minations made with aqueous solutions of crude gliadin. It is completely absent from purified gliadin. Like gliadin, wheat globulin has a wide stability range. It is stable from *p*H 11 through *p*H 2.23, the lowest *p*H studied.

TABLE IV

AQUEOUS SOLUTIONS OF CRUDE GLIADIN. SEDIMENTATION VELOCITY MEASUREMENTS

Experimental conditions as in Table III

*p*H		$s_{20} \times 10^{13}$ Wheat globulin			Gliadin
2.40	6.0	11.0	15.3	..	2.0
2.23	7.0	10.3	..	..	2.2
2.23	...	..	18.0	26.7	2.3

The expenses connected with this investigation have been defrayed by grants from the Rockefeller Foundation, the Nobel Foundation, and the foundation Therese och Johann Anderssons Minne.

Summary

1. An ultracentrifugal study has been made of the proteins extracted from wheat flour by 0.5 *N* solutions of the potassium halides. These solutions, in the order potassium fluoride, potassium chloride, potassium bromide and potassium iodide, extract mixtures of proteins of increasing average molecular weight.

2. Wheat flour was extracted successively with 0.5 *N* solutions of potassium fluoride and potassium chloride; the extracts were further fractionated by dialysis. The material precipitated by dialysis probably consists largely of gliadin and wheat globulin. The globulin, which ordinarily has a sedimentation constant of 11, polymerizes in concentrated solutions to molecules of sedimentation constant 17 and 25, and dissociates on dilution. The material not precipitated by dialysis, probably a mixture of leucosin and proteose, consists of small molecules and is inhomogeneous with respect to molecular weight.

UPSALA, SWEDEN RECEIVED MAY 6, 1935

[CONTRIBUTION NO. 155 FROM THE EXPERIMENTAL STATION, E. I. DU PONT DE NEMOURS AND COMPANY]

Molecular Weight Analysis of Mixtures by Sedimentation Equilibrium in the Svedberg Ultracentrifuge[1]

BY WILLIAM D. LANSING AND ELMER O. KRAEMER

Although Svedberg and his co-workers[2] have found that many of the proteins, when properly purified, are uniform in molecular weight (*i. e.*, are monodisperse), macromolecular substances are generally non-homogeneous in this respect. The typical high polymers, for instance, consist of a mixture of molecules of different sizes, representing different degrees of polymerization. In such cases the usual methods of measuring molecular weight yield average values, but as has been pointed out in a previous paper of this series,[3] different methods may yield different kinds of "average molecular weights" which are not directly comparable. Unfortunately, this fact has not yet been properly considered in previous attempts to compare the results of various investigators, so that a great deal of confusion has resulted.

The authors have already pointed out that at least two "average molecular weights" must be distinguished. The usual physico-chemical methods involve counting of the molecules, and thus yield a "number-average" molecular weight. On the other hand, Staudinger's viscosity method,[4] when applicable, yields a "weight-average" molecular weight. In the present paper a third quantity is introduced, the "*Z*-average" molecular weight, which may be obtained from ultracentrifuge data.

These three different averages are defined by the following equations

$$\text{number-average } M_n = \frac{\Sigma n_i M_i}{\Sigma n_i} = \frac{\Sigma w_i}{\Sigma w_i/M_i} \quad (1)$$

$$\text{weight-average } M_w = \frac{\Sigma n_i M_i^2}{\Sigma n_i M_i} = \frac{\Sigma w_i M_i}{\Sigma w_i} \quad (2)$$

$$\text{and } Z\text{-average } M_Z = \frac{\Sigma n_i M_i^3}{\Sigma n_i M_i^2} = \frac{\Sigma w_i M_i^2}{\Sigma w_i M_i} \quad (3)$$

where n_i is the number of molecules of molecular weight M_i, while w_i is the total weight of that molecular species. For polymeric materials con-

(1) Paper No. III on the Molecular Weight of Linear Macromolecules by Ultracentrifugal Analysis.

(2) Svedberg, *Chem. Rev.*, **14**, 1 (1934); also many papers in THIS JOURNAL, 1926–1934. For complete bibliography see *Naturwissenschaften*, **22**, 225 (1934).

(3) Kraemer and Lansing, *J. Phys. Chem.*, **39**, 153 (1935).

(4) Staudinger, "Die hochmolekularen organischen Verbindungen," J. Springer, Berlin, 1932.

taining a single type of structural unit, the values of M_i are restricted to whole-number multiples of the weight of the unit (162 for cellulose, 104 for polystyrene, etc.). The equations are valid for any sort of mixture, however.

