Arrangement of Substituents in Cellulose Derivatives[1]

BY H. M. SPURLIN

In the course of our consideration of published work and our own results on cellulose derivatives, members of our organization frequently have had occasion to take account of the method of arrangement of the substituents on partially substituted cellulose derivatives. In order to visualize as well as possible the differences that might exist between cellulose derivatives in which the manner of arrangement of the substituents is determined by different influences, several series of calculations have been carried out and the results plotted. While there is nothing original in the considerations on which these calculations were made, they are contributed in order to save others working on the same subject the labor involved, and possibly to bring to the attention of those who have not given the matter much thought the rather interesting conclusions that may be drawn from this study.

The necessity for considering cellulose as having large units of structure in which the substituents are distributed among the available positions according to the laws of chance seems to have been brought out first by Cross and Bevan.[2] Since that time, this conception of a statistical distribution seems to have been before most cellulose investigators at all times. However, a quantitative treatment of the subject was not possible until Denham and Woodhouse's[3] demonstration that substitution on cellulose occurs only on the 2,3- and 6-positions of the glucose units of which cellulose is composed (Fig. 1). This treatment does not seem to have been made, although Thiersch[4] has shown the behavior of the somewhat similar case of coupled reactions.
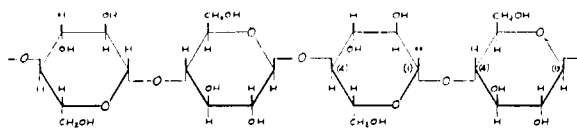

Fig. 1.—Theoretical configuration of a portion of a cellulose molecule.

(1) Delivered at the Baltimore meeting of the American Chemical Society. Cellulose Section. April, 1939.

(2) C. F. Cross and E. J. Bevan, "Researches on Cellulose," Longmans. Green and Co.. London, 1913. p. 47. Vol. II, 1900–1905, 2nd ed.

(3) W. S. Denham and H. Woodhouse. J. Chem. Soc., 111, 244 (1917); see also W. N. Haworth. "The Constitution of Sugars." Edward Arnold and Co., London, 1929. p. 56.

(4) Fr. Thiersch. Z. physik. Chem., 111, 175 (1924).

Basis of the Calculations.—Before beginning the calculations, it is necessary to decide what simplifying assumptions may be made. To begin with, cellulose is considered as a chain of glucose units sufficiently long for the end-groups to be neglected. Other disturbances in the structure, such as partial oxidation of some of the glucose units, may be neglected or may be considered as a form of substitution.

The most important assumption that we make is that the ratio of the reactivities of the three sorts of hydroxyl groups is independent of the degree of substitution of the cellulose chain as a whole or of the state of substitution of neighbor hydroxyls. Of these two influences, we may surmise that the degree of substitution of the chain as a whole will influence more the solubility of the chain, and thereby its rate of reaction, than it will influence the ratios of the reactivities of its hydroxyls. Of course, we know that the influence of substitution on the absolute value of the rate of reaction may be very large. Lorand[5] has shown that this is the case with etherification in the presence of dilute alkali. It must, therefore, be emphasized that all that is assumed in the present work is that the rates of reaction of the three sorts of hydroxyls maintain the same ratios of reactivity constants (i. e., equilibrium constants and reaction rate constants) to one another throughout the course of the reaction, however much the reactivity as a whole may vary.

It is not possible to justify on the basis of any quantitative data the assumption that the substitution of one hydroxyl, in, say, the 2-position, will not influence the reactivity of its neighbor, in the 3-position. However, from what is known of action at a distance in a molecule, it appears that so long as neither group bears an electric charge there will be little influence of the 2- and 3-hydroxyls on one another (they are trans to one another). There should be no interaction between the secondary and the primary hydroxyls.

Briefly, then, what has been done is as follows. A certain reactivity has been assigned to each sort of hydroxyl, the fraction of each sort that has reacted calculated for a given degree of substitution,

(5) E. J. Lorand. Ind. Eng. Chem., 31, 891 (1939).

and the probability that two hydroxyls on the same glucose unit are substituted is set equal to the product of the reacted fractions of those two hydroxyls. In the same manner, the fraction of the glucose units that contains all three hydroxyls in the substituted state is equal to the product of the fractions reacted of the three individual hydroxyls. Obviously, the fraction reacted in the 2- and 3-positions only is equal to the total fraction with substituents on the 2- and 3-positions minus the 2,3,6-trisubstituted fraction. In the same manner, the fraction substituted only in the 2-position is obtained by subtracting the 2,3-, the 2,6- and the 2,3,6-portions from the fraction that contains substituents in the 2-position. Analogous methods are used for the other possible manners of reaction.

