

Evaluation of Lipoprotein Size-Density Distributions from Sedimentation Coefficient Distributions Obtained at Several Solvent Densities. I. Theory

J. L. ONCLEY, *Biophysics Research Division, Institute of Science and Technology, and The Departments of Chemistry and Biological Chemistry, University of Michigan, Ann Arbor, Michigan 48104*

Synopsis

A method is presented for the treatment of distribution functions obtained by ultracentrifugal sedimentation velocity experiments on the plasma lipoproteins, so that adequate corrections can be made for the effects of diffusion and concentration dependence of the sedimentation coefficient. The method involves the use of modified equations derived from those presented in 1952 by Gosting for the transform from $g^*(S)$ to $g(S)$ distributions. The evaluation of the molecular parameters of lipoprotein density and size, from a set of $g(S, \rho)$ distribution functions obtained in solvents of varying density, is also outlined. Preparations of the β -lipoprotein of human plasma are found to follow a bivariate normal distribution that requires evaluation of five parameters, \bar{x}_1 , σ_1 , \bar{x}_2 , σ_2 and p . The first two represent the mean and standard deviation of the lipoprotein density, the third and fourth the same quantities for the size, expressed as a "logarithmic diameter parameter", and the final value is a measure of the correlation between these two variables.

The evaluation of heterogeneity of a biopolymer with respect to the sedimentation coefficient by the methods proposed by Baldwin and Williams¹ has been used effectively for the study of many systems. Recent reviews of this method have been published by Williams, van Holde, Baldwin and Fugita,² Williams,³ and Baldwin and van Holde.⁴ However, there have been few attempts to apply this method to the evaluation of sedimentation distribution functions for the plasma lipoproteins, even though the ultracentrifuge has probably been the most useful tool for lipoprotein purification and characterization.

A distribution function $g^*(s)$ can be obtained rather easily from the measured refractive index gradients $dn(r)/dr$ [transform I, eq. (1) below] obtained in a sedimentation velocity experiment. To convert this $g^*(s)$ function to a useful quantity, however, we must transform it for corrections due to diffusion (transform II) and for the effects of concentration dependence on the sedimentation coefficient (transform III):

$$dn(r)/dr \xrightarrow{\text{I}} g^*(s) \xrightarrow{\text{II}} g(s) \xrightarrow{\text{III}} g(s^\circ). \quad (1)$$

Transform II is the most troublesome. The effects of diffusion are far from negligible for nearly all of the lipoprotein fractions one wishes to

study. The diffusion coefficient, corrected to water at 20°C, for the β -lipoprotein of human plasma is 1.85×10^{-7} cm²/sec (Toro-Goyco⁵). This is about 30% of the value for serum albumin, in good agreement with the observation that the molecular weight of β -lipoprotein is of the order of 35 times that of serum albumin. Because of the extensive heterogeneity of the lipoproteins, there is usually only a small difference in $1/t$ between times when the boundary is just resolved from the starting boundary and times when the plateau region has almost disappeared. This makes it difficult to extrapolate to $1/t = 0$ in order to make the transform II.

Use of the Gosting Transform for Diffusion Effects

We have found the equations proposed by Gosting⁶ for this correction to be of value. Gosting treated the case of a system with a distribution of sedimentation coefficients but a constant diffusion coefficient, and related the observed $g^*(s)$ distribution to the $g(s)$ with an equation requiring values for the first four derivatives of $g(s)$ with respect to s :

$$\begin{aligned} g^*(S) = & g(S) - g'(S)\Delta\zeta[1 + (\frac{5}{8})\delta + (\frac{1}{2})\delta^2 + \dots] + g''(S)\Delta\zeta^2[1 \\ & + \delta + (\frac{2}{3})\delta^2 + (\frac{1}{3})\delta^3 + \dots]/4 - g'''(S)\Delta^2\zeta^3[1 + (\frac{5}{8})\delta + \dots]/4 \\ & + g''''(S)\Delta^2\zeta^4[1 + 2\delta + (\frac{7}{3})\delta^2 + \dots]/32 - \dots \end{aligned} \quad (2)$$

Here we have reformulated eq. (38) in the Gosting paper, using S (expressed in svedberg units) = $10^{13}s$, and the parameters

$$[\Delta = 4Dt/r]^2; \delta = s\omega^2t = \ln(r/r^0); \zeta = 10^{13}/(\omega^2t) = S/\delta. \quad (3)$$

D is the diffusion coefficient in cm/sec, ω the ultracentrifugal speed in radians/sec, t the corrected time in sec, and r and r^0 the radial distances for the observed gradient and the starting boundary (meniscus for sedimentation and cell bottom for flotation) in cm. We can best use the dimensionless quantities Δ , δ , and $\zeta^n g^n(S)$ (where $g^n(S)$ represents the four derivatives of $g(S)$ with respect to S) for calculating the transform II. In the usual ultracentrifuge rotors, δ varies between extremes of 0 and 0.2, ζ from 10^5 to 10^2 , and Δ is a small quantity (less than 10^{-3} for experiments with β -lipoprotein). Thus we are troubled with the convergence of this Gosting equation only at relatively short times, or at positions very near the starting boundary.

