

The Inertia-Equivalent Ellipsoid: a Link Between Atomic Structure and Low-Resolution Models of Small Globular Proteins Determined by Small-Angle X-ray Scattering

BY JÜRGEN J. MÜLLER AND HANNELORE SCHRAUBER

Institut für Molekularbiologie, Robert-Rössle-Strasse, 10, O-1115 Berlin, Germany

(Received 7 January 1991; accepted 25 September 1991)

Abstract

Low-resolution three-parameter models of the shape of a biopolymer in solution can be determined by a new indirect method from small-angle X-ray scattering without contrast-variation experiments. The basic low-resolution model employed is a triaxial ellipsoid – the inertia-equivalent ellipsoid (IEE). The IEE is related to the tensor of inertia of a body and the eigenvalues and eigenvectors of this tensor can be calculated directly from the atomic coordinates and from the homogeneous solvent-excluded body of a molecule. The IEE defines a mean molecular surface (like the sea level on earth) which models the molecular shape adequately if the IEE volume is not more than 30% larger than the dry volume of the molecule. Approximately 10 to 15% of the solvent-excluded volume is outside the ellipsoid; the radii of gyration of the IEE and of the homogeneous molecular body are identical. The largest diameter of the IEE is about 5 to 15% (~0.2–0.8 nm) smaller than the maximum dimension of globular molecules with molecular masses smaller than 65 000 daltons. From the scattering curve of a molecule in solution the IEE can be determined by a calibration procedure. 29 proteins of known crystal structure have been used as a random sample. Systematic differences between the axes of the IEE, calculated directly from the structure, and the axes of the scattering-equivalent ellipsoids of revolution, estimated from the scattering curve of the molecule in solution, are used to derive correction factors for the axial dimensions. Distortions of model dimensions of 20 to 40% (up to 1 nm), caused by misinterpretation of scattering contributions from electron density fluctuations within the molecule, are reduced to a quarter by applying these correction factors to the axes of the scattering-equivalent ellipsoids of revolution. In a computer experiment the axes of the inertia-equivalent ellipsoids have been determined for a further nine proteins with the same accuracy. The automated estimation of the IEE from the scattering curve of a molecule in solution is realized by the Fortran77 program *AUTOIEE*.

I. Introduction

Small- and intermediate-angle X-ray scattering are well established methods for structure characterization of macromolecules in solution (Glatter & Kratky, 1982). In general, no straightforward algorithm exists for a structure determination from the experimental scattering curve. Usually trial-and-error procedures are necessary to estimate structure models with increasing resolution. Therefore, various theoretical methods have been developed for calculation of scattered intensities for models of any resolution level – simple geometrical bodies (Mittelbach & Porod, 1961, 1962), aggregates of small spheres (Rolbin, Feigin & Scedrin, 1971), weighted groups of atoms (Müller, 1983) and molecules with known atomic coordinates (Fedorov, Ptitsyn & Voronin, 1972) and van der Waals radii (Müller, 1983). For models with atomic resolution and models consisting of scattering centres corresponding to atomic groups, the relation between the model and the real structure is evident. This is not the case for low-resolution models, *e.g.* shape models, which are very often the first structure model of a molecule in solution. In the literature, low-resolution models mostly neglect inner electron density fluctuations or are constructed without explanation of the algorithm even if atomic coordinates are the basis. Furthermore, the trial-and-error-estimation procedure to obtain shape models from the scattering curve is mostly performed without sufficient correction of the scattering contributions of the inner electron density fluctuations (Stuhrmann & Kirste, 1965, 1967; Kratky, 1963). Therefore, models with scattering equivalent to that of a molecule in a restricted angular region may not be equivalent to the structure of the macromolecule.

Here, for globular proteins with molecular masses smaller than 65 000 daltons, the relation between atomic structure and the type of ellipsoid model (shape-scattering-equivalent ellipsoid, SSEE; scattering-equivalent ellipsoid, SEE; inertia-equivalent ellipsoid, IEE) will be discussed. The atomic coordinates of 29 selected proteins were taken from the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977). As the most

convenient ellipsoid, the triaxial inertia-equivalent ellipsoid, originally used by Taylor, Thornton & Turnell (1983) for detection of antigenic determinants at protein surfaces, is introduced here to correlate geometrically the real atomic structure with a low-resolution shape model of a globular molecule.

II. Theory and theoretical results

1. The shape at atomic resolution

The algorithm for the calculation of the scattering curve of a macromolecule in solution takes into account the atomic coordinates of a molecule, the atomic structure factors (Fedorov, Ptitsyn & Voronin, 1972) and the van der Waals radii of the atoms (Müller, 1983). The scattering intensity of the crystallographically characterized molecule is then calculated by

$$I(s) = \langle |F_{\text{vac}}(\mathbf{s}) - \varphi_{\text{EV}}(\mathbf{s})|^2 \rangle_{\Omega} \quad (1)$$

$$\varphi_{\text{EV}}(\mathbf{s}) = 8\rho_0 [\sin(s_x E) \sin(s_y E) / s_x s_y] \sum_{j=1}^N \sin(s_z E l_j) / s_z \quad (2)$$

$$F_{\text{vac}}(\mathbf{s}) = \sum_{i=1}^M f_i(s) \exp(i\mathbf{s} \cdot \mathbf{r}_i) \quad (3)$$

F_{vac} is the molecular structure factor in vacuum and φ_{EV} is the structure factor for the volume of the macromolecule homogeneously filled with the constant electron density ρ_0 of the solvent. The symbol $\langle \rangle_{\Omega}$ denotes a spherical average in reciprocal space, \mathbf{s} is the scattering vector with $|\mathbf{s}| = s = 4\pi(\sin \theta) / \lambda$ (2θ : scattering angle; λ : X-ray wavelength) and $\mathbf{r}_i = (x_i, y_i, z_i)$ is the real-space vector of the i th non-hydrogen atom. The homogeneous solvent-excluded volume of the molecule is determined by a modified rolling-sphere procedure (Müller, 1983) which was described originally by Richards (1977). It is constructed of cubes of edge length E and the cubes are collected in N parallelepipeds with edge lengths E , E and $l_j E$. These parallelepipeds define the shape of a molecule at the atomic structure level.

The left insert in Fig. 1 shows a projection of the protein leghemoglobin, and the right insert presents the same projection of its shape in aqueous solution. The shape is defined by the cubes with sufficient exactness for all small-angle and intermediate-angle scattering experiments if cube edges of 0.05 nm or smaller are used for the construction of the solvent-excluded volume (Müller, Pavlov & Fedorov, 1983).

The scattered intensities of the leghemoglobin in solution and of its shape (Fig. 1) are calculated from (1) to (3). The discrepancies between the two curves are representative of all investigated proteins with

molecular masses less than 65 000 daltons. They are caused by the scattering contributions of the inner electron density fluctuations and by the cross term of shape and fluctuation scattering (Stuhrmann & Kirste, 1965, 1967). The theoretically calculated scattering curve of the molecule in solution corresponds to the experimental small-angle X-ray curve. The scattering curve of the shape can be experimentally estimated only by extensive contrast-variation measurements (Stuhrmann & Kirste, 1965, 1967).

