

The Graph Theory for a Description of the Configurational Statistics of Branched Polycondensate Polymers

S. I. KUCHANOV, S. V. KOROLEV, and M. G. SLIN'KO

*Karpov Institute of Physical Chemistry, ul. Obukha 10,
107120 Moscow B-120, USSR*

(Received December 8, 1982)

ABSTRACT: An approach was developed to describe configurational statistics of branched polymers, with recognition of the difference in monomer units caused by the difference in the number of their reacted functional groups. By this approach, configurational statistics are characterized in terms of fractions of arbitrary sequences (*k*-ads) of *k* monomer units. Each *k*-ad is a molecule fragment consisting of *k* interconnected monomer units. Unlike sequences used in the case of linear copolymers, the *k*-ads, introduced in this article, recognize the topology of links between fragment units. Relations are given between *k*-ad fractions with different values of *k* that can be used for processing NMR spectroscopy data. An algorithm is formulated for calculating the probabilities of arbitrary sequences using the theory of branching random processes.

KEY WORDS Branched Polymers / Configurational Statistics / Graph Theory / Equilibrium Polycondensation / NMR Spectroscopy / Network Formation /

Branched macromolecule configuration is determined by the distribution of links between monomer units. With the same number of links, various topological isomers can exist, as in the case of the isomerism of alkanes. With typical values for the degree of polymerization of macromolecules, the number of isomers is so high that a description of a polymer sample by specifying the concentration of chemically individual compounds, as is done in the traditional organic chemistry, makes no sense even for homopolymers.

A similar problem arises in the case of linear copolymers because of the isomerism which results from differences in unit distributions in their macromolecules. In this case, a configurational description of a copolymer is obtained by specifying the probabilities $P(U_k)$ of various sequences (U_k) of *k* units.¹⁻⁴ With increasing *k* the information on the configurational structure of copolymers becomes increasingly detailed and an exhaustive description of such a structure implies designing an algorithm for calculating the probabilities of any sequences. The objective of this paper is to extend this approach to branched polymers. Many physico-

chemical properties of branched macromolecules are determined by their configurational and conformational structures. Some of these, the mean molecular dimensions, the particle light scattering factor, the average hydrodynamic Stokes radius, and the second virial coefficient depend on sequences of all sizes⁵; others, so-called additive structure-borne properties (glass-transition temperature, formation enthalpy, combustion enthalpy, density, and entropy), depend only on the number of some molecule fragments of small size such as atoms or bonds of a certain type and pairs or threes of atoms or bonds.⁶⁻⁹ Essentially all the literature proposals for predicting such measurable quantities are schemes for linear combinations of graph-theoretical invariants.¹⁰ Additive structure-borne properties of oligomer hydrocarbons can be predicted but in the literature there is no reference to the properties of a mixture of numerous individual compounds. This is a problem typical of polymer systems. Therefore an algorithm for calculating the probabilities of various sequences would permit deriving the above physico-chemical properties.

CONFIGURATIONAL STATISTICS
OF POLYMERS

The macromolecule topology is characterized by its molecular graph which is akin to the structural formula in essence. Use of graph theory methods in the polymer chemistry has been developed by Gordon and co-workers in their papers (see, *e.g.* ref 11). It is an important feature of polycondensation polymers that the number of branching units is usually comparable to the total number of units, which is essentially different from polymers of the crosslinked rubber type where chain segments between two network branchpoints contain large numbers of units. While in showing the vulcanizate structure, the graph vertices represent cross-links and the edges, internodal chains of random length, it is better to also show bifunctional monomer units in the molecular graphs of polycondensates. In this case, the graph edges stand for the links of fixed length and the configurational structure of the molecule is completely determined by the graph topology. Typical polymers whose configurational statistics can be described by the theory to be discussed below are phenol-, urea-, and melamine-formaldehyde resins.

Thus, nodes of the molecular graph (vertices of degree higher than 1) represent monomer units, the edges represent links, and vertices of degree 1 (endpoints), unreacted functional groups.^{1,11} To denote the differences in the chemical nature of

these fragments, the associated graph elements are colored. By a sequence (U_k) referred to as a k -ad, we shall mean a connected subgraph of the molecular graph that consists of k nodes with all outgoing edges. Some of these, internal edges, connect the nodes of the sequence (U_k) between themselves and the endpoints while the remaining connecting edges (edges of attachment¹²), connect these nodes with others not included in (U_k) . Figure 1 shows all monads (U_1) , dyads (U_2) and triads (U_3) not containing cycles for the case of the homopolycondensation of a monomer of the 1,3,5-trimethylol benzene type with three identical groups. The idea that such sequences can be used for description of the topology of branched polymers has been suggested by Ziabicki and Walasek.¹³ However, no new useful results in calculating the configurational statistics of macromolecules have been obtained using this idea. In surveying the thermodynamic measurable properties of alkanes, Gordon and Kennedy¹⁰ considered subgraphs of molecular graphs which, unlike those introduced above, do not contain connecting edges. Furthermore, they looked into relations between numbers of subgraphs of a certain fixed graph rather than between numbers of subgraphs in a random mixture of graphs.

