[CONTRIBUTION FROM THE DEPARTMENT OF BIOCHEMISTRY, SCHOOL OF MEDICINE, DUKE UNIVERSITY]

The Apparent Shape of Protein Molecules[1]

BY HANS NEURATH

It has been recognized for some time that the term *globular proteins* does not necessarily imply spherical shape. Measurements of the double refraction of flow,[2-5] dielectric dispersion,[6,7] and viscosity[8-11] of protein solutions have in many instances proved that protein molecules may be unsymmetric. Whether a protein molecule is spherical may also be determined from combined sedimentation and diffusion measurements, as Svedberg[12] showed many years ago. The application of Svedberg's theory to the calculation of the relative dimensions of a number of representative protein molecules was described by us on a previous occasion, incidental to a discussion of the diffusion of proteins.[10]

In the present paper, further studies are reported on the *relative* and *absolute* dimensions of protein molecules and their respective dissociation and aggregation products. After a brief outline of the theory and its implications and limitations, the results are presented, and discussed from the general viewpoint of protein structure.

Theoretical

Svedberg[12,13] has introduced the term *dissymmetry constant* as an indirect measure for the deviations of molecules from spherical shape. The ratio f/f_0, where f is the measured molar frictional coefficient, and f_0 the theoretical molar frictional coefficient for a spherical molecule of the same molecular weight, ought to be unity for spherical molecules, and greater than unity for elongated molecules. The dissymmetry constant is related to the sedimentation constant, s, the molecular weight, M, the diffusion constant, D, the partial specific volume of the solute, V, and the density of the solution, ρ, according to equation (1).

$$f/f_0 = \frac{M(1 - V\rho)/s}{6\pi\eta N(3MV/4\pi N)^{1/3}} \qquad (1)$$

From the way in which this equation was derived it follows that, in the ideal case, f/f_0 must be identical with the ratio D_0/D, where D is the observed diffusion constant, and D_0 the diffusion constant of a spherical molecule of the same molecular weight. Hence, f/f_0 may be calculated merely from the molecular weight, the partial specific volume, and the observed diffusion constant, according to equation (2)

$$f/f_0 = \frac{kT}{6\pi\eta N(3MV/4\pi N)^{1/3}}\frac{1}{D} \qquad (2)$$

where k is the Boltzmann constant, η, the viscosity of the solvent, and N, the Avogadro number.

The strict validity of equations (1) and (2) requires the use of such dilute solutions that interactions between solute molecules become negligible. While this requirement may be fulfilled by sedimentation measurements, diffusion measurements necessarily have to be confined to more concentrated solutions. It is only when the experimentally determined value for f, calculated from either sedimentation equilibrium or diffusion measurements, is found to be independent of the protein concentration, that an unambiguous value may be ascribed to f/f_0.[14]

Difficulties also arise in assigning a true value to f_0. It is known that molecular weight data may be in error by as much as 10%. Although the molecular weight enters into the calculation of f_0 only as the cube root, yet small variations in f/f_0 may result in significant variations in shape factors calculated therefrom.

Another source of uncertainty in the calculation of shape factors relates to the influence of hydration.[16,17] While the theoretical molar frictional coefficient, f_0, is calculated from the molecular weight and partial specific volume of the unhy-

(1) Acknowledgment is made to the Lederle Laboratories, Pearl River, New York, for a grant-in-aid.

(2) Muralt and Edsall, *J. Biol. Chem.*, **89**, 351 (1930).

(3) Böhm and Signer, *Helv. Chim. Acta*, **14**, 1370 (1931).

(4) Lauffer and Stanley, *J. Biol. Chem.*, **123**, 507 (1938).

(5) Mehl, *Symposia Quant. Biol.*, **6**, 218 (1938).

(6) Oncley, THIS JOURNAL, **60**, 1115 (1938); Ferry and Oncley, *ibid.*, **60**, 1123 (1938).

(7) Elliott and Williams, *ibid.*, **61**, 718 (1939).

(8) Daniel and Cohn, *ibid.*, **58**, 415 (1936).