The most satisfactory method so far developed for the determination of the molecular weights of macromolecular materials is by means of sedimentation equilibrium in the Svedberg ultracentrifuge. It has already been shown[5] that the equations for sedimentation equilibrium in the ultracentrifuge can be derived from thermodynamic considerations only. Such a derivation leaves out all reference to the shape of the molecule, and requires only that the solution be sufficiently dilute so that it behaves like an ideal solution. It has also been shown experimentally for one case[6] that the molecular weight of a linear macromolecule can be correctly determined by sedimentation equilibrium.

In addition to the advantages of a satisfactory theoretical background and a relatively small experimental error, the sedimentation equilibrium method is unique in that it gives information regarding the uniformity of the material studied. It will be shown below that all three average molecular weights defined above may be calculated from sedimentation equilibrium. This enables a comparison to be made betwen ultracentrifuge data and such measurements of molecular weight as are given by osmotic pressure and viscosity. It is expected that such a comparison will throw definite light on the usefulness and applicability of Staudinger's viscosity method. The amount by which the three average molecular weights differ will also be shown to give a numerical measure of non-uniformity.

Sedimentation Equilibrium

In general, the concentration of a solution subjected to a centrifugal or a gravitational field varies in the direction of the field, and at equilibrium the resulting concentration gradient is such that at every point in the solution the chemical potential is equal but opposite in sign to the gravitational potential. In ultracentrifugal analysis this concentration gradient is experimentally determined either by a light-absorption method or by a refractive-index method. The light-absorption method involves photographic photometry, and gives directly the *concentrations* at all points along the cell in the ultracentrifuge. In the refractive-index method a transparent scale is photographed through the solution; and the distortion of the scale measured. As shown by Lamm,[7] the distortion of the scale is proportional to the refractive index gradient, which in turn is proportional to the concentration gradient for the low concentrations used in practice. Thus the refractive-index method gives directly the values of the *concentration gradient* at all points along the cell.

For an ideal solution of a homogeneous solute, Rinde[8] has shown that the concentration at any point in the cell may be expressed by means of the equation

$$c_x = c_0 \frac{AM(b^2 - a^2)}{1 - e^{-AM(b^2-a^2)}} e^{-AM(b^2-x^2)} \qquad (4)$$

in which c_x is the concentration by weight at point x, c_0 is the initial uniform concentration, b is the distance from the center of rotation to the bottom of the cell, a is the corresponding distance to the meniscus, x is the distance to the point in question and $A = (1 - V\rho)\omega^2/2RT$ where V is the partial specific volume of the solute, ρ is the density of the solution, ω is the angular velocity of the cell, R is the gas constant, and T is the absolute temperature.

The equation commonly used for calculating molecular weight from sedimentation equilibrium by the absorption method is

$$M = \frac{2RT \ln [c_2/c_1]}{(1 - V\rho)\omega^2(x_2^2 - x_1^2)} = \frac{1}{A} \times \frac{\ln [c_2/c_1]}{x_2^2 - x_1^2} \qquad (5)$$

The corresponding equation for the refractive-index method is as follows

$$M = \frac{2RT \ln [Z_2 x_1 / Z_1 x_2]}{(1 - V\rho)\omega^2(x_2^2 - x_1^2)} = \frac{1}{A} \frac{\ln \left[\frac{dc_2}{dx} x_1 \Big/ \frac{dc_1}{dx} x_2\right]}{x_2^2 - x_1^2} \qquad (6)$$

where Z, the scale displacement, was shown by Lamm to be given by the equation

$$Z = GtL \frac{d\mu}{dx} = k \frac{dc}{dx} \qquad (7)$$

G being the magnification of the displacement by the camera, t the cell thickness, L the distance between the scale and the cell, $d\mu/dx$ the gradient of refractive index, k a constant of proportionality, and dc/dx the concentration gradient at the x-position in question.

If the material is heterogeneous with respect to molecular weight, we may assume that each

(5) Svedberg, *Z. physik. Chem.*, **A121**, 65 (1926); Pedersen, *ibid.*, **A170**, 41 (1934).

(6) Kraemer and Lansing, THIS JOURNAL, **55**, 4319 (1933).

(7) Lamm, *Z. physik. Chem.*, **143**, 177 (1929).

(8) Rinde, "The Distribution of the Sizes of Particles in Gold Sols," Dissertation, Upsala, 1928, p. 176.

molecular species sets up its own equilibrium according to equation (4), independent of the others. In that case, for the mixture as a whole

$$c_x = \Sigma c_{xi} = A(b^2 - a^2)\Sigma \frac{c_{0i} M_i}{1 - e^{-AM_i(b^2-a^2)}} e^{-AM_i(b^2-x^2)} \quad (8)$$

where c_{0i} is the original concentration of the component of molecular weight M_i, and c_{xi} is the concentration of the same component at x after the attainment of equilibrium, M_i representing in turn each and every molecular species in the mixture. The concentration gradient at any point, obtained by differentiation of equation (8), is given by the equation

$$\frac{dc}{dx} = 2A^2x(b^2 - a^2)\Sigma \frac{c_{0i}M_i^2}{1 - e^{-AM_i(b^2-a^2)}} e^{-AM_i(b^2-x^2)} \quad (9)$$