Unlike linear homopolymers, branched ones can be regarded as consisting of units of different kinds. For instance, let us refer to a unit as of the i -th kind if it is connected with other units through i links. Thus the monads of Figure 1 with i external edges

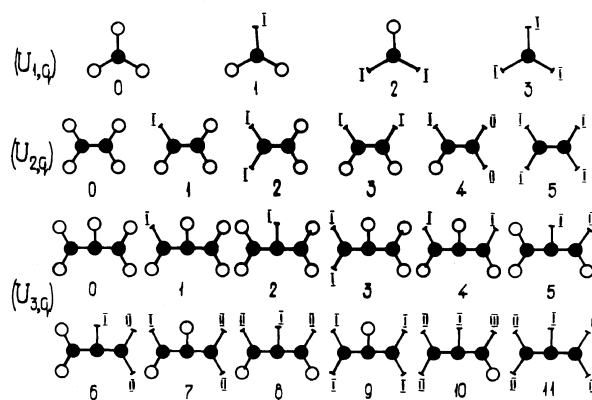


Figure 1. All different monads $(U_{1,q})$, dyads $(U_{2,q})$, and triads $(U_{3,q})$ without cycles which are fragments of molecules resulting from three-functional polycondensation. The arabic numbers denote their ordinal numbers q and the Roman numbers, the equivalence classes of connecting edges. Black vertices denote monomer units and white vertices, unreacted functional groups.

are associated with units of the i -th kind. The simplest possible way to describe the topological structure of a polymer is to specify fractions λ_i of units of the i -th kind. Further levels of detailing the descriptions are obtained by specifying the fractions $P(U_{2,q}), P(U_{3,q}), \dots$ of various dyads ($U_{2,q}$), triads ($U_{3,q}$), \dots which are enumerated in an arbitrary way by the subscript q separately for each set of sequences ($U_{k,q}$) with a given k . The theory of stationary random processes for linear copolymers helps to calculate, with $P(U_{k,q})$ known, the probabilities of all sequences with downwards of $k-1$ units, as is done when describing the tacticity of homopolymers.³ End effects can be neglected and formulae are used that were derived for macromolecules with an infinite number of units. Analogous relations between $P(U_{k-1,q})$ and $P(U_{k,q})$ for branching polymers will be seen unachievable unless fractions of $(k-1)$ -ads which do not contain external edges (such as $(U_{1,o}), (U_{2,o})$ and $(U_{3,o})$ in Figure 1) can be neglected. For polycondensation systems, this condition does not generally hold. Therefore a better technique is to consider the numbers $c(U_{k,q})$ of sequences ($U_{k,q}$) per monomer unit in the polymer. Similar relations can be obtained for these which are stoichiometric and hold for any polymer sample, no matter in what way it is formed. Like analogous relations for copolymers, they can be used in processing NMR spectroscopy data³ in the cases where proper sequences can be resolved spectroscopically.^{14,15}

If there is a kinetic scheme of the polycondensation process, the values of $c(U_{k,q})$ can be rather easily found only with small k . However, with increasing k , the calculation of $c(U_{k,q})$ by the straightforward kinetic method becomes too cumbersome. On the other hand, these values determine the above polymer properties which depend on its configurational structure. In the case of linear copolymers, $c(U_{k,q})$ is calculated with using the theory of random processes, the macromolecules being represented as realizations of such processes. The well-developed tools of the theory of Markov chains are used in most cases.⁴ For basic widely used models of chain growth in linear copolymerization and copolycondensation, this description has been found theoretically and confirmed experimentally.¹ In other cases, this theory is generally inapplicable; nevertheless, configurational statistics of such non-Markov copolymers can be approxi-

mately described in terms of first-, second-, etc. order Markov chains.¹⁶ The order increases with introduction of increasingly higher order correlations between chain units into the model of chain growth. The accuracy of approximate description increases as does the number of probabilistic parameters of the random process and the kinetic parameters which determine them. An analog of Markov chains in the case of branched polymers is a random branching process.¹⁷ Their applicability for describing an ideal¹⁸ (*i.e.* random) polycondensation (both equilibrium and nonequilibrium) and an equilibrium polycondensation with the first shell substitution effects¹⁹ has been rigorously proved.¹⁹⁻²² In other cases, a hierarchical sequence of such processes can probably be constructed to describe increasingly more accurately the statistics of branched polycondensation products. This requires, however, a certain extension of the existing theory which would be analogous with that of Markov chains of an order higher than the first because the usual branching process does not recognize correlation between the reproductive behaviours of various individuals of this process. Such correlations should be recognized in calculating the probabilities of various configurations of branched macromolecules in the framework of all kinetic models allowing for short-range correlation effects¹⁸ other than the above equilibrium polycondensation with first shell substitution effects. In the framework of an arbitrary model of the substitution effects neglecting long-range correlation effects¹⁸ the equilibrium systems probably can be described using such extended branching processes. For arbitrary polycondensation systems, the degree of approximation in describing them in terms of the above hierarchical sequence of random processes should be found anew in each specific case.