(9) Polson, *Nature*, **137**, 740 (1936).

(10) Neurath, *Symposia Quant. Biol.*, **6**, 196 (1938).

(11) Fahey and Green, THIS JOURNAL, **60**, 3039 (1938).

(12) Svedberg, *Z. physik. Chem.*, **127**, 51 (1925).

(13) Svedberg and Sjögren, THIS JOURNAL, **51**, 3594 (1929).

(14) Lamm and Polson,[15] for instance, reported an increase of the diffusion constants of egg albumin, hemoglobin and serum albumin in dilute protein solutions. In the case of egg albumin, serum albumin, insulin and pepsin, for instance, the molar frictional coefficient calculated from diffusion, f_d, differs from that calculated from sedimentation, f_s, by about 10 p. c., whereas equation (1) assumes equality of f_d and f_s.

(15) Lamm and Polson, *Biochem. J.*, **30**, 528 (1936).

(16) Adair and Adair, *Proc. Roy. Soc.* (London), **B120**, 422 (1936).

(17) Mehl and Williams, *Symposia Quant. Biol.*, **6**, 216 (1938).

drated molecule, the measured molar frictional coefficient, f, is determined by the hydrated molecule of higher partial specific volume, and molecular weight. If deviations of the dissymmetry constant from unity were to be ascribed exclusively to effects of hydration, a value of $f/f_0 = 1.1$ would correspond to $V = 0.82$ (assuming 20% hydration), whereas higher values of f/f_0 would result in V greater than unity. While it cannot be denied that the influence of hydration would tend to decrease the apparent dissymmetry constant to some extent, it is evident that hydration cannot account for values of f/f_0 greater than 1.1, even if the high values for the partial specific volume and hydration, suggested by Adair and Adair,[16] are used. As an approximation, we have assumed in this paper that the influence of hydration on dissymmetry constants is of second order. Specifically, the dissymmetry constants have been related to the apparent shape of the *unhydrated* protein molecules.

A quantitative relation between the relative dimensions of proteins and the dissymmetry constant may be obtained by combination of equation (2) with Perrin's[18] equation for the linear diffusion of ellipsoidal molecules. This equation, derived from hydrodynamic considerations, relates the ratio of the short to long axis of a prolate ellipsoid of revolution, a/b, to the ratio of the diffusion constants, D/D_0.[19]

$$\frac{D}{D_0} = \frac{(a/b)^{2/3}}{\sqrt{1-(a/b)^2}} \ln \frac{1+\sqrt{1-(a/b)^2}}{a/b} \quad (3)$$

The absolute values of the length of the short and long axis, a and b, may be estimated by combining equation (3) with the equation for the volume of a prolate ellipsoid of revolution. If M is the molecular weight, V the partial specific volume of the protein, and N the Avogadro number, then

$$MV = a^2 b \pi N/6 \quad (4)$$

It has to be emphasized that the results of such calculations may be subject to considerable error. The true values of b/a depend on the effective shape of the molecules, on the nature of their surface, and on the extent of interaction between solute and solvent. If the surface of the solute molecules is irregular and if there is marked interaction between solute and solvent (hydration), there will be, in addition to the frictional resistance due to the molecular shape, a frictional resistance produced by the surface of the molecules. Perrin's equation is for the specific case of a molecule having the shape of a prolate ellipsoid of revolution and devoid of any interaction with the solvent.

Further uncertainties are due to the fact that the dissymmetry constants recorded in the literature are exact only to one significant place, whereas variations in the second significant place may produce marked changes in the absolute dimensions, b and a. The true values of these constants are, in addition, largely dependent on the exact molecular shape. Thus, if one assumes as effective shape a cylinder, for instance, the short axis of the molecule would be one-third shorter than that of an ellipsoid of revolution of equal length. As, however, the shape of an ellipsoid of revolution is used for the calculation of the *relative* dimensions, the use of the same model for calculating the *absolute* dimensions will most likely give the best results.

