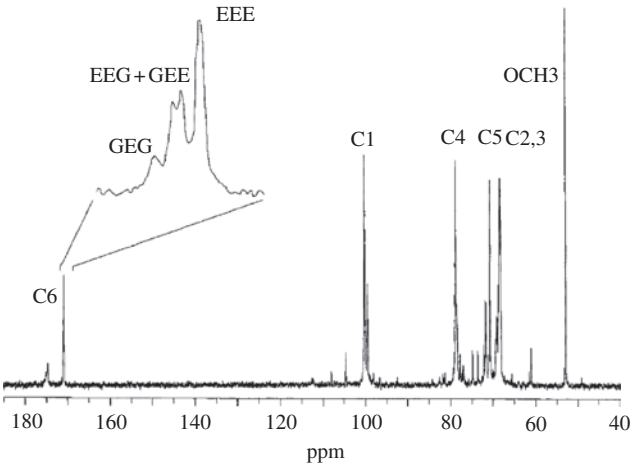

Fig. 1: ^{13}C NMR spectrum of high-methoxy pectin (reproduced from Ref. [18]).

is also rich in sequence information and can be combined with ^{13}C to provide more detailed G/E sequence distributions.

In previous work [18, 21], the present author was involved with a research program to use NMR to determine the sequence distribution of G and E residues in a series of pectin samples. For example, the NMR triad data for two fractions of a pectin sample extracted from citrus peel are given in Table 3. These fractions were obtained through a proprietary process, where fraction C2 was expected to contain a more blocky distribution of unesterified GalA units than fraction C1.

The analysis can be carried out with conventional, perturbed, and multicomponent Bernoullian models. The fit with the **B** model gives relatively large mean deviations (Table 3). The results obtained for the perturbed **B-EMG** model give smaller mean deviations; the non-zero values obtained for σ and τ indicate that both fractions are heterogeneous in nature. The σ and τ values also permit the chemical composition distribution (CCD) curves to be calculated as shown in Fig. 2.

For the multicomponent model, since both fractions C1 and C2 originate from the same pectin sample, their NMR data can be combined and analyzed simultaneously, thereby increasing the degree of freedom for the analysis. As shown in Table 3, the data can be fitted to a three-component **B/B/B** model with DE values of 77 %, 67 % and 2 %. Components 1 and 2 are random in methyl ester sequence distribution while component 3 is an almost homopolymer block of galacturonan. As expected, the two fractions C1 and C2 comprise different proportions of these three components, as indicated in Table 3 and (as bar graphs) in Fig. 2. Fraction C2 contains a higher amount of component 3 than fraction C1, indicating a somewhat blockier microstructure.

Thus, in agreement with the current understanding of pectin structure [2], the NMR data analysis confirms that pectin is compositionally heterogeneous and that it contains blocks of GalA. The combination of NMR and fractionation is very helpful because it permits the three-component model to be tested in this case.

Table 3: Observed triad sequence intensities for two pectin fractions C1 and C2, together with results fitted by 1st order Markov, perturbed and multicomponent models.[a]

Triad	I_{obsd} for fraction C1	I_{obsd} for fraction C2
EEE	43.4	32.6
EEG	27.2	27.5
GEG	3.2	8.5
EGE	16.1	10.3
GGE	4.9	11.8
GGG	5.2	9.4
Bernoullian model (**B**)		
P_E	0.757	0.688
MD	2.0	2.7
Bernoullian exponentially modified Gaussian model (**B-EMG**)		
P_E	0.773	0.675
σ	0.090	0.152
τ	− 0.021	− 0.028
MD	1.1	0.7
Three-component Bernoullian model (**B/B/B**), for both fractions		
Component 1	$P_E = 0.77$ (higher ester, random)	
w_1	0.92	0.29
Component 2	$P_E = 0.67$ (slightly lower ester, random)	
w_2	0.04	0.64
Component 3	$P_E = 0.02$ (almost all acid block)	
w_3	0.04	0.07
MD	1.2	

[a]Adapted from Ref. [18]. I_{obsd} denotes observed NMR intensities (in %). MD denotes mean deviation between observed and calculated triad intensities (in %).

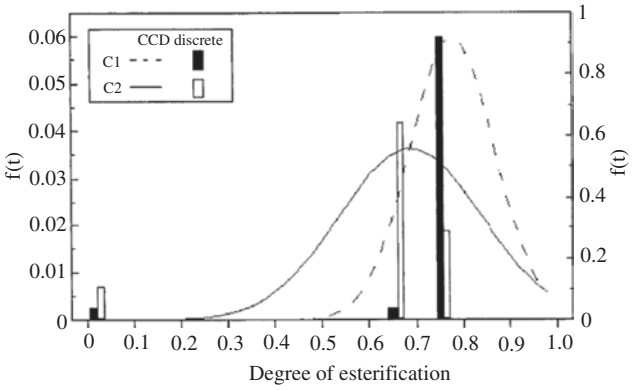

Fig. 2: Calculated chemical composition distribution curves (from perturbed model) and the compositions of the three components (from three-component model) for pectin fractions C1 and C2 (reproduced from Ref. [20]).

