A New Model for the Substitution Patterns in the Polymer Chain of Polysaccharide Derivatives

Petra Mischnick*

Braunschweig University of Technology, Institute of Food Chemistry, Schleinitzstrasse 20, D-38106 Braunschweig, Germany

Christian Hennig

University of Hamburg, Center for Model Building and Simulation, Faculty of Mathematics, Bundesstrasse 55, D-20146 Hamburg, Germany

Received September 13, 2000; Revised Manuscript Received November 27, 2000

A new mathematical model is presented for the analysis of the substituent distribution in the polymer chain of polysaccharide derivatives. For the first time, the influence of substitution on the reactivity of neighbored monomer units is taken into account. The model was applied to various cellulose and amylose ethers and esters and excellently fits the experimental data.

Introduction

Chemical modification of polysaccharides, especially cellulose and starch, has a long tradition and is still of great interest in the development of new functional polymeric materials.[1] The properties of these glucan derivatives are determined by several structural parameters as the gross structure of the polymer backbone, the molecular weight distribution, the type of the substituent (nonionic, anionic, cationic, hydrophilic, hydrophobic), the degree of substitution (DS), and the distribution of substituents. To analyze the latter is a complex task, since various structural levels can be differentiated. Many methods have been developed to determine the regioselectivity of the reaction within the glucosyl unit, e.g. by NMR spectroscopy, the fractions of un-, mono-, di- and trisubstituted monomers, e.g., by HPLC of the hydrolyzed sample, or the total monomer composition by means of NMR, HPAEC, and preferably GLC and GLC−MS after degradation of the polymer and, if required, appropriate derivatization.[2] For many properties, such as solubility, viscosity, formation of supramolecular structures, or biodegradability, the distribution of substituents in the polymer chain is more decisive. Strategies for the analysis of these patterns use the selectivity of enzymes[3−5] or a random degradation and subsequent mass spectrometry.[2,6−10] Furthermore, efforts have been made to get insight in the topochemistry of the derivatization procedure, especially for starch, which is usually reacted in the granular form with its typical layered architecture of amylose and amylopectin.[11−13]

To find out whether the substituent pattern in the glucosyl unit is random without any dependence on the status of the other hydroxy groups, Spurlin[14] developed a mathematical model for cellulose derivatives containing the following assumptions:

1. The relative rates of the reaction at O-2, O-3, and O-6 (k_2, k_3, and k_6) remain constant over the course of the reaction; that means that there is no influence of primary substitution on the reactivity of other OH groups on the same glucosyl unit.

2. All glucosyl units are equally accessible.

3. Terminal groups of the cellulose chain can be neglected.

For a given set of partial DS values x_2, x_3, and x_6 (DS = $x_2 + x_3 + x_6$) the monomer composition can be calculated and compared with the experimental data. Agreement indicates that the model assumptions are fulfilled. This Spurlin I model was then further differentiated, taking into account that for some cellulose ethers the reactivity of O-3 increased when the 2-position of the same glucosyl moiety has been substituted. A further reactivity constant $k'_3$was introduced for this enhanced reactivity. This model usually fits well for methyl and hydroxyalkyl celluloses.[15,16]

In 1995, Arisz et al.[6] determined the substituent distribution in trimers obtained from methyl cellulose after partial random hydrolysis and deuteriomethylation by FAB−MS. This pattern was compared to a theoretical distribution which can be calculated under the assumption of independence of the state of a monomer from neighboring monomer units ("independence model", for dimers see the section "A Model for Neighbor Dependence"). We used a similar approach for methyl[7−8] and silyl ethers,[2] cellulose sulfates,[9] and acetates[10] and found the following patterns: homogeneous (i.e., as to be expected under independence), more heterogeneous, more regular and bimodal, depending on the reaction conditions.[7]

Although Arisz et al.[6] calculated the standardized total deviation from the model and introduced it as a heterogeneity parameter H_n (n denoting the number of monomers in the oligomer fraction considered), the interpretation follows a more qualitative course. Therefore, we developed a new mathematical model taking into account the influence of substitution of a glucosyl residue on the probability of substitution at the neighbored monomer units. We now report on this model and its first applications.

