Principles of Protein–Protein Association

Online at: https://doi.org/10.1088/2053-2563/ab19ba

About the Series

The Biophysical Society and IOP Publishing have forged a new publishing partnership in biophysics, bringing the world-leading expertise and domain knowledge of the Biophysical Society into the rapidly developing IOP ebooks program.

The program publishes textbooks, monographs, reviews, and handbooks covering all areas of biophysics research, applications, education, methods, computational tools, and techniques. Subjects of the collection will include: bioenergetics; bioengineering; biological fluorescence; biopolymers *in vivo*; cryo-electron microscopy; exocytosis and endocytosis; intrinsically disordered proteins; mechanobiology; membrane biophysics; membrane structure and assembly; molecular biophysics; motility and cytoskeleton; nanoscale biophysics; and permeation and transport.

Principles of Protein–Protein Association

Harold P Erickson
Department of Cell Biology, Duke University, Durham, North Carolina, USA

IOP Publishing, Bristol, UK

ISBN 978-0-7503-2412-0 (ebook)
ISBN 978-0-7503-2410-6 (print)
ISBN 978-0-7503-2411-3 (mobi)

DOI 10.1088/2053-2563/ab19ba

Version: 20190601

IOP Expanding Physics
ISSN 2053-2563 (online)
ISSN 2054-7315 (print)

British Library Cataloguing-in-Publication Data: A catalogue record for this book is available from the British Library.

Published by IOP Publishing, wholly owned by The Institute of Physics, London

IOP Publishing, Temple Circus, Temple Way, Bristol, BS1 6HG, UK

US Office: IOP Publishing, Inc., 190 North Independence Mall West, Suite 601, Philadelphia, PA 19106, USA

Contents

Preface

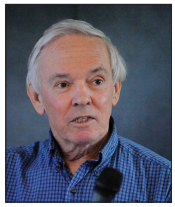
This volume originated from lectures I have been giving to graduate students. The students are mostly first and second year graduate students from the Duke University Program in Cell and Molecular Biology. I presume only a basic knowledge of biochemistry. I highly recommend that students review basic principles of protein structure prior to the course. Excellent sources are the texts: *Molecular Biology of the Cell*, by Alberts *et al*, chapter 3, 'Proteins'; or *Cell Biology* by Pollard and Earnshaw, chapter 2 'Molecular Structures.'

I also highly recommend that students download a protein structure viewer and use it to image on their own computer the structures displayed in the figures. Recommended viewers are Chimera, Pymol and KING.

In recent years the course has comprised six sessions of 80 minutes each, where I present background material and then lead discussion of the assigned papers. The chapters included here have evolved from my notes for these class sessions. These notes may be useful for faculty organizing similar classes, and/or for self-instruction of students and researchers who find a need to understand principles of protein–protein association.

Author biography

Harold P Erickson



Harold P Erickson received his PhD in Biophysics from Johns Hopkins University and did postdoctoral work in Cambridge, England. He joined the faculty at Duke University Medical Center in 1971, and is currently James B. Duke Professor in the Departments of Cell Biology, Biochemistry and Biomedical Engineering. His research has spanned two broad areas of cell biology: cytoskeleton (microtubules and the bacterial tubulin homolog, FtsZ); and extracellular matrix (fibrinogen, tenascin and fibronectin). He has contributed to several areas of electron microscopy (image processing, negative stain and rotary shadowing) and has theoretical work on the thermodynamics of cooperative assembly and diffusion-limited kinetics of protein–protein association.

IOP Publishing

Principles of Protein–Protein Association

Harold P Erickson

Chapter 1

Size and shape of protein molecules at the nm level determined by sedimentation, gel filtration and electron microscopy

This chapter was published in 2009 *(Biol. Proced. Online* 11:32–51*). It presents methods and calculations that are fundamentally important to determining the size and stoichiometry of protein complexes. It is reprinted here to have these resources readily available.*

An important part of characterizing any protein molecule is to determine its size and shape. Sedimentation and gel filtration are hydrodynamic techniques that can be used for this medium resolution structural analysis. This review collects a number of simple calculations that are useful for thinking about protein structure at the nm level. Readers are reminded that the Perrin equation is generally not a valid approach to determine the shape of proteins. Instead, a simple guideline is presented, based on the measured sedimentation coefficient and a calculated maximum *S*, to estimate if a protein is globular or elongated. It is recalled that a gel filtration column fractionates proteins on the basis of their Stokes radius, not molecular weight. The molecular weight can be determined by combining gradient sedimentation and gel filtration, techniques available in most biochemistry laboratories, as originally proposed by Siegel and Monte. Finally, rotary shadowing and negative stain electron microscopy are powerful techniques for resolving the size and shape of single protein molecules and complexes at the nm level. A combination of hydrodynamics and electron microscopy is especially powerful.

1.1 Introduction

Most proteins fold into globular domains. Protein folding is driven largely by the hydrophobic effect, which seeks to minimize contact of the polypeptide with solvent. Most proteins fold into globular domains, which have a minimal surface area.

doi:10.1088/2053-2563/ab19bach1

Peptides from 10–30 kDa typically fold into a single domain. Peptides larger than 50 kDa typically form two or more domains that are independently folded. However, some proteins are highly elongated, either as a string of small globular domains, or stabilized by specialized structures such as coiled coils or the collagen triple helix. The ultimate structural understanding of a protein comes from an atomic-level structure obtained by x-ray crystallography or NMR. However, structural information at the nm level is frequently invaluable. Hydrodynamics, in particular sedimentation and gel filtration, can provide this structural information, and it becomes even more powerful when combined with electron microscopy (EM).

One guiding principle enormously simplifies the analysis of protein structure. The interior of protein subunits and domains consists of closely packed atoms [1]. There are no substantial holes, and almost no water molecules in the protein interior. As a consequence of this, proteins are rigid structures, with a Young's modulus similar to that of Plexiglas [2]. Engineers sometimes categorize biology as the science of 'soft wet materials.' This is true of some hydrated gels, but proteins are better thought of as hard dry plastic. It is obviously important for all of biology, to have a rigid material with which to construct the machinery of life. A second consequence of the close-packed interior of proteins is that all proteins have approximately the same density, about 1.37 g cm^{-3}. For most of the following we will use the partial specific volume, v_2, which is the reciprocal of the density. v_2 varies from 0.70 to 0.76 for different proteins, and there is a literature on calculating or determining the value experimentally. For the present discussion we will ignore these variations and assume the average $v_2 = 0.73\ cm^3\ g^{-1}$.

1.2 How big is a protein molecule?

Assuming this partial specific volume ($v_2 = 0.73\ cm^3\ g^{-1}$), we can calculate the volume occupied by a protein of mass M in Da as follows.

$$\begin{aligned} V(nm^3) &= \frac{(0.73\ cm^3\ g^{-1}) \times (10^{21}\ nm^3\ cm^{-3})}{6.023 \times 10^{23}\ Da\ g^{-1}} \times M(Da) \\ &= 1.212 \times 10^{-3}(nm^3\ Da^{-1}) \times M(Da) \end{aligned} \tag{1.1}$$

The inverse relationship is also frequently useful: M (Da) = 825 V (nm^3).

What we really want is a physically intuitive parameter for the size of the protein. If we assume the protein has the simplest shape, a sphere, we can calculate its radius. We will refer to this as R_{min}, because it is the minimal radius of a sphere that could contain the given mass of protein

$$R_{min} = (3V/4\pi)^{1/3} = 0.066M^{1/3}\ (\text{For } M \text{ in Da}, R_{min} \text{ in nm}) \tag{1.2}$$

Some useful examples for proteins from 5000 to 500 000 Da are given in table 1.1.

