Interpreting Size-Exclusion Data for Highly Branched Biopolymers by Reverse Monte Carlo Simulations

Christopher J. C. Watts,[†] Angus Gray-Weale,[‡] and Robert G. Gilbert*,[†]

Key Centre for Polymer Colloids, School of Chemistry, University of Sydney, Sydney, New South Wales 2006, Australia

Received May 27, 2006; Revised Manuscript Received November 19, 2006

Size-exclusion chromatography with multiple detection provides data on the distributions of various properties in a branched polymer sample, for example, distributions of the number, average mass, mean-squared mass, and branching fraction against hydrodynamic volume. A method is developed that provides a basis to use such data for obtaining structural and biosynthetic information on highly branched polymers, such as amylopectin. We generate by simulation a reference distribution of randomly branched polymers from the experimental distribution of debranched chains of the target polymer. We then select from these simulated chains a set with the same number (or other) distribution as the actual polymer sample, using reverse Monte Carlo simulations. Properties of these model polymers are used to interpret the differences with experiment as due to correlations in branching structure. The same methodology can be applied to data from other separation techniques such as field-flow fractionation and high-performance anionic exchange chromatography.

Introduction

Size-exclusion chromatography (SEC) separates by size, and to an acceptable approximation for many systems (both linear and branched), this size is the hydrodynamic volume (V_h) of the polymer molecule.[1−4] For linear polymers, there is a direct relation between size and molecular weight, and SEC data for linear polymers can be plotted with minimal model assumptions and can reveal considerable information on the polymers' (bio)-synthetic process (see, for example, refs 5−10). Branched polymers differ because there is no one-to-one relationship between size and molecular weight: A sample wherein all chains have the same hydrodynamic volume will contain many molecular weights. It is therefore difficult to obtain information from SEC without extensive model-based assumptions which ipso facto are hard to refute.

One approach might be to assume a set of reactions (including, for biopolymers such as starch, influences such as crystallinity and location within the grain[11,12]) describing the (bio)synthetic process, modeling the kinetics, and thus generating a size distribution. Such a procedure however probably cannot refute a mechanistic hypothesis, since the rate parameters in the reaction mechanism are not certain and are adjusted to fit experiment. Extracting unambiguous mechanistic information from SEC data is particularly difficult for starch: Even though it is a homopolymer, its amylopectin component is very highly branched.

For a given sample of a branched polymer, multiple in-line SEC detectors, using, for example, differential refractive index, fluorescence, light scattering, viscosity, and NMR, give the mass of the polymer, number of chains, viscosity, and branching fraction. It is thus possible to obtain a number of different distributions, for example, $P(V_h)$ and $W(V_h)$, defined, respectively, as the number and weight of chains with elution volume V_h. The number distribution can be obtained by two separate methods: placing a single fluorophore on each chain (by fluorescent labeling of the reducing end of each starch chain) or combining in-line viscometric and differential refractive index detection.[2,13,14] The weight distribution is that from conventional differential refractive index detection. For example, multiple detection data have been reported for starch by Takeda and coworkers.[15−17] However, the availability of multiple types of distributions does not in itself provide an assumption-free means of obtaining useful mechanistic and structural information from SEC data.

The objective of this paper is to develop a theoretical tool that enables SEC V_h distributions for branched biopolymers and synthetic polymers, particularly starch, to be interpreted with as few assumptions as possible, which has the potential to lead to qualitative and quantitative understanding of the structure and the (bio)synthetic pathways. The new tool is adapted from the reverse Monte Carlo[18] method, usually used to generate arrangements of atoms consistent with partial structural information such as structure factors measured in scattering experiments.[19]

We note at this point that it is important to differentiate between the instantaneous and cumulative molecular weight or hydrodynamic volume distributions. The instantaneous distribution is the distribution of chains formed at any moment, while the cumulative distribution obtained from a polymer sample is the accumulation of the instantaneous distributions over the entire synthetic process. Any experiment perforce samples the cumulative distribution. The instantaneous distribution can be obtained experimentally by taking out samples at short time intervals as the polymer is synthesized and subtracting suitably normalized distributions from successive time intervals.[20,21] Under many circumstances, specifically where the rates controlling chain growth and stoppage do not change significantly during the course of the synthesis of most of the polymer present, the instantaneous and cumulative distributions are very

* Author to whom correspondence should be addressed. Current address: Hartley Teakle Building, University of Queensland, Brisbane, Queensland 4072, Australia. Fax: +61 7 3365 1188. E-mail: b.gilbert@uq.edu.au.

[†] Key Centre for Polymer Colloids, School of Chemistry.

[‡] School of Chemistry.

10.1021/bm0605199 CCC: $37.00
Published on Web 01/10/2007

similar. For notational convenience, we take that to be the case here, noting that the aforementioned subtraction technique to obtain so-called pseudo-instantaneous distributions has wide applicability.[7,9,20,21]

The new tool developed in the present paper is analogous to the following methodology[9,22] for interpreting SEC results on linear polymers (where the distribution in V_h can be uniquely converted to one in molecular weight and thus degree of polymerization N). Consider, for a linear chain, the instantaneous number distribution $P(N)$ of chains with degree of polymerization N. In a linear polymer, many of the steps for chain growth and stoppage can be random. For example, in free-radical polymerization, this applies to the processes of propagation (growth) and both transfer and termination by combination or disproportionation; in each case, $P(N)$ has the functional form[22]

$$P(N) = \exp\left(-\frac{k_{stop}}{k_{growth}}N\right) \quad (1)$$

where k_{stop} and k_{growth} are the rate coefficients for stoppage and growth, respectively. (An arbitrary normalization constant has been omitted.) To interpret SEC data with minimal model-based assumptions, one starts with the very general hypothesis that some parts of the (bio)synthetic process might involve random growth and stoppage; eq 1 then implies that this length distribution should be linear if plotted logarithmically. Because this is a plausible simplest-case scenario, this log plot, the "ln P method", provides a good means of displaying the data. (In quantitative application of this methodology to unbranched chains, it is necessary to take SEC band broadening into account,[8,9,23] and doubtless the same will be true for the adaptation of the new method developed here for branched polymers; fortunately, a general means of implementing this has been devised.[23])

This methodology has been applied to unbranched polymers in a number of cases, including[8] to the length distribution of the enzymatically debranched (i.e., linear) chains of starch. For unbranched chains, regions showing linearity in ln P can be assumed to obey the random growth/stoppage hypothesis, and in regions showing nonlinearity, the ln P plot can be interpreted in terms of deviations from random growth and stoppage. Applied to starch, this has led to a number of insights into biosynthesis.[8]

The means developed in the present paper for a corresponding approach to branched polymers is that a similar assumption can be made for branching—that the simplest possible scenario is when the branches are joined randomly. A crucial component of the new method is that the distributions of unbranched chains from which the randomly branched simulated polymer is assembled has been measured and denoted $P_{debr}(N)$. As with the debranched chains, this randomness is not assumed to be uniformly applicable, or indeed occurring at all, but merely a useful starting point for data interpretation. The simulation then provides the means of seeing if the type of data in question either is consistent with random branching or is able to distinguish whether or not the system is randomly branched.