The theory of branching processes permits finding probabilities of sequences ($U_{k,q}$) of an arbitrary size k using a small number of probabilistic parameters which, in turn, are expressed in terms of $c(U_{k,q})$ with small k . The latter, as noted above, can be measured experimentally or calculated theoretically by the kinetic method. Note that calculation of probabilities of arbitrary sequences has not been done before. Knowledge of these probabilities enables calculating the above macroscopic characteristics of a polymer sample. Since for some kinetic

schemes, applicability of the theory of branching processes for description of the configurational statistics of associated polymers has been proved, comparison of theoretical fractions of various k -ads for initial values of k with their experimental NMR values reveals whether the model is a good representation of the real polycondensation system.

TOPOLOGICAL STOICHIOMETRY

To derive the relations between probabilities of various sequences, let us examine a certain $(k-1, r)$ -ad $(U_{k-1,r})$ containing $L(k-1, r)$ connecting edges. If $(U_{k-1,r})$ does not coincide with the $(k-1)$ -mer molecule, *i.e.*, $L(k-1, r) \neq 0$, then it is a component (subgraph) of $L(k-1, r)$ various k -ads. Each of the latter contains, in addition to the $(k-1)$ nodes of the sequence $(U_{k-1,r})$, another node of the molecular graph with which $(U_{k-1,r})$ is connected by its connecting edge. Consequently, the number $c(U_{k-1,r})$ of such $(k-1)$ -ads per monomer unit in all the molecules is one $L(k-1, r)$ -th fraction of the number of various k -ads if each of these is counted as many times as there are subgraphs of the type $(U_{k-1,r})$ it contains. Thus,

$$c(U_{k-1,r}) = \sum_q \kappa(k-1, r; k, q) c(U_{k,q}) / L(k-1, r) \quad (1)$$

Here the coefficient of topological stoichiometry $\kappa(k-1, r; k, q)$, or the weak lattice constant,¹² denotes the number of subgraphs $(U_{k-1,r})$ in $(U_{k,q})$, or the number of ways in which one of the nodes of $(U_{k,q})$ can be removed by breaking the proper bond so that the remaining $(k-1)$ nodes make a sequence $(U_{k-1,r})$. Thus by removing either extreme node of the triad $(U_{3,0})$ (Figure 1) we have a dyad $(U_{2,1})$ and so $\kappa(2, 1; 3, 0) = 2$. From the triad $(U_{3,1})$ we have

dyads $(U_{2,1})$ and $(U_{2,3})$, etc. (see Table I).

What is important is that the numbers $c(U_{k,q})$ of sequences with the same k and different q are generally mutually dependent in that between them some linear relations may exist which are obtained by considering the topological stoichiometry of the sequences $(U_{k,q})$. Each contains some of the summands on the right-hand side of eq 1 and is derived by generalization (1) in the following way. Let $L(k-1, r)$ connecting edges of the sequence $(U_{k-1,r})$ generate $A(k-1, r)$ equivalence classes with $\sigma_\alpha(k-1, r)$ elements in the α -th class. Two edges (or vertices) of a graph belong to the same equivalence class if they can be mapped one on the other with certain automorphism of the graph.¹² In Figure 1, the equivalence classes of connecting edges of the graphs are shown in Roman numbers. A certain sequence $(U_{k-1,r})$ is a subgraph of $\sigma_\alpha(k-1, r)$ various k -ads, each obtained when its $(k-1)$ nodes are supplemented with another molecular graph node connected with $(U_{k-1,r})$ by its connecting edge from the equivalence class α . Note that such a k -ad $(U_{k,q})$ was not obtained by adding a node to a connecting edge from a different equivalence class $\beta \neq \alpha$. Consequently, all k -ads $(U_{k,q})$ with nonzero values of the coefficient $\kappa(k-1, r; k, q)$ are divided into $A(k-1, r)$ nonintersecting classes. The number $c(U_{k-1,r})$ is $\sigma_\alpha(k-1, r)$ times smaller than the total number of k -ads from the α -th class (counted with due regard to the multiplicity of inclusion of $(U_{k-1,r})$ into $(U_{k,q})$, as in eq 1)

$$c(U_{k-1,r}) = \sum_\alpha \kappa(k-1, r; k, q) c(U_{k,q}) / \sigma_\alpha(k-1, r) \quad (2)$$

where the subscript α of the summations symbol denotes the summation of $(U_{k,q})$ only from the α -th class. Consequently, each sequence with $A(k-1, r)$

Table I. Values of $\kappa(2, r; 3, q)$ for the dyads and triads of Figure 1

r	q											
	0	1	2	3	4	5	6	7	8	9	10	11
0	0	0	0	0	0	0	0	0	0	0	0	0
1	2	1	0	1	0	0	0	0	0	0	0	0
2	0	0	2	0	0	1	1	0	0	0	0	0
3	0	1	0	0	2	0	0	1	0	0	0	0
4	0	0	0	1	0	1	0	1	2	2	1	0
5	0	0	0	0	0	0	1	0	0	0	1	2

equivalence classes of connecting edges is associated with $A(k-1, r) - 1$ relations between the numbers $c(U_{k,q})$. A single such relation for the triads of Figure 1 is,

$$c(U_{2,4}) = c(U_{3,3}) + c(U_{3,7}) + 2c(U_{3,9}) \\ = (c(U_{3,5}) + 2c(U_{3,8}) + c(U_{3,10}))/2$$

since only in $(U_{2,4})$ do the connecting edges make more than one equivalence class. For 32 tetrads of three-functional nodes not containing cycles, we have seven such relations because each of the triads $(U_{3,5})$, $(U_{3,6})$, $(U_{3,7})$, $(U_{3,8})$, and $(U_{3,11})$ contains two equivalence classes of connecting edges and $(U_{3,10})$ contains three such classes. The total number of various tree-like k -ads with arbitrary numbers of functionalities and types of nodes can be calculated by the graph theory but with small k it can be easily found by exhaustive search.