Extent of Reaction Governed by an Equilibrium.—The following symbols will be employed in the general discussion: x_2, x_3, x_6 are the fractions of the glucose units that contain substituents in the 2, 3 and 6 positions, respectively; s_2, s_3 and s_6, the fractions with substituents *only* in the positions named. In the same manner, $s_{2,6}$ indicates the fraction of the glucose units containing substituents only in the two and six positions. To indicate the fraction of the glucose units that are unsubstituted, c_0 is employed; monosubstituted, c_1; disubstituted, c_2; trisubstituted, c_3. The average number of substituents per glucose unit, S, is obviously equal to the sum of x_2, x_3 and x_6.

If the extent of the reaction is determined by an equilibrium between some substituting medium and the hydroxyl groups, as in nitration, we may write (considering only reaction of one of the hydroxyls, say that in the 2-position, while those in the 3- and 6-position may or may not be reacting at the same time)

$$C_6H_{10}O_5 + RX \rightleftarrows C_6H_9O_5R + HX$$
$$\frac{(C_6H_9O_5R)}{(C_6H_{10}O_5)} = \frac{(RX)}{(HX)} K$$
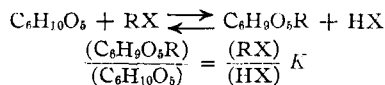
or
$$x_2/(1 - x_2) = AK_2;\ x_2 = AK_2/(1 + AK_2)$$
Similarly
$$x_3 = AK_3/(1 + AK_3)$$
$$x_6 = AK_6/(1 + AK_6) \tag{1}$$

A is a constant denoting the substituting power of the reaction mixture, while K_2, K_3 and K_6 are the equilibrium constants for the substitution reactions of the three sorts of hydroxyls

$$c_3 = s_{2,3,6} = x_2 x_3 x_6 =$$
$$\frac{A^3K_2K_3K_6}{(1 + AK_2)(1 + AK_3)(1 + AK_6)} = \frac{A^3K_2K_3K_6}{D} \tag{2}$$

setting D equal to the denominator of the above expression. Also

$$s_{2,3} = x_2 \cdot x_3 - s_{2,3,6} = A^2K_2K_3/D$$
$$s_{2,6} = A^2K_2K_6/D \tag{3}$$
$$s_{3,6} = A^2K_3K_6/D$$
$$s_2 = x_2 - s_{2,3} - s_{2,6} - s_{2,3,6} = AK_2/D$$
$$s_3 = AK_3/D \tag{4}$$
$$s_6 = AK_6/D$$

Collecting the equations for the fractions of unsubstituted, mono-, di- and trisubstituted glucose, we have

$$c_0 = 1 - s_2 - s_3 - s_6 - s_{2,3} - s_{2,6} - s_{3,6} - s_{2,3,6} = 1/D \tag{5}$$
$$c_1 = A(K_2 + K_3 + K_6)/D \tag{6}$$
$$c_2 = A^2(K_2K_3 + K_2K_6 + K_3K_6)/D \tag{7}$$
$$c_3 = A^3K_2K_3K_6/D \tag{2}$$

The substitution S, equal to $x_2 + x_3 + x_6$, is calculated most easily from the relationship

$$S = 3 - 3c_0 - 2c_1 - c_2 \tag{8}$$

Extent of Substitution Governed by Rates of Reaction.—Many reactions of cellulose, and in particular acylation with an acid anhydride and etherification, are characterized by the fact that they do not stop until complete substitution is reached, the reagent is exhausted, or the reaction is interrupted; there is no back reaction.