The principle of the new method is to generate a population of randomly branched chains from $P_{debr}(N)$ subject to the constraints that this branched population belongs to one or more observed distributions, for example, that it belongs to the observed number distribution $P(V_h)$. One can then compare another type (or types) of distribution calculated from this random population with experiment, for example, comparing the calculated randomly branched weight distribution $W_{random}(V_h)$ with the observed distribution $W_{obs}(V_h)$. Where the calculated and observed distributions agree, then we might assume that the chains are randomly branched. Alternatively, the simulations might suggest that the experiment in question is incapable of distinguishing whether or not the chain is randomly branched. Assumptions other than random branching could be tested in this way, and we construct the simulation method to be as flexible as possible. The approach is illustrated schematically in Figure 1, which compares the approaches for linear and for branched polymers.

While the methodology here is specifically developed for SEC data, it can equally well be applied to data for branched polymers using other separation techniques, particularly field-flow fractionation[24−26] and high-performance anion exchange chromatography.[27−29]

We would emphasize that simulating an ensemble of randomly branched polymers is not an attempt to simulate an actual starch molecule, but rather to provide a reference to see if appropriate SEC data for an actual ensemble of starch molecules have regions of hydrodynamic volume where the signal cannot be distinguished from that predicted for a randomly branched polymer. It is possible, but highly unlikely, that data on the structure of starch will prove consistent with this random hypothesis over a broad range of hydrodynamic volumes. Generating this hypothetical randomly branched structure is important exactly because we do not yet know what a randomly branched polymer's distributions would look like. One of the most useful results of the new methodology is that it can identify whether or not a particular type of SEC data (e.g., using different kinds of detectors) can actually reveal structurally and/or mechanistically interesting information about the branched polymer, either that the polymer is randomly branched (which has mechanistic inferences) or that the data are unable to distinguish random and nonrandom branching.

Some synthetic polymers exhibiting long-chain branching could in fact be randomly branched, e.g., poly(butyl acrylate) and poly(vinyl acetate). The present method, which requires the debranched distribution, would be directly applicable to some types of long-chain branches in poly(vinyl esters), which can be debranched by hydrolysis,[30,31] but there are no obvious means of debranching polyacrylates, a point to which we return later.

The type of data that can be analyzed by the new method can be obtained using existing SEC equipment. Unfortunately, there appears to be no such data reported yet in the literature for amylopectin, which is a prime candidate for the new method. The closest data at present are those reported by Takeda and co-workers,[15−17] who reported distributions for debranched chains and number and mass distributions, but the latter two data sets are reported in terms of elution time, without the calibration in terms of standards with known hydrodynamic volume (e.g., whose Mark−Houwink parameters in the particular solvent system are established) necessary to convert elution time to hydrodynamic volume. These data are used to exemplify the new procedure by assuming a calibration based on simulated hydrodynamic volumes.

The implementation of the new interpretation technique has two requirements: (i) the statistical-mechanical and computational tools to generate a population of randomly branched chains where the branches are selected randomly from a given $P_{debr}(N)$ and to calculate the hydrodynamic volume of each chain and (ii) the means of ensuring that the selected chains are from the chosen target experimental distribution, which for definiteness we take to be the observed number distribution $P(V_h)$. The second of these tasks is implemented using the reverse Monte

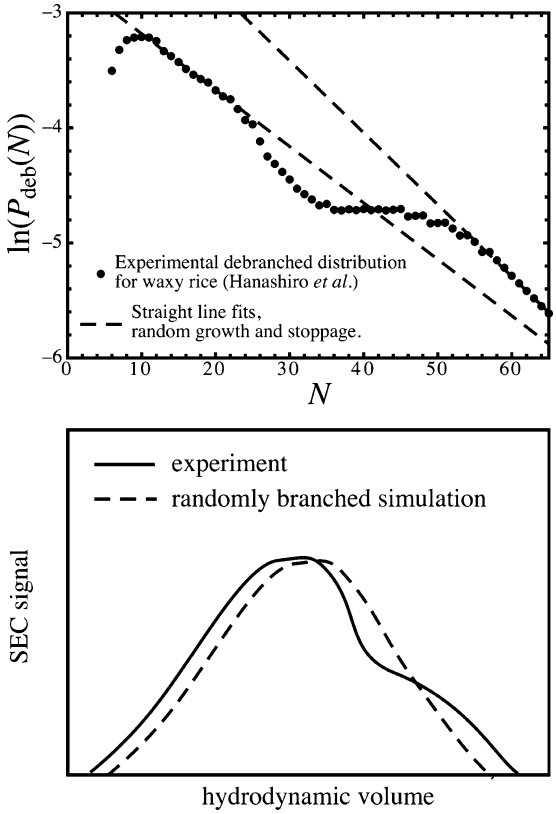

Figure 1. Comparison of assumption-free means of inferring information on the structure and biosynthetic processes for unbranched chains and branched chains. The top panel shows the experimental debranched MWD of waxy rice, taken from the molar-based distribution combined with the calculated DP in Figure 2 (subgraph labeled "Whole") of the work of Takeda et al.[15] The dashed lines show regions N in which $\ln(P(N))$ is linear, indicating random growth and stoppage. The bottom panel shows a schematic of the analogous treatment of branched polymers. Where the experimental SEC distribution resembles or is parallel to the simulation with random branching, we might assume that the polymers are randomly branched or at least whatever correlations are present are not perceived by SEC.

Carlo[18] (RMC) method. Most computer simulations of condensed matter are based on an approximate Hamiltonian that describes interactions, and the quality of the simulation then depends a great deal on that Hamiltonian. The great advantage of the RMC method is that it needs no model for interactions but instead uses the Metropolis algorithm[32] to select model structures consistent with the measured distributions (usually structure factors). Reverse Monte Carlo is developed here in a new direction, demonstrating for the first time its use with SEC experiments, the particular target being starch. In place of the structure factors, we match our model structures to distribution data from SEC experiments.