Our first task in making this Gosting equation useful for computation is the substitution of the four derivatives $g^{*n}(S)$ for the $g^n(S)$ set which he used. This can be done by a series of approximations, and leads to the result:

$$\begin{aligned} g(S) = & g^*(S) + g^{*'}(S)\Delta\zeta[1 + (\frac{5}{8})\delta + (\frac{1}{2})\delta^2 + \dots] - g^{*''}(S)\Delta\zeta^2[1 \\ & + \delta + (\frac{2}{3})\delta^2 + \dots]/4 - g^{*'''}(S)\Delta^2\zeta^3[1 + (\frac{5}{8})\delta + \dots]/8 \\ & + g^{*''''}(S)\Delta^2\zeta^4[1 + 2\delta + (\frac{7}{3})\delta^2 + \dots]/32 - \dots \end{aligned} \quad (4)$$

Introduction of the $\ln (S)$ Parameter

We have found that the distribution functions are much more nearly normally distributed if we use the parameter

$$Z = \log (|S|) = M \ln (|S|) \tag{5}$$

rather than the parameter S . We will call this new parameter the "logarithmic sedimentation parameter." The quantity M is, of course, $\log (e) = 0.4342945$. This substitution simply recognizes the fact that a difference of 1 svedberg for a lipoprotein with a sedimentation coefficient of 6 svedbergs is of much more importance than a similar difference for a lipoprotein with s equal to 60 svedbergs. This logarithmic parameter will, of course, be double valued as we go from positive sedimentation coefficients to negative, and become $-\infty$ for $s = 0$. This is not of much practical significance, since it is experimentally difficult, if not impossible, to work with refractive index gradients in ultracentrifugal experiments where some material sediments while the remainder floats. There would be an imaginary component; $j\pi M$ for a complex parameter Z when the sedimentation coefficient is negative, and this will always be balanced by a similar imaginary component in the Archimedes factor, which we will write as $\log (|x_1 - \rho_i|)$.

If we wish to transform the distribution functions $g(S)$ and $g^*(S)$ to $g(Z)$ and $g^*(Z)$, we must multiply by the quantity $dS/dZ = |S/M|$. (The absolute value sign is used, since we usually consider $g(S)$ and $g^*(S)$ to be positive quantities and thus to be really $(1/c^\circ)(|dc/dS|)$ when we deal with flotation patterns in the ultracentrifuge.) If we put the Gosting transform in terms of $g^*(Z)$ and $g(Z)$ we get the somewhat more complex expression:

$$\begin{aligned} g(Z) = & g^*(Z) \{1 - \Delta/(2\delta^2) [1 + 3\delta + (\frac{7}{3})\delta^2 + \dots] + 3\Delta^2/(4\delta^2) [1 \\ & + 3\delta + 4\delta^2 + \dots]\} + Mg^*(Z) \{3\Delta/(4\delta^2) [1 + (\frac{7}{3})\delta + (\frac{16}{9})\delta^2 + \dots] \\ & - 25\Delta^2/(16\delta^4) [1 + (\frac{7}{3})\delta + (\frac{16}{9})\delta^2 + \dots]\} - M^2g''(Z) \{\Delta/(4\delta^2) [1 \\ & + \delta + (\frac{2}{3})\delta^2 + \dots] - 35\Delta^2/(32\delta^4) [1 + (\frac{8}{3})\delta + (\frac{7}{3})\delta^2 + \dots]\} \\ & - M^3g'''(Z) \{5\Delta^2/(16\delta^4) [1 + (\frac{16}{9})\delta + 3\delta^2 + \dots]\} \\ & + M^4g''''(Z) \{\Delta^2/(32\delta^4) [1 + 2\delta + (\frac{7}{3})\delta^2 + \dots]\} + \dots \tag{6} \end{aligned}$$

Evaluation of an Analytic Function for $g^*(S)$ or $g^*(Z)$

In order to obtain the four derivatives of a somewhat "noisy" $g^*(S)$ or $g^*(Z)$ function, it is necessary to have some adequate smoothing function. If this is an analytic function, then it is much easier to obtain the necessary derivatives. A natural function to use for this purpose is the Gram-Charlier series which expresses the function in terms of a normal distribution function and a series of terms involving the derivatives of that function.

The use of an orthonormal Hermite polynomial is advantageous in this regard, and we choose to represent $g^*(S)$ or $g^*(Z)$ by the equation:

$$g^*(X) = [\Psi_0/\sigma] [1 + \sum_{i=3}^n B_i \text{Hen}(U)] \quad (7)$$

where

$$U = (X - \bar{X})/\sigma \quad (8)$$

$$\Psi_0 = (1/\sqrt{2\pi}) \exp \{-U^2/2\} \quad (9)$$

and the orthonormal Hermite polynomials are defined by the first two polynomials and a recurrence equation

$$\text{Hen}_1(U) = U; \text{Hen}_2(U) = (U^2 - 1)/\sqrt{2}; \text{Hen}_i(U) = (U/\sqrt{i})\text{Hen}_{i-1}(U) - \sqrt{(1 - 1/i)}\text{Hen}_{i-2}(U) \quad (10)$$

Here X is taken to represent either S or Z , and we require a value for its mean, \bar{X} , and its standard deviation, σ . Evaluation of the B_i parameters in order to obtain a "best fit" to our observed distribution function $g^*(X)$, will be considered in a later section. The parameters B_1 and B_2 are omitted, since they should be zero (or at least very small) if the proper \bar{X} and σ are used for the transform of X to U (eq. 8).