10.1021/bm000093c CCC: $20.00
Published on Web 12/21/2000

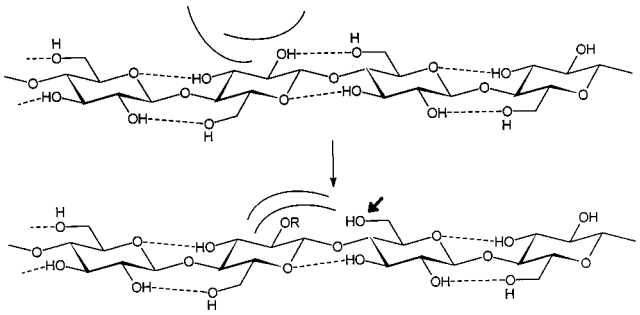

Figure 1. Interruption of the intramolecular hydrogen bonding pattern of cellulose I (arrow) and local change of solvation (symbolized by bended lines) by primary substitution in a heterogeneous reaction.

A Model for Neighbor Dependence

Deviations from a pattern of independence of substituents in the polymer chain of starch or cellulose derivatives especially occur for those prepared under kinetic control in heterogeneous reactions, as is the case for most of the industrially produced ethers. The first reaction of a glucosyl unit changes the hydrogen-bonding pattern and the hydrophilicity at this monomer unit and its near neighborhood. Figure 1 shows the hydrogen pattern of cellulose and its interruption by a primary substitution. Therefore, it could be expected that a mathematical model including this conditional probability would enable an improved fit of the experimental data.

To define such a model, we assume that dimers are observed. We denote the probability that the first monomer unit of a pair is *i*-substituted, and the second one is *j*-substituted, by $P(i, j)$, $i, j = 0, 1, 2, 3$. By the standard definition of conditional probability, we have

$$P(i, j) = p_i p_{j|i} \tag{1}$$

where p_i is the probability that the first monomer unit is *i*-substituted and $p_{j|i}$ is the conditional probability that the second monomer unit is *j*-substituted, given that the first monomer unit is *i*-substituted. Note that, by observation, we do not distinguish between the "first" and "second" monomer unit of a dimer. Therefore, p_i can be interpreted as the probability that an arbitrary monomer unit is *i*-substituted, which could be the second unit of a dimer as well. That is, for $i = 0, 1, 2, 3$

$$p_i = p_0 p_{i|0} + p_1 p_{i|1} + p_2 p_{i|2} + p_3 p_{i|3} \tag{2}$$

In previous work,[6−10] only the case was considered that the probability of substitution is independent of the state of the neighbor monomer unit. This means that $p_{j|i} = p_j$ and

$$P(i, j) = p_i p_j \quad \text{for all } i, j \tag{3}$$

We call this model M_0 (the "independence model").

Given a dimer of an *i*- and a *j*-substituted monomer unit, we only observe the distribution of the sum of its substituents, i.e., $i + j$. We denote as C_0, ..., C_6 the relative frequencies of 0−6 substituents of a dimer. Further we observe the distribution of the number of substituents of the single monomer units. That is, we may estimate all p_i as the observed relative frequencies c_0, c_1, c_2, c_3 of un-, mono-, di-, and trisubstituted units, while the conditional probabilities $p_{j|i}$ cannot be observed. That is, M_0 can be applied directly, unlike a general model of the form (1). But the values C_0, ..., C_6 can be used to estimate the conditional probabilities in such a way that the substitution of a monomer unit may depend on the substitution of its neighbor as discussed above.

We introduce a dependence between the neighbors by assuming that for a given *j* there can be two distinct values for $p_{j|i}$, depending on *i*. We consider two dependence models:

1. Model M_1 distinguishes between unsubstituted and substituted neighbors. $p_{j|0}$ is the probability for a *j*-substituted monomer unit, given that its neighbor is unsubstituted. $p_{j|+}$ denotes the probability for a *j*-substituted unit, given that its neighbor is substituted somehow, i.e.

$$p_{j|1} = p_{j|2} = p_{j|3} = p_{j|+} \tag{4}$$

2. Model M_2 distinguishes between trisubstituted and less-substituted neighbors. $p_{j|3}$ is the probability for a *j*-substituted monomer unit, given that its neighbor is trisubstituted. $p_{j|-}$ denotes the probability for a *j*-substituted unit, given that its neighbor has less than three substituents, i.e.