This may also be written

$$\frac{dc}{dx} = 2Ax\,\Sigma c_{xi}M_i \quad (10)$$

In experimental work it is customary to apply equation (5) or (6) to the experimental data, regardless of whether the material is homogeneous or not, and it is assumed that the light absorption, index of refraction, and partial specific volume of the solute are independent of molecular weight when these quantities are calculated per gram of solute. There is good experimental evidence to show that this is normally the case for high polymers. In what follows, we shall also assume the absence of dissociative or associative equilibria of all kinds between the molecular species of the solute.

If the values of molecular weight as calculated by equations (5) or (6) vary with the distance along the cell, it is inferred that the material being centrifuged is non-uniform.[9] It may be shown that these values of molecular weight represent, respectively, the weight-average and Z-average molecular weights for the material present at the point of the cell in question, and that from these values the weight-average and Z-average molecular weights of the solute mixture as a whole may be calculated. In order to show these relationships, it is necessary to convert equations (5) and (6) to the differential form; *i. e.*

$$M_{wx} = \frac{dc}{dx} \times \frac{1}{2Axc_x} \quad (11)$$

$$M_{Zx} = \left(\frac{dZ}{dx} \times \frac{1}{2AxZ}\right) - \frac{1}{2Ax^2} \quad (12)$$

where M_{wx} and M_{Zx} refer to the molecular weights calculated for the point x in the cell.

(9) Svedberg, *Z. physik. Chem.*, **121**, 65 (1926).

Substitution of values of c_x and dc/dx from equations (8) and (10) into the above equations gives

$$M_{wx} = \frac{\Sigma c_{xi}M_i}{\Sigma c_{xi}} \quad (13)$$

and

$$M_{Zx} = \frac{\Sigma c_{xi}M_i^2}{\Sigma c_{xi}M_i} \quad (14)$$

Since the concentrations are expressed by weight, these equations correspond exactly to the definitions as given by equations (2) and (3) for the weight-average and Z-average molecular weights of the material at the level x.

It is apparent that if the values of x_1 and x_2 in equations (5) and (6) are chosen sufficiently close together, the values of M obtained thereby will approach as close as we please to those calculated from the differential forms of the equations. Numerical examples show that the usual distance of 0.05 centimeter between x_1 and x_2 is sufficiently close for practical purposes.

In order to obtain average values of the molecular weight for all the material in the cell, integration must be carried out over the entire cell length. It is apparent that these averages may be obtained by means of the following integrations

$$M_w = \frac{\int_a^b M_{wx}xc_x dx}{\int_a^b xc_x dx} \quad (15)$$

$$M_Z = \frac{\int_a^b M_{Zx}Z dx}{\int_a^b Z dx} \quad (16)$$

since substitution into these integrals gives precisely our previous definitions for weight-average and Z-average molecular weights.

The number-average molecular weight is not so readily obtained. By definition

$$M_{nx} = \frac{\Sigma c_{xi}}{\Sigma c_{xi}/M_i} \quad (17)$$

It can be shown that

$$\Sigma \frac{c_{xi}}{M_i} = 2A \int_a^x xc_x dx + \Sigma \frac{c_{ai}}{M_i} \quad (18)$$

which gives us a method of determining number-average molecular weight at the point x, providing the constant of integration can be determined. The number-average of the entire solute may be determined by means of the equation

$$M_n = \frac{\int_a^b xc_x dx}{\int_a^b \frac{xc_x dx}{M_{nx}}} \quad (19)$$

Uniformity

The equations developed above hold for any conceivable distribution of molecular weights. No matter what the distribution, average molecular weights can be calculated from the above expressions with an accuracy limited only by the conditions of the experiment and the methods of integration. It is a more difficult problem, however, to determine a distribution curve from sedimentation equilibrium data. Rinde[3,10] has developed a method involving the solution of a series of simultaneous equations. Due to unavoidable inaccuracies in the experimental data, this method may give results that are quite unreasonable, such as a negative number of molecules of a particular size class. In a second method of calculation Rinde overcame this objection by using a very difficult polynomial. Both methods are very tedious to use in practice, and the mathematics is rather involved.

For ordinary purposes it is perhaps as useful and in any case much simpler to determine from sedimentation equilibrium data an average molecular weight and some number which may be called "non-uniformity coefficient." This average molecular weight and the non-uniformity coefficient may very well be parameters in some simple distribution function.