2. The shape-scattering-equivalent ellipsoid (SSEE) – a low-resolution shape model

For discussion of the overall structure of a molecule and of distances between domains in macromolecules or between different macromolecules, structure dimensions must be used – usually shape dimensions. The largest diameter L of a molecule can be determined directly from the solution scattering curve, without any distortions by scattering from inner electron density fluctuations (Damaschun, Kley, Müller & Pürschel, 1968). The other dimensions of the shape can only be estimated from the shape scattering curve (Fig. 1). For low resolution, a spheroidal model was chosen and the dimensions of its axes were calculated by a non-linear least-squares curve-fitting procedure (Marquardt, 1963; Reich,

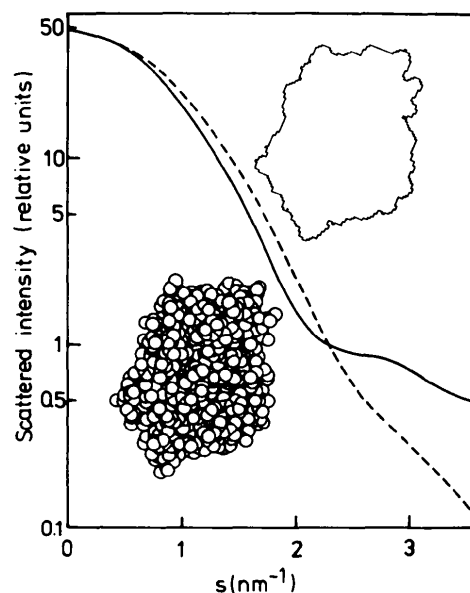


Fig. 1. Theoretical small-angle X-ray scattering curves of leghemoglobin (1LH4), calculated by the improved cube method from atomic coordinates and van der Waals radii. — Scattering curve of the molecule in solution (insert lower left: projection of the molecule); --- scattering curve of the homogeneously filled body of the solvent-excluded volume (insert upper right: projection of the shape).

Wangermann, Falck & Rohde, 1972; Müller, Damaschun & Hübner, 1979). The scattering curves of the ellipsoids are fitted to the shape scattering curve within a scattering-vector region $s \leq 3\pi/L$. This fitting interval of three sampling points (Damaschun & Pürschel, 1971; Damaschun, Müller & Pürschel, 1968) guarantees that the information content in the curve fragment is high enough for determination of three model parameters, the three axes of the shape-scattering-equivalent ellipsoid SSEE. Such a unique ellipsoid exists if the differences between the model scattering curve and the theoretical shape scattering curve of the molecule are smaller than 1% throughout the fitting region, and if the variances of the estimated model parameters are less than 1%. The result of this fitting procedure is shown for leghemoglobin in Fig. 2, and the half axes of ellipsoids modelling the homogeneous bodies of a further 28 proteins are summarized in Table 1.

Geometrical parameters of the molecular shape, such as solvent-excluded volume V_{HOM} , largest diameter L and radius of gyration $R_{\text{G,HOM}}$ (Table 1), are directly calculated from the parallelepipeds which built up the homogeneous molecular body. The comparison of the structural characteristics of the molecule and of the SSEE proves its formal suitability as a low-resolution shape model:

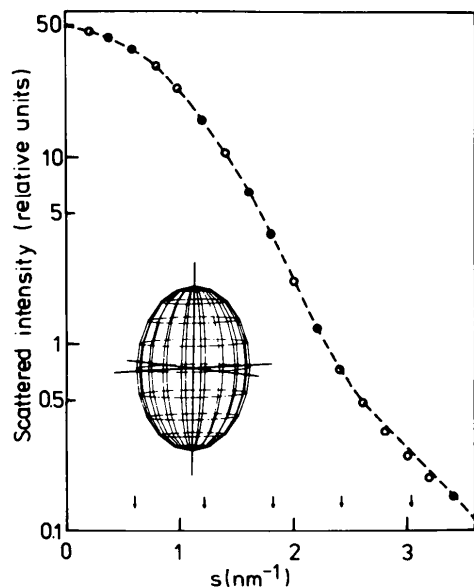


Fig. 2. Low-resolution shape determination by curve fitting in the angular region of the first three sampling points. --- Theoretical shape scattering of the molecule leghemoglobin (1LH4) in aqueous solution; $\circ\circ\circ$ scattering curve of the shape-scattering-equivalent ellipsoid (SSEE). \downarrow marks the sampling points. Insert: SSEE, calculated by curve fitting corresponding to Marquardt (1963); half-axes $A = 2.46$, $B = 1.91$, $C = 1.2$ nm.

(i) the radii of gyration of the SSEE and the homogeneously filled body of the molecule are nearly identical;

(ii) the largest diameter of the SSEE is only about 5 to 10% (~ 0.2 – 0.5 nm) smaller than the maximum dimension of the molecules in the random sample in Table 1;

(iii) the volume of the SSEE is only 10 to 20% larger than the dry volume of the molecules.

For some molecules no triaxial SSEE is equivalent in scattering in the fitting region (human deoxy-hemoglobin A, 2HHB/A; subtilisin novo, 2SBT; concanavalin A, 2CNA), but the two axes of ellipsoids of revolution can be determined with low variances in such cases. [The four-character code used in connection with the atomic coordinate sets follows the Brookhaven Protein Data Bank convention (Bernstein *et al.*, 1977).]

However, no algorithm exists for direct calculation of the SSEE from a known atomic structure. Because, furthermore, the experimental estimation of the shape scattering is only possible by X-ray or neutron contrast-variation experiments, the SSEE is rejected as a low-resolution shape model.

3. The scattering-equivalent ellipsoid (SEE)

In many small-angle scattering experiments, macromolecules are investigated in aqueous solution. During the low-resolution modelling procedure, the scattering contributions of the inner-electron density fluctuations at small angles are ignored or partially 'corrected' by subtraction of a constant term (Kratky, 1963). As can be seen in Fig. 1, the intensity scattered by the homogeneous molecular body is significantly changed by the scattering from inner electron density fluctuations and by cross terms from shape and fluctuation scattering. The shown scattering curves of the leghemoglobin molecule and of its shape are typical for globular proteins with molecular masses smaller than 65 000 daltons. Thus the additional scattering terms can be neither neglected nor corrected by subtraction of a constant from the experimental solution scattering curve.

In the first case, the scattering curve of the molecule in solution would be interpreted as pure shape scattering. The dimensions of ellipsoidal models determined by curve-fitting procedures using the observed scattered intensities will be significantly different from the dimensions of the SSEE. The calculation of triaxial scattering-equivalent ellipsoids (SEE) was impossible for the proteins collected in Table 1 when using Marquardt's procedure within the first three sampling points of the scattering curve. The fitting region cannot be further enlarged due to the dominance of the scattering contributions of the inner electron density fluctuations at larger angles.