NMR spectroscopy data lead, rather than to numbers $c(U_{k,q})$ of various k -ads, to their fractions $P(U_{k,q}) = c(U_{k,q})/c_k$ where c_k is a total number of various k -ads per monomer unit. Note that the number of $(k-1)$ -mers, or $(k-1)$ -ads without connecting edges (with subscript 0 in Fig. 1) cannot be expressed in terms of numbers $c(U_{k,q})$. Therefore formulae (1) and (2) make it possible to find the probabilities $P(U_{k-1,r})$ of all the remaining sequences from the available $P(U_{k,q})$ only up to an uncertain factor S :

$$P(U_{k-1,r}) = (L(k-1, r))^{-1} \sum_q \kappa(k-1, r; k, q) P(U_{k,q}) / S \\ = (\sigma_\alpha(k-1, r))^{-1} \sum_q \kappa(k-1, r; k, q) P(U_{k,q}) / S \quad (3)$$

To determine the factor S , it is necessary to know also the fraction of $(k-1)$ -mers among all $(k-1)$ -ads. Thus if only the gel fraction, *i.e.* an infinite network, is considered, then the contribution of finite molecules is zero. In this case the relations (3) can be supplemented with normalization conditions whereby the sum of $P(U_{k,q})$ over q with any k is equal to unity and so

$$S = \sum_{q,r} (\kappa(k-1, r; k, q) P(U_{k,q}) / L(k-1, r)) \quad (4)$$

From (3), formulae are obtained which are extensions of well-known relations earlier derived for linear polymers,³ assuming the statistical stationarity of the random process of motion along the

chains.² This stationarity condition is not, however, necessary because in deriving the relation of $P(U_{k-1,r})$ with $P(U_{k,q})$, only infinity of macromolecules is used in this article. Indeed, in the case of linear infinite polymers of different units, we have for all sequences $L(k, q) = 2$. Since from each sequence $(U_{k,q})$ exactly two $(k-1)$ -ads (identical or different) can be obtained, the sum of the coefficients $\kappa(k-1, r; k, q)$ over r is equal to 2 and according to (4), the factor $S = 1$. Consequently, from (3) and (4) we have the limit (for infinite molecules) relations

$$P(U_{k-1,r}) = \frac{1}{2} \sum_q \kappa(k-1, r; k, q) P(U_{k,q}) \\ = (\sigma_\alpha(k-1, r))^{-1} \sum_q \kappa(k-1, r; k, q) P(U_{k,q}) \quad (5)$$

If, furthermore, the sequence $(U_{k-1,r})$ is nonsymmetrical, then its two external edges belong to two different equivalence classes with $\alpha = 1$ and $\alpha = 2$. Equating the right-hand sides of eq 5 with these values of α , we have exactly one relation between the probabilities $P(U_{k,q})$. There are as many such relations as are nonsymmetrical sequences $(U_{k-1,r})$. Thus, in the case of a binary copolymer, the dyad (RS) enters two times in the triads (RSR) and (SRS) and once in (RSS) and (RRS). Then from eq 5 immediately follow the well-known relation

$$P(\text{RS}) = 2P(\text{RSR}) + P(\text{RSS}) \\ = P(\text{RRS}) + 2P(\text{SRS})$$

as earlier obtained³ by the theory of random stationary processes. Equation 5 easily leads to all analogous relations without using this theory.

For linear copolycondensations the limit expression (eq 5) holds since the processes are carried out up to nearly complete conversions, $p = 1$. The bulk of monomer units is included in reasonably long macromolecules. This is not the case of branched polycondensation, at least until the gel-point is reached where the limit expressions (eq 4 and 5) should be replaced by more general formulae (eq 1 and 2). For instance, in random f -functional polycondensation, the fraction $\lambda_0 = P(U_{1,0}) = (1-p)^f$ of monomers is not small until the gel-point is reached, $p < (f-1)^{-1}$, so that the contribution of monomers to the normalizing

sum S (eq 4). In the gel fraction, limit relations naturally become accurate. All the above formulae are valid for describing the configuration statistics of polymers containing cycles that form during the synthesis. In addition to tree-like k -ads (Figure 1), there are cycle-containing sequences. In counting the numbers $L(k, q)$ or $\sigma_d(k, q)$ for such sequences, all edges which connect $(U_{k, q})$ with a certain molecular graph node not belonging to $(U_{k, q})$ should be counted as one connecting edge. Since cycles lead to great complications in calculating the configurational statistics of the polymer by the theory of branching random processes, our discussion is confined to tree-like polymers without cycles.