Most distribution functions involve the concept of a continuous variation in molecular weight, whereas actual polymeric materials contain components differing in molecular weight by definite steps. However, since the steps are ordinarily small compared to the molecular weights involved, it is probable that little error is introduced by using a continuous distribution function, and the mathematical methods are considerably simplified. For most cases of experimental interest, it is justifiable to limit consideration to curves with a single maximum. In general, polymeric materials contain very little if any low molecular weight material, so it is desirable that the distribution function be so chosen as to have a substantially zero number of molecules at low molecular weights.

One of the simplest distribution functions meeting these conditions is a logarithmic number distribution curve, corresponding to the equation

$$dn = \frac{N}{\sqrt{\pi}} e^{-y^2} dy \qquad (20)$$

in which

$$y = (\ln M/M_0)/\beta \qquad (21)$$

N is the total number of molecules and dn is the number corresponding to the interval between y and $y + dy$. When this distribution is written on the weight basis it becomes

$$dw = \frac{W}{M_n \beta \sqrt{\pi}} e^{-y^2} dM \qquad (22)$$

in which dw is the weight of the material having a molecular weight between M and $M + dM$, W is the total weight of material, M_0 is the molecular weight at the maximum value of dw/dM, and β is what we shall call the non-uniformity coefficient. For the logarithmic distribution we have the following expressions for the three average molecular weights discussed above, and for the most probable molecular weight, M_p

$$\begin{aligned} M_0 &= M_p e^{0.5\beta^2} \\ M_n &= M_p e^{0.75\beta^2} \\ M_w &= M_p e^{1.25\beta^2} \\ M_z &= M_p e^{1.75\beta^2} \end{aligned} \qquad (23)$$

It is apparent that values of M_p, M_0 and β can readily be obtained if we know any two of the average values. Table I illustrates how these quantities are related numerically.

TABLE I

β	M_p	M_0	M_n	M_w	M_z
0	1	1.00	1.00	1.00	1.00
.2	1	1.02	1.03	1.05	1.07
.4	1	1.08	1.13	1.22	1.32
.8	1	1.38	1.62	2.23	3.06
1.2	1	2.05	2.93	6.05	12.4
1.6	1	3.60	6.83	24.5	88.2
2.0	1	7.39	20.1	148	1097

The logarithmic number distribution function is not the only possible distribution curve fitting a given set of average molecular weights. Theoretically there are many such curves. For single-maximum curves, none deviates very much from the logarithmic curve, and in general the form of these other curves is unknown. It is justifiable, then, to use the logarithmic function for purposes of illustration. Although the molecular weight distribution of a given polymer is not thus determined unequivocally, the non-uniformity coefficient β may be defined by means of equations (23) without reference to the type of distribution. A perfectly homogeneous material is characterized by β equal to zero, while the more non-uniform a material is, the larger the value of β.

Practical Methods of Calculation.—An examination of the theory of sedimentation equi-

(10) Svedberg and Nichols, THIS JOURNAL, **48**, 3081 (1926), especially page 3090.

librium given above shows that several integrations must be carried out during the calculations. One of the simplest methods for approximate integration of experimental data is the trapezoidal rule, in which the curve between the two points (x_1, c_1) and x_2, c_2) is replaced by a straight line. Since both the concentration and the Z curves are concave upward, this method always gives high results.

A more accurate method is to replace the curve between the two points (x_1, c_1) and (x_2, c_2) with the arc of a sedimentation equilibrium curve for a uniform material. If we combine all the terms in equation (4) that do not depend on x, we get

$$c_x = Ke^{AMx^2} \qquad (24)$$

When K and M are chosen as arbitrary constants, it is always possible to pass this curve through any two points (x_1, c_1) and x_2, c_2). In this case M takes the value of M_{wx} obtained by substituting the given values of x and c in equation (5). It can readily be shown that, upon the basis of the above assumptions

$$\int_{x_1}^{x_2} xc_x dx = \frac{1}{2AM_{wx}}(c_2 - c_1) \qquad (25)$$

By an entirely similar argument, we can show that

$$\int_{x_1}^{x_2} Z dx = \frac{1}{2AM_{Zx}}(Z_2/x_2 - Z_1/x_1) \qquad (26)$$

The values of the integrals $\int_a^b xc_x dx$ and $\int_a^b Z dx$, required in equations (15) and (16), will then be the sum of the integrals of the several segments as determined by equations (25) and (26). Since we usually calculate molecular weights at x-intervals of 0.05 cm., the same intervals can be used for carrying out the integration.