Although the main focus of the present paper is amylopectin (the highly branched type of starch), the method can also be applied to other polymers, as long as a debranched distribution is available. The most natural next application is to glycogen, the more randomly structured animal analog of starch, or to a synthetic hyperbranched polymer such as polyglycerol. (We have preliminary results interpreting experimental data[33] on the latter.) In the case of a synthetic polymer it is more difficult to measure the debranched distribution experimentally because it is usually very hard to snip the branches without breaking other bonds, but in fact the debranched molecular weight distribution (MWD) can be reliably calculated given the polymerization mechanism (e.g., free-radical polymerization) and an experimental measurement of branching frequency (e.g., using NMR). Thus, for example, in a free-radical polymerization in which extensive long-chain branching occurs, the number distribution of debranched chains is an exponential in the degree of polymerization, whose single parameter is the product of the total (undebranched) degree of polymerization and the branching fraction.

For this first treatment we confine our attention to small hydrodynamic volumes, so avoiding the computational complexity of the self-avoiding polymer chain. This approximation confers a benefit: By comparison with experiment we see when repulsive forces begin to affect conformation.

Random Walk Models and the Mark−Houwink Parameters

The basic element of our model structures is a random walk. As is frequently done in such calculations, we group a number of monomer units together and treat them as a single entity; the number of monomer units in this entity is the so-called persistence length s. The mean-squared radius of gyration of a linear polymer of degree of polymerization (DP) N in a θ solvent is given by

$$\langle R_g^2 \rangle_{RW} = \frac{C_\infty N}{6} sb^2 \qquad (2)$$

where C_∞ is the expansion factor, b is the distance between monomer units (which for vinyl polymers is twice the C−C bond length), and RW refers to s random walk. For simplicity, our model structures are assumed to be in a θ solvent, a restriction that is discussed further below (and can be avoided as will be shown in later work from our group). The value of s could be obtained from a small-scale rotational isomeric state

(RIS) calculation or from experimental data on a linear polymer. Here it is treated as a parameter whose value is determined as part of the "calibration" between elution time and hydrodynamic volume, as discussed later.

One of the input parameters to our random walk simulation is the size of a monomer unit (b in eq 2). Ideally, we would obtain this value from experiment, but our method would be much more powerful if it did not require such a measurement. We examine this issue here. Linear polymers have a simple relationship between hydrodynamic volume and DP, though this may vary between solvents and monomers. This can be found from the Mark−Houwink relation

$$[\eta] = KM^{a} \tag{3}$$

where $[\eta]$ is the intrinsic viscosity, K and a are the Mark−Houwink parameters (for a θ solvent, $a = {}^1/_2$), and M is the molecular weight; when the dimensions of the quantities in eq 3 are considered, it is important to note that this molecular weight is dimensionless (being defined as a ratio), whereas "molar mass" has the same numerical value but has units of g mol^{-1}. This empirical relation is usually valid for all except short chains, but surprisingly for linear starch there are indications that it is also valid for oligomers of low DP.[34] Flory[35] showed that the intrinsic viscosity may be eliminated to obtain

$$V_h = \frac{4}{3}\pi \frac{K(M_0 N)^{a+1}}{6^{3/2}\Phi} \tag{4}$$

where M_0 is the molar mass of a monomer, N is the DP, and $\Phi \approx 2.6 \times 10^{23}$ mol^{-1} is the Flory constant.[36−38] Equation 4 is useful for the calibration of elution time against hydrodynamic volume using linear standards of a polymer with known Mark−Houwink parameters; alternatively V_h can be measured directly with in-line viscometric detection.[14]

Sheridan et al.[39] have shown from simulations that for hyperbranched polymers the mean-squared radius of gyration $\langle R_g^2 \rangle$ and hydrodynamic volume are related to an acceptable approximation by

$$V_h \approx \frac{4}{3}\pi \langle R_g^2 \rangle^{3/2} \tag{5}$$

This is not exactly true,[38,40] but it is expected on dimensional grounds that at least $V_h \propto R_g^3$; extensions of the method given here using better approximations are straightforward but require increased computational resources.

Equations 4 and 5 relate the radius of gyration to the degree of polymerization, as does eq 2. If we compare the two relations in a ratio, then

$$\frac{\langle R_g^2 \rangle_{hyd}}{\langle R_g^2 \rangle_{RW}} = \frac{(\Phi^{-1}K)^{2/3} M_0}{sC_\infty b^2} \tag{6}$$

where $\langle R_g^2 \rangle_{hyd}$ refers to the mean-squared radius of gyration calculated from eqs 4 and 5, derived from the Mark−Houwink and Flory relations.

Evaluation of the ratio of eq 6 using literature values of C_∞ and of the Mark−Houwink parameter K for polymer/solvent systems where $a = {}^1/_2$ (i.e., θ conditions) for polymers such as styrene and methyl methacrylate yields ratios of the "hydrodynamic" and "random walk" radii of gyration that are quite close to (e.g., 30% less than) unity. However, doing this for the Mark−Houwink value of K inferred[34] for linear starch, in a solvent (ammonium acetate solution in water) where $a \approx 0.48$, yields ratios that are ~2.5 times smaller than that predicted by any plausible random walk model, e.g., with the persistence length taken as 5 monomers and the monomer length taken as 0.5 nm. This anomaly cannot be ascribed to an incorrect single set of Mark−Houwink parameters, because those cited[34] are near those measured for starch by a variety of methods. It is noted in this context that it has been reported that the empirical Mark−Houwink relationship is inaccurate for other polysaccharides.[41]

This anomaly, which is unexplained, poses for the present purpose the problem of finding the hydrodynamic volume of a chain found by a random walk simulation so as to construct the target distribution. The problem is simply avoided by using Mark−Houwink parameters calculated from simulated linear polymers with $b = 0.5$ nm. As long as we construct these linear polymer models with the same method as the hyperbranched ones, this approach is consistent. In fact, if we were to choose a different monomer length, then both experimental and simulated volumes would be scaled, because the monomer size is the only quantity with the dimensions of length input to the calculation. We choose also a θ solvent for the present modeling. If the Mark−Houwink parameter a (see eq 3) for linear starch were different from 0.5, then this would scale the $\log(V_h)$ axes in our results. To account consistently and fully for solvent interactions we would need a Mark−Houwink-like relation for a hyperbranched polymer, and none is available yet. The fact that our simulations do agree with the experiment at low molecular weight suggests that this approximation is not too severe. We are working on a mean-field theory of randomly hyperbranched polymers that allows us to study exactly this solvent effect. We can consistently make comparisons without a good knowledge of the monomer size. Experiment can then determine precisely the magnitude of the volumes involved.