It is important to emphasize that this is the minimum radius of a smooth sphere that could contain the given mass of protein. Since proteins have an irregular surface, even ones that are approximately spherical will have an average radius larger than the minimum.

Table 1.1. R_{min} for proteins of different mass.

Protein M (kDa)	5	10	20	50	100	200	500
R_{min} (nm)	1.1	1.42	1.78	2.4	3.05	3.84	5.21

Table 1.2. Distance between molecules as function of concentration.

Concentration	1 M	1 mM	1 μM	1 nM
Distance between molecules (nm)	1.18	11.8	118	1180

1.3 How far apart are molecules in solution?

It is frequently useful to know the average volume occupied by each molecule, or more directly, the average distance separating molecules in solution. This is a simple calculation based only on the molar concentration.

In a 1 M solution there are 6×10^{23} molecules/liter, = 0.6 molecules/nm^3, or inverting, the volume per molecule is $V = 1.66$ nm^3/molecule at 1 M. For a concentration C, the volume per molecule is $V = 1.66/C$.

We will take the cube root of the volume per molecule as an indication of the average separation.

$$d = V^{1/3} = 1.18/C^{1/3}, \tag{1.3}$$

where C is in molar, and d is in nm. Table 1.2 gives some typical values.

Two interesting examples are hemoglobin and fibrinogen. Hemoglobin is 330 mg ml^{-1} in erythrocytes, making its concentration 0.005 M. The average separation of molecules (center to center) is 6.9 nm. The diameter of a single hemoglobin molecule is about 5 nm. These molecules are very concentrated, near the highest physiological concentration of any protein (the crystallins in lens cells can be at >50% protein by weight).

Fibrinogen is a large, rod-shaped molecule that forms a fibrin blood clot when activated. It circulates in plasma at a concentration of around 2.5 g l^{-1}, about 9 μM. The fibringogen molecules are therefore about 60 nm apart, comparable to the 46 nm length of the rod-shaped molecule.

1.4 The sedimentation coefficient and frictional ratio. Is the protein globular or elongated?

Biochemists have long attempted to deduce the shape of a protein molecule from hydrodynamic parameters. There are two major hydrodynamic methods that are used to study protein molecules—sedimentation and diffusion (or gel filtration, which is the equivalent of measuring the diffusion coefficient).

The sedimentation coefficient, S, can be determined in an analytical ultracentrifuge. This was a standard part of the characterization of proteins in the 1940s and

1950s, and values of $S_{20,w}$ (sedimentation coefficient standardized to 20 °C in water) are collected in references such as the *CRC Handbook of Biochemistry* [3]. Today S is more frequently determined by zone sedimentation in a sucrose or glycerol gradient, by comparison to standard proteins of known S. 5%–20% sucrose gradients have been most frequently used, but we prefer 15%–40% glycerol gradients in 0.2 M ammonium bicarbonate, because this is the buffer used for rotary shadowing EM (section 1.8). The protein of interest is sedimented in one bucket of the swinging bucket rotor, and protein standards of known S (table 1.5) are sedimented in a separate (or sometimes the same) gradient. Following sedimentation, the gradient is eluted into fractions and each fraction is analyzed by SDS–PAGE to locate the standards and the test protein. Figure 1.1 shows an example determining the sedimentation coefficient of BsSMC (the SMC protein from *Bacillus subtilis*).

The sedimentation coefficient of a protein is a measure of how fast it moves through the gradient. Increasing the mass of the protein will increase its sedimentation, while increasing its size or asymmetry will decrease its sedimentation. The relationship of S to size and shape of the protein is given by the Svedberg formula:

$$S = M(1 - v_2\rho)/N_o f = M(1 - v_2\rho)/\ (N_o 6\pi\eta R_s) \tag{1.4}$$

M is the mass of the protein molecule in Da; N_o is Avogadro's number, 6.023×10^{23}; v_2 is the partial specific volume of the protein, typical value is 0.73 $cm^3\ g^{-1}$; ρ is the density of solvent (1.0 g cm^{-3} for H_2O); η is the viscosity of the solvent (0.01 g cm^{-1} for H_2O).

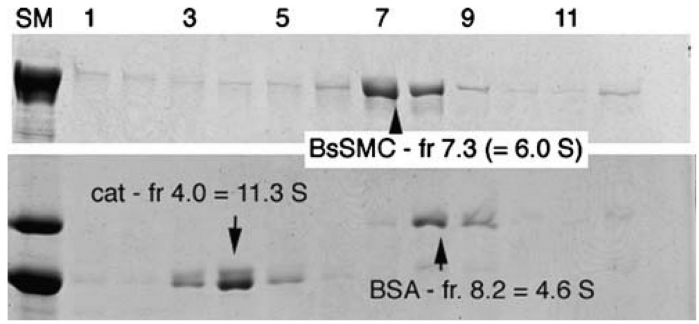


Figure 1.1. Glycerol gradient sedimentation analysis of SMC protein from *Bacillus subtilis* (BsSMC) (upper panel) and sedimentation standards catalase and bovine serum albumin (lower panel). A 200 µl sample was layered on a 5.0 ml gradient of 15%–40% glycerol in 0.2 M ammonium bicarbonate, and centrifuged in a Beckman SW55.1 swinging bucket rotor, 16 h, 38 000 rpm, 20 °C. 12 fractions of 400 µl each were collected from a hole in the bottom of the tube and each fraction was run on SDS–PAGE. Lane SM shows the starting material, and fraction 1 is the bottom of the gradient. The bottom panel shows that the 11.3 s catalase eluted precisely in fraction 4, while the 4.6 s BSA eluted mostly in fraction 8, with some in fraction 9. We estimated the BSA to be centered on fraction 8.2. Experiments with additional standard proteins have demonstrated that the 15%–40% glycerol gradients are linear over the range 3–20 s, so a linear interpolation is used to determine S of the unknown protein. BsSMC is in fractions 7 and 8, estimated more precisely at fraction 7.3. Extrapolating from the standards we determine a sedimentation coefficient of 6.0 s for BsSMC. Other experiments gave an average value of 6.3 s for BsSMC [19].

A critical factor in the equation is the frictional coefficient, f (dimensions g s^{-1}) which depends on both the size and shape of the protein. For a given mass of protein (or given volume), f will increase as the protein becomes elongated or asymmetrical. (f can be replaced by an equivalent expression containing R_s, the Stokes radius, to be discussed later.) S has the dimensions of time (seconds). For typical protein molecules S is in the range of 2–20 $\times$ 10^{-13} s, and the value 10^{-13} s is designated a Svedberg unit, S. Thus typical proteins have sedimentation coefficients of 2–20 s.

From the above definition of parameters it is clear that S depends on the solvent and temperature. In classical studies the solvent-dependent factors were eliminated and the sedimentation coefficient was extrapolated to the value it would have at 20 ° C in water (for which ρ and η are given above). This is referred to as $S_{20,w}$. In the present treatment we will be referring mostly to standard proteins that have already been characterized, or unknown ones that will be referenced to these in gradient sedimentation, so our use of S will always mean $S_{20,w}$.