CONFIGURATIONAL STATISTICS OF TREE-LIKE POLYMERS

To employ the theory of random processes in describing the configurational statistics of linear polymers consisting of different units (copolymers and homopolymers with due regard to tacticity), it is necessary to choose one of two possible directions of motion along the macromolecule. At this point, the notion of directed sequences $\{U_{k, q}\}$ is introduced; each of these is obtained from a k -ad $(U_{k, q})$ choosing one of its boundary units to serve as the initial one.² The convenience offered by using this notion is that probabilities (fractions) of directed sequences, in terms of which $P(U_{k, q})$ are expressed, can be calculated by the theory of random processes. A similar procedure will be used for finding the probabilities of sequences $(U_{k, q})$ of branched polymers. Directed sequences that are associated with some undirected sequence are obtained from the latter by choosing one of its units to serve as the root. To determine the relations between the numbers $c(U_{k, q})$ (or the probabilities

$P(U_{k, q})$) and the probabilities $P\{U_{k, q}\}$ of directed sequences and to calculate the latter using the theory of branching processes, tree-like graphs should be replaced, following Gordon *et al.*,²³ by rooted trees.

From a Molecular Forest to an Ordered Clone

A set of macromolecules is associated with a set of nonrooted trees referred to as a molecular forest. The probability of finding a certain tree is equal to the fraction of associated molecules in the polymer sample. The chemical composition of the molecule is described by a vector \mathbf{l} whose component l_ν is equal to the number of units of the ν -th type in it and a symmetrical matrix \mathbf{B} with elements $b_{ij} = b_{ji}$ which are equal to the number of intramolecular bonds resulting from reactions of the functional groups A_i and A_j . With \mathbf{l} and \mathbf{B} fixed, the molecules may differ in their topology and form a set of isomers differentiated hereafter by the subscript q . The fraction $f_w(\mathbf{l}, \mathbf{B}; q)$ of units in the q -th isomers of (\mathbf{l}, \mathbf{B}) -mers is equal to that of nodes in the molecular forest taken by trees representing that isomer. The function $f_w(\mathbf{l}, \mathbf{B}; q)$ is the weighted molecular configuration distribution (MCD). The set of rooted trees obtained in this way is one clone. From the technique for obtaining a clone, it follows that the probability of finding a certain rooted tree (its share among all trees of the clone) is equal to that of choosing the node associated with the root in a random choice among all nodes of the molecular forest. Consequently, the fraction of rooted trees associated with the $(\mathbf{l}, \mathbf{B}; q)$ -isomer is equal to $f_w(\mathbf{l}, \mathbf{B}; q)$. Let us now associate each rooted tree of the clone with all different ordered, or plane rooted trees,^{1,11} by permuting its vertices in different ways. Different ordered trees obtained from the same rooted tree will be regarded, by definition, as equiprobable and their total probability should equal

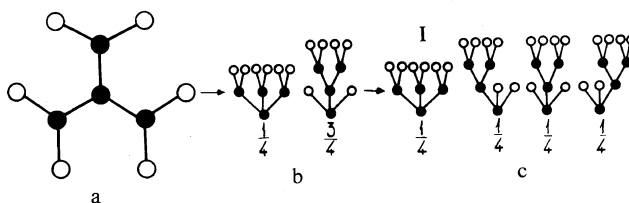


Figure 2. A tree from a molecular forest (a) and the associated trees from a clone (b) and an ordered clone (c). The numbers denote fractions of rooted unordered (b) and ordered (c) trees associated with the isomer (a).

that of the original unordered rooted tree. A set of such ordered trees in Figure 2 will be referred to as an ordered clone. The probabilities of different trees with l , B and q fixed are naturally interrelated. Indeed, in some equivalence class α of nodes of a (l, B, q) -isomer, there are $\sigma_\alpha(l, B, q)$ elements. The probability of a rooted tree with a root from that class (all $\sigma_\alpha(l, B, q)$ such trees are indistinguishable among themselves and differ from trees with roots from another equivalence class) is then equal to $\sigma_\alpha(l, B, q)f_w(l, B, q)/l$ where l is the total number of units of that isomer. Let $D_\alpha(l, B, q)$ denote the number of various ordered trees with a root from the α -th equivalence class that are associated with the (l, B, q) -isomer. They are all equiprobable and so the probability of each is equal to

$$P\{l, B, q\}_\alpha = \sigma_\alpha(l, B, q)f_w(l, B, q)/lD_\alpha(l, B, q) \quad (6)$$

The number $D_\alpha(l, B, q)$ is related to the total number $D(l, B, q)$ of various rooted ordered trees associated with the (l, B, q) -isomer and can be found from the relation²³

$$D_\alpha(l, B, q) = \frac{\sigma_\alpha(l, B, q)f_\alpha D(l, B, q)}{\sum_\beta \sigma_\beta(l, B, q)f_\beta} = \frac{\sigma_\alpha(l, B, q)f_\alpha \prod_\beta ((f_\beta - 1)!)^{\sigma_\beta}}{\mathcal{S}(l, B, q)} \quad (7)$$

where f_α is the degree of a node (equal to the functionality of the associated monomer unit) of the α -th equivalence class and $\mathcal{S}(l, B, q)$ is the order of the automorphism group of the graph of the (l, B, q) -isomer.¹² From eq 6 and 7, it follows that the product of the probability of an ordered tree by the degree of its root is constant for all trees associated with the (l, B, q) -isomer

$$P\{l, B, q\}_\alpha f_\alpha = f_w(l, B, q) \sum_\beta \sigma_\beta(l, B, q)f_\beta / lD(l, B, q) \quad (8)$$