Simulation Method

The simulations consist of two stages. First, a master ensemble of branched polymers with evenly distributed total degrees of polymerization is generated. These represent all possible randomly branched chains, with any properties of interest recorded for each chain, such as the radius of gyration and the number of branch points. Each of these polymers is generated by randomly selecting linear chains from $P_{debr}(N)$, until the total number of monomer units in the selection passes a desired total degree of polymerization N_{tot}. These chains are then randomly attached to each other, with their three-dimensional conformations generated in the present implementation as random walks; more realistic models, particularly the RIS treatment,[42,43] could also be implemented (which would require significantly greater computational resources). The usual neglect of higher-than-pair correlations in RMC corresponds to the assumption of random branching in our work. This approximation becomes for us a powerful tool; comparison of our model structures to real polymers tells us how random the branching is. For these first calculations we allow the chains to overlap,[19] in effect assuming a θ solvent.

A recent evaluation of the Mark−Houwink parameters for linear (debranched) starch reports a Mark−Houwink exponent of $a = 0.48$ in a solvent commonly used for SEC of starch (see eq 3 above and accompanying discussion),[34] and this exponent is close to the value of ${}^1/_2$ obeyed by a random walk. While using a random walk approximation is likely to be less accurate for branched polymers than for linear polymers, we deem it suitable for the first application of the RMC method to SEC;

we discuss its effect further below. The only limitation placed on the branching is that there can be only one branch per monomer; all other physical limitations are ignored.

The simulations here use a non-self-avoiding random walk. The radius of gyration R_g of a particular branched polymer chain is found using

$$R_g^2 = \left\langle \frac{1}{N_{tot}} \sum_{i=1}^{N_{tot}} r_i^2 \right\rangle \quad (7)$$

where r_i is the distance of a monomer from the polymer's center of mass and the angle brackets denote averaging across many random walks corresponding to a particular selection of debranched chains and branch points. This choice of R_g, averaged over the configurations of the chains for each branching structure, is appropriate because the hydrodynamic volume is averaged over the fluctuations in the conformation of the polymer that do not break any bonds.

The second part of the simulation selects from the master ensemble a subensemble of polymers that have some desired (measured) distribution of hydrodynamic volumes, and this is the stage that uses RMC. The hydrodynamic volume distribution of this subensemble is then our reference for the real polymers. Knowing that the model matches the measured distribution of hydrodynamic volumes, we can compare its distribution of some other property to experiment. Any discrepancy is then a measure of nonrandom branching over the applicable range of V_h.

Reverse Monte Carlo resembles conventional Monte Carlo simulation in that the Metropolis method is used to sample a Boltzmann probability distribution

$$p(x) \propto e^{-H(x)/k_B T} \quad (8)$$

where k_B is Boltzmann's constant, T is the temperature, H is the Hamiltonian, and x represents all of the variables describing the model polymers' structures. In RMC, the Hamiltonian is no longer the usual physical one but rather a measure of how closely the ensemble matches the desired distribution in some property. We wish to generate a subensemble of polymers with a specified distribution of hydrodynamic volumes $P_0(V_h)$. If $P(V_h)$ is the distribution of our subensemble, then we choose the "Hamiltonian"

$$H = {}^1/_2 \int_0^\infty (P(V_h) - P_0(V_h))^2 \, dV_h \quad (9)$$

This integral is positive and definite, vanishing when P equals P_0. The "temperature" in eq 8 is not the thermodynamic temperature,but rather a parameter that governs the size of the permitted fluctuations of P about P_0. The simulation starts at high k_BT, with a random selection of polymers from the master ensemble. Moves consisting of swaps of polymers in the subensemble for those in the master ensemble are accepted with the Metropolis criterion.[32] As k_BT is lowered, the fluctuations of P about P_0 become small. This method is closely related to simulated annealing.

The simple quadratic function of eq 9 is easily minimized, and other methods might well be used to match $P(V_h)$ to $P_0(V_h)$, but this calculation is the prototype of more powerful and realistic ones, where RMC is clearly the best strategy. As an illustration, suppose one wanted to construct an ensemble of polymers matching two different distributions, $P(V_h)$ and $Q(N_{tot})$. We would use in place of eq 9

$$H = {}^1/_2 \int_0^\infty (P(V_h) - P_0(V_h))^2 \, dV_h + \frac{\lambda}{2} \int_0^\infty (Q(N) - Q_0(N))^2 \, dN \quad (10)$$

where the parameter λ governs the relative importance of the two distributions. λ is not a physical parameter but rather a numerical one that we choose to study the relationship between the two distributions P_0 and Q_0. If varying λ makes no difference to the ensembles, then we would conclude that the two distributions are not in conflict. A more elaborate choice of the Hamiltonian such as that in eq 10 might be used to find out if the two distributions are consistent with each other, given some particular debranched distribution or other assumption about the branching structure of the polymer.

The other likely modification to our method concerns the choice of Monte Carlo move. In this calculation, we swap a polymer in our subensemble for another in the master ensemble. This simple approach needs storage of the master ensemble in the computer's memory. Other moves could just as easily be used, such as deleting a region of a polymer and replacing it with new random walks.

Models and Parameters

Master Ensemble and Debranched MWD To construct the master ensemble, we first need to know the lengths of the debranched (or simple, linear) chains that make up the branched polymer. We take the debranched MWD from the work of Takeda et al.[15] (Figure 1). The randomly branched structures generated from this debranched distribution and from an exponential distribution, $P_{debr}(N) = \alpha \exp(-\alpha N)$, with the same mean, are compared later.

The construction of a polymer of a given N_{tot} proceeds via several steps. An initial chain is randomly selected from $P_{debr}(N)$, and then another chain is selected from $P_{debr}(N)$ and connected at random. This random addition is repeated until the desired N_{tot} is reached. The debranched chains are small compared to the overall polymer size, so the actual degrees of polymerization are only very slightly more than the target ones. The random walk that gives the three-dimensional structure is performed 500 times, and the radius of gyration for that polymer is obtained by averaging over these structures (see eq 7 above). Details of the master ensemble of polymers are given in Table 1. We take each monomer to contribute a distance 1 nm to the random walk and use persistence lengths of 3, 5, and 7 monomers.