Upon multiplying both sides of equations (25) and (26) with M_{wx} and M_{Zx}, respectively, and adding the integrals of the several segments, we see that

$$\int_a^b M_{wx} xc_x dx = \frac{1}{2A}(c_b - c_a) \qquad (27)$$

and

$$\int_a^b M_{Zx} Z dx = \frac{1}{2A}(Z_b/b - Z_a/a) \qquad (28)$$

It will be observed in this method that we replace the variable molecular weight with an average value in the small interval between x_1 and x_2. This same procedure can be used for determining the integrals in equation (19) used for determining the number-average molecular weight. In this case we divide equation (25) by an average value of M_{nx}, and thus get

$$\int_a^b \frac{xc_x dx}{M_{nx}}$$

by adding the integrals of the several segments.

Since the details of the calculation are different for the light-absorption and the refractive-index methods, we shall discuss each one separately.

In the case of the refractive-index method, the original data are plotted as scale displacements Z against cell distance x. From a smooth curve drawn through the experimental data, values of Z are taken at regular intervals of x, usually at intervals of 0.05 cm. The molecular weight M_{Zx} is calculated by means of equation (6), and the results plotted as a function of cell distance. It must be remembered, of course, that the value of x corresponding to a particular value of M_{Zx} is the mid-point of the interval for which M_{Zx} was determined. Before the calculations can proceed further, it is necessary to extrapolate to the ends of the column of liquid in the cell. This is best done on the molecular weight plot, and the corresponding values of Z are calculated by equation (6) from extrapolated values of M_{Zx}. The Z-average molecular weight, M_Z, may be obtained by means of equation (16), after the two integrals involved have been evaluated with the help of equations (26) and (28).

The concentrations along the cell, which are required for calculating M_w, are obtained by integration of the scale displacement data. Upon integration of equation (7), we obtain

$$GtL\frac{d\mu}{dc}(c_x - c_a) = \int_a^x Z\,dx \qquad (29)$$

This integration can be carried out at the same time that M_Z is calculated. The constant of integration c_a is the concentration at the meniscus, and can be determined by so choosing c_a that

$$\int_a^b xc_x dx = c_0 \int_a^b x\,dx \qquad (30)$$

The first integral corresponds to the total amount of material in the cell after attainment of equilibrium; the second integral is the total amount of material at the beginning of the run before sedimentation has started. From the values of c_x calculated by this method we can obtain M_{wx} by the use of equation (5). The weight-average molecular weight, M_w, may be obtained from these data by means of equations (15), (25) and (27).

In the case of the light absorption method, the original data yield a curve of concentration against cell distance. The molecular weight, M_{wx}, may be calculated by means of equation (5) from values of x and c taken at regular intervals. Here, again, extrapolation to the ends of the column of liquid in the cell is necessary before the calculations can be completed. This is best done on the molecular weight-cell distance plot, exactly as for the refractive index case. The weight-average molecular weight may now be obtained by integration.

In order to obtain M_{Zx} values from light absorption runs, it is necessary to differentiate the concentration-cell distance curve. This may be done very simply by rearranging equation (11) to give

$$dc/dx = 2AM_{wx}xc_x \qquad (31)$$

With the help of equation (7), we obtain

$$Z/x = 2AGtL\frac{d\mu}{dc}(M_{wx}c_x) \qquad (32)$$

In this expression, M_{wx} refers not to the usual values, obtained from equation (5), but to the values defined by equation (11). The latter figures can readily be obtained by graphical interpolation of the usual molecular weight figures.

From the various values of Z/x as determined by equation (32) we can calculate M_{Zx} using equation (6). This calculation, as will be shown below, is not necessary for getting the Z-average molecular weight, M_Z; it does, however, give some information regarding the ideality of the solution.

Upon suitable combination of equations (16), (28), (29) and (32), it can be shown that

$$M_Z = \frac{M_{wb}c_b - M_{wa}c_a}{c_b - c_a} \qquad (33)$$

It is thus simpler to obtain the Z-average molecular weight from concentration data than from scale-displacement data.

All of the calculations and equations given above rest on the implicit assumption that the solution obeys Henry's law throughout the centrifuge cell. Although it is possible for non-ideal cases to derive equations similar to several of those given above by using activities instead of concentrations, it is not possible to determine the average molecular weights. Fortunately, marked deviation from Henry's law reveals itself in the character of the sedimentation equilibrium. Previous discussions[5,11] have indicated a qualitative rule, for which a proof will now be given.

(11) Signer and Gross, *Helv. Chim. Acta*, **17**, 335 (1934).