$$p_{j|0} = p_{j|1} = p_{j|2} = p_{j|-} \tag{5}$$

It is not possible to allow all $p_{j|i}$ to be distinct, since then the 11 observed frequencies c_0, ..., c_3, C_0, ..., C_6 do not suffice to identify all the 16 conditional probability values. (Note that the irrelevance of the order of the two members of a dimer, i.e., $P(i, j) = P(j, i)$, does not imply $p_{j|i} = p_{i|j}$.) Therefore, both models require some of the $p_{j|i}$ to be equal. Thus, we introduced (4) and (5). The estimation of the conditional probabilities is explained in the "Materials and Methods" section.

The choice between the models M_1 and M_2 should be made in such a way that the relation between the relative frequencies of monomers of the two classes of neighbors distinguished by the model is as close as possible to 50:50. For example, if there would be only 5% unsubstituted monomers, 95% of the monomers would be used to estimate $p_{j|+}$, while only 5% would enter in the estimation of $p_{j|0}$, which would result in $p_{j|+} \approx p_j$ and an almost meaningless value of $p_{j|0}$. That is, model M_2 should be used under conditions where there are few unsubstituted but more trisubstituted monomer units (high DS), and model M_1 should be used when there are more unsubstituted monomer units (low DS).

Results and Discussion

The models were tested for various cellulose and amylose derivatives. The oligomer patterns after random degradation were determined by FAB and MALDI mass spectrometry as has been described.[7,9−10]

Derivatives with $c_0 > c_3$. For these derivatives with DS values usually lower than 1.5 (for monomodal distributions), the model M_1 was applied, differentiating between unsubstituted (corresponding to c_0) and substituted (corresponding to $c_1 + c_2 + c_3$) monomer units. Figure 2 (corresponding to Table 1) illustrates the result for a methyl amylose (DS 0.44)

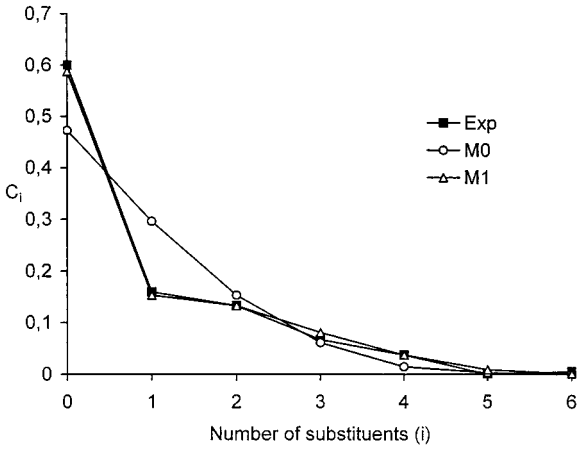

Figure 2. Mole fractions C_i of dimers with i substituents ($i = 0, 1, 2, 3, 4, 5, 6$). Comparison of the experimental data (exp) for a methyl amylose (DS 0.44) with those calculated by models M_0 (random under independence) and M_1 (with conditional probability). Data are given in Table 1.

Table 1. Methyl Amylose with Heterogeneous Distribution of Substituents, DS 0.44, Conditional Probabilites $p_{i|j}$ and Molar Fractions C_i As Calculated from Models M_0 and M_1 (for Illustration See Figure 2)

i	c_i(exptl)	$p_{i\|0}$	$p_{i\|+}$
0	0.6876	0.853749	0.321901
1	0.2151	0.121498	0.421120
2	0.0773	0.024753	0.192958
3	0.0200	0	0.064021

i	C_i(exptl)	$C_i(M_0)$	$C_i(M_1)$
0	0.6000	0.472794	0.587038
1	0.1589	0.295806	0.152783
2	0.1322	0.152571	0.132486
3	0.0669	0.060759	0.080496
4	0.0369	0.014579	0.037109
5	0	0.003092	0.008808
6	0.0051	0.000400	0.001280
std error		0.071621	0.008303