Calculations

In spite of these theoretical uncertainties, we have calculated the relative and absolute dimensions of a large number of proteins, with equations (2), (3), and (4). It was felt that sufficient evidence has accumulated by other experimental methods for the belief that a physical meaning may be ascribed to the great variations in the observed dissymmetry constants. The limits of the error involved in these calculations may be appreciated by considering the assumptions discussed in the preceding paragraphs. Most of the calculations have been based on the values for the dissymmetry constants tabulated by Svedberg.[20] In some instances these have been recalculated from values for the molecular weight and diffusion constant. Where marked discrepancies between our and Svedberg's values have been found, both values have been listed (see Table II). For convenience, $V = 0.75$ has been assumed throughout the calculations.

In Table I, values for f/f_0, b/a, b, and a, are given. Unless stated otherwise, they have been calculated from f_d, and from the molecular weight as determined by diffusion and sedimentation velocity measurements. For reasons to be mentioned in the discussion, the proteins have been arranged according to increasing values of the short diameter of the molecule, a. In Table II the corre-

(18) Perrin, *J. phys. Radium*, **7**, 1 (1936).

(19) A similar relation exists for flat ellipsoids of revolution, but in all probability such a model has no significance for considerations of protein molecules.

(20) Svedberg, *Ind. Eng. Chem.*, **30**, 113 (1938); *Kolloid Z.*, **85**, 119 (1938).

sponding values for proteins and their respective dissociation and aggregation compounds are listed. The values for the hemocyanins have been taken from the work by Eriksson-Quensel and Svedberg.[21]

TABLE I

APPROXIMATE RELATIVE AND ABSOLUTE DIMENSIONS OF PROTEIN MOLECULES[a]

M = molecular weight. f/f_0 = dissymmetry constant. a = short diameter. b = long diameter.

	Protein	M	f/f_0	b/a	a, Å.	b, Å.
1	Zein	35,000	2.0	20.1	16	322
2	Cytochrome C	15,600	1.3	5.8	18	98
3	Gliadin	26,000	1.6	11.1	18	196
4	Hordein	27,500	1.6	11.1	18	196
5	Erythrocruorin (*Chironimus*)[b]	31,400	1.6	11.1	19	208
6	Serum albumin, urea denatured[22]	67,100	1.98	19.4	20	359
7	Lactalbumin α	17,500	1.2	4.3	21	91
8	Erythrocruorin (*Lampetra*)	17,100	1.2	4.3	22	94
9	Bence-Jones β	37,700	1.3	5.8	25	144
10	Myoglobin	17,200	1.1	2.9	24	70
11	Crotoxin	30,000	1.2	4.3	25	109
12	Concanavalin B	42,000	1.3	5.8	26	149
13	Tuberculin protein	32,000	1.2	4.3	26	112
14	Lactoglobulin	41,800	1.2	4.3	28	122
15	Pepsin	35,500	1.08	2.7	31	84
16	Insulin	40,900	1.13	3.3	31	102
17	Egg albumin[b]	40,500	1.1	2.9	32	91
18	Hemoglobin (horse)	69,000	1.24	4.8	32	155
19	Serum albumin (horse)[b]	67,100	1.2	4.3	34	145
20	Yellow ferment	82,800	1.2	4.3	36	152
21	Canavalin	113,000	1.3	5.8	36	207
22	Serum globulin	167,000	1.4	7.5	37	280
23	Diphtheria toxin[23]	72,000	1.2	4.3	34	145
24	Antipneumococcus serum globulin (rabbit)	157,000	1.4	7.5	37	274
25	Antipneumococcus serum globulin (man)	195,000	1.5	9.2	37	338
26	Concanavalin A	96,000	1.1	2.9	43	124
27	Erythrocruorin (*Arc a*)[b]	33,600	1.0	1	43	43
28	Bence-Jones α[b]	35,000	1.0	1	43	43
29	Catalase	248,000	1.3	5.8	46	297
30	Antipneumococcus serum globulin (horse)	920,000	2.0	20.1	47	950
31	Phycoerythrin (*Ceramium*)	290,000	1.2	4.3	54	232
32	Amandin	329,000	1.3	5.8	51	291
33	Thyroglobulin	628,000	1.5	9.2	54	498
34	Edestin	309,000	1.2	4.3	55	237
35	Excelsin	294,000	1.1	2.9	62	179
36	Urease	483,000	1.2	4.3	64	274
37	Hemocyanin (*Palinurus*)	446,000	1.2	4.3	62	268

[a] Unless otherwise stated, the values on which these calculations are based have been taken from Svedberg's tables.[17]

[b] Calculated from sedimentation equilibrium values.