Target Distribution of Hydrodynamic Volumes. Figure 2 shows the distribution of elution times for branched and debranched starch molecules reported by Takeda and coworkers.[16] The distributions for branched polymers are reported in terms of elution time, which ideally would be converted to V_h. Unfortunately, the calibration data required for this are not available. (We note that during the preparation of the present paper, more precise data from Takeda and colleagues appeared in the literature,[17] but the new data, although sufficient for the goals of the authors of that paper, still did not provide the calibration needed for the present application.) It turns out that we can make a consistent and meaningful comparison of our simulations to experiment without knowing the magnitude of the hydrodynamic volumes involved (see previous section). Takeda's data do include elution times for linear starch, but Mark−Houwink parameters for the particular solvent system used by these authors are not available. As described above, we estimate these Mark−Houwink parameters from simulations

Table 1. Input Parameters for Monte Carlo Generation and Simulated Annealing of Polymers with Various Persistence Lengths

	segment length (monomers)		
	3	5	7
R_g iterations		500	
population size		10^4	
number of brackets		400	
start V_h (Å^3)	1.8×10^7	3.8×10^7	6.1×10^7
end V_h (Å^3)	3.9×10^8	1.2×10^9	1.28×10^9
initial k_BT		120.0	
total iterations		6×10^6	
change k_BT		4×10^5	
polymers used	1.97×10^5	9.72×10^5	1.99×10^5
polymers out of range	2541	8159	1264
maximum DP	1.0×10^5	1.0×10^6	1.0×10^5

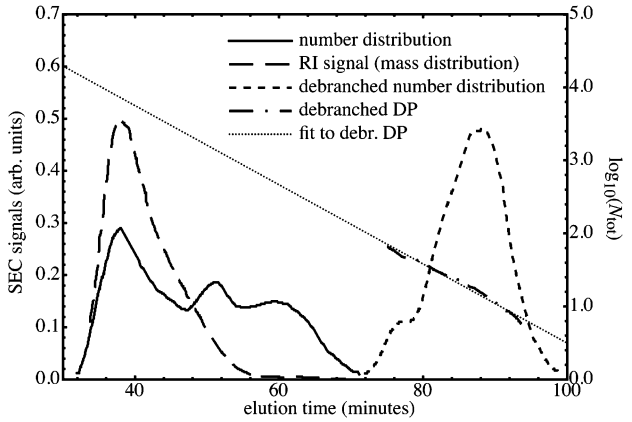

Figure 2. Experimental SEC data for waxy rice showing both branched and debranched polymers, taken from Figure 1 (subgraph labeled "Waxy Rice") of Takeda et al.[15] The debranched number distribution (short dashed line) is used to generate $P_{debr}(N)$ shown in Figure 1, the number distribution for branched polymers (solid line) gives us our target $P(V_h)$, and the refractive index signal gives us the experimental data to which we compare our simulated polymers.

of linear starch. Provided that the same persistence length and monomer size (s and b in eq 2) are used to estimate the Mark−Houwink parameters as to generate the master ensemble, this approach is consistent. It would be preferable to have the experimental Mark−Houwink parameters and so to be able to determine the persistence length. Instead, we examine the results for three reasonable choices of persistence length. Given the Mark−Houwink parameters and the linear-equivalent DP of the branched starch molecules, we obtain the target number distribution of hydrodynamic volumes, $P(V_h)$. To implement this, the range of V_h was divided up into linearly spaced boxes and assigned an initial target population, and for computational convenience the target distribution was fitted to a sum of four Gaussians.

Figure 3 shows the experimental dependence of t_{el} on log N for linear starch from the data presented by Takeda et al. and the simulated variation of log V_h with log N. These two data sets, extrapolated linearly, provide the conversion between the experimental elution time and V_h for branched polymers.

Simulations of simple linear chains with these step lengths, combined with the experimental data on debranched polymers shown in Figure 2, allow us to construct the target distribution of hydrodynamic volumes appropriate to each step length. Note that if we took the monomer length to be other than 1 nm, then all volumes would be scaled, but the comparison of simulation to experiment would be unaffected.

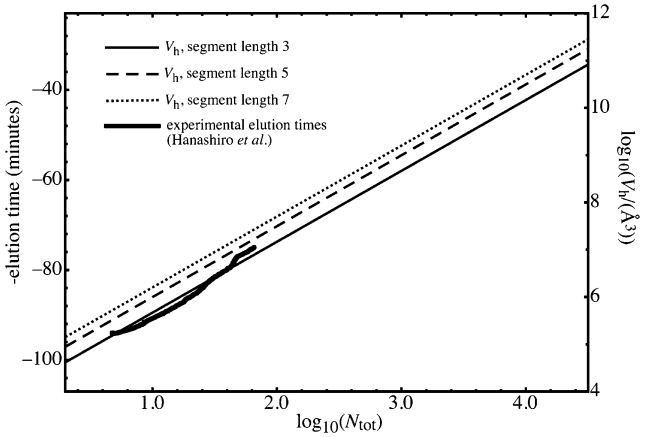

Figure 3. Simulated and experimental data on linear starch. Left-hand axis: Experimental relation between elution time t_{el} and degree of polymerization N from the data of Takeda et al.[15] Right-hand axis: Simulated relation between V_h and N. Mark−Houwink parameters were taken from the random walk simulations of linear polymers.

Cooling Sequence. The initial population was randomly chosen from the ensemble of available polymers. For each iteration a "move" was performed by randomly selecting a polymer from the current population and replacing it with a polymer randomly selected from the larger ensemble. The new population was then accepted if it satisfied the criterion

$$\exp(-\Delta H/k_BT) > \chi \tag{11}$$

where ΔH is the change in H in that move and χ is a random number between 0 and 1. After every 4×10^5 iterations, the temperature, k_BT in eq 8, was lowered by 40%. The simulation was terminated when either the total number of iterations reached 6×10^6 or $H = 0$.

Results

Effect of the Debranched Distribution on Polymer Conformation. Once the master ensemble is constructed using the method described above, the variation of various properties of the model polymers with degree of polymerization and hydrodynamic volume can be examined. The upper plot in Figure 4 shows the radius of gyration squared (eq 7) as a function of the degree of polymerization. Results for the three different choices of segment length are shown. For segment lengths of 3 and 7, R_g^2 is shown with "error" bars indicating the range of radii in the ensemble. For the segment length of 5, results are shown for two different choices of the debranched distribution $P_{debr}(N)$. The first is the experimentally measured distribution[16] (Figure 1, top panel), and the second is an exponential distribution, chosen so that its mean matches the mean of the experimental debranched distribution. The lower plot in Figure 4 shows the mean-squared fluctuation in the number of branch points for the two choices of debranched MWD.