prepared under heterogeneous conditions (water/dioxan/NaOH/CH_3I).[17] The molar ratios of dimers with 0−6 substituents (C_0−C_6) as determined by mass spectrometry, as calculated by the independence model M_0 and by the extended model M_1, are compared in this graphic. The methyl amylose sample shows a very high heterogeneity with a much higher amount of two neighbored unsubstituted glucosyl units (DP2, $n(CH_3) = 0$, C_0) than calculated from the monomer composition by model M_0. The standard error (per parameter, see the "Materials and Methods" section) of the pattern under independence is 0.072. In contrast, the new model M_1 excellently fits to the experimental data. The probability that a glucosyl unit remains free is 2.65 times higher beside an unsubstituted residue than in the neighborhood of an already derivatized one ($p_{0|0}/p_{0|+}$). At the same time the reactivity of a glucosyl unit beside a substituted one is enhanced by a factor of 4.64, compared to those next to an unsubstituted neighbor ($\sum_{j=1}^{3} p_{j|+}/\sum_{j=1}^{3} p_{j|0}$). The standard error per parameter for M_1 is only 0.008.

The second example is a cellulose sulfate (DS 0.35, see Table 2) which was prepared from cellulose acetate ($DS_{Ac} = 2.5$) in DMF with chlorosulfonic acid.[9] The determination

Table 2. Cellulose Sulfate Prepared from a Cellulose Acetate, DS 0.35, Conditional Probabilites $p_{i|j}$ and Molar Fractions C_i As Calculated from Models M_0 and M_1 (for Illustration See Figure 3)

i	c_i(exptl)	$p_{i\|0}$	$p_{i\|+}$
0	0.6910	0.719289	0.625715
1	0.2696	0.259275	0.293035
2	0.0367	0.021437	0.071573
3	0.0027	0	0.009677

i	C_i(exptl)	$C_i(M_0)$	$C_i(M_1)$
0	0.4935	0.476527	0.496532
1	0.3446	0.372395	0.347754
2	0.1139	0.123786	0.116966
3	0.0270	0.024078	0.032012
4	0.0160	0.002983	0.006134
5	0.0050	0.000222	0.000572
6	0	0.000009	0.000029
std error		0.013918	0.004931

of the substituent distribution in the fraction of dimers comprised the following steps: permethylation, sulfate−acetyl exchange, acetyl−deuteriomethyl exchange, partial methanolysis (random degradation), again deuteriomethylation, and FAB−MS analysis as described.[9] Compared to the model M_0, the cellulose sulfate showed a slight deviation from a pattern under independence (standard error per parameter 0.014). Again the extended model excellently fits the experimental pattern (standard error per parameter 0.005). Looking to the dimers obtained by partial random degradation of the polymer the average DS of monomer units beside an unsubstituted glucosyl residue is 0.30 (calculated from the $p_{j|0}$, $j = 1, 2, 3$), while the average value for monomer units beside a substituted moiety is 0.46 (calculated from the $p_{j|+}$, $j = 1, 2, 3$). The probability of substitution for an unsubstituted glucosyl moiety is enhanced by a factor of 1.33 under the condition that the neighbored monomer unit is already derivatized in at least one position ($\sum_{j=1}^{3} p_{j|+}/\sum_{j=1}^{3} p_{j|0}$). So there is again a positive intermonomeric effect, increasing the reactivity in the neighbored unit by primary substitution. Comparison of experimental data for C_0−C_6 with models M_0 and M_1 are shown in Figure 3 for the cellulose sulfate.

Derivatives with $c_0 < c_3$. For samples of higher DS values, model M_2 is more appropriate, which differentiates between un-, mono-, and disubstituted units as a group (corresponding to $c_0 + c_1 + c_2$), and trisubstitituted monomers (corresponding to c_3), since the best adaption of the experimental data can be expected for a ratio as close as possible to 50:50. As a first example the results for an commercial cellulose acetate (DS 2.44) are presented[10] in Table 3. Figure 4 shows the molar composition of dimers (C_0−C_6) as determined by mass spectrometry and as estimated from the monomer fractions by models M_0 and M_2. The standard error of the experimental and the calculated data is 0.022 for model M_0 and 0.003 for model M_2, indicating the high efficiency of the latter. The probability that a glucosyl moiety located beside a trisubstituted one becomes also trisubstituted ($p_{3|3}$) is 1.28 times higher than the probability for it to occur beside a less substituted neighbor ($p_{3|-}$). Comparable results are obtained for a commercial methyl cellulose (DS 1.82) with an enhanced