A useful concept is the minimum value of f, which would obtain if the given mass of protein were packed into a smooth, unhydrated sphere. As we have discussed in section 1, the radius of this sphere will be $R_{\min} = 0.066\ M^{1/3}$ (equation (1.2)). In about 1850 G G Stokes calculated theoretically the frictional coefficient of a smooth sphere (note that the equation is similar to that for the Stokes radius, to be discussed later, but the parameters here are different):

$$f_{\min} = 6\pi\eta R_{\min} \tag{1.5}$$

We have now designated $f_{\min}$ as the minimal frictional coefficient for a protein of a given mass, which would obtain if the protein were a smooth sphere of radius $R_{\min}$.

The actual f of a protein will always be larger than $f_{\min}$ because of two things. First, the shape of the protein normally deviates from spherical, to be ellipsoidal or elongated; closely related to this is the fact that the surface of the protein is not smooth but rather rough on the scale of the water molecules it is traveling through. Second, all proteins are surrounded by a shell of bound water, 1–2 molecules thick, which is partially immobilized or frozen by contact with the protein. This water of hydration increases the effective size of the protein, and thus increases f.

1.4.1 The Perrin equation does not work for proteins

If one could determine the amount of water of hydration and factor this out, there would be hope that the remaining excess of f over $f_{\min}$ could be interpreted in terms of shape. Algorithms have been devised for estimating the amount of bound water from the amino acid sequence, but these generally do not distinguish between buried residues, which have no bound water, and surface residues which bind water. Some attempts have been made to base the estimate of bound water based on polar residues, which are mostly exposed on the surface. 0.3 g H_2O per g protein is a typical estimate, but in fact this kind of guess is almost useless for analyzing f.

In the older days, when there was some confidence in these estimates of bound water, physical chemists calculated a value called f_o, which was the frictional coefficient for a sphere that would contain the given protein, but enlarged by the

estimated shell of water. (Other authors use f_o to designate what we term f_{min} [3, 4]. We recommend using f_{min} to avoid ambiguity.) The measured f for proteins was almost always larger than f_o, suggesting that the protein was asymmetrical or elongated. A very popular analysis was to model the protein as an ellipsoid of revolution, and calculate the axial ratio from f/f_o, using an equation first developed by Perrin. This approach is detailed in most classical texts of physical biochemistry. In fact the Perrin analysis always overestimates the asymmetry of the proteins, typically by a factor of two to five. It should not be used for proteins.

The problem is illustrated by an early collaborative study of phosphofructokinase, in which the laboratory of James Lee did hydrodynamics and our laboratory did EM [5]. We found by EM that the tetrameric particles were approximately cylinders, 9 nm in diameter and 14 nm long. The shape was therefore like a rugby ball, with an axial ratio of 1.5 for a prolate ellipsoid of revolution. The Lee group measured the molecular weight and sedimentation coefficient, determined f and estimated water of hydration and f_o. They then used the Perrin equation to calculate the axial ratio. The ratio was five, which would suggest that the protein had the shape of a hot dog. The EM structure (which was later confirmed by x-ray crystallography) shows that the Perrin equation overestimated the axial ratio by a factor of 3.

Teller *et al* [6] summarized the situation: 'Frequently the axial ratios resulting from such treatment are absurd in light of the present knowledge of protein structure.' They explained that the major problem with the Perrin equation is that it treats the protein as a smooth ellipsoid, when in fact the surface of the protein is quite rough. Teller *et al* went on to show how the frictional coefficient can actually be derived from the known atomic structure of the protein, by modeling the surface of the protein as a shell of small beads of radius 1.4 Å. The shell coated the surface of the protein, modeling its rugosity, and increasing the size of the protein by the equivalent of a single layer of bound water. This analysis has been extended by Garcia de la Torre and colleagues [7].

1.4.2 Interpreting shape from $f/f_{min} = S_{max}/S$

If the Perrin equation is useless, is there some other way that shape can be interpreted from f? The answer is yes, at a semiquantitative level. We have discovered simple guidelines where the ratio f/f_{min} can provide a good indication of whether a protein is globular, somewhat elongated or very elongated.

Instead of proceeding with the classical ratio f/f_{min}, where f is in non-intuitive units, we will reformulate the analysis directly in terms of the sedimentation coefficient, which is the parameter actually measured. We will define a value S_{max} as the maximum possible sedimentation coefficient, corresponding to f_{min}. S_{max} is the S value that would be obtained if the protein were a smooth sphere with no bound water. These two ratios are equal: $f/f_{min} = S_{max}/S$. Combining equations (1.2), (1.4) and (1.5), we have

$$S_{max} = 10^{13} M(1 - v_2\rho)/N_o(6\pi\eta R_{min}) = M[2.378 \times 10^{-4}]/R_{min} \qquad (1.6a)$$

Table 1.3. S_{max} calculated for proteins of different mass.

Protein M_r (kDa)	10	25	50	100	200	500	1000
S_{max} Svedbergs	1.68	3.1	4.9	7.8	12.3	22.7	36.1

Table 1.4. S_{max}/S values for representative globular and elongated proteins.

Globular Protein Standards Dimensions are from pdb files					
Protein	Dimensions nm	Mass	S_{max}	S	S_{max}/S
Phosphofructokinase	14 × 9 × 9	345 400	17.77	12.2	1.46
Catalase	9.7 × 9.2 × 6.7	230 000	13.6	11.3	1.20
Serum albumin	7.5 × 6.5 × 4.0	66 400	5.9	4.6	1.29
Hemoglobin	6 × 5 × 5	64 000	5.78	4.4	1.32
Ovalbumin	7.0 × 3.6 × 3.0	43 000	4.43	3.5	1.27
FtsZ	4.8 × 4 × 3	40 300	4.26	3.4	1.25
Elongated Protein Standards—Tenascin fragments [27, 28]; heat repeat [29, 30]					
Protein	**Dimensions nm**	**Mass**	S_{max}	S	S_{max}/S
TNfn1–5	14.7 × 1.7 × 2.8	50 400	4.94	3.0	1.65
TNfn1–8	24.6 × 1.7 × 2.8	78 900	6.64	3.6	1.85
TNfnALL	47.9 × 1.7 × 2.8	148 000	10.1	4.3	2.36
PR65/A HEAT repeat	17.2 × 3.5 × 2.0	60 000	5.53	3.6	1.54
fibrinogen	46 × 3 × 6	390 000	19.3	7.9	2.44

$$S_{max} = 0.00361 M^{2/3} \tag{1.6b}$$

The leading factor of 10^{13} in (1.6*a*) converts S_{max} to Svedberg units. The numbers in brackets in (1.6*a*) are calculated using $v_2 = 0.73$ cm^3 g^{-1}, $\rho = 1.0$ g cm^{-3}, $\eta = 0.01$ g cm^{-1} s^{-1} $= 10^{-9}$ g nm^{-1} s^{-1}. The final expression, equation (1.6*b*) expresses S_{max} in Svedbergs for a protein of mass M in Daltons.

Some typical numerical values of S_{max} for proteins from 10 000 to 1 000 000 Da are given in table 1.3.

We have surveyed values of S_{max}/S for a variety of proteins of known structure. Table 1.4 presents S_{max}/S for a number of approximately globular proteins and for a range of elongated proteins, all of known dimensions. It turns out that S_{max}/S is an excellent predictor of the degree of asymmetry of a protein. From this survey of known proteins we can propose a number of general principals.

- No protein has $S_{max}/S = f/f_{min}$ smaller than ~1.2.
- For approximately globular proteins:
  S_{max}/S is typically between 1.2 and 1.3.
- For moderately elongated proteins:
  S_{max}/S is in the range of 1.5–1.9.