For these types of reactions, if the concentration of reagents in the reaction medium were kept constant, the rate of disappearance of the three hydroxyls would be given by an equation of the type

$$dx_2/dt = k_2'(1 - x_2)$$
$$x_2 = 1 - e^{-k_2't} \tag{9}$$

We are not restricting ourselves to the case where the rates remain constant. However, if the rates of the three reactions retain a constant ratio to one another at all times, the ratios of the substituted fractions will bear the same ratio to one another as if the above equation had been followed. In order to avoid confusion, the exponent in equation (9) has been changed from $-k_2't$ to $-Bk_2$, and we have (k_2 and k_2' are reaction rate constants

$$x_2 = 1 - e^{-Bk_2}, x_3 = 1 - e^{-Bk_3}, x_6 = 1 - e^{-Bk_6} \tag{10}$$
$$c_3 = s_{2,3,6} = (1 - e^{-Bk_2})(1 - e^{-Bk_3})(1 - e^{-Bk_6}) \tag{11A}$$
$$s_{2,3} = (1 - e^{-Bk_2})(1 - e^{-Bk_3})e^{-Bk_6}$$
$$s_{2,6} = (1 - e^{-Bk_2})(1 - e^{-Bk_6})e^{-Bk_3}$$
$$s_{3,6} = (1 - e^{-Bk_3})(1 - e^{-Bk_6})e^{-Bk_2} \tag{12}$$
$$s_2 = (1 - e^{-Bk_2})e^{-B(k_3 + k_6)}$$
$$s_3 = (1 - e^{-Bk_3})e^{-B(k_2 + k_6)}$$
$$s_6 = (1 - e^{-Bk_6})e^{-B(k_2 + k_3)} \tag{13}$$
$$c_0 = e^{-B(k_2 + k_3 + k_6)} \tag{14}$$
$$S = 3 - e^{-Bk_2} - e^{-Bk_3} - e^{-Bk_6} \tag{15}$$

Grouping the equations for the fraction of the glucose units that are unsubstituted, mono-, di- and trisubstituted, and putting in a form suitable for easy calculation from a table of exponentials, we have

$$c_0 = e^{-B(k_2 + k_3 + k_6)} \qquad (14)$$

$$c_1 = e^{-B(k_2 + k_3)} + e^{-B(k_2 + k_6)} + e^{-B(k_3 + k_6)} - 3e^{-B(k_2 + k_3 + k_6)} \qquad (16)$$

$$c_2 = e^{-Bk_2} + e^{-Bk_3} + e^{-Bk_6} - 2e^{-B(k_2 + k_6)} - 2e^{-B(k_3 + k_6)} + 3e^{-B(k_2 + k_3 + k_6)} + e^{-B(k_2 + k_6)} \qquad (17)$$

$$c_3 = 1 - e^{Bk_2} - e^{Bk_3} - e^{-Bk_6} + e^{-B(k_2 + k_3)} + e^{-B(k_3 + k_6)} - e^{-B(k_2 + k_3 + k_6)} \qquad (11B)$$

Choice of Constants in the Equations.—In order to plot the above equations, it is necessary to assume ratios of $k_2 : k_3 : k_6$ or $K_2 : K_3 : K_6$, and calculate corresponding values of S and the other quantities for a sufficient number of values of A or B to determine the course of the curves. An explicit solution of the c's and s's in terms of S is very difficult in the general case. It will be noted that these curves will not depend on the absolute value of the K's and k's but only on their ratios.

clusively substituted in the 2-position. From all that we know, it appears likely that the primary hydroxyl has an esterification equilibrium constant less than twice that of the secondary hydroxyls, and we might assume $K_2 : K_3 : K_6 = 1 : 1 : 2$. However, the curves so obtained would not depart noticeably from those with a ratio of $1:1:1$, if the content of unsubstituted, mono-, di- and trisubstituted glucose units were plotted against the average substitution, in the manner of Fig. 2. Therefore, curves have been plotted with ratios of the constants of $1:1:1$, $1:4:16$ and $1:1:10$ (Fig. 2). Curves for s_2, $s_{2,3}$, etc., have not been plotted, as these quantities always bear a constant ratio to c_1 and c_2, easily derivable from Equations (3) and (4). For example

$$\frac{s_2}{c_1} = \frac{K_2}{K_2 + K_3 + K_6} \qquad (18)$$

$$\frac{s_{2.3}}{c_2} = \frac{K_2 K_3}{K_2 K_3 + K_2 K_6 + K_3 K_6} \qquad (19)$$

While this relationship shows that the disubstituted glucose unit derived from the two hydroxyls with the smallest constants is present in relatively small quantity, it will always hold its own. This is not the case with the rate controlled distribution, as will be shown later.