Differentiation of equations (11) and (12) with respect to x and substitution of values of c_x and dc/dx from equations (8) and (10) yield the equations

$$\frac{dM_{wx}}{dx} = 2Ax\left[\frac{\Sigma c_{xi}M_i^2}{\Sigma c_{xi}} - \left(\frac{\Sigma c_{xi}M_i}{\Sigma c_{xi}}\right)^2\right] \qquad (34)$$

and

$$\frac{dM_{Zx}}{dx} = 2Ax\left[\frac{\Sigma c_{xi}M_i^3}{\Sigma c_{xi}M_i} - \left(\frac{\Sigma c_{xi}M_i^2}{\Sigma c_{xi}M_i}\right)^2\right] \qquad (35)$$

For any distribution whatever the quantities in the square brackets are positive when Henry's law is obeyed (except in the special case of a uniform material, when these quantities are zero). Stated in words, the molecular weights calculated by means of equations (5) and (6) must always increase toward the more concentrated end of the cell. A maximum in the molecular weight–cell distance curve is impossible as long as Henry's law is obeyed. The presence of such a maximum, then, is definite evidence of non-ideality of the solution. Experimentally, it has been found that the scale displacement method is more sensitive than the absorption method in detecting deviations from Henry's law.

The above method is qualitative in that it gives no data regarding the magnitude of the activity coefficient of the solute. Moreover, it is possible that a small departure from Henry's law would not cause a maximum in the curve. It is tacitly assumed above that the deviation is in the direction of a "swelling pressure" added to the osmotic pressure, that is, that the activity coefficient becomes greater than unity. A deviation from Henry's law in the opposite direction would not be detected by this test. Such a deviation would, however, be anomalous for a macromolecular material. In any case, the only safe procedure is to employ concentrations so low that the results are independent of concentration.

Both the Z-average molecular weight and the weight-average molecular weight may be determined as outlined above with a fair degree of accuracy. The errors involved are mainly experimental, and rarely exceed 10%, except in extremely unfavorable cases. Calculation of the number-average molecular weight, on the other hand, involves a constant of integration which cannot be evaluated from ultracentrifuge data independently of the values of M_{nx}. For this reason number-average molecular weights are not considered as reliable as the other two averages. It is, however, important to calculate M_n be-

TABLE II

MOLECULAR WEIGHT OF GELATIN

M_Z = 151,000. M_w = 36,000. M_n = 9750. M_0 = 4200. M_p = 1000. β = 1.7

x, cm.	c_x	M_{wx}	$2A\int_a^x xcdx$	$\Sigma c_{xi}/M_i$	M_{nx}	kZ/x	M_{Zx}
5.53	0.212[a]	8,300[a]	0.00 × 10^{-6}	26.50 × 10^{-6}	8,000	1.71 × 10^3	19,100
5.59	.222	9,070	1.20	27.70	8,000	1.91	31,200
5.64	.232	10,280	2.30	28.80	8,050	2.22	47,400
5.69	.244	12,840	3.47	29.97	8,150	2.80	72,000
5.74	.260	19,110	4.71	31.21	8,350	4.00	95,000
5.79	.286	29,630	6.07	32.57	8,800	6.43	122,000
5.84	.332	44,900	7.62	34.12	9,750	11.6	141,000
5.89	.417	73,000	9.51	36.01	11,600	23.8	170,000
5.94	.606	100,000[a]	12.20	38.70	15,700	57.0	224,000
5.95	.673[a]		12.86	39.36	17,100	71.6	

[a] Extrapolated.

cause it is the only average value given by the usual physico-chemical methods, such as osmotic pressure.

The number-average molecular weight may be obtained by means of equations (17), (18), and (19). The integral in equation (18) may be determined at the same time that M_w and the integral in equation (30) are evaluated. Before molecular weights can be calculated, it is necessary to determine the value of $\Sigma c_{ai}/M_i$, the constant of integration, which physically is equal to the concentration at the meniscus divided by the number-average molecular weight at the same point. The number-average at the meniscus should be less than the values of M_{wx} and M_{Zx} at the same point. An equation similar to equations (34) and (35) can be derived from M_{nx}, which shows that M_{nx} also must always increase toward the more concentrated end of the cell. This condition also serves as a criterion for the choice of the constant of integration $\Sigma c_{ai}/M_i$. By means of these two tests, it is possible to choose a reasonable value for this constant. From the values of M_{nx}, as has already been pointed out above, the number-average molecular weight may be determined by integration.

The non-uniformity coefficient β can be obtained from these three average molecular weights by means of equations (23). If the values of β obtained by using different pairs of these equations do not agree, most weight is to be given the value determined by means of the Z-average and the weight-average molecular weights. However, a disagreement definitely greater than that likely due to experimental error is an indication that the distribution of molecular weight does not correspond to the logarithmic function upon which β is based. Notwithstanding the ambiguity that may sometimes arise on this account, we believe that at present the simplest and most informative way of expressing the results of sedimentation equilibrium analyses on a non-uniform material is to give the weight-average molecular weight, M_w, and the non-uniformity coefficient, β.

Molecular Weight and Uniformity of Gelatin.—As a specific example of the method of analysis of sedimentation equilibrium data outlined above, we have recalculated the results on gelatin obtained by Krishnamurti and Svedberg.[12] The original published data will be found in the first three columns of Table II. The values of M_{wx} were calculated from the concentration data by means of equation (5) and are shown in Fig. 1, Curve II.