- For highly elongated proteins (tropomyosin, fibrinogen, extended fibronectin): S_{max}/S is in the range of 2.0–3.0.
- For very long, thread-like molecules like collagen, or huge extended molecules like the tenascin hexabrachion (not shown): S_{max}/S can range from 3 to 4 or more.

Apart from indicating the shape of a protein, S_{max}/S can often give valuable information about the oligomeric state, if one has some idea of the shape. For example, if one knows that the protein subunit is approximately globular (from EM for example), but finds $S_{max}/S = 2.1$, this would suggest that the protein in solution is a actually a dimer. On the other hand if one thinks a protein is a dimer, but finds $S_{max}/S < 1.0$ for the dimer mass, the protein is apparently sedimenting as a monomer.

The use of S_{max}/S to estimate protein shape has been described briefly in [8].

1.5 The Kirkwood/Bloomfield calculation

The understanding of how protein shape affects hydrodynamics is elegantly extended by an analysis originally developed by Kirkwood [9], and later extended by de la Torres and Bloomfield [10–12]. In its simplest application it calculates the sedimentation coefficient of a rigid oligomeric protein composed of subunits of known S and known spacing relative to each other. In more complex applications, a protein of any complex shape can be modeled as a set of non-overlapping spheres or beads. See Byron [13] for a comprehensive review of the principals and applications of hydrodynamic bead modeling of biological macromolecules.

The basis of the Kirkwood/Bloomfield analysis is to account for how each bead shields the others from the effect of solvent flow, and thereby determine the hydrodynamics of the ensemble from its component beads. Figure 1.2 shows a simple example of the bead modeling approach, and provides an instructive look at how size and shape affect sedimentation. There are several important conclusions.

- A rod of three beads has about a two-fold higher S than a single bead.
- S_{max}/S is 1.18 for the single bead (the effect of the assumed shell of water); 1.34 for the three-bead rod; 1.93 for the straight 11-bead rod. This is consistent with the principals given in section 3 for globular, somewhat elongated and very elongated particles.
- Bending the rod at 90° in the middle causes only a small increase in S. Bending it into a U-shape with the arms about one bead diameter apart increases S a bit more. Bending this same 11-bead structure more sharply, so the two arms are in contact, causes a substantial increase in S, from 5.05 to 5.58. The guiding principle is that folding affects S when one part of the molecule is brought close enough to another to shield it from water flow.

1.6 Gel filtration chromatography and the Stokes radius

'Gel filtration chromatography is widely used for determining protein molecular weight.' This quote from Sigma–Aldrich bulletin 891A is a widely held

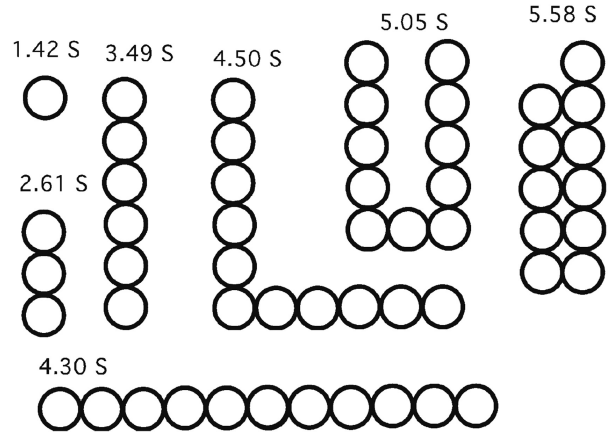


Figure 1.2. Each bead models a 10 kDa domain, with an assumed sedimentation coefficient of 1.42 s. The radius of the bead is 1.67 nm, using $R_{min} = 1.42$ nm, and adding 0.25 nm for a shell of water. The beads are an approximation to FN-III or Ig domains, which are ~1.7 × 2.8 × 3.5 nm. The sedimentation coefficients of multi-bead structures were calculated by the formula of Kirkwood/Bloomfield.

misconception. The fallacy is obscurely corrected by a later note in the bulletin that 'Once a calibration curve is prepared, the elution volume for a protein of similar shape, but unknown weight, can be used to determine the MW.' The key issue is 'of similar shape.' Generally the calibration proteins are all globular, and if the unknown protein is also globular the calibrated gel filtration column does give a good approximation of its molecular weight. The problem is that the shape of an unknown protein is generally unknown. If the unknown protein is elongated it can easily elute at a position twice the molecular weight of a globular protein.

The gel filtration column actually separates proteins not on their molecular weight, but on their frictional coefficient. Since the frictional coefficient, f, is not an intuitive parameter, it is usually replaced by the Stokes radius R_s. R_s is defined as the radius of a smooth sphere that would have the actual f of the protein. This is much more intuitive since it allows one to imagine a real sphere approximately the size of the protein, or somewhat larger if the protein is elongated and has bound water.

As mentioned above for equation (1.5), Stokes calculated theoretically the frictional coefficient of a smooth sphere to be:

$$f = 6\pi\eta R_s \tag{1.7}$$

The Stokes radius R_s is larger than R_{min} because it is the radius of a smooth sphere whose f would match the actual f of the protein. It accounts for both the asymmetry of the protein and the shell of bound water. More quantitatively, $f/f_{min} = S_{max}/S = R_s/R_{min}$.

Siegel and Monte [4] argued convincingly that the elution of proteins from a gel filtration column correlates closely with the Stokes radius, R_s, presenting experimental data from a wide range of globular and elongated proteins. The Stokes radius is known for large number of proteins, including ones convenient for

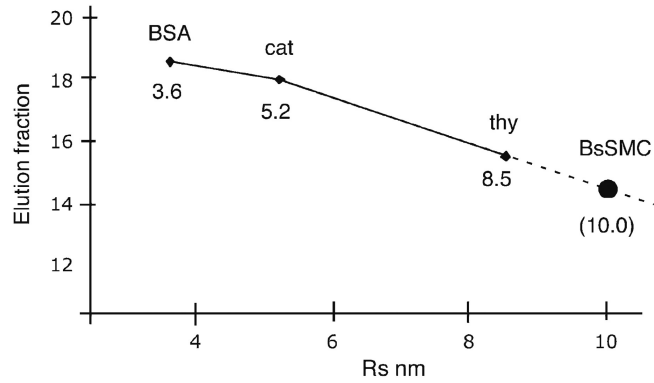


Figure 1.3. Determination of R_s of BsSMC by gel filtration. The column was calibrated by running standard proteins BSA, catalase and thyroglobulin over the column, then BsSMC. BsSMC eluted in fraction 14.2, which corresponds to an R_s of 10 nm on the extrapolated curve. In repeated experiments the average R_s was 10.3 nm [19].

calibrating gel filtration columns (table 1.5). Figure 1.3 shows an example where the R_s of the unknown protein BsSMC was determined by gel filtration.

The standard proteins should span R_s values above and below that of the protein of interest (but in the case of BsSMC a short extrapolation to a larger value was used). The literature generally recommends determining the void and included volumes of the column and plotting a partition coefficient K_{AV} [4]. However, we have found it generally satisfactory to simply plot elution position versus R_s for the standard proteins. This generally gives an approximately linear plot, but otherwise it is satisfactory to draw lines between the points and read the R_s of the protein of interest from its elution position on this standard curve.