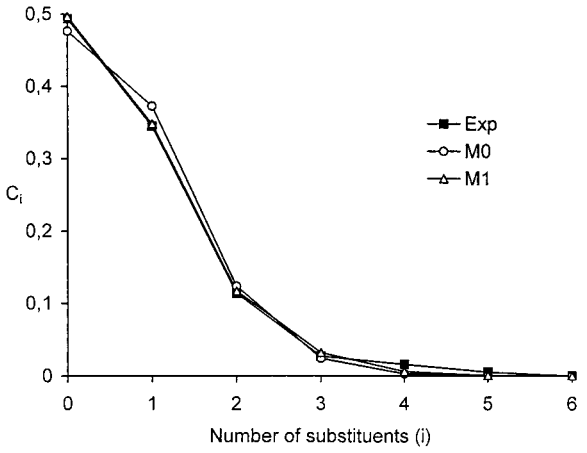

Figure 3. Mol fractions C_i of dimers with i substituents ($i = 0, 1, 2, 3, 4, 5, 6$). Comparison of the experimental data (exp) for a cellulose sulfate (DS 0.35) with those calculated by models M_0 (random under independence) and M_1 (with conditional probability). Data are given in Table 2.

Table 3. Cellulose Acetate, Commercial Sample, DS 2.44, Conditional Probabilites $p_{i|j}$ and Molar Fractions C_i As Calculated from Models M_0 and M_2 (for Illustration See Figure 4)

i	c_i(exptl)	$p_{i\|3}$	$p_{i\|-}$
0	0.0123	0.010940	0.013923
1	0.0791	0.051119	0.112510
2	0.3643	0.333651	0.400835
3	0.5443	0.604290	0.472733

i	C_i(exptl)	$C_i(M_0)$	$C_i(M_2)$
0	0	0.000151	0.000171
1	0	0.001945	0.002485
2	0.0248	0.015216	0.018900
3	0.0864	0.071010	0.084455
4	0.2114	0.218795	0.211226
5	0.3521	0.396571	0.353821
6	0.3253	0.296312	0.328943
std error		0.021395	0.002931

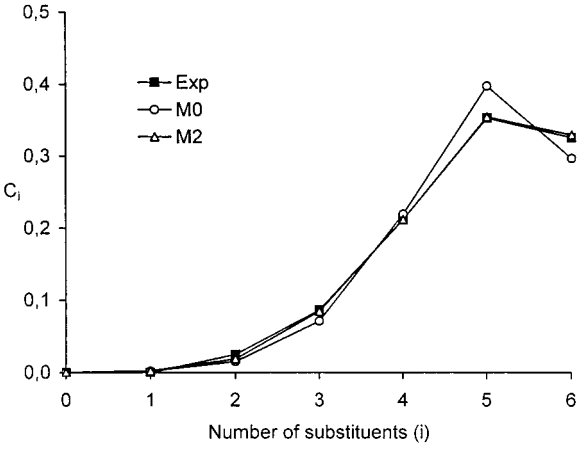

Figure 4. Mol fractions C_i of dimers with i substituents ($i = 0, 1, 2, 3, 4, 5, 6$). Comparison of the experimental data (exp) for a cellulose acetate (DS 2.44) with those calculated by models M_0 (random under independence) and M_2 (with conditional probability). Data are given in Table 3.

probability of the sequence of two permethylated monomers of 1.62 (details not shown).

The model even works for samples showing a bimodal substitution pattern with high ratios of un- and trisubstituted

Table 4. Methyl Amylose with Bimodal Distribution of Substituents, DS 1.72, Conditional Probabilites $p_{i|j}$ and and Molar Fractions C_i As Calculated from Models M_0 and M_2 (for Illustration See Figure 5)

i	c_i(exptl)	$p_{i\|3}$	$p_{i\|-}$
0	0.0918	0	0.139217
1	0.4389	0	0.665605
2	0.1287	0.111256	0.137710
3	0.3406	0.888744	0.057467

i	C_i(exptl)	$C_i(M_0)$	$C_i(M_2)$
0	0.0166	0.008427	0.012780
1	0.1072	0.080582	0.122205
2	0.3500	0.216263	0.322693
3	0.1288	0.175507	0.151380
4	0.0545	0.315542	0.042946
5	0.0368	0.087670	0.045290
6	0.3061	0.116008	0.302706
std error		0.135075	0.015659