Curves.—The curves of Fig. 2 show that, within the usual limits of error in working with such substances, the amounts of unsubstituted, mono-, di- and trisubstituted glucose units will be the same for the three sets of ratios of constants for the equilibrium controlled case. Detection of differences of reactivity of the three hydroxyls will always be possible if it is possible to determine the ratios of the three possible mono- or disubstituted glucose units to one another. For this purpose, it is recommended that attention be directed to products of either very low or very high substitution in order to avoid the difficulties of separating mono- and disubstituted glucoses from one another.

In the case of the rate controlled substitution, we have even fewer data for estimating the ratios of the constants. A survey of the data in Landolt–Börnstein leads to the conclusion that esterification reactions of primary alcohols proceed at about ten times the rate of esterification of sec-

Fig. 2.

Average no. of substituents per glucose unit.

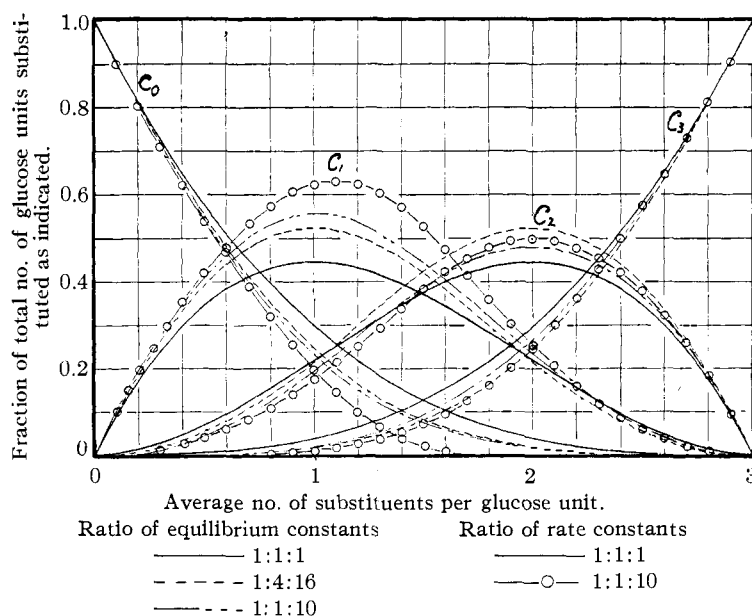| Ratio of equilibrium constants	Ratio of rate constants
——— 1:1:1	——— 1:1:1
- - - - - 1:4:16	——o— 1:1:10
—— - - - 1:1:10	

The equilibrium controlled condition will be realized only in the case of addition compounds and esters of cellulose. The data of Hatch and Adkins[6] indicate that there is little difference in the esterification equilibria for primary and secondary hydroxyls. Lieser[7] has abandoned his contention[8] that cellulose xanthogenate is ex-

(6) G. B. Hatch and H. Adkins, This Journal, **59**, 1694 (1937).

(7) Th. Lieser, Papier-Fabrikant, **36**, 272 (1938).

(8) Th. Lieser, Ann., **470**, 104 (1929).

ondary alcohols. Some old data on etherification[9] indicate even slighter differences. Not too
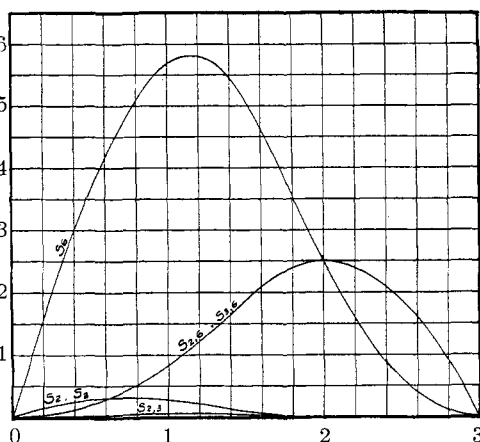
Extent of reaction governed by rate of monomolecular reaction with
$$k_2:k_3:k_6 = 1:1:10.$$