Values for either S or Z corresponding to a given r value at a time t are calculated in the usual manner:

$$S = (10^{13}/\omega^2 t) \ln (r/r^0); Z = \log (|S|) = M \ln (|S|) \quad (11)$$

The mean and standard deviation are calculated in the usual manner, using the concentration increments as an integrating factor:

$$\begin{aligned} c^0/K &= \int_{r^0}^{r^\infty} y(r)(r/r^0)^2 dr; & \bar{X} &= (K/c^0) \int_{r^0}^{r^\infty} Xy(r)(r/r^0)^2 dr \\ \bar{X}^2 &= (K/c^0) \int_{r^0}^{r^\infty} X^2y(r)(r/r^0)^2 dr; & \sigma &= \sqrt{\bar{X}^2 - (\bar{X})^2} \end{aligned} \quad (12)$$

Here $y(r)$ is the displacement of the schlieren diagram at position r . It is taken as positive for either flotation or sedimentation experiments. K is the parameter necessary for converting these y values into concentration units (including the refractive index increment, dn/dc , the cell thickness, and the various magnification factors, optical distances and schlieren angle). Having values for \bar{X} and σ , the calculation of U and the various orthonormal Hermite polynomials can be carried out.

Evaluation of the Parameters B_i in the Gram-Charlier Series

The parameters B_i can be evaluated by a number of methods. Perhaps the most easily applied is to make use of the orthonormal properties of the Hen_i polynomials; that is

$$\int_{-\infty}^{\infty} \Psi_0 \text{Hen}_i(U)\text{Hen}_k(U)dU = \begin{cases} 0 & \text{for } j \neq k \\ 1 & \text{for } j = k \end{cases} \quad (13)$$

Thus we can multiply both sides of eq. (7) by $\text{Hen}_k(U)$ and integrate over X from the starting boundary to the plateau region, and find that

$$\int_{X^0}^{X^\infty} g^*(X)\text{Hen}_k(U)dX = B_k \tag{14}$$

If we investigate this method, however, we find that the result is a "best fit in the least squares sense" with a weighting function, $W(U)$.

To evaluate $W(U)$ for this method, let us write

$$\int_{X^0}^{X^\infty} \left\{ \sigma g^*(X) - \Psi_0 \left[1 + \sum_{i=3}^n B_i \text{Hen}_i(U) \right] \right\}^2 W^2 dX = F(B_3, B_4, \dots, B_k, \dots, B_n) \tag{15}$$

For a "best fit in the least squares sense", $F(B_3, B_4, \dots, B_k, \dots, B_n)$ should be a minimum with respect to $B_3, B_4, \dots, B_k, \dots, B_n$. Then

$$\begin{aligned} \partial F / \partial B_k = -2 \int_{X^0}^{X^\infty} \left\{ \sigma g^*(X) - \Psi_0 \left[1 \right. \right. \\ \left. \left. + \sum_{i=3}^n B_i \text{Hen}_i(U) \right] \right\} W^2 \Psi_0 \text{Hen}_k(U) dX = 0 \end{aligned} \tag{16}$$

and

$$\begin{aligned} \int_{X^0}^{X^\infty} g^*(X) W^2 \Psi_0 \text{Hen}_i(U) dX = \int_{-\infty}^{\infty} \Psi_0 W^2 \text{Hen}_k(U) dU \\ + \sum_{i=3}^n B_i \int_{-\infty}^{\infty} \Psi_0 W^2 \text{Hen}_i(U) \text{Hen}_k(U) dU \end{aligned} \tag{17}$$

If the final integral is to resemble eq. (13), we must make

$$W = 1/\sqrt{\Psi_0} \tag{18}$$

With this weighting factor, $W(U)$, we find that eq. (17) is similar to eq. (14). Accordingly, we can describe the orthonormal relation, eq. (14), as yielding a "best fit in the least squares sense" with the weighting factor of eq. (18), when we consider the radial distance from the starting boundary to the plateau area of the ultracentrifuge cell transformed to the coordinate X . Since a first approximation to $\sigma g^*(X)$ is Ψ_0 , we can describe $\sigma g^*(X)/\sqrt{\Psi_0}$ as approximately $\sqrt{\sigma g^*(X)}$. Application of the orthonormal relationship thus greatly favors a fit at the edges of the distribution function and leads to larger deviations at the center of the distribution. Since our later use of this function primarily considers the shape of the distribution curve about this center, the fit by the use of eq. (14) will be rather poor.

An alternative method for determining the B_i parameters is provided by considering the set of m equations formed by values of $\sigma g^*(X)$, Ψ_0 and $\text{Hen}_i(U)$ at each observed point, and evaluating the set of m equations of the form of eq. (7) with the $n - 2$ quantities $B_3, B_4, \dots, B_k, \dots, B_n$ as

variables. Equation (19) illustrates such a set of m equations in $n - 2$ unknowns:

$$\begin{aligned}
 [\sigma g^*(X_1) - \Psi_0(U_1)] &= B_3 \Psi_0(U_1) \text{Hen}_3(U_1) + B_4 \Psi_0(U_1) \text{Hen}_4(U_1) \\
 &\quad + \dots + B_n \Psi_0(U_1) \text{Hen}_n(U_1) \\
 [\sigma g^*(X_2) - \Psi_0(U_2)] &= B_3 \Psi_0(U_2) \text{Hen}_3(U_2) + B_4 \Psi_0(U_2) \text{Hen}_4(U_2) \\
 &\quad + \dots + B_n \Psi_0(U_2) \text{Hen}_n(U_2) \quad (19) \\
 &\dots\dots\dots \\
 [\sigma g^*(X_m) - \Psi_0(U_m)] &= B_3 \Psi_0(U_m) \text{Hen}_3(U_m) \\
 &\quad + B_4 \Psi_0(U_m) \text{Hen}_4(U_m) + \dots + B_n \Psi_0(U_m) \text{Hen}_n(U_m)
 \end{aligned}$$