This formula is independent of the polymer formation conditions and enables expressing the MCD $f_w(l, B, q)$ in terms of $D(l, B, q)$, $\sigma_\alpha(l, B, q)$, f_α which depend only on the topology of the (l, B, q) -isomer,

and the probability of the ordered tree $P\{l, B, q\}_\alpha$ which depends on the polycondensation mechanism and kinetics. When the probabilities $P\{l, B, q\}_\alpha$ can be calculated using the theory of branching random processes, the distribution $f_w(l, B, q)$ is characterized in terms of a small number of probabilistic parameters of such processes. These parameters are, in turn, dependent on the probabilities $P\{U_{k,q}\}$ of directed sequences with a low number k of units and, through these, with the number $c(U_{k,q})$ of undirected k -ads. The relationship between $P\{U_{k,q}\}$ and $c(U_{k,q})$ is dictated by the above replacement of nonroot trees by ordered trees.

Number of the k -Ads

The sequence $(U_{k,q})$ from a molecular forest is associated in an ordered clone with $D(k, q)$ various k -ads $\{U_{k,q}\}_r$ ($r = 1, 2, \dots, D(k, q)$) obtained from $(U_{k,q})$ by successive choices, as the roots, of all k nodes and permutations of the remaining vertices in various ways as shown in Figure 3. Because choice of the root specifies the direction ("from the root") of all the tree's branches, such k -ads $\{U_{k,q}\}_r$ are directed in a manner similar to the directed sequences of random process describing linear copolymers.² By choosing nodes from a certain α -th equivalence class of nodes in $(U_{k,q})$ to serve as the root we have $D_\alpha(k, q)$ various $\{U_{k,q}\}_r$. Expression 7 also applies to an arbitrary sequence $(U_{k,q})$ if the argument (l, B, q) , related to the q -th isomer of the (l, B) -mer, is replaced by (k, q) . Let us consider $D(k, q)$ families of ordered trees whose r -th includes all trees starting with $\{U_{k,q}\}_r$ (i.e. having $\{U_{k,q}\}_r$ as a subgraph). The probability $P\{U_{k,q}\}_r$ of the directed k -ad $\{U_{k,q}\}_r$ is equal to the total probability of trees from the r -th family. Note that the sum of probabilities of all families is not equal to unity because, first, the families do not include trees associated with (l, B) -mers with $l < k$ and secondly, a tree of an ordered clone can be a member of several such families, as many as there are non-coinciding $(U_{k,q})$, where its node which is a root is included. Thus, tree I in Figure 2 is included in three families that start with the first three dyads of Figure 3 and its root is included in three noncoinciding dyads

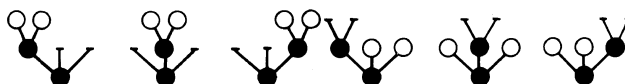


Figure 3. All different directed diads associated with the undirected dyad $(U_{2,2})$ of Figure 1.

($U_{2,2}$) (Figure 1). Consequently, the total probability of trees of all families is equal to the total concentration of monomer units included into the k -ads ($U_{k,q}$) if each of these units is counted as many times as there are k -ads containing it. Since the number $c(U_{k,q})$ of the k -ads ($U_{k,q}$) is one k -th of the total number of the units included, taking multiplicity of their inclusion into consideration, we have

$$kc(U_{k,q}) = \sum_r P\{U_{k,q}\}_r \quad (9)$$

In each family, there are the same number of ordered trees associated with some $(I, B; q)$ -isomer since no matter what root of a certain k -ad is chosen and in what order the vertices in it are placed, the remaining vertices of the $(I, B; q)$ -isomer can be permuted in the same number of ways. Since the product of the probability of any such tree by the degree of its root is constant, the same holds for the probabilities of directed k -ads. Summing in eq 9 first all $D_\alpha(k, q)$ of equal addends $P\{U_{k,q}\}_r = P\{U_{k,q}\}_\alpha$ corresponding to the root from the α -th equivalence class of the nodes ($U_{k,q}$) and summing all α (with expression 7 considered), eq 8 is generalized as,

$$\begin{aligned} kc(U_{k,q}) &= \sum_\alpha D_\alpha(k, q) P\{U_{k,q}\}_\alpha \\ &= kP\{U_{k,q}\}_\alpha f_\alpha D(k, q) / \sum_\beta \sigma_\beta(k, q) f_\beta \\ &= kP\{U_{k,q}\}_\alpha D_\alpha(k, q) / \sigma_\alpha(k, q) \end{aligned} \quad (10)$$