Fig. 3.—Proportions of the six different mono- or disubstituted glucoses, if the nature of substitution is determined by rates of reaction of the three hydroxyls in the ratio 1:1:10.

much confidence can be given the application of these results to carbohydrates, but a rate ratio of 1:1:10 has been selected as a probable limit, with a ratio of 1:1:1 as the other limit. The curves in the latter case are identical with the similar curves for the equilibrium controlled substitution. The fourth set of curves of Fig. 2 gives the distribution of mono-, di-, tri-, and unsubstituted glucose units for the ratio of rate constants 1:1:10. These curves are much less symmetrical than in the case of the equilibrium curves. However, all four sets of curves are so close together that careful experimental work would be required to distinguish between actual cases with the constants used for these curves.

Again, as in the case of the equilibrium curves, an examination of the proportions of the individual monosubstituted or disubstituted glucose units gives us much more information. As no simple ratio exists between these, they have been plotted for the case $k_2:k_3:k_6 = 1:1:10$ in Fig. 3. It is evident that with this ratio of constants, practically all of the groups that are substituted at all are substituted in the 6-position after the early stages of the reaction. Conversely, if examination of a cellulose derivative of substitution above 2, whose degree of substitution is governed by the reaction time, shows that more than one sort of monosubstituted glucose is present, it is

(9) W. Sagrebin. *Z. physik. Chem.*, **34**, 149 (1900).

obvious that the two highest reaction rate constants must be close to one another. In order to form an idea of the relative size of the two smallest constants, the relative amounts of the three monosubstituted glucoses at the beginning of the reaction or of the two disubstituted glucoses chiefly present at some intermediate stage would have to be determined.

As the recent data of Heddle and Percival[9a] on the monomethyl glucoses from a methyl cellulose of substitution 0.5–0.6 indicate that the 2- and 3-positions are much more readily methylated than position 6, and are approximately equal in reactivity, the curves for a ratio of reaction rate constants of 1:10:10 were plotted (Fig. 4). It is evident that the proportions of mono-, di- and unsubstituted glucose for substitutions less than 1.2 are similar to those for the types of reaction plotted in Fig. 2. In this case, as well as in those previously considered, mere determination of the amount of unsubsti-
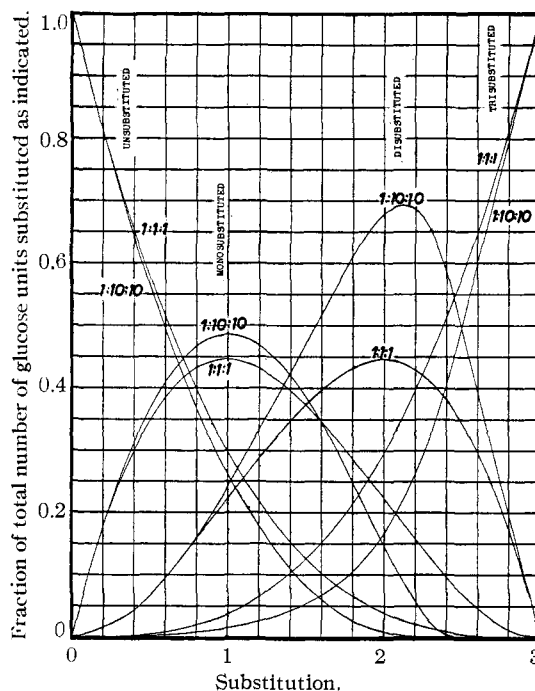

Fig. 4.—Proportions of unsubstituted, mono-, di- and tri-substituted glucose units if the nature of substitution is determined by rates of reaction of the three hydroxyls in the ratio 1:10:10 or 1:1:1.

(9a) W. J. Heddle and E. G. V. Percival, *J. Chem. Soc.*, 1690 (1938).

tuted and mono- and disubstituted glucose in a derivative of substitution less than one will not give any information about the relative reactivity of the three hydroxyls. Only determination of the relative percentages of the three monosubstituted glucoses, as attempted by Heddle and Percival, will give this information.[10]

It may be a surprise to some to note that a monosubstituted cellulose derivative may contain less than 45% of monosubstituted glucose units and still be perfectly uniform, in the sense that each cellulose molecule has the same degree of substitution.