For the case where m is equal to or larger than n , we can solve this set of equations for a best least square fit by various matrix inversion methods. We can also multiply both sides of the set of eq. (19) by any weighting function that we wish to introduce in order to favor the fit of any particular part of the distribution function.

We have investigated various weighting functions. As already mentioned, we find the function $1/\sqrt{\Psi_0}$ to favor the extreme edges of the distribution function. A factor Ψ_0^2 was found to weight the center of the distribution function by such a large factor that the approximation at the edges of the distribution bore no resemblance to the observed function. Investigations of factors between these two extremes are still underway, in order to obtain a most favorable fit for the evaluation of lipoprotein size-density distribution functions. The matrix method with $W(U) = 1$ seems to be a good compromise.

After obtaining satisfactory B_i parameters, integration and differentiation of the $g^*(X)$ function is easily carried out:

$$\begin{aligned}
 c &= c^\circ \left[\Psi_{-1} + \frac{1}{2} - \sum_{i=3}^n B_i \Psi_0 / (\sigma \sqrt{i}) \text{Hen}_{i-1}(U) \right] \quad \text{where } \Psi_{-1} = \int_0^U \Psi_0 dU \\
 g^{*'}(X) &= - (\Psi_0 / \sigma^2) \left[\text{Hen}_1(U) + \sum_{i=3}^n B_i \sqrt{(i+1)} \text{Hen}_{i+1}(U) \right] \\
 g^{*''}(X) &= (\Psi_0 / \sigma^3) \left[\sqrt{2} \text{Hen}_2(U) + \sum_{i=3}^n B_i \sqrt{(i+1)(i+2)} \text{Hen}_{i+2}(U) \right] \\
 &\hspace{20em} (20) \\
 g^{*'''}(X) &= - (\Psi_0 / \sigma^4) \left[\sqrt{6} \text{Hen}_3(U) \right. \\
 &\quad \left. + \sum_{i=3}^n B_i \sqrt{(i+1)(i+2)(i+3)} \text{Hen}_{i+3}(U) \right] \\
 g^{*''''}(X) &= (\Psi_0 / \sigma^5) \left[\sqrt{24} \text{Hen}_4(U) \right. \\
 &\quad \left. + \sum_{i=3}^n B_i \sqrt{(i+1)(i+2)(i+3)(i+4)} \text{Hen}_{i+4}(U) \right]
 \end{aligned}$$

Concentration Dependence Transform

Corrections for the concentration dependence of the sedimentation coefficients [transform III in eq. (1)] have involved the use of the equation:

$$S = S^\circ / (1 + kc_u) \quad (21)$$

where S° is the sedimentation coefficient corrected to zero concentration, c_u is the concentration (uncorrected for the sector shape of the cell), and k is the parameter describing the concentration dependence of the sedimentation. In terms of the logarithmic sedimentation parameter, we have

$$Z = Z^\circ - M \ln (1 + kc_u) \quad (22)$$

Differentiation of eqs. (21) and (22) yields

$$dS/dS^\circ = 1 / \{1 + k[c_u + r\delta(dc_u/dr)]\}$$

and

$$dZ/dZ^\circ = (1 + kc_u) / \{1 + k[c_u + r\delta(dc_u/dr)]\} \quad (24)$$

Here we have introduced the actual concentration (uncorrected for the sector shape of the cell or for diffusion effects). Jullander⁷ has pointed out from theoretical considerations that the uncorrected concentration should be used in this relationship. It is interesting that, in the analysis of the lipoprotein data, we have found that only this function will keep the areas under the $g(Z)$ and $g(Z^\circ)$ curves constant.

Values of the parameter, k , defining the concentration dependence of the sedimentation coefficient must, of course, be experimentally determined. In heterogeneous materials, k might be a function of the composition, and need to be introduced as some sort of function of $g(X)$. In the case of the β -lipoprotein of human plasma, however, Toro-Goyco⁵ has shown that a single value of k ($= 0.12$ dl/g) can be used for a number of subfractions of the total preparation. The sedimentation coefficient of these subfractions, calculated from the maximum of the concentration gradient curve (S°)_m, varied from -4.2 to -7.8 (in a solvent of density 1.062 at 25°C.)

The Bivariate Normal Distribution

The distribution functions $g(S^\circ)$ or $g(Z^\circ)$ do not reveal much information concerning the molecular parameters of the lipoprotein fraction under study. We wish now to transform these distribution functions into a distribution function with the molecular parameters of lipoprotein density and size. We can choose the size parameter as the diameter d of a hydrodynamically equivalent sphere, or in terms of a "logarithmic diameter parameter", $x_2 = \log(d) = M \ln(d)$. We will express d in Å rather than cgs units, and thus $d = 2 \times 10^8 R$, where R is the effective lipoprotein radius in cgs units. The lipoprotein density parameter will be called x_1 ,

and is introduced into the equation relating the weight of a single lipoprotein molecule to its molecular volume:

$$M = (4\pi/3)R^3Nx_1 = (\pi/6)d^3N \times 10^{-24}x_1 \quad (25)$$

where N represents Avagadro's number.