Table 1. Geometrical parameters of the homogeneous solvent-excluded body of the molecules and of the shape-scattering-equivalent ellipsoid (SSEE)

Volume V_{HOM} , radius of gyration $R_{\text{G,HOM}}$ and largest diameter L are calculated by using the improved cube method (Müller, 1983). A , B , C , V_{SSEE} are the half-axes and the volume of the ellipsoid equivalent in scattering to the homogeneous body. $\langle \rangle$ indicates that no unique three-axial ellipsoid exists. A , B in the code name designate the chain.

1FB4: only amino acids 1 to 108 of the light chain. MRD is the mean relative difference between the lengths of axes of the SSEE and the IEE, $\text{MRD} = (1/N') \sum_1^{N'} (X_{\text{SSEE}} - X_{\text{IEE}}) \times 100\% / X_{\text{SSEE}}$ and σ is the corresponding standard deviation of MRD.

Code	V_{HOM} (nm ³)	$R_{\text{G,HOM}}$ (nm)	L (nm)	SSEE				
				A (nm)	B (nm)	C (nm)	V_{SSEE} (nm ³)	
I. Oxygen-binding proteins								
Leghemoglobin (deoxy)	1LH4	21.11	1.49	5.17	2.46	1.91	1.20	23.62
Myoglobin (sperm whale, met)	1MBN	21.10	1.49	4.72	2.37	2.03	1.12	22.57
Myoglobin (sperm whale, oxy)	1MBO	21.56	1.50	5.00	2.39	2.02	1.19	24.06
Hemoglobin V (lamprey, met)	2LHB	20.27	1.46	4.71	2.42	1.80	1.28	23.36
Hemoglobin (human, deoxy, A)	2HHB/A	18.81	1.43	4.46	<2.11	2.11	1.15	21.45
	2HHB/B	19.67	1.49	5.05	2.46	1.92	1.14	22.55
Hemoglobin (horse, met)	2MHB/B	19.95	1.47	4.96	2.43	1.88	1.18	22.58
MRD _A ; σ_A (%)					3.2; 0.8			
MRD _B ; σ_B (%)						-1.1; 1.8		
MRD _C ; σ_C (%)							-10.2; 3.4	
II. Proteases								
Chymotrypsinogen	1CHG	27.76	1.60	5.20	2.45	2.17	1.45	32.29
Trypsinogen (bovine)	1TGN	27.38	1.61	4.95	2.48	2.18	1.40	31.70
Trypsin	1TPO	27.61	1.60	5.13	2.51	2.10	1.44	31.79
Chymotrypsinogen A (bovine)	2CGA/B	30.74	1.65	5.16	2.48	2.21	1.58	36.27
γ -Chymotrypsin	2GCH	29.47	1.61	5.08	2.50	2.06	1.58	34.08
Kallikrein A (porcine)	2PKA	29.95	1.64	5.24	2.51	2.22	1.50	35.01
Proteinase II (rat)	3RP2	29.49	1.63	5.28	2.54	2.15	1.50	34.31
α -Chymotrypsin (bovine)	5CHA/B	29.30	1.62	5.24	2.48	2.13	1.55	34.30
α -Lytic protease	2ALP	23.55	1.49	4.94	2.36	1.86	1.44	26.48
Subtilisin novo	2SBT	32.74	1.66	5.30	<2.37	2.37	1.60	37.64
Proteinase A (<i>S. griseus</i>)	2SGA	20.86	1.42	4.57	2.23	1.79	1.41	23.58
Proteinase B (<i>S. griseus</i>)	3SGB	21.70	1.47	4.71	2.35	1.88	1.32	24.43
MRD _A ; σ_A (%)					2.7; 1.5			
MRD _B ; σ_B (%)						3.2; 2.5		
MRD _C ; σ_C (%)							-12.9; 3.2	
III. Other proteins								
Cytochrome B562 (<i>E. coli</i> , oxy)	156B	13.50	1.40	5.07	2.56	1.43	0.99	15.18
Cytochrome C (bonito)	1CYC	13.82	1.26	4.04	2.02	1.50	1.29	16.37
Carbonic anhydrase C (human)	1CAC	33.86	1.74	5.72	2.60	2.36	1.65	42.41
Ca-binding parvalbumin	1CPV	13.43	1.26	3.87	1.95	1.63	1.20	15.98
Immunoglobulin Fab (λ , Kol)	1FB4	13.42	1.33	4.43	2.31	1.56	1.02	15.37
Ribonuclease A	1RN3	15.72	1.41	4.70	2.38	1.84	0.99	18.16
Carbonic anhydrase B (human)	2CAB	33.34	1.70	5.85	2.59	2.26	1.61	39.48
Concanavalin A	2CNA	30.47	1.69	5.95	<2.91	1.71	1.71	35.64
Carboxypeptidase A (bovine)	5CPA	40.88	1.78	5.91	2.79	2.17	1.84	46.66
Lysozyme (egg white)	5LYZ	16.43	1.38	4.92	2.37	1.54	1.26	19.26
MRD _A ; σ_A (%)					2.7; 2.9			
MRD _B ; σ_B (%)						1.9; 7.4		
MRD _C ; σ_C (%)							-14.6; 10.8	

Unequivocal results with variances smaller than 1% for the free parameters (half-axes) were only obtained if the fitting region was restricted to the first two sampling points $s \leq 2\pi/L$. According to information theory, only two parameters are then estimable and the ellipsoids, equivalent in scattering to the molecule in the interval $0 \leq s \leq 2\pi/L$, are two-axial ellipsoids of revolution. In general, no distinction is possible

between oblate and prolate ellipsoids (Fig. 3). The dimensions of the semi-axes of the SEEs for the proteins of the random sample are summarized in Table 2. A comparison between the axial dimensions of the SEEs and of the shape-scattering-equivalent ellipsoids shows that there exist differences of up to 50%, e.g. a 1 nm length difference between the smallest axes of the prolate SEE and of the SSEE for

Table 2. Geometrical parameters of the inertia-equivalent ellipsoid (IEE) and of prolate and oblate ellipsoids equivalent in scattering up to the second sampling point with the macromolecules in solution (SEE)

A , B , C , V_{IEE} are the half-axes and the volume of the inertia-equivalent ellipsoid, V_{HOM} external is the percentage of the homogeneous molecular volume outside the surface defined by the IEE; o – oblate shape, p – prolate shape. $C_{X,Y}$ are the factors by which the axes of the SEE have to be multiplied to derive the axes X of the IEE within the group Y . C_A , C_B , C_C are the average factors over all the groups. For the protein names, see Table 1.