Variation of Substitution with Time.—As pointed out above, the constant B in the equations for the rate controlled substitution is of the nature of a time. In order to enable a picture of the relative times required for a given degree of reaction when the constants have differing ratios, Table I is given.

TABLE I

Substitu-tion, S	$k_1:k_2:k_6 =$ 1:1:1	Relative time, B $k_2:k_3:k_6 =$ 1:1:10	$k_3:k_2:k_6 =$ 1:10:10
0.1	0.034	0.010	...
.5	.18	.050	0.027
1.0	.40	.14	.064
1.5	.69	.34	.12
2.0	1.1	.69	.23
2.5	1.8	1.4	.70
2.7	2.3	1.9	1.25
2.8	2.7	2.3	1.7
2.9	3.4	3.0	2.4

The time for reaching a high degree of substitution depends mainly on the rate of reaction of the slowest hydroxyl. It will be noted from Table I that the time required to reach a substitution of 2.5 is from four to ten times the time required to reach a substitution of 1.0, even if the concentration of reagents remains constant. The rate of etherification of cellulose seems to be some high power of the concentration of alkali, which of course drops during the reaction if it is not replenished. It is evident from this why so many investigators have had difficulty in reaching a high degree of substitution on etherification. Actually this is not difficult if the alkali concentration is kept high.

Comparison with the Literature.—A few papers have appeared bearing on the distribution of substituents in cellulose. Of these, the most in-

formative are those of Compton,[11] Traube, Piwonka and Funk[12] and Schenck.[13] In Compton's paper, the yield of unsubstituted glucose from an apparently very uniform methyl cellulose with an average substitution of about 0.9 was given as 30–35%. The highest curve of Fig. 2 indicates 34%, the lowest 25%, for this substitution. From the other data in Compton's article, it would appear that the contents of mono-, di- and trisubstituted glucose were substantially as indicated in Fig. 2 or 4. It is evident that these data are in agreement with the calculations for a statistical grouping of the substituents and a low ratio of the reactivities of the most and the least reactive hydroxyl group.

Traube and his co-workers[12,13] have attempted to calculate the content of mono-, di- and trisubstituted glucose from the alkoxyl content of fractions obtained on distilling the hydrolysis products of partially alkylated cellulose. This is, of course, only justified if very good fractionation is obtained. As there is no evidence that any sort of a column was used, it appears doubtful that this separation was very good. The results reported therefore possess only qualitative value. They are sufficiently extensive, however, to justify discussion of the point whether the copper present in the reaction mixtures exerted a directive influence on the position of substitution.

The first section of Table II gives the results of Traube, Piwonka and Funk on methyl celluloses. The first four samples represent materials obtained by methylating alkaline copper complexes of cellulose, the fifth sample is a commercial methyl cellulose, and the sixth is a remethylated sample. The second section of the table gives calculated results for equal reactivities of the hydroxyls, while the third section gives the calculated results if one hydroxyl has ten times the reactivity of the other two. Table III is arranged similarly and based on Schenck's values for ethyl cellulose.[13]

It is apparent that the experimental results in Table II, third, fourth and sixth columns, are close checks of the calculated values for ratios of 1:1:10 for the reactivities of the hydroxyls. The values in the first two columns show a higher content of monosubstituted glucose than calculated. This indicates that under the conditions of these

(10) Destruction of the 6-monosubstituted glucose units of the cellulose by selective oxidation at the 2,3-position with periodic acid or lead tetraacetate should be a powerful tool for detecting the presence of other modes of monosubstitution in a highly substituted product.

(11) J. Compton, THIS JOURNAL, **60**, 2823 (1938).

(12) W. Traube, R. Piwonka and A. Funk, *Ber.*, **69**, 1483 (1936).

(13) H.-J. Schenck, "Zur Kenntniss der Alkylcellulosen," Dissertation, Berlin, 1936.