Perhaps the simplest distribution function for the two variables x_1 and x_2 which should be considered is the bivariate normal distribution.* If two variables v and w are stochastically independent and normally correlated with parameters (\bar{v}, σ_v) and (\bar{w}, σ_w) the probability that the variables lie inside the rectangle defined by the intervals $v \pm dv/2$ and $w \pm dw/2$ will be

$$q(v,w) = (2\pi\sigma_v\sigma_w)^{-1} \exp \left\{ -\frac{1}{2}[(v - \bar{v})^2/\sigma_v^2 + (w - \bar{w})^2/\sigma_w^2] \right\} \quad (26)$$

This bivariate normal distribution, $h = q(v,w)$ is represented by a surface in the (v,w,h) space, and $q(v,w) dv dw$ by a volume element of height h and base area $dv dw$. The maximum value of the distribution function is $h^0 = (2\pi\sigma_v\sigma_w)^{-1}$ for $(v,w) = (\bar{v},\bar{w})$, and the function tends to zero as $|(v - \bar{v})/\sigma_v|$ and $|(w - \bar{w})/\sigma_w|$ become large.

The intersection of the distribution surface and any vertical plane is a curve of the same shape as the normal distribution curve. The intersection of the distribution surface and a horizontal plane is an ellipse, the equation $h = q(v,w) = \text{constant}$ being identical with the equation

$$(v - \bar{v})^2/\sigma_v^2 + (w - \bar{w})^2/\sigma_w^2 = \chi^2 = -2 \ln (h/h^0). \quad (27)$$

Equation (27) represents an ellipse with center in (\bar{v},\bar{w}) , semiaxes of length $(\chi\sigma_v, \chi\sigma_w)$ and axes that are parallel to the coordinate axes. An ellipse of this kind is called a contour ellipse. The total volume under the bivariate distribution function is unity. The volume enclosed by the distribution surface and an elliptical cylinder with a contour ellipse as directrix represents the fraction, P , of the lipoprotein preparation with parameters v and w within the contour ellipse, and is equal to:

$$P = 1 - \exp \{ -(\chi^2/2) \} \quad (28)$$

Table I shows χ values for various fractions P of the normalized bivariate distribution.

It can be seen from this table [or from eqs. (27) and (28)] that a series of contour ellipses with equal intervals between P values will also represent the intersections of the distribution surface with a series of planes ($h = \text{constant}$) with equal intervals between h values.

When we have coordinates x_1 and x_2 which are not stochastically independent, it is necessary to introduce another parameter, the correlation coefficient, p , which characterizes the dependence between x_1 and x_2 . This parameter is introduced into a product term $-2p(x_1 - x_1)(x_2 - x_2)/\sigma_1\sigma_2$ to be added to the exponential factors in the distribution function,

*This discussion of the bivariate normal distribution follows the fine discussion of Hald³, Chapter 19.

TABLE I

P	x	h/h°	\bar{P}	x	h/h°
0.0000	0.000	1.000	0.800	1.794	0.200
.100	.459	.900	.900	2.146	.100
.200	.668	.800	.950	2.448	.050
.300	.845	.700	.980	2.797	.020
.400	1.011	.600	.990	3.035	.010
.500	1.177	.500	.995	3.255	.005
.600	1.354	.400	.999	3.717	.001
.700	1.552	.300	[.39347	1.000	.60653]

giving

$$q(x_1, x_2) = (2\pi\sigma_1\sigma_2 \sqrt{1 - p^2})^{-1} \exp \left\{ - \left[(x_1 - \bar{x}_1)^2/\sigma_1^2 - 2p(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)/\sigma_1\sigma_2 + (x_2 - \bar{x}_2)^2/\sigma_2^2 \right] / [2(1 - p^2)] \right\} \quad (29)$$

The correlation coefficient p can take any value from -1 to $+1$, and when $p = 0$ this function [eq. (29)] reduces to the previously considered one [eq. 24].

The parameters σ_1, σ_2, p of the correlated coordinate system (x_1, x_2) are related to $\sigma_v, \sigma_w, \alpha$ of the stochastically independent coordinate system by the relations

$$\begin{aligned} \sigma_v\sigma_w &= \sigma_1\sigma_2 \sqrt{1 - p^2} \\ \sigma_v^2 + \sigma_w^2 &= \sigma_1^2 + \sigma_2^2 \\ \tan 2\alpha &= 2p\sigma_1\sigma_2/(\sigma_1^2 - \sigma_2^2) \quad \text{for } \sigma_1 \neq \sigma_2 \\ &= \pi/4 \quad \text{for } \sigma_1 = \sigma_2 \end{aligned} \quad (30)$$