	IEE					Prolate SEE		Oblate SEE	
	A (nm)	B (nm)	C (nm)	V_{IEE} (nm ³)	V_{HOM} external (%)	A (nm)	$B = C$ (nm)	$A = B$ (nm)	C (nm)
I. Oxygen-binding proteins									
1LH4	2.41	1.90	1.30 <i>o</i>	24.94	12	2.75	1.65	2.39	1.23
1MBN	2.28	2.05	1.31 <i>o</i>	25.65	12	2.64	1.66	2.31	1.31
1MBO	2.31	2.04	1.31 <i>o</i>	25.86	12	2.63	1.69	2.32	1.34
2LHB	2.36	1.78	1.40 <i>p</i>	24.64	12	2.55	1.66	2.25	1.34
2HHB/A	2.21	1.91	1.32 <i>o</i>	23.34	13	2.48	1.65	2.20	1.36
2HHB/B	2.36	1.98	1.24 <i>o</i>	24.27	13	2.72	1.62	2.35	1.23
2MHB/B	2.34	1.94	1.27 <i>o</i>	24.15	13	2.70	1.63	2.35	1.22
$C_{A,I}$						0.880 (8)		1.006 (7)	
$C_{B,I}$						1.176 (21)		0.841 (15)	
$C_{C,I}$						0.792 (10)		1.012 (13)	
II. Proteases									
1CHG	2.41	2.06	1.67 <i>o</i>	34.73	13	2.61	1.89	2.37	1.65
1TGN	2.45	2.14	1.53 <i>o</i>	33.60	12	2.70	1.83	2.41	1.52
1TPO	2.46	2.07	1.55 <i>o</i>	33.00	12	2.69	1.83	2.40	1.54
2CGA/B	2.37	2.19	1.78 <i>o</i>	38.70	13	2.64	1.99	2.43	1.78
2GCH	2.39	2.04	1.76	35.94	13	2.61	1.94	2.38	1.71
2PKA	2.48	2.12	1.69 <i>p</i>	37.26	13	2.72	1.93	2.45	1.66
3RP3	2.50	2.03	1.73 <i>p</i>	36.78	13	2.71	1.90	2.44	1.62
5CHA/B	2.40	2.07	1.74	36.21	12	2.69	1.89	2.44	1.57
2ALP	2.29	1.80	1.64 <i>p</i>	28.32	12	2.44	1.72	2.19	1.47
2SBT	2.31	2.23	1.88 <i>o</i>	40.49	12	2.66	2.04	2.45	1.83
2SGA	2.11	1.80	1.57	24.90	12	2.29	1.68	2.09	1.47
3SGB	2.30	1.73	1.58 <i>p</i>	26.33	12	2.44	1.66	2.18	1.39
$C_{A,II}$						0.909 (7)		1.004 (10)	
$C_{B,II}$						1.084 (11)		0.858 (10)	
$C_{C,II}$						0.897 (12)		1.036 (12)	
III. Other proteins									
156B	2.58	1.35	1.14 <i>p</i>	16.63	14	2.78	1.25	-	-
1CYC	1.91	1.63	1.31	17.08	12	2.01	1.45	1.83	1.25
1CAC	2.58	2.36	1.89 <i>o</i>	48.20	12	2.88	2.08	2.61	1.81
1CPV	1.85	1.69	1.28 <i>o</i>	16.76	12	2.01	1.58	1.87	1.43
1FB4/L108	2.28	1.41	1.28 <i>p</i>	17.24	14	2.45	1.33	2.09	0.84
1RN3	2.39	1.59	1.32 <i>p</i>	21.01	14	2.60	1.44	2.23	0.95
2CAB	2.51	2.12	1.91 <i>p</i>	42.57	14	2.80	2.00	2.53	1.73
2CNA	2.74	2.00	1.67 <i>p</i>	38.33	13	3.16	1.84	2.71	1.39
5CPA	2.57	2.31	1.96 <i>o</i>	48.90	12	2.92	2.14	2.65	1.89
5LYZ	2.32	1.52	1.38 <i>p</i>	20.35	12	2.48	1.46	2.14	1.08
$C_{A,III}$						0.907 (8)		1.020 (15)	
$C_{B,III}$						1.086 (10)		0.814 (30)	
$C_{C,III}$						0.914 (13)		1.104 (51)	
C_A						0.901 (5)		1.010 (7)	
C_B						1.104 (10)		0.840 (12)	
C_C						0.880 (11)		1.052 (18)	

the oxygen-binding proteins and for ribonuclease A (1RN3). The volumes of the SEEs are 30 to 60% larger than the volumes V_{HOM} of the molecules (Tables 1 and 2), whereby a 'solvation' or 'degree of hydration' of 30 to 60% by volume of the dry molecule is simulated (Pilz, 1982; Damaschun *et al.*, 1978). Therefore, by the

electron density fluctuations effected 'shape distortion', the meaning of the geometrical dimensions of the SEE related to the real structure is unclear and the ellipsoids cannot be used directly as shape models.

In the second case, if the scattering contributions from inner electron density fluctuations are 'correc-

ted' by a procedure emphasized by Kratky (1963) and very often applied to experimental data to reach convergency in data treatment, the geometrical distortions are enlarged for all protein models of the random sample discussed above. For example, the half-axes of the ellipsoid, equivalent in scattering to the calculated 'corrected' scattering curve of leghe-moglobin (1LH4) within two sampling intervals, would be determined as $A = 2.37$, $B = 2.37$, $C = 1.32$ nm. Thus the smallest axis of the oblate SEE (Table 2) is further enlarged by about 10% (~ 0.2 nm) as is, consequently, the model volume.

In the following, a new ellipsoid – the inertia-equivalent ellipsoid – will be proposed as an undistorted low-resolution shape model, which is directly related to the atomic structure of a molecule and which can be deduced from the scattering-equivalent ellipsoids of revolution.

4. The inertia-equivalent ellipsoid (IEE)

4.1. *Calculation of the IEE from the atomic structure.* The straightforward determination of an inertial ellipsoid and of the related inertia-equivalent ellipsoid (IEE) of a body is possible using well known methods of classical mechanics (Sommerfeld, 1962). A

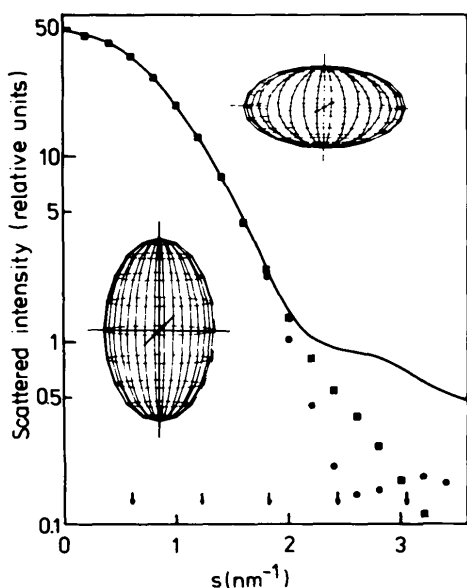


Fig. 3. Direct low-resolution shape determination by curve fitting according to Marquardt (1963) in the angular region of the first two sampling points. — Theoretical scattering curve of leghemoglobin (1LH4) in aqueous solution; ■ ■ ■ scattering curve of the scattering-equivalent oblate ellipsoid of revolution (SEE); ● ● ● scattering curve of the scattering-equivalent prolate ellipsoid of revolution (SEE). ↓ marks the sampling points. Insert upper right: oblate SEE with half-axes $A = B = 2.39$, $C = 1.23$ nm. Insert lower left: prolate SEE with $A = 2.75$, $B = C = 1.65$ nm.

description of the procedure applied to molecules composed of unit point masses was given by Taylor, Thornton & Turnell (1983). Here, for the purpose of shape modelling, we use the homogeneous solvent-excluded-volume body of the macromolecule for computation of the inertial ellipsoid.