A gel filtration column can determine R_s relative to the R_s of the standard calibration proteins. The R_s of these standards was generally determined from experimentally measured diffusion coefficients. Some tabulations of hydrodynamic data list the diffusion coefficient, D, rather than R_s, so it is worth knowing the relationship:

$$D = kT/f = kT/(6\pi\eta R_s) \tag{1.8}$$

where $k = 1.38 \times 10^{-16}$ g cm^2 s^{-2} K^{-1} is Boltzman's constant and T is the absolute temperature. k is given here in cgs units because D is typically expressed in cgs; R_s will be expressed in cm in this equation. Typical proteins have D in the range of 10^{-6}–10^{-7} cm^2 s^{-1}. Converting to nm and for $T = 300$ K and $\eta = 0.01$:

$$R_s = (1/D)\ 2.2 \times 10^{-6}, \tag{1.9}$$

where R_s is in nm and D is in cm^2 s^{-1}.

Simply knowing R_s is not very valuable in itself, except for estimating the degree of asymmetry, but this would be the same analysis developed above for S_{max}/S. However, if one determines both R_s and S, this permits a direct determination of

molecular weight, which cannot be deduced from either one alone. This is described in the next section.

1.7 Determining the molecular weight of a protein molecule—combining S and R_s à la Siegel and Monte

With the completion of multiple genomes and increasingly good annotation, the primary sequence of almost any protein can be found in the databases. The molecular weight of every protein subunit is therefore known from its sequence. But an experimental measure is still needed to determine if the native protein in solution is a monomer, dimer or oligomer, or if it forms a complex with other proteins. If one has a purified protein the molecular weight can be determined quite accurately by sedimentation equilibrium in the analytical ultracentrifuge. This technique has made a strong comeback with the introduction of the Beckman XL-A analytical ultracentrifuge. There are a number of good reviews [14, 15], and the documentation and programs that come with the centrifuge are very instructive.

What if one does not have an XL-A centrifuge, or the protein of interest is not purified? In 1966, Siegel and Monte [4] proposed a method that achieves the results of sedimentation equilibrium, with two enormous advantages. First, it requires only a preparative ultracentrifuge for sucrose or glycerol gradient sedimentation, and a gel filtration column. This equipment is available in most biochemistry laboratories. Second, the protein of interest need not be purified; one needs only an activity or an

Table 1.5. Standards for hydrodynamic analysis.

Protein	M_r aa seq	$S_{20,w}$	Smax/S	R_s(nm)	Source	M_r S–M
Ribonuclease A beef pancreas	14 044	2.0[a]	1.05[a]	1.64	HBC	13 791
Chymotrypsinogen Abeef pancreas	25 665	2.6	1.21	2.09	HBC	22 849
Ovalbumin hen egg	42 910 s	3.5	1.27	3.05	HBC	44 888
Albumin beef serum	69 322	4.6[a]	1.33	3.55	S–M,HBC	68 667
Aldolase rabbit muscle	157 368	7.3	1.45	4.81	HBC	147 650
Catalase beef liver	239 656	11.3	1.21	5.2	S–M	247 085
Apo-ferritin horse spleen	489 324	17.6	1.28	6.1	HBC	451 449
Thyroglobulin bovine	606 444	19	1.37	8.5	HBC	679 107
Fibrinogen, human	387 344	7.9	2.44	10.7	S–M	355 449

Gel filtration calibration kits, containing globular proteins of known molecular weight and R_s, are commercially available (GE healthcare, Sigma–Aldrich). These same proteins can be used for sedimentation standards. The proteins in these kits are included in the table along with some others that we have found useful. The values for M_r given in the first column are from amino acid sequence data. Values for $S_{20,w}$ and R_s are from the Siegel–Monte paper (indicated S–M under source), or the *CRC Handbook of Biochemistry* [3] (indicated HBC). They agree with the values listed for R_s in the GE Healthcare gel filtration calibration kit, with the exception of ferritin. The 'M_r *calc*' in the last column was obtained by our simplification of the Siegel–Monte calculation ($M = 4205\ s\ R_s$). Note that the worst disagreement with 'M_r *aa seq*' is about 10%.

[a] S for ribonuclease A is questionable because of the low S_{max}/S (1.05). S values for bovine serum albumin vary in the literature from 4.3 to 4.9. Many sources use 4.3, but we find that 4.6 gives a better fit with other standards (note that the standard curve in figure 1.5 used 4.3, but 4.6 would have placed it closer to the line).

antibody to locate it in the fractions. This is a very powerful technique, and should be in the repertoire of every protein biochemist.

The methodology is very simple. The protein is run over a calibrated gel filtration column to determine R_s, and hence f. Separately the protein is centrifuged through a glycerol or sucrose gradient to determine S. One then uses the Svedberg equation (equation (1.4)) to obtain M as a function of R_s and S.

$$M = SN_o(6\pi\eta R_s)/\ (1 - v_2\rho) \tag{1.10a}$$

setting $\eta = 0.01$, $v_2\rho = 0.73$, converting S to Svedberg units and R_s to nm, we can simplify further:

$$M = 4205\ (SR_s) \tag{1.10b}$$

where S is in Svedberg units, R_s is in nm and M is in Daltons.

This is pretty simple! Importantly, in typical applications this method gives the protein mass within about ±10%. This is more than enough precision to distinguish between monomer, dimer or trimer.

Application to BsSMC. In the sections above we showed how S of the SMC protein from *B. subtilis* was determined to be 6.3 Svedberg units from glycerol gradient sedimentation, and R_s was 10.3 nm, from gel filtration. Putting these values in equation (1.10*b*) we find that the molecular weight of BsSMC is 273 000 Da. From the amino acid sequence we know that the molecular weight of one BsSMC subunit is 135 000 Da. The Siegel–Monte analysis finds that the BsSMC molecule is a dimer.

Knowing that BsSMC is a dimer with molecular weight 270 000 Da, we can now determine its S_{max}/S. S_{max} is 15.1 (equation (1.6*b*)) so S_{max}/S is 2.4. The BsSMC molecule is thus expected to be highly elongated. EM (see below) confirmed this prediction.

1.8 Electron microscopy of protein molecules

Since the early 1980s electron microscopy has become a powerful technique for determining the size and shape of single protein molecules, especially ones larger than 100 kDa. Two techniques available in most EM laboratories, rotary shadowing and negative stain, can be used for imaging single molecules. Cryo EM is becoming a powerful tool for protein structural analysis, but it requires special equipment and expertise. For a large number of applications rotary shadowing and negative stain provide the essential structural information.

For rotary shadowing a dilute solution of protein is sprayed on mica, the liquid is evaporated in a high vacuum, and platinum metal is evaporated onto the mica at a shallow angle. The mica is rotated during this process, so the platinum builds up on all sides of the protein molecules. The first EM images of single protein molecules were obtained by Hall and Slayter using rotary shadowing [16]. Their images of fibrinogen showed a distinctive trinodular rod. However, rotary shadowing fell into

disfavor because the images were difficult to reproduce. Protein tended to aggregate and collect salt, rather than spread as single molecules. In 1976 James Pullman, then a graduate student at the University of Chicago, devised a protocol with one simple but crucial modification—he added 30% glycerol to the protein solution. For reasons that are still not understood, the glycerol greatly helps the spreading of the protein as single molecules.

Pullman never published his protocol, but two labs saw his mimeographed notes and tested out the effect of glycerol, as a part of their own attempts to improve rotary shadowing [17, 18]. They obtained reproducible and compelling images of fibrinogen (the first since the original Hall and Slayter study, and confirming the trinodular rod structure) and spectrin (the first ever images of this large protein). The technique has since been used in characterizing hundreds of protein molecules.