TABLE II

CONTENT OF MONO-, DI-, TRI- AND UNSUBSTITUTED GLUCOSE IN METHYL CELLULOSE (FROM TRAUBE, PIWONKA AND FUNK[12])

Sample	1	2	3	4	5	6
			Experimental Results			
S	0.73	0.79	0.91	0.94	1.44	2.15
c_0	30.2	25.6	23.1	23.9	14.7	0
c_1	61.8	63.2	62.3	57.3	34.7	21
c_2	8.0	11.1	14.6	18.8	41.2	52.5
c_3	0	0	0	0	9.3	26.5
			Calculated, $k_2:k_3:k_6 = 1:1:1$			
c_0	44	40	34	33	14	2
c_1	42	43	44	44	39	18
c_2	13	15	19	20	36	44
c_3	1	2	3	3	11	36
			Calculated, $k_2:k_3:k_6 = 1:1:10$			
c_0	37	32	25	23	3	0
c_1	54	57	61	61	56	18
c_2	9	10.5	14	15	36	49
c_3	0	0.5	1	1	6	33

TABLE III

CONTENT OF MONO-, DI-, TRI- AND UNSUBSTITUTED GLUCOSE IN ETHYL CELLULOSE (SCHENCK, DISSERTATION[13])

Sample	1	2	3
		Experimental Results	
S	0.52	1.01	1.17
c_0	53.8	32	27.4
c_1	38.9	38	36.6
c_2	7.3	27	29.8
c_3	0	3.1	6.2
	Calculated, $k_2:k_3:k_6 = 1:1:1$		
c_0	57	29	22.5
c_1	35	45	43.5
c_2	7.5	22	27.5
c_3	0.5	4	6.5
	Calculated, $k_2:k_3:k_6 = 1:1:10$		
c_0	54	19	11
c_1	42	62	62
c_2	5	17	24
c_3	0	2	3

experiments, the reactivity of one of the hydroxyls was more than ten times that of the next most reactive hydroxyl. These five samples were all made in the presence of copper oxide. The sample in column five, with 1.44 substitution, had a manner of substitution closer to the calculated values for reactivity ratios 1:1:1. This sample, therefore, does not show the same tendency for one hydroxyl to be more reactive than the others. This is undoubtedly connected with the fact that a sodium–copper derivative was not used in the preparation of Sample 5. It appears from the data of Table II that Traube and his co-workers are justified in their belief that the sodium–copper

complex of cellulose is especially reactive in one of the three positions. However, examination of the data of Table III throws doubt on this conclusion, as all three of the reported results fall fairly near the calculated values for equal reactivities of the three hydroxyls.

There is no evidence to indicate that any of the products made from sodium–copper cellulose is composed of a mixture of higher and lower substitution products. However, it is interesting to observe that the composition of the product of substitution 1.44 (Table II) may be calculated fairly exactly if a mixture of products of substitution 1.2 to 1.6 is assumed, with the 2- and 3-positions about twice as reactive as position 6.

Sakurada[14] and Shorygin[15] have attempted to determine the ratio of primary to total free hydroxyls in cellulose acetates of substitution 2.3, with substantial agreement that this ratio is about 0.3. Secondary cellulose acetate probably occupies an intermediate position between the rate controlled and the equilibrium controlled distributions considered in this paper. This would lead us to believe that in the hydrolysis reaction leading to secondary acetate, the three hydroxyls are losing their acetate at very nearly the same rates.

The work of Piwonka[16] would tend to indicate that in his special conditions the 3-hydroxyl reacts preferentially. The work of Heddle and Percival[9a] indicates that both 2 and 3 hydroxyls react under these conditions, but that the 6 position does not react.

Conclusions

The available experimental evidence is entirely in harmony with the viewpoint developed in this paper: namely, that the substituents in a cellulose derivative are distributed among the hydroxyls according to the laws of chance, and the three sorts of hydroxyls are substituted to an extent relative to one another that is determined by the relative reactivities of the hydroxyls (determined by their inherent properties as well as by the reaction medium) and the nature of the reaction (proceeding to an equilibrium at a partially substituted state, or with a tendency to proceed to complete substitution, but stopped at a partially substituted state).

(14) I. Sakurada and T. Kitabatake. *J. Soc. Chem. Ind. Japan,* **37**, 604B (1934).

(15) P. P. Shorygin, A. E. Veitsman and N. N. Makarova-Zemlyanskaya, *J. Gen. Chem. U. S. S. R.,* **7**, 430 (1937).

(16) R. Piwonka, *Ber.,* **69**, 1965 (1936).