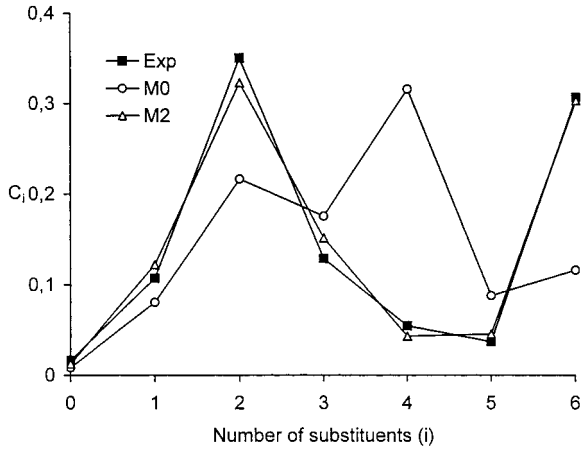

Figure 5. Mol fractions C_i of dimers with i substituents ($i = 0, 1, 2, 3, 4, 5, 6$). Comparison of the experimental data (exp) for a methyl amylose (DS 1.72) with those calculated by models M_0 (random under independence) and M_2 (with conditional probability). Data are given in Table 4.

units. These are usually obtained when methylation is performed in an aprotic solvent with solid sodium hydroxide as the base as a result of two competing reactions: in solution, and on the solid surface in adsorbed domains of the originally dissolved polysaccharide.[7,17−18] For a methyl amylose prepared in DMSO with $NaOH/CH_3I$ (DS 1.72, see Table 4), the ratio $p_{3|3}/p_{3|-}$ is 15.5 compared to 1 for the independence model M_0 in agreement with the proposed reaction mechanism. The standard error per parameter of models M_0 and M_2 from the experimental data are 0.135 and 0.016, respectively. The comparison of models M_0 and M_2 with the experimental data is illustrated in Figure 5.

The very good fit to the experimental data by the new models confirms that primary substitution of a glucosyl unit in the polymer chain of cellulose or amylose mainly influences the accessibility and reactivity of its direct neighbors. Like the formation of a nucleus, primary substitution might effect a change in the molecular structure as for example hydrogen-bonding interactions, conformational rigidity, or solvation, e.g., by the ionic or lipophilic nature of the primary substituent, and therefore the local accessibility and reactivity. The probabilities of new "nucleus formation"

and propagation of the reaction around this "nucleus" determine the outcome of polymer analogous modifications under kinetic control.

Conclusion

A mathematical model has been developed that allows the estimation of the substitution pattern in dimeric sequences of a polysaccharide derivative taking into account the influence of substitution of one monomer unit on the reactivity of the neighbored glycosyl moieties. This intermonomeric effect can be quantified by the new model.

Materials and Methods

Analytical data for amylose and cellulose derivatives have been determined as described elsewhere.[7,9,10,17]

The conditional probabilities of the models M_1 and M_2 were calculated by the method of least squares.

In detail: The values C_0, ..., C_6 can be interpreted as estimators for the theoretical probabilities P_0, ..., P_6 that the total number of substituents of a dimer is 0, ..., 6. These theoretical probabilities can be calculated from the probabilities $P(i, j)$ of pairs of i-, and j-substituted monomers as follows:

$$P_0 = P(0, 0)$$

$$P_1 = P(0, 1) + P(1, 0)$$

$$P_2 = P(0, 2) + P(1, 1) + P(2, 0)$$

$$P_3 = P(0, 3) + P(1, 2) + P(2, 1) + P(3, 0)$$

$$P_4 = P(1, 3) + P(2, 2) + P(3, 1)$$

$$P_5 = P(2, 3) + P(3, 2)$$

$$P_6 = P(3, 3)$$

The $P(i, j)$ can be replaced by $c_i p_{j|i}$ by use of (1), where the p_i are estimated by the observations c_i. With that, the $p_{j|i}$, i, $j = 0, 1, 2, 3$ can be estimated by minimization of the sum of squared errors

$$\mathrm{SSE} = \sum_{i=0}^{6} (C_i - P_i)^2$$

under the following constraints:

1. All conditional probabilities must lie between 0 and 1.
2. (2) has to be fulfilled.
3. $\sum_{i=0}^{3} p_{i|j} = 1$ for $j = 0, 1, 2, 3$.
4. (4) must hold for model M_1 and (5), respectively, for model M_2. Note that the numerical calculation for both models is the same, since the role of $p_{j|0}$ for M_1 is analogous to that of $p_{j|3}$ for M_2.