The homogeneous body is built up from N parallelepipeds with edge lengths E , E and El_j and mass centres at (x_j, y_j, z_j) as discussed by Müller (1983). E was chosen to be 0.05 nm. The components $T(k)$ of the symmetric inertia tensor of the molecular shape are given by

$$T(1) = E^3 \sum_{j=1}^N [l_j y_j^2 + l_j z_j^2 + (1/12)l_j(l_j^2 - 1)] \quad (4)$$

$$T(2) = -E^3 \sum_{j=1}^N l_j x_j y_j \quad (5)$$

$$T(4) = -E^3 \sum_{j=1}^N l_j x_j z_j \quad (6)$$

$$T(5) = -E^3 \sum_{j=1}^N l_j y_j z_j \quad (7)$$

$$T(6) = E^3 \sum_{j=1}^N l_j(x_j^2 + y_j^2) \quad (8)$$

$$T(3) = E^3 \sum_{j=1}^N [l_j x_j^2 + l_j z_j^2 + (1/12)l_j(l_j^2 - 1)]. \quad (9)$$

The axes of the inertial ellipsoid are found by solving the eigenvalue problem for this tensor. The inertial ellipsoid is then defined by

$$\lambda_1 x^2 + \lambda_2 y^2 + \lambda_3 z^2 = 1 \quad (10)$$

with $\lambda_1, \lambda_2, \lambda_3$ being the eigenvalues,

$$\lambda_1 = (B^2 + C^2)/5; \quad \lambda_2 = (A^2 + C^2)/5; \quad (11)$$

$$\lambda_3 = (A^2 + B^2)/5.$$

A, B, C are the semi-axes of the related inertia-equivalent ellipsoid, which is described by

$$x^2/A^2 + y^2/B^2 + z^2/C^2 = 1. \quad (12)$$

The eigenvectors determine the orientation of the ellipsoid relative to the molecule. The centre of the ellipsoid is localized in the mass centre of the solvent-excluded body. For solution of the eigenvalue problem the Fortran77 program *INERTCUBE* was developed.

4.2 *Structural characteristics of the IEE.* The IEE has been calculated according to equations (4)–(12) from the solvent-excluded molecular bodies of the 29 proteins contained in Table 1. The length of the half-axes and the volumes of the IEEs are summarized in Table 2. The result for the molecule

leghemoglobin (1LH4) is shown in Fig. 4. By comparison of the directly calculated molecular parameters – radius of gyration, largest diameter and volume (Table 1) – with the corresponding parameters of the IEE, it follows that

- (i) the radii of gyration of the homogeneous solvent-excluded volume and of the IEE are identical,
- (ii) the largest diameters of the molecules are about 5–15% (~ 0.2 – 0.8 nm) larger than the maximum axes of the IEEs,
- (iii) the volume of the IEE is 20 to 30% larger than the molecular dry volume and
- (iv) 12 to 14% of the molecular volume is outside the IEE, like continents above the sea level on the earth.

The IEE is defined for each structure and can be deduced from each scattering curve as will be shown below, but it is not necessarily the best existing homogeneous low-resolution shape model. A heuristic criterion is that IEEs whose volumes exceed the molecular dry volumes by more than 30% do not describe the low-resolution shape very well. For ribonuclease (1RN3), immunoglobulin Fab (1FB4) and human carbonic anhydrase B (2CAB) and C (1CAC), the IEE should not be used as a shape model. In the case of a difference larger than 30%, the molecules are mostly composed of clearly distinguishable domains or subunits and the IEE encloses clefts and holes. In Fig. 5 the approximation of the two-domain molecule calmodulin (3CLN) by the IEE is demonstrated. The molecule has not been included in the random sample set of proteins in Table 1 because the heuristic criterion for an acceptable model is strongly violated. The volume of the IEE is

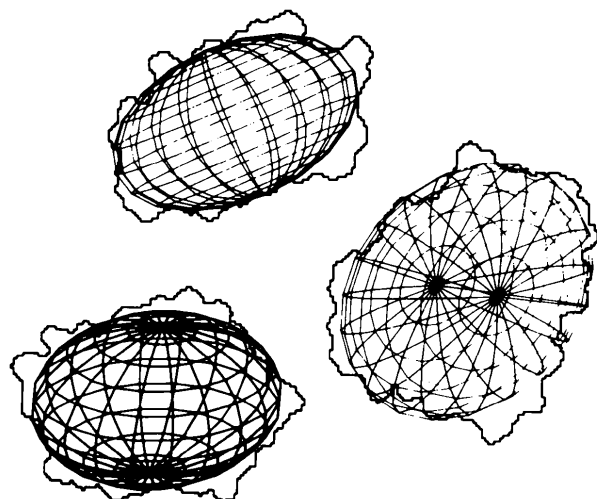


Fig. 4. Projections of the highly resolved shape of the molecule leghemoglobin (1LH4) and the corresponding views of the inertia-equivalent ellipsoid (IEE). The IEE is directly calculated from the solvent-excluded-volume body of the molecule.

in this case twice as big as the dry volume and the ellipsoid should be replaced by a model with more degrees of freedom. The IEE for the molecule leghemoglobin (1LH4) is an adequate model for the real shape and its scattering curve fits that of the protein shape in the region of the first three sampling points (Fig. 6). Consequently, the axes of the IEE agree approximately with the corresponding axes of the SSEE; the smallest axis of the SSEE is about 10% smaller than the corresponding axis of the IEE (Tables 1 and 2). In absolute units this difference amounts to 0.2 nm. The agreement between the axes of the IEE and the SSEE of the other molecules in the random sample is also excellent. The mean relative difference between the lengths of the two major axes is about 2 to 5% (~ 0.05 – 0.3 nm). The minor axis of each IEE is 10 to 20% (~ 0.2 – 0.6 nm) larger than the corresponding axis of the SSEE.