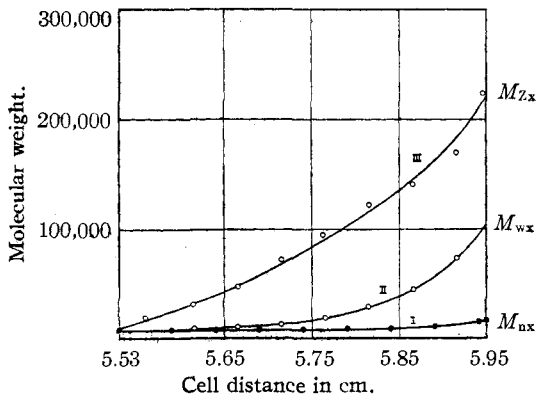


Fig. 1.—Variation of molecular weight with distance from axis of rotation: I, number-average molecular weight; II, weight-average molecular weight (data of Krishnamurti and Svedberg); III, Z-average molecular weight.

The extrapolation of the concentrations to the cell bottom and to the meniscus was carried out

(12) Krishnamurti and Svedberg, THIS JOURNAL, **52**, 2897 (1930).

by plotting log M_{wx} against cell distance and obtaining the extrapolated values for M_{wx} which are given in Table II. The weight-average molecular weight can be obtained from these data by means of equation (15).

The number-average molecular weight was determined, as may be seen from Table II, by means of equations (17), (18) and (19). In this case the constant of integration was so chosen that the number-average molecular weight at the meniscus was slightly lower than the weight-average molecular weight at the same point (see Curve I, Fig. 1).

The Z-average molecular weight was determined by differentiation of the concentration–distance data. Values of Z/x given in column 7 of Table II were obtained by means of equation (32). The Z-average molecular weights corresponding to different heights in the cell were obtained by means of equation (6) and are plotted as Curve III on Fig. 1. These values of M_{Zx} should have been experimentally obtained if the run had been carried out by the refractive-index method rather than the light-absorption method. The Z-average molecular weight of the material was calculated by means of equation (33).

From the three values of M_Z, M_w and M_n, given in Table II, it is possible to determine the non-uniformity coefficient β by means of equations (23). The values of M_Z and M_w are probably determined fairly accurately by means of the given data. On the other hand, the number-average molecular weight M_n depends a great deal on the constant of integration. In this particular case the same value of β is obtained from any two of these three average molecular weights, which gives increased weight to the value of M_n used above. The number-average molecular weight (9750) is what one would expect to obtain by means of osmotic pressure measurements properly carried out.

The non-uniformity coefficient, as may be seen from Table II, came out 1.7. This is the largest value of β that we have so far found. The equivalent logarithmic distribution of molecular weights for this sample of gelatin is given in Table III.

Although the results in Table III adequately illustrate the application of the theory of molecular weight analysis, their specific validity, as quantitative information concerning gelatin, deserves some remarks. In using the logarithmic distribution, we assume that there are a very large number of molecular species, differing in molecular weight by only small amounts, whereas, as a matter of fact, the sedimentation equilibrium method is not capable of distinguishing between distributions consisting of a very large or of a few molecular species. However, sedimentation velocity runs in the ultracentrifuge fill this gap, and by this means Krishnamurti and Svedberg showed that the particular gelatin studied contained more than a few molecular species. It should in any case be remembered that the three average molecular weights comprise the experimental data, and the large spread in their values constitutes unequivocal evidence that the gelatin studied was dispersed in particles of very different size. The logarithmic distribution is inferred from the experimental values of the average molecular weights and simply represents an equivalent degree of non-uniformity.

TABLE III

EQUIVALENT MOLECULAR-WEIGHT DISTRIBUTION OF GELATIN

Molecular weight	Per cent. by weight of total
Less than 5000	13
5,000 to 10,000	17
10,000 to 20,000	23
20,000 to 30,000	13
30,000 to 50,000	14
50,000 to 100,000	12
100,000 to 200,000	6
200,000 to 400,000	2

Question may also be raised concerning the propriety of referring to the "molecular weights" of the gelatin. As a matter of fact, the ultracentrifugal analysis only yields information on the sizes of the dissolved units, whether they be single or associated molecules. Just as in the case of conventional molecular-weight methods (boiling point, osmotic pressure, etc.) the decision as to whether the particle weight represents the molecular weight must rest on other considerations. Since the ultracentrifuge run on the gelatin was made at a temperature of 20°, *i. e.*, below the gelatin point, it may be suspected that the very large particles in the solution represent associated molecules or aggregates. Notwithstanding this fact, the weights of the associated molecules should be correctly given by the sedimentation equilibrium method, provided the solution at all points in the cell was sufficiently low to permit the applicability of Henry's law. Experiments at still lower concentrations would have to be carried out to prove this point, but it

is to be expected that deviations from Henry's law would apparently reduce the non-uniformity of the solute.