The minimization problem under constraints 2−4 can be solved by use of the method of Lagrange factors. If such a solution violates condition 1, a minimum on the border of the set of possible solutions can be found by successively turning suitable inequalities of condition 1 to equations and applying the Lagrange factors method again.

The quality of the model fits can be compared by means of the standard error per parameter $(\mathrm{SSE}/7)^{1/2}$.

Acknowledgment. We gratefully acknowledge financial support of our work by the Deutsche Forschungsgemeinschaft (DFG-Mi-398/4-2). We would like to thank Simon Karrenberg, University of Hamburg, who wrote the software for the calculations of the "Materials and Methods" section, using MATHEMATICA (registered trademark of Wolfram Research).

References and Notes

(1) Klemm, D.; Philipp, B.; Heinze, T.; Heinze, U.; Wagenknecht, W. *Comprehensive Cellulose Chemistry*; Wiley-VCH: Weinheim, Germany, 1998; Vols. 1 and 2.
(2) Mischnick, P.; Heinrich, J.; Gohdes, M.; Wilke, O.; Rogmann, N. *Macromol. Chem. Phys.* **2000**, *201*, 1985.
(3) Wirick, M. G. *J. Polym. Sci. A-1: Polym. Chem.* **1968**, *6*, 1705 and 1965.
(4) Gelman, R. A. *J. Appl. Polym. Sci.* **1982**, *27*, 2597.
(5) Horner, S.; Puls, J.; Saake, B.; Klohr, E.-A.; Thielking, H. *Carbohydr. Polym.* **1999**, *40*, 1.
(6) Arisz, P: W. F.; Kauw, H. J. J.; Boon, J. J. *Carbohydr. Res.* **1995**, *271*, 1.
(7) Mischnick, P.; Kühn, G. *Carbohydr. Res.* **1996**, *290*, 199.
(8) Kern, H.; Choi, S. W.; Wenz, G.; Heinrich, H.; Ehrhardt, L.; Mischnick, P.; Garidel, P.; Blume, A. *Carbohydr. Res.* **2000**, *326*, 67.
(9) Gohdes, M.; Mischnick, P. *Carbohydr. Res.* **1998**, *309*, 109.
(10) Heinrich, J.; Mischnick, P. *J. Polym. Sci. A* **1999**, *37*, 3011.
(11) Steeneken, P. A. M.; Woortmann, A. J. J. *Carbohydr. Res.* **1994**, *258*, 207.
(12) Van der Burgt, Y. E. M.; Bergsma, J.; Bleeker, I. P.; Mijland, P. J. H. C.; Van der Kerk-van Hoof, A.; Kamerling, J. P.; Vliegenthart, J. F. G. *Carbohydr. Res.* **1998**, *312*, 201.
(13) Richardson, S.; Nilsson, G. S.; Bergquist, K.-E.; Gorton, L.; Mischnick, P. *Carbohydr. Res.* **2000**, *328*, 365.
(14) Spurlin, H. M. *J. Am. Chem. Soc.* **1939**, *61*, 2222.
(15) Reuben, J. *Macromolecules* **1984**, *17*, 156.
(16) Reuben, J. *Carbohydr. Res.* **1986**, *157*, 201. Reuben, J.; Conner, H. T. *Carbohydr. Res.* **1983**, *115*, 1.
(17) Mischnick, P. Analytische Charakterisierung von Polysacchariden und Polysaccharidderivaten mittels chemischer, enzymatischer und massenspektrometrischer Methoden. Habilitationsschrift, Hamburg 1996.
(18) Liebert, T.; Heinze, T. In *Cellulose Derivatives, Modification, Characterization, and Nanostructures*; Heinze, T. J., Glasser, W. G., Eds.; ACS Symposium Series 688; American Chemical Society: Washington DC, 1998; p 61.

BM000093C