A few comments may be made as to the utility of the NMR methodology. First, at least some of the NMR peaks for a copolysaccharide spectrum need to show splittings due to sequence distribution; this is a prerequisite for NMR sequence analysis. If the splittings are small, higher magnetic field, better spectral resolution, and sometimes higher probe temperatures can improve peak separation. Secondly, the peak intensities should optimally be split to the triad (or higher) sequence level. Thirdly, the precision of the peak intensities must be optimized. It is recommended that peak integration be repeated at least two to three times and an average value taken. Curve deconvolution can also help to obtain more accurate intensities when the peaks are overlapped. Finally, other analytical techniques (such as fractionation, SEC, HPLC, IR, and mass spectrometry) may be combined with NMR to maximize the overall information content. For example, these techniques may offer additional proofs of the heterogeneity of the polysaccharide in question. Alternatively, if the polysaccharide can be separated into different fractions (e.g. through fractionation or chromatography), the NMR data from the fractions can be analyzed simultaneously to provide further information.

Conclusions

This article aims to review the theoretical basis and the methodologies that have been developed for the NMR analysis of heterogeneous polymer systems. Three examples have been given to show how these techniques can be applied to the analysis of NMR data of polysaccharides. In these cases, the sequence intensities have been tested by either the perturbed or multicomponent models for the presence of compositional heterogeneity. As advances in technology enable the magnetic field frequency to increase (now at 1 GHz for ^{1}H NMR), the NMR spectra of more copolysaccharides may show sensitivity to sequence effects, and the sequence intensities may also be measured at greater precision. These developments will hopefully permit more copolysaccharides to be analyzed in the future by the NMR methodologies described herein.

Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

References

[1] G. M. Hallman, S. G. Whittington. *Macromolecules* **6**, 386 (1973).
[2] A. G. J. Voragen, G. Coenen, R. P. Verhoef, H. A. Schols. *Struct. Chem.* **20**, 263 (2009).
[3] M. H. Ottoy, K. M. Varum, O. Smidsrod. *Carbohydr. Polym.* **29**, 17 (1996).

[4] K. I. Draget, S. T. Moe, G. Skjak-Braek, O. Smidsrod. Alginates, in *Food Polysaccharides and their Applications*, A. M. Stephen, G. O. Phillips, P. A. Williams (Eds.), pp. 289–334, Taylor and Francis, Boca Raton (2006).
[5] Y. Inoue, Y. Itabashi, R. Chujo, Y. Doi. *Polymer* **25**, 1640 (1984).
[6] J. F. Ross. *J. Macromol. Sci.-Chem.* **A23**, 1451 (1986).
[7] H. N. Cheng. *J. Appl. Polym. Sci.* **35**, 1639 (1988).
[8] H. N. Cheng. *Macromolecules* **24**, 1724 (1991).
[9] H. N. Cheng. S. B. Tam, L. J. Kasehagen. *Macromolecules* **25**, 3779 (1992).
[10] H. N. Cheng. *J. Appl. Polym. Sci.: Appl. Polym. Symp.* **51**, 21 (1992).
[11] H. N. Cheng. *Macromolecules* **25**, 2351 (1992).
[12] H. N. Cheng. *Makromol. Chem.: Theory Simul.* **1**, 415 (1992).
[13] H. N. Cheng. *Makromol. Chem.: Theory Simul.* **2**, 901 (1993).
[14] H. N. Cheng. *Makromol. Chem.: Theory Simul.* **2**, 561 (1993).
[15] H. N. Cheng. *Macromolecules* **30**, 4117 (1997).
[16] H. N. Cheng. *Int. J. Polym. Anal. Charact.* **4**, 71 (1997).
[17] H. N. Cheng, M. Miri. *ACS Symp. Ser.* **1077**, 371 (2011).
[18] T. G. Neiss, H. N. Cheng, P. J. H. Daas, H. A. Schols. *ACS Polym. Prepr.* **39**(1), 688 (1998).
[19] H. N. Cheng. *Polym. Bull.* **43**, 247 (1999).
[20] T. G. Neiss, H. N. Cheng, P. Daas, H. A. Schols. *Macromol. Symp.* **140**, 165 (1991).
[21] T. G. Neiss, H. N. Cheng. *ACS Symp. Ser.* **834**, 382 (2002).
[22] H. N. Cheng. *ACS Polym. Prepr.* **44**(1), 261 (2003).
[23] K. M. Varum, M. W. Anthonsen, H. Grasdalen, O. Smidsrod. *Carbohydrate Res.* **211**, 17 (1991).
[24] K. M. Varum, M. W. Anthonsen, H. Grasdalen, O. Smidsrod. *Carbohydrate Res.* **217**, 19 (1991).
[25] H. Grasdalen, B. Larsen, O. Smidsrod. *Carbohydr. Res.* **89**, 179 (1981).
[26] H. Grasdalen. *Carbohydr. Res.* **118**, 255 (1983).
[27] B. T. Stokke, O. Smidsrod, P. Bruheim, G. Skjak-Braek. *Macromolecules* **24**, 4637 (1991).
[28] H. Grasdalen, O. E. Bakoy, B. Larsen. *Carbohydr. Res.* **184**, 183 (1988).
[29] E. Westerlund, P. Aman, R. Andersson, R. E. Andersson, S. M. M. Rahman. *Carbohydr. Polym.* **15**, 67 (1991).
[30] E. Westerlund, P. Aman, R. E. Andersson, R. Andersson. *Carbohydr. Polym.* **14**, 179 (1991).
[31] A. K. Andersen, B. Larsen, H. Grasdalen. *Carbohydr. Res.* **273**, 93 (1995).