Summary

The theory for the determination of average molecular weights for mixtures has been developed and the need for distinguishing various kinds of "averages" has been discussed. It has been shown how number-average, weight-average, and Z-average molecular weights may be calculated from data on sedimentation equilibrium in the Svedberg ultracentrifuge, and the methods have been applied to experimental data on gelatin. A numerical measure for the non-uniformity of mixtures with respect to molecular weight has been proposed. The significance of the results in connection with macromolecular materials, such as cellulose, rubber, proteins, etc., has been explained and the correct method for comparing molecular weights by osmotic pressure or equivalent methods and Staudinger's viscosity method has been elucidated.

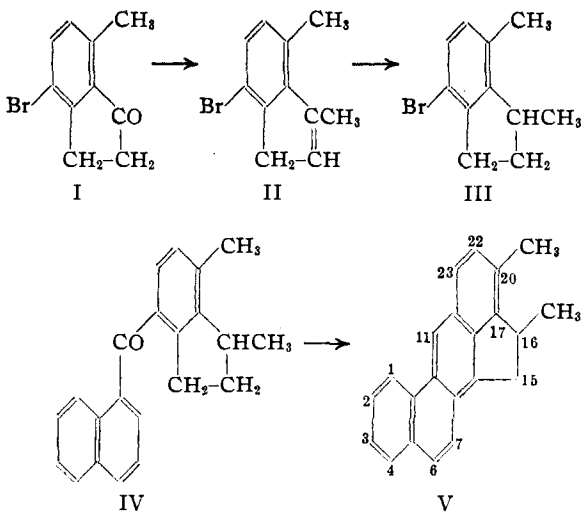
WILMINGTON, DELAWARE RECEIVED APRIL 8, 1935

[CONTRIBUTION FROM THE CHEMICAL LABORATORY OF HARVARD UNIVERSITY]

16,20-Dimethylcholanthrene

BY LOUIS F. FIESER AND ARNOLD M. SELIGMAN

With the idea that optically active substances capable of producing cancer in test animals would afford a very interesting field for investigation, we have as a first step synthesized a hydrocarbon, V, having an asymmetric carbon atom and of a type likely to possess carcinogenic activity. Starting with 7-methyl-4-bromohydrindone-1 (I)[1] the synthesis was accomplished through the series of transformations I ⟶ V, the ketone IV being obtained by the interaction of the Grignard reagent from III with α-naphthoyl chloride. The very faintly yellow hydrocarbon resulting from the pyrolysis of IV has the composition and properties consistent with formula V[1a]. Since the compound is a homolog of the carcinogenically active[2] methylcholanthrene, and since it would be expected from the work of Barry and others[3] that the branching produced by the added methyl group in the aliphatic chain attached to the 1,2-benzanthracene nucleus at the favorable C_5-position would enhance the potency, the new hydrocarbon may be a favorable case for study. Tests for activity are being made with the *dl*-compound before attempting a resolution.



Since the name "cholanthrene" has been applied[4] to the unsubstituted pentacyclic structure, it is convenient to refer to the new hydrocarbon as a dimethylcholanthrene. To provide a system of numbering capable of application both to cholanthrene and to its hydro-derivatives, we propose to use the sterol numbering without modification, even though this has the awkward feature of assigning the numbers 20 and 22 to ortho positions in a benzene ring. According to the suggested system (see formula V), the carcinogenic hydrocarbon obtained from bile acids acquires the specific name of 20-methylcholanthrene and the new hydrocarbon is the 16,20-dimethyl derivative of the parent hydrocarbon.

(1) Fieser and Seligman, THIS JOURNAL, **57**, 942 (1935).

(1a) Dr. Egon Lorenz reports that the absorption spectrum of the hydrocarbon is practically identical in intensity and in the positions of the bands with that of methylcholanthrene. The absorption spectrum of cholanthrene, the parent hydrocardon recently synthesized by Cook, Haslewood and Robinson [*J. Chem. Soc.*, 667 (1935)], is also of a similar pattern; as compared with methylcholanthrene, the bands are somewhat sharper and there is a general shift of about 10 Å. in the direction of shorter wave length. A synthesis of cholanthrene by the modified Elbs reaction will be reported shortly.

(2) Cook and Haslewood, *J. Chem. Soc.*, 428 (1934).

(3) Barry, Cook, Haslewood, Hewett, Hieger and Kennaway, *Proc. Roy. Soc.* (London), **B117**, 318 (1935).

(4) Wieland and Dane, *Z. physiol. Chem.*, **219**, 240 (1933).