Discussion

It is striking that, with a few exceptions, protein molecules are not spherical, and most exhibit high molecular dissymmetry.

Also it is well known that the apparent shape of macromolecules frequently reflects in some manner their stereo-chemical structure. If proteins are considered to consist of polypeptide chains, folded or fully extended, the long diameter of the molecule may be thought of as being determined by the length of the chains, whereas the short diameter would reflect the number of chains within the molecule. Led by such considerations, we have arranged, in Table I, the proteins according to their absolute dimensions. The length of the short diameter has been chosen as the criterion for our classification. It is interesting to note that proteins which belong to different groups as regards their molecular weight become members of the same group if arranged in this manner. Thus, cytochrome C, gliadin, and erythrocruorin (*Chironimus*), with the molecular weights of 15,600, 26,000 and 31,400, respectively, are found to have a common short diameter of about 18 Å. Likewise proteins of molecular weight 35,000, 96,000, 248,000 and 920,000, respectively, fall into one group. Although the calculated absolute dimensions may be subject to considerable error, we feel that their value for comparative purposes is unquestionable.

TABLE II

APPROXIMATE RELATIVE AND ABSOLUTE DIMENSIONS OF PROTEINS AND THEIR RESPECTIVE DISSOCIATION AND AGGREGATION COMPONENTS

M = molecular weight. f/f_0 = dissymmetry constant. a = short diameter of molecule. b = long diameter of molecule. The arrows indicate the direction of cleavage: → parallel to long axis; ↓ parallel to short axis.

Protein	M	f/f_0	b/a	a, Å.	b, Å.	Cleavage
Hemoglobin						
Main component	69,000	1.24[a]	4.8	32	155	
Dissociation component[24]	39,000	1.20[a]	4.3	28	118	Probably ↓
Phycocyan (*Ceramium*)[25]						
Main component	272,000	1.26[a]	5.3	49	261	
		1.2	4.3	53	228	
Dissociation component	131,000	1.4	7.5	34	258	→
Hemocyanin (*Eledone*)[21]						
Main component	2,791,000	1.4	7.4	95	715	
Dissociation component	457,000	1.92	18.2	39	707	→
Hemocyanin (*Helix pomatia*)[21]						
Main component	6,630,000	1.2	4.3	153	660	
Dissociation component	814,000	1.66[a]	12.4	54	663	→
		1.9	17.8	48	845	
Dissociation component	502,000	1.81[a]	15.7	42	661	→
		1.5	9.2	50	463	↓
Hemocyanin (*Busycon*)[21]						
Aggregation component	9,980,000	1.27[a]	5.4	159	922	
		1.3	..	...	...	
Main component	6,814,000	1.2	4.3	155	665	↓
Dissociation component	379,000	1.4	7.5	57	430	↓→

[a] f/f_0 recalculated from the values for D and M given by the authors.

(21) Eriksson-Quensel and Svedberg, *Biol. Bull.*, **71**, 488 (1936).

(22) Neurath and Saum, *J. Biol. Chem.*, **126**, 347 (1939).

(23) Lundgren, Pappenheimer and Williams, THIS JOURNAL, **61**, 533 (1939).

(24) Steinhardt, *J. Biol. Chem.*, **123**, 543 (1938).

(25) Eriksson-Quensel, *Biochem. J.*, **32**, 585 (1938).