4.3. *Determination of the IEE from the solution scattering curve.* From the theoretical (or experimental) scattering curve of a molecule in solution without contrast variation only a two-parametric homogeneous shape model may be deduced, *e.g.* an ellipsoid of revolution (SEE) with distorted dimensions. This SEE cannot simply be reduced with constant axial ratios to approximate the IEE. To estimate the three-parameter inertia-equivalent ellipsoid of the shape from the first two sampling points

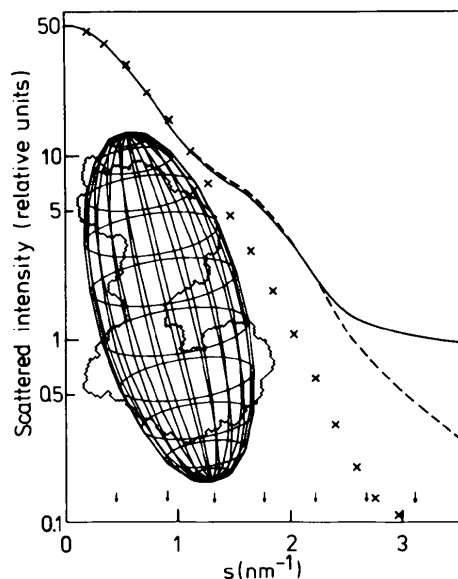


Fig. 5. Theoretical small-angle X-ray scattering curves of the protein calmodulin (3CLN) and of the IEE. — Scattering curve of the molecule in aqueous solution; --- scattering curve of the homogeneously filled body of the solvent-excluded volume; $\times \times \times$ scattering curve of the inertia-equivalent ellipsoid. \downarrow marks the sampling points. Insert: shape projection and IEE. The IEE is directly calculated from the solvent-excluded-volume body; $A = 4.39$, $B = 1.75$, $C = 1.29$ nm.

of the solution scattering curve, additional information is needed. This information can be retrieved from the crystal structure and from the theoretically calculated scattering curves of selected proteins.

4.3.1. *Calibration of ellipsoids on the basis of structure families.* The protein molecules are first allocated to families which exhibit similar structural features, such as similar molecular mass or volume, similar parts of secondary structures or an identical number of subunits or domains. The relative scattering contribution of the inner electron density inhomogeneities are then comparable for molecules belonging to such a group and should be correctable. The investigated proteins (Tables 1, 2 and 3) have been allocated to a group of oxygen-binding proteins, a serine protease family and a functionally nonrelated heterogeneous group of proteins. All of the oxygen-binding proteins are one-domain proteins of molecular mass between 16 000 and 19 000 daltons with an α -helical content of nearly 80%. The molecular masses of the proteases are around 23 000 daltons. The α -helical content is about 25% and the β -pleated sheet contribution to the H bridges is 35% on average. Only the proteinase B (3SGB) shows a slightly bisected form; the others are quite compact. The third group is much more heterogeneous in molecular mass, ranging from 11 500 to 35 000 daltons and are pure α -helical proteins (Ca-binding parvalbumin, 1CPV; cytochrome B562,

156B), β -proteins (concanavalin A, 2CNA; immunoglobulin Fab, 1FB4), or are of mixed type (human carbonic anhydrase B, 2CAB; ribonuclease A, 1RN3). The shape varies more than in the other two groups, from banana-like (1RN3; lysozyme, 5LYZ) and binodular (2CAB) to quite compact (cytochrome C, 1CYC; 1CPV).

For all proteins of the random sample the IEE has been calculated from the structure and the oblate and prolate SEEs have been estimated by curve fitting, using the angular region of two sampling points of the theoretically calculated scattering curves of the molecules in solution. For the molecules with differences of less than 30% between the volume of the IEE and the molecular dry volume, the relative differences between the corresponding axial dimensions of the IEE and SEE are calculated and the mean relative deviation is estimated within each group. For the serine protease family the major axis A of the prolate ellipsoid is 9% (~ 0.45 nm) larger than that of the IEE, on average, and the minor axes B and C of the prolate SEE are 8% (~ 0.3 nm) smaller and 10% (~ 0.4 nm) larger, respectively, than the corresponding IEE axes (Table 2). The small standard deviation for the correction factors given in Table 2 proves that the mean relative deviation between the axes of IEE and SEE is a systematic rather than a statistical difference, particularly for the prolate scattering-equivalent ellipsoids. The result is the same even for the proteins in the structurally heterogeneous third group in Table 2. The prolate ellipsoids are favoured over the oblate ones as basic models because the σ values of the correction terms are lower for the first model, particularly in the last group.

The second step is the correction of the systematic distortions of the axial dimensions of the prolate SEE. Table 3 contains the half-axial dimensions of the IEEs which are calculated by multiplication of the half-axial lengths of the prolate SEEs by the group- and molecule-specific factors (Table 2). The remaining standard deviation, calculated from the differences between the half-axes of the structure-based IEE and of the prolate SEE-based IEE is smaller than 5%. The comparison of these errors with those of the correction factors for prolate ellipsoid axes (Table 2) confirms the hypothesis of the statistical origin of the remaining errors.

In a computer experiment, the procedure is tested for molecules which do not belong to the random sample used in Tables 1, 2 and 3. The result is shown in Table 4. The molecules have been associated with the three different groups used in this paper for a structural classification. For sperm whale carbonmonoxymyoglobin (1MB5), deoxymyoglobin (1MBD), elastase (1EST), thermolysin (3TLN) and cytochrome C2 (3C2C), the spheroidal low-resolution shape model is acceptable, corresponding to the

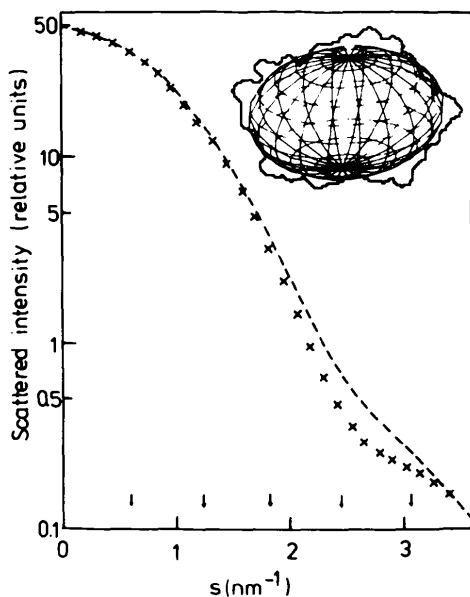


Fig. 6. Theoretical scattering curves of the solvent-excluded-volume body --- and of the inertia-equivalent ellipsoid $\times \times \times$ of leghemoglobin (1LH4). \downarrow marks the sampling points. Insert: shape projection and IEE. The IEE is directly calculated from the solvent-excluded-volume body, $A = 2.41$, $B = 1.90$, $C = 1.3$ nm.

Table 3. Geometrical semi-axes of the IEE, calculated from the oblate and prolate SEE by multiplication of the axes with the group specific factors $C_{X,Y}$ given in Table 2

σ is the percentage standard deviation of the difference between the length of semi-axes of the IEE calculated from the SEE and the semi-axes of the IEE calculated directly from the atomic structure. For the protein names see Table 1.