Here α is the angle (positive in quadrant I) measured from the coordinates (x_1, x_2) (v, w) . Equations (30) can be transformed into a more useful set for practical calculations:

$$\begin{aligned} (\sigma_v + \sigma_w)^2 &= \sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2 \sqrt{1 - p^2} \\ (\sigma_v - \sigma_w)^2 &= \sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2 \sqrt{1 - p^2} \\ \tan \alpha &= [\sigma_2^2 - \sigma_1^2 + \sqrt{(2p\sigma_1\sigma_2)^2 + (\sigma_2^2 - \sigma_1^2)^2}] / (2p\sigma_1\sigma_2) \end{aligned} \quad (31)$$

Transform of the (x_1, x_2) coordinate system to the (v, w) system involves a simple rotation by the angle α about (\bar{x}_1, \bar{x}_2) . If we choose (\bar{v}, \bar{w}) to be the point $(0, 0)$ in the (v, w) system, then

$$\begin{aligned} x_1 - \bar{x}_1 &= v \cos \alpha - w \sin \alpha \\ x_2 - \bar{x}_2 &= v \sin \alpha + w \cos \alpha \end{aligned} \quad (32)$$

and

$$\begin{aligned} v &= (x_1 - \bar{x}_1) \cos \alpha + (x_2 - \bar{x}_2) \sin \alpha \\ w &= - (x_1 - \bar{x}_1) \sin \alpha + (x_2 - \bar{x}_2) \cos \alpha \end{aligned} \quad (33)$$

Since the transformation consists only of a parallel displacement and a rotation of the coordinate axes,

$$dx_1 dx_2 = dv dw \quad (34)$$

The marginal distributions

$$q(x_1) = \int_{-\infty}^{\infty} q(x_1, x_2) dx_2 \quad (35)$$

and

$$q(x_2) = \int_0^{\infty} q(x_1, x_2) dx_1 \quad (36)$$

are also useful quantities. They are both normally distributed. The function $q(x_1)$ represents the distribution of the lipoprotein with respect to density alone, and would be determined at the equilibrium condition of a stable density gradient ultracentrifugation experiment. The function $q(x_2)$ represents the distribution with respect to the logarithmic diameter parameter, and would be approximated by a proper treatment of the sedimentation distribution function obtained at a solvent density so far removed from \bar{x}_1 that we could approximate the $(x_1 - \rho)$ factor by $(\bar{x}_1 - \rho)$ where ρ is the solvent density. Experiments leading to the marginal distribution with respect to x_1 have been reported by Toro-Goyco⁵ and by Adams and Schumaker.⁹ Adams and Schumaker¹⁰ also make use of the marginal distribution with respect to a size parameter, by studying β -lipoprotein in a solvent of density 1.48. Both of these distributions can be calculated from a $q(x_1, x_2)$ distribution, by carrying out the integrations indicated in eq. (35) and (36).

Evaluation of the Parameters \bar{x}_1 , \bar{x}_2 , σ_1 , σ_2 , and p

In order to obtain the parameters x_1 , x_2 , σ_1 , σ_2 , and p , we must obtain a number of sedimentation distributions $g(S_i^0)$ or $(g(Z_i^0))$ at different solvent densities ρ_i . We can locate curves in the $(x_1 x_2)$ plane representing a constant S_i^0 or Z_i^0 from consideration of the equation

$$S_i^0 = d^2(x_1 - \rho_i)/(18\eta_i) = [(x_1 - \rho_i)/(18\eta_i)] \exp \{2x_2/M\} \quad (37)$$

or its logarithmic equivalent

$$Z_i^0 = 2x_2 + M \ln (x_1 - \rho_i) - M \ln (\eta_i) - M \ln (18) \quad (38)$$

where η_i is the solvent viscosity. Equations (37) and (38) hold when S_i^0 , d , and η are in cgs units, or in svedberg, Å, and millipoise units, and we will use this convenient latter set of units.

Figure 1 shows a set of curves for $Z_i^0 = \text{constant}$ for a solvent of density ρ_i . Curves to the right of ρ_i represent sedimentation (positive S_i^0), and those to the left represent flotation (negative S_i^0). Each curved line represents a constant S_i^0 or Z_i^0 . The curved lines are drawn for equal

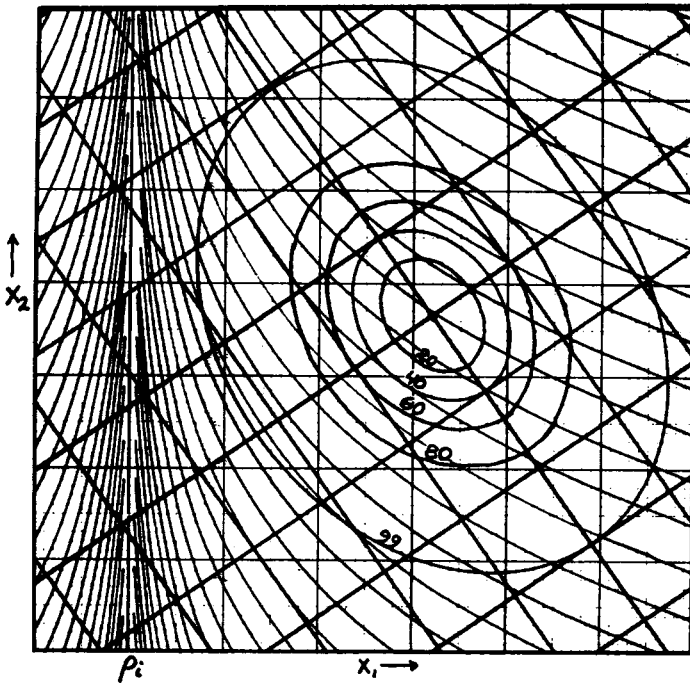


Figure 1.