Figure 1.4 shows rotary shadowed BsSMC, fibrinogen and hexabrachion (tenascin). BsSMC is highly elongated, consistent with its high S_{max}/S discussed above [19]. The fibrinogen molecules show the trinodular rod, but these images also resolved a small fourth nodule next to the central nodule [20], not seen in earlier studies. The central nodule is about 50 kDa, and the smaller fourth nodule is about 20 kDa. The 'hexabrachion' tenascin molecule [21] illustrates the power of rotary shadowing at two extremes. First, the molecule is huge. Each of its six arms is made up of ~30 repeating small domains, totaling ~200 000 Da. At the larger scale the EM shows that each arm is an extended structure, matching the length expected if the repeating domains are an extended string of beads. At the finer scale, the EM can distinguish the different sized domains. The inner segment of each arm is a string of 3.5 kDa EGF domains, seen here as a thinner segment. A string of 10 kDa FN-III domains is clearly distinguished as a thicker outer segment. The terminal knob is a single, 22 kDa fibrinogen domain. The R_{min} of these domains are 0.8, 1.7 and 2.8 nm, and these can be distinguished by rotary shadowing. Rotary shadowing EM can visualize single globular domains as small as 10 kDa (3.5 nm diameter), and elongated molecules as thin as 1.5 nm (collagen).

Negative stain is another EM technique capable of imaging single protein molecules. It is especially useful for imaging larger molecules with a complex internal structure, which appear only as a large blob in rotary shadowing. Importantly, non-covalent protein–protein bonds are sometimes disrupted in the rotary shadowing technique [8], but uranyl acetate, in addition to providing high

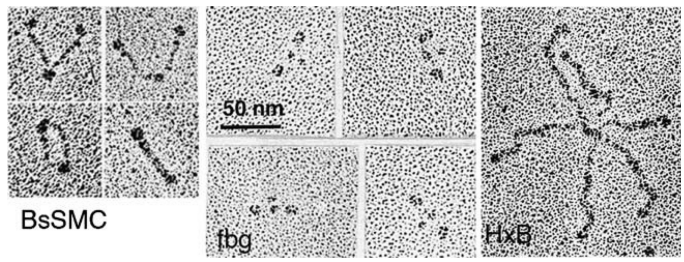


Figure 1.4. Rotary shadowing EM of three highly elongated protein molecules: the SMC protein from *B. subtilis* [19], fibrinogen [20], and the hexabrachion protein, tenascin [21]. Reprinted with permission from the indicated references.

resolution contrast, fixes oligomeric protein structures in a few milliseconds [22]. An excellent review of modern techniques of negative staining, with comparison to cryo EM, is given in [23].

The simple picture of the molecule produced by EM is frequently the most straightforward and satisfying structural analysis at the 1–2 nm resolution. When the structure is confirmed by hydrodynamic analysis the interpretation is even more compelling.

1.9 Hydrodynamic analysis and EM applied to large multi-subunit complexes

The text box above showed the application of the Siegel–Monte analysis to BsSMC, which had only one type subunit and was found to be a dimer. Similar hydrodynamic analysis can be used to analyze multi-subunit protein complexes. There are many examples in the literature; I will show here an elegant application to DASH/Dam1.

The protein complex called DASH or Dam1 is involved in attaching chromosomal kinetochores to microtubules in yeast. DASH/Dam1 is a complex of ten proteins that assemble into a particle containing one copy of each subunit. These complexes further assemble into rings that can form a sliding washer on the microtubule [24, 25]. The basic ten-subunit complex has been purified from yeast, and has also been expressed in *E. coli* and purified (this required the heroic effort of expressing all ten proteins simultaneously [24]). Figure 1.5 shows the hydrodynamic characterization of the purified protein complex, and illustrates several important features.

- For both the gel filtration (size exclusion chromatography, figure 1.5(a)) and gradient sedimentation, figure 1.5(b), two calibration curves of known protein standards are shown, green and black. These are independent calibration runs. In this study the gel filtration column was calibrated in terms of the reciprocal diffusion coefficient, $1/D$, which is proportional to R_s (equation (1.7)).
- The fractions were analyzed by western blot for the location of two proteins of the complex, Spc34p and Hsk3p. Methods notes that 1 ml fractions from gel filtration were precipitated with perchloric acid and rinsed with acetone prior to SDS–PAGE, an essential amplification for the dilute samples of yeast cytoplasmic extract. These two proteins eluted together in both gel filtration and sedimentation, consistent with their being part of the same complex.
- The profiles of the two proteins were identical when analyzed in their native form in yeast cytoplasmic extract, and as the purified complex expressed in *E. coli*. This is strong evidence that the expression protein is correctly folded and assembled.
- There is minimal trailing of any subunits. This means that there is no significant dissociation during the tens of minutes for the gel filtration, or the 12 h centrifugation. The complex is held together by very high affinity bonds, making it essentially irreversible.
- Combining the $R_s = 7.6$ nm from $1/D = 0.35 \times 10^{-7}$, and $S = 7.4$, equation (1.10b) gives a mass of $M = 236$ kDa, close to the 204 kDa obtained from adding the mass of the ten subunits. S_{max} is 12.6 giving $S_{max}/S = 1.7$, suggesting a moderately elongated protein.

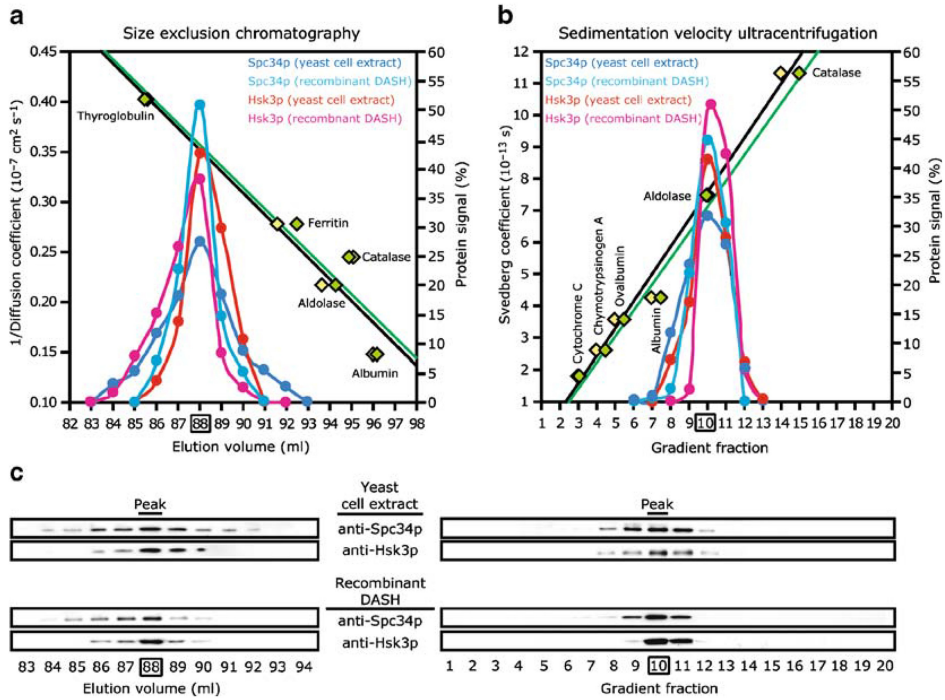


Figure 1.5. Hydrodynamic analysis of the DASH/Dam1 complex. Gel filtration is shown in a and sucrose gradient sedimentaion in b. Independent calibration curves using standard proteins are shown in black and green. Dark and light blue show Spc34p in yeast cytoplasmic extract and in the purified recombinant protein. Red and purple show Hsk3p. The proteins were identified and quantitated by western blot of the fractions, shown in c. The four protein bands eluted together at $1/D = 0.35 \times 10^7$, corresponding to $R_s = 7.6$ nm, and at 7.4 *S*. Reproduced from Miranda *et al* [24] with permission from Springer Nature.