Molecule	IEE derived from oblate SEE			IEE derived from prolate SEE		
	A (nm)	B (nm)	C (nm)	A (nm)	B (nm)	C (nm)
I. Oxygen-binding proteins						
1LH4	2.40	2.01	1.25	2.42	1.95	1.30
1MBN	2.32	1.94	1.33	2.32	1.96	1.31
1MBO	2.33	1.95	1.36	2.31	1.99	1.34
2LHB	2.26	1.89	1.36	2.24	1.96	1.31
2HHB/A	2.21	1.85	1.38	2.18	1.95	1.30
2HHB/B	2.36	1.97	1.25	2.39	1.91	1.28
2MHB/B	2.36	1.97	1.24	2.38	1.92	1.29
σ (%)	1.9	4.7	3.3	2.4	5.0	3.2
II. Proteases						
1CHG	2.38	2.03	1.71	2.37	2.05	1.70
1TGN	2.42	2.07	1.58	2.45	1.98	1.64
1TPO	2.41	2.06	1.60	2.44	1.98	1.64
2CGA/B	2.44	2.09	1.84	2.40	2.16	1.79
2GCH	2.39	2.04	1.77	2.37	2.10	1.74
2PKA	2.46	2.10	1.72	2.47	2.09	1.73
3RP2	2.45	2.09	1.68	2.46	2.06	1.70
5CHA/B	2.45	2.09	1.63	2.45	2.05	1.70
2ALP	2.20	1.88	1.52	2.22	1.86	1.54
2SBT	2.46	2.10	1.90	2.42	2.21	1.83
2SGA	2.10	1.79	1.52	2.08	1.82	1.51
3SGB	2.19	1.87	1.44	2.22	1.80	1.49
σ (%)	3.5	4.1	4.2	2.9	2.2	4.4
III. Other proteins						
156B				2.52	1.36	1.14
1CYC	1.87	1.49	1.38	1.82	1.58	1.33
1CAC	2.66	2.12	2.00	2.61	2.26	1.90
1CPV	1.91	1.52	1.58	1.82	1.72	1.44
1FB4/L108	2.13	1.70	0.93	2.22	1.45	1.22
1RN3	2.27	1.82	1.05	2.36	1.56	1.32
2CAB	2.58	2.06	1.91	2.54	2.17	1.83
2CNA	2.76	2.21	1.54	2.87	2.00	1.68
5CPA	2.70	2.16	2.09	2.65	2.32	1.96
5LYZ	2.18	1.74	1.19	2.25	1.59	1.33
σ (%)	4.5	12.5	15.6	3.1	2.7	5.0

heuristic limit of 30% for the difference between the dry volume and the IEE volume. For the lysozyme-water complex (all water molecules indexed in the Protein Data Bank are included), superoxide dismutase (2SOD), phosphoglycerate kinase (3PGK) and calmodulin (3CLN), the inertia-equivalent ellipsoid has to be rejected as a low-resolution model. In reality, each of the latter three molecules consists of two clearly separated domains (for 3CLN see Fig. 5) and the adequate low-resolution structure models have more than three parameters. For all molecules, regardless of whether or not the ellipsoidal model has

been accepted, the differences between the axes of the IEE, indirectly deduced from the prolate ellipsoid of revolution and the corresponding axes of the IEE, calculated directly from the structure, are smaller than 10%, and usually smaller than 5%. For the oxygen-binding proteins, 1MB5 and 1MBD, and for the protease 1EST, allocation to their structure family rendered it possible to determine indirectly the axes of the IEEs with errors smaller than 6% (≤ 0.2 nm).

4.3.2 *General calibration of ellipsoids.* If an investigated protein has unknown structure and does not belong to a family with known crystal structure, the correction factors have to be calculated in a modified manner from the random sample discussed in Table 1. The correction factors are calculated from the differences between the corresponding axes of the SEE and the IEE of all proteins without allocating the molecules to structural families. The factors for the axes A , B and C are then 0.901(5), 1.104(10) and 0.880(11) to change the distorted axes of the scattering-equivalent prolate ellipsoid of revolution into the axes of the inertia-equivalent ellipsoid. The differences between the axes of the IEE, calculated from the atomic coordinates, and the IEE derived from the scattering curve by calibration are mostly larger than in the first family-related method. The largest differences between the sets of axes exist for the oxygen-binding proteins, mostly clear oblate molecules, which have been approximated by prolate ellipsoids in this algorithm. The maximum difference is about 10–13% (~ 0.3 – 0.4 nm) for the smallest axis. For the other axes and proteins the differences are less than 10% and usually less than 5% (not explicitly shown in the tables). The results of the computer experiment with molecules not in the random sample are listed in Table 4. Also for these molecules the differences between the corresponding axes of the structure-based and the scattering-based IEE are about 10% or less (≤ 0.6 nm for the largest axis, ≤ 0.5 nm for the middle and ≤ 0.4 nm for the smallest axis).

4.4. *Determination of the IEE from experimental scattering curve.* The indirect method for determination of the IEE dimensions from the scattering has been tested with experimental data. For sperm whale myoglobin the axes of the prolate ellipsoid of revolution deduced from the experimental solution scattering curve (Stuhrmann, 1973) are 2.84 and 1.83 nm. The inertia-equivalent ellipsoid, derived from these dimensions *via* the correction procedure using the factors given in Table 2 for the group of oxygen-binding proteins, is somewhat larger than the IEEs of crystalline myoglobins 1MBN and 1MBO. The length of the half-axes of the IEE in solution, $A = 2.50$, $B = 2.15$ and $C = 1.45$ nm (or by the general calibration $A = 2.56$, $B = 2.02$ and

Table 4. Computer experiment for molecules not belonging to the random sample in Table 3

Geometrical semi-axes of the IEE, calculated from the prolate SEE by multiplication of the axial lengths of the SEE by group-specific or general factors given in Table 2. (na: not accepted as shape model because of the 30%-limit violation.)

Molecule		Prolate SEE		IEE derived from prolate SEE by: (a) group-specific factors (b) general factors			IEE derived directly from the structure		
		A (nm)	B = C (nm)	A (nm)	B (nm)	C (nm)	A (nm)	B (nm)	C (nm)
I. Oxygen-binding proteins									
Myoglobin-CO	1MB5	2.579	1.660	a 2.272	1.955	1.313	2.303	2.073	1.31
				b 2.324	1.833	1.461			
Myoglobin (deoxy)	1MBD	2.611	1.679	a 2.300	1.978	1.328	2.332	2.057	1.325
				b 2.353	1.854	1.477			
II. Proteases									
Elastase	1EST	2.721	1.893	a 2.479	2.044	1.707	2.495	2.055	1.808
				b 2.452	2.090	1.666			
Thermolysin	3TLN	3.612	1.948	a 3.291	2.104	1.757	3.455	2.056	1.699
				b 3.255	2.151	1.714			
III. Other proteins									
Lysozyme/water	1LZT	2.757	1.663	a 2.548	1.808	1.505	2.333	1.816	1.497 na
				b 2.484	1.836	1.463			
Superoxide dismutase	2SOD	3.868	1.774	a 3.574	1.928	1.605	3.602	1.939	1.746 na
				b 3.485	1.959	1.561			
Phosphoglycerate kinase	3PGK	4.506	2.180	a 4.164	2.369	1.973	4.428	2.309	2.039 na
				b 4.060	2.407	1.918			
Cytochrome C2	3C2C	2.082	1.356	a 2.006	1.474	1.227	2.034	1.558	1.324
				b 1.876	1.497	1.193			
Calmodulin	3CLN	4.513	1.510	a 4.170	1.641	1.367	4.388	1.745	1.291 na
				b 4.067	1.667	1.329			

$C = 1.61$ nm) support the hypothesis of a slight loosening of the structure (Stuhrmann, 1973) or a rearrangement of the large helices H/G (Fedorov & Denesyuk, 1978) induced by the crystal-solution transition.