intervals of 0.1 in Z_i^0 , and increases towards the top of the figure at any constant x_1 value. The coordinates at 0 and 90° represent the (x_1, x_2) set, and the stochastically independent coordinates (v, w) are shown at angles of about 35 and 125°. A few of the contour ellipses are also shown; these contours represent the situation when 20, 40, 60, 80, and 99% of the distribution lie within the contour ellipse. If we consider the three-dimensional figure in (x_1, x_2, h) space, discussed in the previous section, then these contour ellipses would represent cross sections cut by planes perpendicular to the h axis, and equally spaced at $h^0/5$ units except for the 99% contour (at $h^0/100$).

For our lipoprotein distribution, $(1/c^0)(d^2c/dx_1dx_2) = q(x_1, x_2)$, and we can consider the weight fraction of the preparation with density between x_1 and $x_1 + dx_1$ and logarithmic diameter parameter between x_2 and $x_2 + dx_2$ to be $q(x_1, x_2)dx_1dx_2$. Geometrically, this is the volume element in (x_1, x_2, h) space with base dx_1dx_2 in the $h = 0$ plane and height $h = q(x_1, x_2)$. The total volume between the $q(x_1, x_2)$ surface and the $h = 0$ plane will be unity. It is also useful to consider a generalized cylinder* containing the

*We use the word *cylinder* in the general mathematical sense that refers to any surface generated or swept out by a straight line moving along a plane curve and remaining parallel to a given line. The generating curve is called the *directrix* of the cylinder. A circular cylinder or an elliptical cylinder are special cases of this generalized definition, where the directrix is a circle or an ellipse. A serpentine wall represents a generalized cylinder with a sin curve as directrix.

curve $Z_i^0 = \text{const.}$ in the $h = 0$ plane as directrix and parallel to the h axis. Such a cylinder will cut the volume between the $q(x_1, x_2)$ surface and the $h = 0$ plane into two parts, the volume below and to the left representing the fraction of the lipoprotein with logarithmic sedimentation parameter less than Z_i^0 . This fraction could also have been defined by the integral distribution function

$$G(Z_i^0) = \int_{-\infty}^{Z_i^0} g(Z_i^0) dZ_i^0 = \int_0^{\infty} \left[\int_{-\infty}^{\kappa} q(x_1, x_2) dx_2 \right] dx_1 \quad (39)$$

where we introduce the new variable $\kappa = \kappa(Z_i^0, x_1)$ in order to define all x_2 values along a curve of constant Z_i^0 as a function of x_1 :

$$\kappa = Z_i^0/2 + (M/2) \ln(18\eta_i) - (M/2) \ln(|x_1 - \rho_i|) \quad (40)$$

To relate the function $q(x_1, x_2)$ to $g(Z_i^0)$ we can differentiate eq. (39) with respect to Z_i^0 :

$$g(Z_i^0) = \int_0^{\infty} q(x_1, \kappa) (\partial\kappa/\partial Z_i^0)_{x_1} dx_1 = \left(\frac{1}{2}\right) \int_0^{\infty} q(x_1, \kappa) dx_1 \quad (41)$$

Here we have obtained $(\partial\kappa/\partial Z_i^0)_{x_1} = \frac{1}{2}$ by differentiation of eq. (40). We could equally well have reversed the order of integration in eq. (39) and introduced another variable $\xi = \xi(Z_i^0, x_2)$ to obtain

$$g(Z_i^0) = \int_{-\infty}^{\infty} q(\xi, x_2) (\partial\xi/\partial Z_i^0)_{x_2} = \int_{-\infty}^{\infty} q(\xi, x_2) (18\eta_i/M) \exp [(Z_i^0 - 2x_2)/M] dx_2 \quad (42)$$

with

$$\xi = \rho_i + 18\eta_i \exp [(Z_i^0 - 2x_2)/M] = \rho_i + 18\eta_i S_i^0/d^2 \quad (43)$$

The rather complex eqs. (41,42) can be simplified by the use of the standardized variables $u_1 = (x_1 - \bar{x}_1)/\sigma_1$ and $u_2 = (x_2 - \bar{x}_2)/\sigma_2$, and either x_1 or x_2 removed by substitution of eq. (37) or (38) with constant S_i^0 or Z_i^0 .

To find numerical values for the parameters $\bar{x}_1, \bar{x}_2, \sigma_1, \sigma_2$, and p , we need to use quantitative data from the various $g(S_i^0)$ or $g(Z_i^0)$ data sets. The most easily used characteristics of these curves are the coordinates, $[(X_i^0)_m, g(X_i^0)_m]$ representing the maximum value of the $g(X_i^0)$ distribution. [Here we again use X_i^0 as representing either S_i^0 or Z_i^0 .] It seems obvious that the maximum of the $g(X_i^0)$ distribution will lie close to the point (\bar{x}_1, \bar{x}_2) where we have a maximum value of $q(x_1, x_2)$. There will usually be some difference between the maximum of the $g(S_i^0)$ distribution and the $g(Z_i^0)$ distribution, and both of these maxima should be investigated. Either should give a reasonable first approximation. A graph of the usual sort of $(S_i^0)_m \eta_i$ against ρ_i can give us values for \bar{d} and \bar{x}_1 by using eq. (37) with mean values for sedimentation coefficient, lipoprotein density, and lipoprotein diameter. Equation (38) is slightly more difficult to apply to date for $(Z_i^0)_m + \log(\eta_i)$ as a function of ρ_i . Perhaps the easiest method