This expression clearly relates the number $c(U_{k,q})$ of undirected k -ads and the probabilities $P\{U_{k,q}\}_\alpha$ of associated directed sequences with roots from any α -th equivalence class. The quantities $D(k, q)$, $D_\alpha(k, q)$, $\sigma_\alpha(k, q)$ and f_α are topological characteristics of the k -ad ($U_{k,q}$) while the multiplier $P\{U_{k,q}\}_\alpha$ (the same for all equivalence classes α with the same root degree) represents the individuality of the configurational structure of the polymer sample. The numbers D and D_α should be found for each such sequence separately. With small k , this is done by exhaustive search; with large k , eq 7 and the algorithm for finding the tree automorphism group order²⁴ should prove useful. In a particular case of nodes of the same degree ($f_\alpha = f$ for all α) the probabilities of all trees associated with the same isomer are identical; thence $P\{U_{k,q}\}_r = P\{U_{k,q}\}$ for all r and

$$kc(U_{k,q}) = D(k, q) P\{U_{k,q}\} \quad (11)$$

Molecular Graph Representation

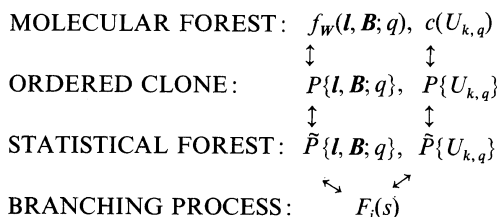
Molecular graphs and their subgraphs ($U_{k,q}$) can be shown without endpoints representing unreacted functional groups. Such trees are referred to as T -trees, as opposed to D -trees with unreacted groups discussed thus far. The degree of a vertex of a T -tree is equal to the number of its links with other units, rather than to the functionality of the unit. In particular, the degree of such a vertex may be equal to unity and so, in dealing with T -trees, there is no need to distinguish endpoints from all other vertices (nodes). All the above findings concerning D -trees are extendible to T -trees. The set of nodes becomes the set of vertices and all the reasoning concerning the former applies to the latter. In eq 6—11 the numbers D and D_α of various D -trees should be replaced by corresponding numbers T and T_α of T -trees. These numbers are related in a simple way since to obtain all D -trees corresponding to a certain T -tree, it is necessary to supplement in all ways possible the missing edges which lead to endpoints independently for each vertex of the T -tree. These endpoints correspond to unreacted groups of various types whose number is unambiguously restored by coloring of edges and vertices of the T -tree. Thus, in the case of f -functional units, the root of the i -th kind is associated with $C_f^i = fC_{f-1}^{i-1}/i$ ways to place $(f-i)$ unreacted groups among its i connecting edges and any other vertex of the i -th kind, with C_{f-1}^{i-1} ways (because one of the edges connects it to the preceding generation). Therefore, if the $(I, B; q)$ -isomer contains l'_i units of the i -th kind and i_α denotes the degree of the root which belongs to the α -th equivalence class, then

$$D_\alpha(I, B; q) = T_\alpha(I, B; q) \prod_i (C_{f-1}^{i-1})^{l'_i} \quad (12)$$

From eq 6 it follows that the probability of the T -tree is $D_\alpha(I, B; q)/T_\alpha(I, B; q)$ times higher than From formula (6) it follows that the probability of the T -tree is $D_\alpha(I, B; q)/T_\alpha(I, B; q)$ times higher than that of the corresponding D -tree. Replacement of D -trees by T -trees reduces thus to obtaining simple combinatorial multipliers which in some cases simplify expressions for $P\{U_{k,q}\}_r$ calculated using the theory of branching processes.

CONCLUSION

In summary, the above findings suggest an algorithm for describing the configurational statistics of branching polymers. Any set of macromolecules which make a polymer sample with a certain molecular configurational distribution $f_w(\mathbf{l}, \mathbf{B}; q)$ is associated with an ordered clone with relations (8) of probabilities $P(\mathbf{l}, \mathbf{B}; q)$ of ordered rooted trees in it. The numbers $c(U_{k,q})$ of various sequences ($U_{k,q}$) in the molecular forest are related as eq 10 with the probabilities of directed sequences $P\{U_{k,q}\}$ in the clone through topological characteristics $D(k, q)$, $\sigma_\alpha(k, q)$, and f_α of the k -ad ($U_{k,q}$) which are independent of the conditions under which the sample was obtained and may be found by the graph theory alone. On the other hand, any branching process characterized by probability generation functions $F_i(s)$ of the number of children of individuals of different types i is also associated with a certain set of ordered trees (a statistical forest²³) with specific values $\tilde{P}\{\mathbf{l}, \mathbf{B}; q\}$ of probabilities. A branching process also specifies the probabilities $\tilde{P}\{U_{k,q}\}$ of arbitrary directed sequences



Sometimes a branching process may be selected in a way such that the probabilities $\tilde{P}\{\mathbf{l}, \mathbf{B}; q\}$ of its associated statistical forest with all $\mathbf{l}, \mathbf{B}, q$ are equal to the probabilities $P\{\mathbf{l}, \mathbf{B}; q\}$ of an ordered clone associated with a specified molecular forest. In such cases, the configurational statistics of the polymer sample represented by this molecular forest can be calculated by the theory of random branching processes. The probabilities $\tilde{P}\{U_{k,q}\}$ with arbitrary k and q can then be expressed in a standard way through probabilistic parameters of the branching process determined in turn by probabilities $\tilde{P}\{U_{k,q}\}$ with small k . Since from the equalities $\tilde{P}\{\mathbf{l}, \mathbf{B}; q\} = P\{\mathbf{l}, \mathbf{B}; q\}$ for all $\mathbf{l}, \mathbf{B}, q$ follow the equalities $\tilde{P}\{U_{k,q}\} = P\{U_{k,q}\}$ for all k, q and the quantities $P\{U_{k,q}\}$ and $c(U_{k,q})$ are clearly interrelated, the number $c(U_{k,q})$ with arbitrary values of k, q is

expressed through the number of k -ads having a low number of monomer units. The latter can be measured experimentally by spectroscopic methods or calculated theoretically if an appropriate kinetic model of polymer formation is available.