For the crystallographically noncharacterized molecule of bovine serum albumin, the half-axes of the IEE have been determined from the experimental solution scattering curve (Damaschun, Müller & Pürschel, 1967) to be $A = 5.28$, $B = 2.6$ and $C = 2.19$ nm. The structural relevance of these dimensions is documented by a true prediction of hydrodynamic parameters of the molecule. The harmonic mean of the rotational diffusion correlation time of the macromolecule was calculated on the basis of this triaxial ellipsoid to be 49.5 ns. This value is in conformity with the 47.6–54.9 ns measured with fluorescence depolarization and dielectric relaxation experiments (Moser, Squire & O'Konski, 1966; Wahl, 1966; Steiner, 1953). The prediction algorithm has been discussed recently (Müller, 1991).

III. Concluding remarks

The structure-related inertia-equivalent ellipsoid of a molecule can be deduced from the solution scattering

curve of the polymer by the calibration method discussed. The errors in the axial lengths of the triaxial ellipsoid are about 5% if the molecule under investigation belongs to a structure family of molecules with known atomic structure. If such a relationship is unknown, the errors increase to 10–15%. For molecules with a volume of the inertia-equivalent ellipsoid no more than 30% larger than the molecular dry volume, the IEE is an adequate shape model and can be automatically estimated from the scattering curve without an experimental contrast-variation procedure. Otherwise, if the scattering contributions of the inner electron density fluctuations have been neglected or changed by subtraction of a constant term from the scattering intensity, distortions of the geometrical dimensions can be expected of up to 50% for proteins with molecular masses below 65 000 daltons.

The investigations will be continued to enlarge the basis for calibration. Both the number of structurally related protein families and that of the members in each group will be increased and the results will be stored in a data bank together with other structural and hydrodynamic parameters, scattered intensities and real-space functions.

The authors are grateful to J. M. Thornton for providing the program for calculation of an inertia ellipsoid from the atoms of a molecule.

References

- BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F., BRICE, M. D., RODGERS, J. R., KENNARD, O., SHIMANOCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535–542.
- DAMASCHUN, G., DAMASCHUN, H., DEMBO, A. T., KAYUSHINA, R. L., KRÖBER, R., MOSHKOW, K. A., MÜLLER, J. J., NEIFAKH, S. A., ROLBIN, J. A., SHAVLOVSKY, M. M. & ZIRWER, D. (1978). *Stud. Biophys.* **71**, 53–76.
- DAMASCHUN, G., KLEY, G., MÜLLER, J. J. & PÜRSCHEL, H. V. (1968). *Acta Biol. Med. Ger.* **20**, 409–412.
- DAMASCHUN, G., MÜLLER, J. J. & PÜRSCHEL, H.-V. (1967). *Biol. Rundsch.* **5**, 180–181.
- DAMASCHUN, G., MÜLLER, J. J. & PÜRSCHEL, H.-V. (1968). *Monatsh. Chem.* **99**, 2343–2348.
- DAMASCHUN, G. & PÜRSCHEL, H.-V. (1971). *Acta Cryst.* **A27**, 193–197.
- FEDOROV, A. A., PTITSYN, O. B. & VORONIN, L. A. (1972). *FEBS Lett.* **28**, 188–190.
- FEDOROV, B. A. & DENESYUK, A. I. (1978). *FEBS Lett.* **88**, 114–117.
- GLATTER, O. & KRATKY, O. (1982). *Small-Angle X-ray Scattering*. New York: Academic Press.
- KRATKY, O. (1963). *Progr. Biophys. Mol. Biol.* **13**, 105–173.
- MARQUARDT, D. W. (1963). *J. Soc. Ind. Appl. Math.* **2**, 431–469.
- MITTELBACH, P. & POROD, G. (1961). *Acta Phys. Austriaca*, **14**, 185–211, 405–439.
- MITTELBACH, P. & POROD, G. (1962). *Acta Phys. Austriaca*, **15**, 122–147.
- MOSER, P., SQUIRE, P. G. & O'KONSKI, C. T. (1966). *J. Phys. Chem.* **70**, 744–756.
- MÜLLER, J. J. (1983). *J. Appl. Cryst.* **16**, 74–82.
- MÜLLER, J. J. (1991). *Biopolymers*, **31**, 149–160.
- MÜLLER, J. J., DAMASCHUN, G. & HÜBNER, G. (1979). *Acta Biol. Med. Ger.* **38**, 1–10.
- MÜLLER, J. J., PAVLOV, M. YU. & FEDOROV, B. A. (1983). *Stud. Biophys.* **97**, 121–128.
- PILZ, I. (1982). In *Small-Angle X-ray Scattering*, edited by O. GLATTER & O. KRATKY. New York: Academic Press.
- REICH, J. G., WANGERMANN, G., FALCK, M. & ROHDE, K. (1972). *Eur. J. Biochem.* **26**, 368–379.
- RICHARDS, F. M. (1977). *Annu. Rev. Biophys. Bioeng.* **6**, 151–176.
- ROLBIN, J. A., FEIGIN, L. A. & SCEDRIN, B. M. (1971). *Appar. Metody Rentgenovskogo Anal.* **9**, 46–50.
- SOMMERFELD, A. (1962). *Mechanik*. Leipzig: Akademische Verlagsgesellschaft Geest & Portig K.-G.
- STEINER, R. G. (1953). *Arch. Biochem. Biophys.* **46**, 291–311.
- STUHRMANN, H. B. (1973). *J. Mol. Biol.* **77**, 363–369.
- STUHRMANN, H. B. & KIRSTE, R. G. (1965). *Z. Phys. Chem. (Frankfurt am Main)*, **46**, 247–250.
- STUHRMANN, H. B. & KIRSTE, R. G. (1967). *Z. Phys. Chem. (Frankfurt am Main)*, **56**, 334–337.
- TAYLOR, W. R., THORNTON, J. M. & TURNELL, W. G. (1983). *J. Mol. Graph.* **1** 30–38.
- WAHL, M. P. (1966). *C. R. Acad. Sci.* **263D**, 1525–1528.