to apply is to obtain a first approximation to \bar{x}_1 by the method outlined just above, calling this value a_1 . A plot of $(Z_i^0)_m + \log(18\eta_i) - \log(|a_1 - \rho_i|)$ versus $1/(|a_1 - \rho_i|)$ then permits evaluation of \bar{x}_1 and \bar{x}_2 by use of the equation:

$$\bar{Z}_i^0 + \log(18\eta_i) - \log(|a_1 - \rho_i|) = 2\bar{x}_2 - M(a_1 - \bar{x}_1)/(|a_1 - \rho_i|) \quad (44)$$

With these approximations for \bar{x}_1 and \bar{x}_2 , we can now calculate the remaining three parameters, σ_1 , σ_2 and p from data for $[g(S_i^0)]_m$ or $[g(Z_i^0)]_m$. Defining \bar{Z}_i^0 by the relationship $g(\bar{Z}_i^0) = [g(Z_i^0)]_m$, and evaluating it from eq. (38) as being very nearly $Z_i^0(\bar{x}_1, \bar{x}_2)$, we can obtain $\kappa(\bar{Z}_i^0, x_1)$ from eq. (40):

$$\kappa(\bar{Z}_i^0, x_1) = \bar{x}_2 - (M/2) \ln [(x_1 - \rho_1)/(\bar{x}_1 - \rho_1)] \quad (45)$$

This equation can be approximated for small values of $(x_1 - \bar{x}_1)/(\rho_i - \bar{x}_1)$ by expanding the \ln term, yielding

$$\begin{aligned} \kappa(\bar{Z}_i^0, x_1) = \bar{x}_2 + (x_1 - \bar{x}_1)H_i + (x_1 - \bar{x}_1)^2 H_i^2 + 4(x_1 \\ - \bar{x}_1)^3 H_i^3 / (3M^2) + \dots \end{aligned} \quad (46)$$

where $H_i = (M/2)/(\rho_i - \bar{x}_1)$. Using only the first two terms of this series, we can calculate the integral in eq. (41) and obtain an approximation to $[g(Z_i^0)]_m$:

$$[g(Z_i^0)]_m = 1/[2\sqrt{2\pi(\sigma_2^2 - 2p\sigma_1\sigma_2H_i + \sigma_1^2H_i^2)}] \quad (47)$$

or

$$1/\{8\pi[g(Z_i^0)]_m^2\} = \sigma_2^2 - 2p\sigma_1\sigma_2H_i + \sigma_1^2H_i^2 \quad (48)$$

If we have data with n solvents of different densities, and hence different H_i values, the set of n equations of the form eq. (48) can be solved for the best least squares fit, and the two standard deviations and the correlation coefficient can be obtained.

We have used this method with a number of lipoprotein preparations, and then revised our estimates of the five resulting parameters of the $g(x_1, x_2)$ distribution by calculating the predicted distribution curves $g(Z_i^0)$ and $g(S_i^0)$, and comparing them with the observed distributions. Small variations of the parameters can easily be carried out around the position determined by the approximate eq. (48), and a best set determined in this way. Results of such studies will be reported in subsequent publications.

The author wishes to thank a number of his students and colleagues for help in this study. Experimental data for calculations of this type have been supplied by Drs. Efraim Toro-Goyco, Shosaku Numa, Nancy R. Harvie and York Tsang. Computer programs for this work have been written by Mr. Sam Bledsoe, Robert vM. Morse and James B. Bell. Drs. Gordon Nordby and Walter Scheider have been of great help in developing this program, and Dr. Margaret J. Hunter has helped in the preparation of this manuscript. Support for this work on the human plasma lipoproteins has been provided in part by grants from the National Heart Institute, and from the Life Insurance Medical Research Fund.

References

1. R. L. Baldwin and J. W. Williams; *J. Amer. Chem. Soc.*, **72**, 4325 (1950).
2. J. W. Williams, K. E. van Holde, R. L. Baldwin, and H. Fujita; *Chem. Rev.*, **58**, 715 (1958).
3. J. W. Williams, *Prog. Chem. Org. Nat. Prod.*, **18**, 434 (1960).
4. R. L. Baldwin, and K. E. van Holde, *Fortschr. Hochpolym.-Forsch.*, **1**, 451 (1960).
5. E. Toro-Goyco, Ph.D. Dissertation, Harvard University, Cambridge, Massachusetts, 1958.
6. L. J. Gosting, *J. Amer. Chem. Soc.*, **74**, 1542 (1952).
7. I. Jullander, *Ark. Kemi, Mineral., Geol.* **21A**, No. 8 (1945). See page 61-62.
8. A. Hald, *Statistical Theory with Engineering Applications*, Wiley, New York, 1952.
9. G. H. Adams and V. N. Schumaker; *Nature*, **202**, 490 (1964).
10. G. H. Adams, Ph.D. Dissertation, University of Pennsylvania, Philadelphia, Pennsylvania, 1966.