We see that increasing the segment length increases the size of the polymer, as it should, but that the precise shape of the debranched MWD does not greatly affect the conformations of the branched polymers. It is probably sufficient to know the average length of the debranched chains to study the overall structure of the branched polymer for these randomly branched reference sets; it would be of considerable interest to see if the same holds for actual systems, which for starch are not randomly branched.

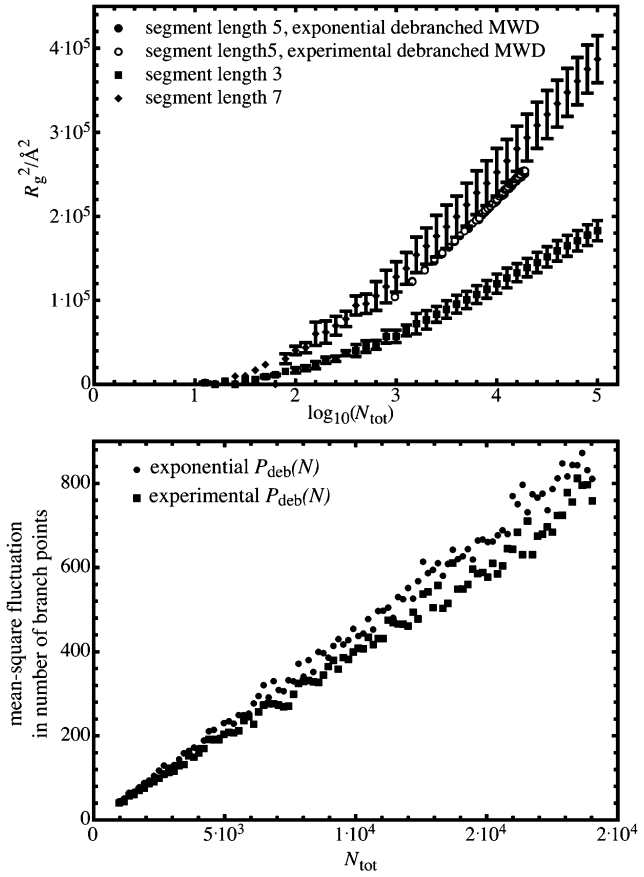

Figure 4. Top panel: Radius of gyration squared of branched chains as a function of DP. Bottom panel: Mean-squared fluctuation in the number of branches, as a function of DP, for branched polymers generated from the experimental $P_{debr}(N)$ of Hanashiro et al.[16] and generated from an exponential $P_{debr}(N)$ with the same mean as the experimental $P_{debr}(N)$ of Hanashiro et al. (symbols).

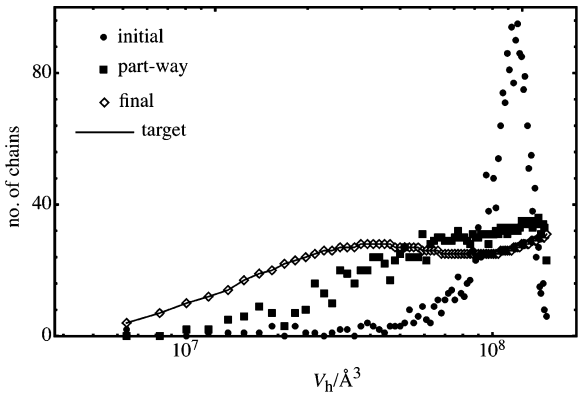

Figure 5. Evolution of the distribution of hydrodynamic volumes as the RMC temperature is lowered. The initial set of randomly chosen polymers (circles) evolves toward the target (line), via the distribution shown with squares, and quickly converges to the final (diamonds). The convergence is fast because the "Hamiltonian" has a simple form. We have shown only every fifth point in each distribution for clarity.

Number Distribution. Figure 5 shows the progress of the hydrodynamic volume distribution for the subensemble toward the target distribution, as simulated annealing progresses. Initially, the distribution of hydrodynamic volumes corresponds to a random choice of polymers from the master ensemble. As the simulation runs and the temperature is lowered, the distribution approached the target, experimental distribution. Table 1 gives the numbers of polymers involved in generating these distributions.

The strength of this approach is that once converged, if we run the Monte Carlo calculation for long enough, virtually all of the polymers in our master ensemble will be sampled. There might be other ways to achieve this but none so elegant as RMC and none that remains so powerful and effective as more and more complicated functions are optimized, for example, as we match our subensemble to more experimental measurements (see discussion above).

Weight Distributions. Having generated from the experimental debranched distribution $P_{debr}(N)$ a randomly branched reference set whose (branched) number distribution $P(V_h)$ matches the observed one, the final step in the methodology illustrated here is to generate the weight distribution $W(V_h)$ from this ensemble and compare it with that observed experimentally, which is obtained using the differential refractive index (DRI) detector.

Figure 6 shows the experimental DRI signal and the simulated $W(V_h)$ obtained by summing the total number of monomers (i.e., the total weight) in each V_h interval. The simulations are performed for three different assumed persistence (segment) lengths. Because the experimental elution time signal is converted to hydrodynamic volume using the linear-equivalent degree of polymerization and the Mark−Houwink parameters estimated from simulations, Figure 6 shows the experimental DRI signal (obtained in terms of t_{el}) three times, once for each assumed persistence length. For convenience, the simulation data are normalized by shifting vertically on the log−log plot to match the simulation.

In each case, the region of agreement is small, but both simulation and experiment show the same growth in refractive index signal with $\sim V_h^{2.8}$. This agreement is shown by the straight line for a segment length of 5. As the hydrodynamic volume increases, the simulated signal for a segment length of 5 shows a sharp increase, because without self-avoiding chains new chains added to the simulated polymer add to the interior rather than the periphery, so the mass increases without much change in volume. (The calculations with segment lengths of 3 and 7 do not show this effect because they were done with master ensembles extending only to a much smaller DP—see Table 1.) The convenient approximation of overlapping chains fails drastically for hyperbranched polymers, as one might expect, but this tells us where overlap effects begin to modify the

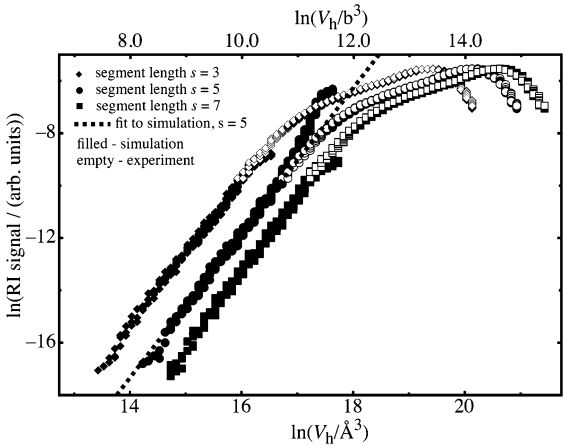

Figure 6. Experimental DRI signal and simulated randomly branched $W(V_h)$. The simulated data are shown for three persistence lengths. Normalization is arbitrary, so each simulated $W(V_h)$ is normalized to have the same maximum. The three different experimental distributions are obtained because the experimental elution time is converted to hydrodynamic volume with Mark−Houwink parameters determined from simulations with three different persistence lengths (Figure 3).