The importance of such considerations becomes even more apparent if the dimensions of protein are compared with those of their respective dissociation and aggregation components.[26] From Table II it may be seen that dissociation of phycocyan (*Ceramium*) for instance, affects only the length of the short axis, when the molecular weight is halved. Similarly, dissociation of hemocyanin (*Eledone*) only decreases the short axis when the molecular weight drops from 2,790,000 to 457,000. Such data immediately suggest that these molecules split in the direction parallel to the major molecular axis. In other cases, only the major axis is shortened during dissociation indicating that splitting takes place at right angles to the major axis. In Table II, the apparent direction of splitting is indicated by arrows, a horizontal arrow indicating cleavage along the major axis, and a vertical arrow cleavage along the minor axis.[27]

Comparison of the absolute dimensions of the anti-pneumococcus serum globulins indicates a similar relationship. These proteins seem to differ chiefly in respect to the length of the molecule, if the species from rabbit, man, and horse are compared with each other (Table I). An exact interpretation of the molecular dimensions in terms of chemical configuration would be premature at present. Only in a few isolated cases, such as zein,[7] or hemoglobin,[29] has yet a satisfactory agreement between the shape factors determined by different independent methods been established. Further experimental work is needed before the values recorded in Tables I and II can be accepted as exact values.

The work of Astbury[30] and of Bergmann and Niemann[31] has furnished much evidence for the hypothesis that the difference between fibrous and globular proteins is merely a quantitative one. The data presented in this paper likewise illustrate the fibrous character of the globular proteins, thus lending support to the theory that both types of proteins are structurally closely related.

(26) Here the influence of hydration is most likely the same for the undissociated protein and its respective dissociation products.

(27) In the case of the dissociation of hemoglobin by urea, this regularity does not seem to hold exactly. It appears from such considerations, as well as from others,[28] that the value for the dissymmetry constant of this protein is somewhat too low.

(28) Neurath, *Symposia Quant. Biol.*, **6**, 23 (1938).

Summary

A method for the estimation of the relative and absolute dimensions of protein molecules from the dissymmetry constant and Perrin's diffusion equation is outlined and its limitations discussed. The data indicate that most proteins are unsymmetric in shape. A certain regularity exists, independent of the molecular weight, when proteins are grouped on the basis of their molecular dimensions. Comparison of the dimensions of proteins with those of their respective dissociation and aggregation products indicates that cleavage occurs only in directions parallel to either the major or minor molecular axis.

(29) Arrhenius, *Physik. Z.*, **39**, 559 (1938).

(30) Astbury, *Nature*, **140**, 968 (1937).

(31) Bergmann and Niemann, *J. Biol. Chem.*, **115**, 77 (1936); **118**, 301 (1937).

DURHAM, NORTH CAROLINA RECEIVED MARCH 31, 1939

[CONTRIBUTION FROM THE RESEARCH LABORATORY OF ORGANIC CHEMISTRY, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, NO. 192]

The Hydroxylation of Unsaturated Substances. V. The Catalytic Hydroxylation of Certain Unsaturated Substances with Functional Groups

BY NICHOLAS A. MILAS, SIDNEY SUSSMAN AND HOWARD S. MASON

Further application of the recent hydroxylation method developed in this Laboratory[1] has led to the production of a number of useful products which hitherto have been rather difficult to procure in satisfactory yields. Of particular interest along this line is the production of glycolaldehyde from divinyl ether in yields as high as 96%. Furthermore, it has been found possible to add hydroxyl groups to substances containing double bonds adjacent to carbonyl groups. Böeseken,[2] for example, has found that ethyl fumarate and ethyl maleate were not appreciably oxidized with perbenzoic acid, and in the case of unsaturated ketones having the same arrangement of groups he reports[3] oxidation of the ketones without the formation of epoxy derivatives but with cleavage

(1) Milas and Sussman, THIS JOURNAL, **58**, 1302 (1936); **59**, 2345 (1937).

(2) Böeseken, *Rec. trav. chim.*, **45**, 838 (1926).

(3) Böeseken and co-workers, *ibid.*, **50**, 827 (1931); **52**, 874 (1933); **55**, 786 (1936).