Figure 1.6 shows EM images of DASH/DAM1 by rotary shadowing (figure 1.6a) and negative stain (figure 1.6b). Rotary shadowing showed irregular particles about 13 nm in diameter [24]. The particles had variable and frequently elongated shapes, but internal structure could not be resolved. A later study used state of the art negative staining and sophisticated computer programs to sort images into classes and average them [26]. These images resolved a complex internal structure. The negative stain showed most of the particles (80%) to be dimers, with 15% monomers and 5% trimers. This contradicts the hydrodynamic analysis of Miranda *et al* [24] showing that the particles were monomers. The reason for this discrepancy is not known.

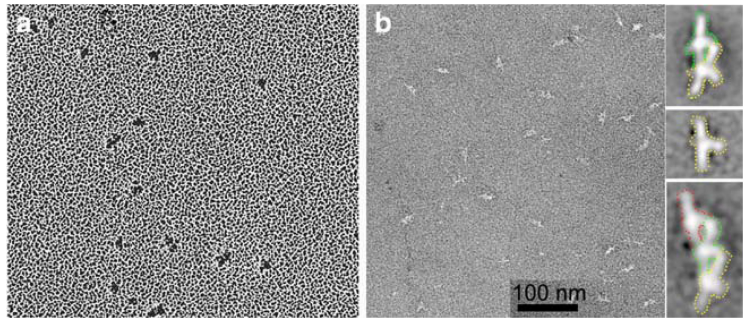


Figure 1.6. EM of DASH/Dam1. (a) Rotary shadowing shows particles roughly 13 nm in size, with irregular shape. (b) State of the art negative stain coupled with single particle averaging shows a complex internal structure of the elongated particles. The scale bar indicates 100 nm for the unprocessed images. The averaged images on the right show a monomer, dimer and trimer. These panels are 14 nm wide. The dimer was the predominant species. Left panel (rotaty shadowing) from Miranda *et al* [24] reprinted with permission from Springer Nature. Right panels (negative stain) reprinted with permission of Wang *et al* [26].

Acknowledgement: Supported by NIH grant CA47056.

References

[1] Richards F M 1974 The interpretation of protein structures: total volume, group volume distributions and packing density *J. Mol. Biol.* **82** 1–14
[2] Gittes F, Mickey B, Nettleton J and Howard J 1993 Flexural rigidity of microtubules and actin filaments measured from thermal fluctuations in shape *J. Cell Biol.* **120** 923–34
[3] Sober H A 1966 *Handbook of Biochemistry* (Cleveland, OH: The Chemical Rubber Co.)
[4] Siegel L M and Monte K J 1966 Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation *Biochim. Biophys. Acta* **112** 346–62
[5] Hesterberg L K, Lee J C and Erickson H P 1981 Structural properties of an active form of rabbit muscle phosphofructokinase *J. Biol. Chem.* **256** 9724–30
[6] Teller D C, Swanson E and De Haen C 1979 The translational friction coefficient of proteins *Methods Enzymol.* **61** 103–24
[7] Garcia De La Torre J, Huertas M L and Carrasco B 2000 Calculation of hydrodynamic properties of globular proteins from their atomic-level structure *Biophys. J.* **78** 719–30
[8] Schürmann G, Haspel J, Grumet M and Erickson H P 2001 Cell adhesion molecule L1 in folded (horseshoe) and extended conformations *Mol. Biol. Cell* **12** 1765–73
[9] Kirkwood J G 1954 The general theory of irreversible processes in solutions of macromolecules *J. Polymer Sci.* **12** 1–14
[10] Bloomfield V, Dalton W O and van Holde K E 1967 Frictional coefficients of multisubunit structures. I. Theory *Biopolymers* **5** 135–48
[11] Carrasco B and Garcia de la Torre J 1999 Hydrodynamic properties of rigid particles: comparison of different modeling and computational procedures *Biophys. J.* **76** 3044–57
[12] Garcia de la Torre J, Llorca O, Carrascosa J L and Valpuesta J M 2001 HYDROMIC: prediction of hydrodynamic properties of rigid macromolecular structures obtained from electron microscopy images *Eur. Biophys. J.* **30** 457–62

[13] Byron O 2008 Hydrodynamic modeling: the solution conformation of macromolecules and their complexes *Methods Cell. Biol.* **84** 327–73
[14] Schuster T M and Toedt J M 1996 New revolutions in the evolution of analytical ultracentrifugation *Curr. Opin. Struct. Biol.* **6** 650–58
[15] Hansen J C, Lebowitz J and Demeler B 1994 Analytical ultracentrifugation of complex macromolecular systems [Review] *Biochemistry* **33** 13155–63
[16] Hall C E and Slayter H S 1959 The fibrinogen molecule: its size, shape, and mode of polymerization *J. Biophys. Biochem. Cytol.* **5** 11–6
[17] Fowler W E and Erickson H P 1979 Trinodular structure of fibrinogen. Confirmation by both shadowing and negative stain electron microscopy *J. Mol. Biol.* **134** 241–49
[18] Shotton D M, Burke B E and Branton D 1979 The molecular structure of human erythrocyte spectrin *J. Mol. Biol.* **131** 303–29
[19] Melby T E, Ciampaglio C N, Briscoe G and Erickson H P 1998 The symmetrical structure of structural maintenance of chromosomes (SMC) and MukB proteins: long, antiparallel coiled coils, folded at a flexible hinge *J. Cell Biol.* **142** 1595–604
[20] Erickson H P and Fowler W E 1983 Electron microscopy of fibrinogen, its plasmic fragments and small polymers *Ann. N. Y. Acad. Sci.* **408** 146–63
[21] Erickson H P and Iglesias J L 1984 A six-armed oligomer isolated from cell surface fibronectin preparations *Nature* **311** 267–69
[22] Zhao F Q and Craig R 2003 Capturing time-resolved changes in molecular structure by negative staining *J. Struct. Biol.* **141** 43–52
[23] Ohi M, Li Y, Cheng Y and Walz T 2004 Negative staining and image classification—powerful tools in modern electron microscopy *Biol. Proced. Online* **6** 23–34
[24] Miranda J J, De Wulf P, Sorger P K and Harrison S C 2005 The yeast DASH complex forms closed rings on microtubules *Nat. Struct. Mol. Biol.* **12** 138–43
[25] Westermann S, Avila-Sakar A, Wang H W, Niederstrasser H, Wong J, Drubin D G, Nogales E and Barnes G 2005 Formation of a dynamic kinetochore–microtubule interface through assembly of the Dam1 ring complex *Mol. Cell* **17** 277–90
[26] Wang H W, Ramey V H, Westermann S, Leschziner A E, Welburn J P, Nakajima Y, Drubin D G, Barnes G and Nogales E 2007 Architecture of the Dam1 kinetochore ring complex and implications for microtubule-driven assembly and force-coupling mechanisms *Nat. Struct. Mol. Biol.* **14** 721–26
[27] Aukhil I, Joshi P, Yan Y and Erickson H P 1993 Cell- and heparin-binding domains of the hexabrachion arm identified by tenascin expression proteins *J. Biol. Chem.* **268** 2542–53
[28] Leahy D J, Axel R and Hendrickson W A 1992 Crystal structure of a soluble form of the human T cell coreceptor CD8 at 2.6 Å resolution *Cell* **68** 1145–62
[29] Chen S C, Kramer G and Hardesty B 1989 Isolation and partial characterization of an M_r 60,000 subunit of a type 2A phosphatase from rabbit reticulocytes *J. Biol. Chem.* **264** 7267–75
[30] Groves M R, Hanlon N, Turowski P, Hemmings B A and Barford D 1999 The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs *Cell* **96** 99–110

Full list of references

Chapter 1

[1] Richards F M 1974 The interpretation of protein structures: total volume, group volume distributions and packing density *J. Mol. Biol.* **82** 1–14

[2] Gittes F, Mickey B, Nettleton J and Howard J 1993 Flexural rigidity of microtubules and actin filaments measured from thermal fluctuations in shape *J. Cell Biol.* **120** 923–34

[3] Sober H A 1966 *Handbook of Biochemistry* (Cleveland, OH: The Chemical Rubber Co.)