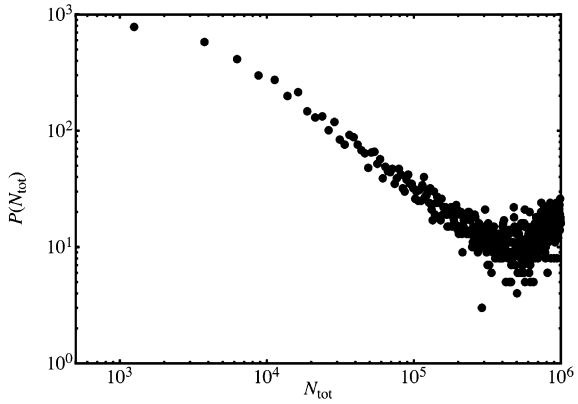

Figure 7. MWD of the simulated, branched polymers for a segment length of 5 monomers. This MWD is expected to be accurate at low molecular weights where the agreement in Figure 6 suggests that the polymers are essentially randomly branched or at least that correlations are not perceptible to SEC measurement. The upturn in the MWD corresponds to the sharp increase in the segment length 5 data in Figure 6. This is where the neglect of self-avoidance ceases to be a good approximation.

conformation of the polymer. We note that the appearance of overlap effects is close to a change in the shape of the experimental refractive index signal. These results suggest that our method reproduces the structure as measured by SEC and so that the branching structure of starch at these low molecular weights is indistinguishable from random by SEC, but these results remain tentative until we extend the range of comparison by using self-avoiding walk simulations. While we suspect that this supposition of random branching is too simplistic, it could be tested by acquisition of data on hyperbranched polymers using additional types of detectors, especially those that detect the $\bar{M}_w$ and fraction branching as functions of elution time. (These detectors being multiangle laser light scattering and NMR, respectively.) The possibility remains that as far as available experimental probes can tell the structure of starch is randomly branched.

A branched MWD from the simulated annealing was created by bracketing N_{tot} into evenly spaced brackets on a linear scale (Table 1). Figure 7 shows this MWD for the case of a segment length of 7 monomers, which gives the best agreement in Figure 6. This quantity is particularly difficult to measure, although in principle techniques such as MALDI can give this up to DP values of 10^5. We suggest that our predicted MWD will be accurate for polymers small enough that the effects of repulsion between simple chains is negligible. For much of the range of the degree of polymerization shown, the MWD decays as $N_{tot}^{-1.7}$.

Conclusions

The methodology developed here gives a method of obtaining mechanistic information on the synthesis and structure of branched polymers from data obtained using multidetector size-exclusion chromatography. While applied here to data from SEC, the general methodology is equally applicable to interpreting data from other separation techniques such as the various types of field-flow fractionation and high-performance anion exchange chromatography.

By using reverse Monte Carlo to fit randomly branched polymers to the experimental hydrodynamic volumes it is possible to compare the simulated and experimental observables to evaluate how randomly branched the polymers are. This provides a means of seeing if particular data can differentiate between "interesting" (i.e., more realistic) biosynthetic processes compared to the "trivial" one where chains are joined randomly. We have found that for polymers small enough that they may be modeled by non-self-avoiding random walks the SEC data are consistent with random branching.

It is emphasized that the simulation is not intended to mimic the actual polymer (although one might expect that long-chain branching in polyacrylates and poly(vinyl esters) might in fact be random) but instead to provide a reference distribution whereby multidetector SEC data on the actual polymer can be compared with what is predicted from the simplest possible system from the same parent distribution of debranched chains, i.e, that ensemble where these chains are randomly branched.

We have illustrated the first use of RMC to analyze the results of SEC experiments and shown that much more information can be obtained from the experiments when they are coupled to simulations of this type that test hypotheses about the microscopic structure of the polymer. This is especially applicable to branched biopolymers such as amylopectin. In particular we have shown that the comparison is meaningful even without precise knowledge of the size of a monomer or the persistence length of the chain. We conclude that the method is useful and worth developing with improved experimental information.

The present paper has the objective of setting out the methodology and applying its numerically simplest implementation to extant literature data. This demonstrative implementation makes many approximations, such as non-self-avoiding walks and various values of the persistence length; these various simplifications can be avoided as more data become available, including self-avoiding walks with b and s obtained from a combination of viscometric data and RIS calculations on linear starch. Other types of reference sets could be used as well, such as randomly assembling the polymer by growing the chain from the inside out, as is commonly done in the simulation of flocs (e.g., ref 44). Similarly, information on hydrodynamic volume (as obtainable with an in-line viscometric detector) were unavailable for the target experimental data used here, again a problem that can readily be solved in the future. Because of these limitations, the present calculations cannot be used to make unambiguous inferences about mechanisms for starch biosynthesis. However, the methodology presented here can be applied with improved numerical techniques and augmented experimental data to provide an almost assumption-free method for understanding highly branched biological and synthetic polymers.

One of the most important results of the new methodology is that it can identify whether or not SEC data using a given combination of detector types can actually reveal structurally and/or mechanistically interesting information about the branched polymer, either that the polymer is randomly branched (which would have mechanistic inferences) or that the data are unable to distinguish random and nonrandom branching.

Acknowledgment. We thank the Australian Research Council for the financial support of a Discovery and of a LIEF grant and the Australian Centre for Advanced Computing and Communications for a grant of computer time. Constructive input from Dr. Jef Castro, Dr. David Lamb, Dr. Patrice Castignolles, Dr. Melissa Fitzgerald, and Doug Willoughby are greatly appreciated.