APPLICATION TO NMR DATA

The theory described above was applied to an experimental study of urea-formaldehyde resins. In ¹³C NMR spectra, one can measure the squares of CH₂, CO, CH₃-group signals in various molecule fragments.^{14,25} The signals of CH₂-groups are clearly separated and permit determination not only different formaldehyde monads but more complicated fragments. However, individual peaks of CO-group signals are only partly resolved,¹⁴ and this is the reason why only the content of free and monosubstituted urea can be determined by reliable experiments. The numbers of symmetric and asymmetric disubstituted, and trisubstituted urea were calculated by eq 2 and CH₂-group signals. The theory of branching processes²⁶ made it possible to find the number $c(U_k)$ of arbitrary k -ads. The details of calculation will be published in the following paper specially devoted to the study of configurational statistics of urea-formaldehyde resins on the basis of the given theory.

REFERENCES

- S. I. Kuchanov, "Metody Kineticheskikh Raschetov v Khimii Polimerov" ("Kinetic Calculation Methods in the Polymer Chemistry"), Khimiya, Moscow, 1978.
- B. D. Coleman and T. G. Fox, *J. Polym. Sci., A*, **1**, 3183 (1963).
- H. L. Frish, C. L. Mallows, and F. A. Bovey, *J. Chem. Phys.*, **45**, 1565 (1966).
- G. G. Lowry, Ed., "Markov Chains and Monte Carlo Calculation in Polymer Science," Marcel Dekker Inc., New York, 1970.
- K. Kajiwara, *J. Chem. Phys.*, **54**, 296 (1971).
- D. W. Van Krevelen, "Properties of Polymers Correlations with Chemical Structure," Elsevier, Amsterdam-London-New York, 1972.
- R. Becker, *Plast. Kautsch.*, **22**, 790 (1975).
- V. M. Tatevskii, V. A. Benderskii, and S. S. Yarovoi, "Rules and Methods for Calculating the Physicochemical Properties of Paraffinic Hydrocarbons," Pergamon, Oxford, 1961.
- S. S. Yarovoi, "Metody Rascheta Fiziko-khimicheskikh Svoistv Uglevodorodov" ("Methods for

- Calculation of Physico-chemical Hydrocarbon Properties"), Khimiya, Moscow, 1978.
10. M. Gordon and J. W. Kennedy, *J. Chem. Soc., Faraday II*, **69**, 484 (1973).
 11. M. Gordon and W. B. Temple, "The Graph-like State of Matter and Polymer Science," in "Chemical Applications of Graph Theory," A. T. Balaban, Ed., Academic Press, New York, 1976, p 300.
 12. J. W. Essam and M. E. Fisher, *Rev. Mod. Phys.*, **42**, 272 (1970).
 13. A. Ziabicki and J. Walasek, *Macromolecules*, **11**, 471 (1978).
 14. I. Ya. Slonim, *et al.*, *Vysokomol. Soedin, Ser. A*, **19**, 776, 793 (1977).
 15. M. Dawbarn, *et al.*, *Polymer*, **19**, 1309 (1978).
 16. N. A. Plate, A. D. Litmanovich, and O. V. Noa, "Makromolekulyarnye Reaktsii" ("Macromolecular Reactions"), Khimiya, Moscow, 1977.
 17. T. E. Harris, "The Theory of Branching Processes," Springer, Berlin, 1963.
 18. S. I. Kuchanov and Ye. B. Brun, *Uspekhi Khimii*, **48**, 297 (1979).
 19. M. Gordon and G. R. Scantlbery, *Trans. Faraday Soc.*, **60**, 604 (1964).
 20. S. V. Korolev, S. I. Kuchanov, and M. G. Slin'ko, *Dokl. Akad. Nauk USSR*, **258**, 1157 (1981).
 21. S. V. Korolev, S. I. Kuchanov, and M. G. Slin'ko, *Dokl. Akad. Nauk USSR*, **263**, 633 (1982).
 22. S. V. Korolev, S. I. Kuchanov, and M. G. Slin'ko, *Dokl. Akad. Nauk USSR*, **262**, 1422 (1982).
 23. M. Gordon, T. G. Parker, and W. B. Temple, *J. Combinatorial Theory, B*, **11**, 142 (1971).
 24. B. Z. Feinberg, *Dokl. Akad. Nauk Bel. SSR*, **13**, 1065 (1969).
 25. I. Ya. Slonim, *et al.*, *Vysokomol. Soedin, Ser. A*, **20**, 2286 (1978).
 26. S. V. Korolev, S. I. Kuchanov, and M. G. Slin'ko, *Polym. J.*, **15**, 785 (1983).