CDV

References and Notes

(1) Gallot-Grubisic, Z.; Rempp, P.; Benoit, H. *J. Polym. Sci., Polym. Lett. Ed.* **1967**, *5*, 753−759.
(2) Kostanski, L. K.; Keller, D. M.; Hamielec, A. E. *J. Biochem. Biophys. Methods* **2004**, *58*, 159−186.
(3) Sun, T.; Chance, R. R.; Graessley, W. W.; Lohse, D. J. *Macromolecules* **2004**, *37*, 4304−4312.
(4) Tobita, H.; Kawai, H. *e-Polym.* **2002**, 48.
(5) Vidal, F.; Guyot, A.; Gilbert, R. G. *Macromol. Chem. Phys.* **1996**, *197*, 1835−1840.
(6) Miller, C. M.; Clay, P. A.; Gilbert, R. G.; El-Aasser, M. A. *J. Polym. Sci., Polym. Chem. Ed.* **1997**, *35*, 989−1006.
(7) Clay, P. A.; Christie, D. I.; Gilbert, R. G. Termination rate coefficients from molecular weight distributions. In *Controlled Radical Polymerization*; Matyjaszewski, K., Ed.; ACS Symposium Series 685; American Chemical Society: Washington, DC, 1998; pp 104−119.
(8) Castro, J. V.; Dumas, C.; Chiou, H.; Fitzgerald, M. A.; Gilbert, R. G. *Biomacromolecules* **2005**, *6*, 2248−2259.
(9) van Berkel, K. Y.; Russell, G. T.; Gilbert, R. G. *Macromolecules* **2005**, *38*, 3214−3224.
(10) Zaragoza-Contreras, E. A.; Navarro-Rodriguez, D.; Maldonado-Textle, H. *J. Appl. Polym. Sci.* **2002**, *84*, 1513−1523.
(11) Smith, A. M. *Biomacromolecules* **2001**, *2*, 335−341.
(12) Ball, S. G.; Morell, M. K. *Annu. Rev. Plant Biol.* **2003**, *54*, 207−233.
(13) Hamielec, A. *Pure Appl. Chem.* **1982**, *54*, 293−307.
(14) Gaborieau, M.; Gilbert, R. G.; Gray-Weale, A.; Hernandez, J. M.; Castignolles, P. *Macromol. Theory Simul.*, **2007**, *16*, 13−28.
(15) Takeda, Y.; Shibahara, S.; Hanashiro, I. *Carbohydr. Res.* **2003**, *338*, 471−475.
(16) Hanashiro, I.; Tagawa, M.; Shibahara, S.; Iwata, K.; Takeda, Y. *Carbohydr. Res.* **2002**, *337*, 1211−1215.
(17) Charoenkul, N.; Uttapap, D.; Pathipanawat, W.; Takeda, Y. *Carbohydr. Polym.* **2006**, *65*, 102−108.
(18) McGreevy, R. L. *J. Phys.: Condens. Matter* **2001**, *13*, R877−R913.
(19) Korobko, A. V.; Jesse, W.; Lapp, A.; Egelhaaf, S. U.; van der Maarel, J. R. C. *J. Chem. Phys.* **2005**, *122*, 024902.
(20) Clay, P. A.; Gilbert, R. G.; Russell, G. T. *Macromolecules* **1997**, *30*, 1935−1946.
(21) Cunningham, B. F.; Mahabadi, H. K. *Macromolecules* **1996**, *29*, 835−841.
(22) Clay, P. A.; Gilbert, R. G. *Macromolecules* **1995**, *28*, 552−569.
(23) Castro, J. V.; van Berkel, K. Y.; Russell, G. T.; Gilbert, R. G. *Aust. J. Chem.* **2005**, *58*, 178−181.
(24) Lee, S.; Nilsson, P. O.; Nilsson, G. S.; Wahlund, K. G. *J. Chromatogr., A* **2003**, *1011*, 111−123.
(25) Roger, P.; Baud, B.; Colonna, P. *J. Chromatogr., A* **2001**, *917*, 179−185.
(26) You, S.; Stevenson, S. G.; Izydorczyk, M. S.; Preston, K. R. *Cereal Chem.* **2002**, *79*, 624−630.
(27) Yoo, S.-H.; Jane, J.-l. *Carbohydr. Polym.* **2002**, *49*, 307−314.
(28) Tomlinson, K. L.; Lloyd, J. R.; Smith, A. M. *Plant J.* **1997**, *11*, 31−43.
(29) Nilsson, G. S.; Richardson, S.; Huber, A.; Torto, N.; Laurell, T.; Gorton, L. *Carbohydr. Polym.* **2001**, *46*, 59−68.
(30) McDowell, W. H.; Kenyon, W. O. *J. Am. Chem. Soc.* **1940**, *62*, 415−417.
(31) Moad, G.; Solomon, D. H. *The Chemistry of Free Radical Polymerization*, 2nd ed.; Elsevier: Amsterdam, 2006.
(32) Metropolis, N.; Ulam, S. *J. Am. Stat. Assoc.* **1949**, *44*, 335−341.
(33) Kainthan, R. K.; Muliawan, E. B.; Hatzikiriakos, S. G.; Brooks, D. E. *Macromolecules* **2006**, *39*, 7708−7717.
(34) Castro, J. V.; Ward, R. M.; Gilbert, R. G.; Fitzgerald, M. A. *Biomacromolecules* **2005**, *6*, 2260−2270.
(35) Flory, P. J. *Principles of Polymer Chemistry*; Cornell University Press: Ithaca, NY, 1953.
(36) Gedde, U. W. *Polymer Physics*; Chapman & Hall: London, 1995.
(37) Pyun, C. W.; Fixman, M. *J. Chem. Phys.* **1966**, *44*, 2107−2115.
(38) Oono, Y.; Kohmoto, M. *J. Chem. Phys.* **1983**, *78*, 520−528.
(39) Sheridan, P. F.; Adolf, D. B.; Lyulin, A. V.; Neelov, I.; Davies, G. R. *J. Chem. Phys.* **2002**, *117*, 7802−7812.
(40) Burchard, W. *Adv. Polym. Sci.* **1999**, *143*, 113−194.
(41) Renaud, M.; Belgacem, M. N.; Rinaudo, M. *Polymer* **2005**, *46*, 12348−12358.
(42) Flory, P. J. *Statistical Mechanics of Chain Molecules*; Interscience: New York, 1969**.**
(43) Bovey, F. A. *Chain Structure and Conformation of Macromolecules*; Academic Press: New York, 1982.
(44) Ansell, G. C.; Dickinson, E. *Faraday Discuss.* **1987**, *83*, 167−177.

BM0605199