[4] Siegel L M and Monte K J 1966 Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation *Biochim. Biophys. Acta* **112** 346–62

[5] Hesterberg L K, Lee J C and Erickson H P 1981 Structural properties of an active form of rabbit muscle phosphofructokinase *J. Biol. Chem.* **256** 9724–30

[6] Teller D C, Swanson E and De Haen C 1979 The translational friction coefficient of proteins *Methods Enzymol.* **61** 103–24

[7] Garcia De La Torre J, Huertas M L and Carrasco B 2000 Calculation of hydrodynamic properties of globular proteins from their atomic-level structure *Biophys. J.* **78** 719–30

[8] Schürmann G, Haspel J, Grumet M and Erickson H P 2001 Cell adhesion molecule L1 in folded (horseshoe) and extended conformations *Mol. Biol. Cell* **12** 1765–73

[9] Kirkwood J G 1954 The general theory of irreversible processes in solutions of macromolecules *J. Polymer Sci.* **12** 1–14

[10] Bloomfield V, Dalton W O and van Holde K E 1967 Frictional coefficients of multisubunit structures. I. Theory *Biopolymers* **5** 135–48

[11] Carrasco B and Garcia de la Torre J 1999 Hydrodynamic properties of rigid particles: comparison of different modeling and computational procedures *Biophys. J.* **76** 3044–57

[12] Garcia de la Torre J, Llorca O, Carrascosa J L and Valpuesta J M 2001 HYDROMIC: prediction of hydrodynamic properties of rigid macromolecular structures obtained from electron microscopy images *Eur. Biophys. J.* **30** 457–62

[13] Byron O 2008 Hydrodynamic modeling: the solution conformation of macromolecules and their complexes *Methods Cell. Biol.* **84** 327–73

[14] Schuster T M and Toedt J M 1996 New revolutions in the evolution of analytical ultracentrifugation *Curr. Opin. Struct. Biol.* **6** 650–58

[15] Hansen J C, Lebowitz J and Demeler B 1994 Analytical ultracentrifugation of complex macromolecular systems [Review] *Biochemistry* **33** 13155–63

[16] Hall C E and Slayter H S 1959 The fibrinogen molecule: its size, shape, and mode of polymerization *J. Biophys. Biochem. Cytol.* **5** 11–6

[17] Fowler W E and Erickson H P 1979 Trinodular structure of fibrinogen. Confirmation by both shadowing and negative stain electron microscopy *J. Mol. Biol.* **134** 241–49

[18] Shotton D M, Burke B E and Branton D 1979 The molecular structure of human erythrocyte spectrin *J. Mol. Biol.* **131** 303–29

[19] Melby T E, Ciampaglio C N, Briscoe G and Erickson H P 1998 The symmetrical structure of structural maintenance of chromosomes (SMC) and MukB proteins: long, antiparallel coiled coils, folded at a flexible hinge *J. Cell Biol.* **142** 1595–604

[20] Erickson H P and Fowler W E 1983 Electron microscopy of fibrinogen, its plasmic fragments and small polymers *Ann. N.Y. Acad. Sci.* **408** 146–63

[21] Erickson H P and Iglesias J L 1984 A six-armed oligomer isolated from cell surface fibronectin preparations *Nature* **311** 267–69
[22] Zhao F Q and Craig R 2003 Capturing time-resolved changes in molecular structure by negative staining *J. Struct. Biol.* **141** 43–52
[23] Ohi M, Li Y, Cheng Y and Walz T 2004 Negative staining and image classification—powerful tools in modern electron microscopy *Biol. Proced. Online* **6** 23–34
[24] Miranda J J, De Wulf P, Sorger P K and Harrison S C 2005 The yeast DASH complex forms closed rings on microtubules *Nat. Struct. Mol. Biol.* **12** 138–43
[25] Westermann S, Avila-Sakar A, Wang H W, Niederstrasser H, Wong J, Drubin D G, Nogales E and Barnes G 2005 Formation of a dynamic kinetochore–microtubule interface through assembly of the Dam1 ring complex *Mol. Cell* **17** 277–90
[26] Wang H W, Ramey V H, Westermann S, Leschziner A E, Welburn J P, Nakajima Y, Drubin D G, Barnes G and Nogales E 2007 Architecture of the Dam1 kinetochore ring complex and implications for microtubule-driven assembly and force-coupling mechanisms *Nat. Struct. Mol. Biol.* **14** 721–26
[27] Aukhil I, Joshi P, Yan Y and Erickson H P 1993 Cell- and heparin-binding domains of the hexabrachion arm identified by tenascin expression proteins *J. Biol. Chem.* **268** 2542–53
[28] Leahy D J, Axel R and Hendrickson W A 1992 Crystal structure of a soluble form of the human T cell coreceptor CD8 at 2.6 A resolution *Cell* **68** 1145–62
[29] Chen S C, Kramer G and Hardesty B 1989 Isolation and partial characterization of an M_r 60,000 subunit of a type 2A phosphatase from rabbit reticulocytes *J. Biol. Chem.* **264** 7267–75
[30] Groves M R, Hanlon N, Turowski P, Hemmings B A and Barford D 1999 The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs *Cell* **96** 99–110

Chapter 2

[1] Doty P and Myers G E 1953 Thermodynamics of the association of insulin molecules *Discuss. Faraday Soc.* **13** 51–8
[2] Chothia C and Janin J 1975 Principles of protein–protein recognition *Nature* **256** 705–8
[3] Erickson H P 1989 Cooperativity in protein–protein association: the structure and stability of the actin filament *J. Mol. Biol.* **206** 465–74
[4] Horton N and Lewis M 1992 Calculation of the free energy of association for protein complexes *Prot. Sci.* **1** 169–81

Chapter 3

[1] Amit A G, Mariuzza R A, Phillips S E and Poljak R J 1986 Three-dimensional structure of an antigen–antibody complex at 2.8 A resolution *Science* **233** 747–53
[2] Fischmann T O, Bentley G A, Bhat T N, Boulot G, Mariuzza R A, Phillips S E, Tello D and Poljak R J 1991 Crystallographic refinement of the three-dimensional structure of the FabD1.3-lysozyme complex at 2.5-Å resolution *J. Biol. Chem.* **266** 12915–20
[3] Oldfield C J, Meng J, Yang J Y, Yang M Q, Uversky V N and Dunker A K 2008 Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners *BMC Genomics* **9** S1
[4] Chothia C and Janin J 1975 Principles of protein–protein recognition *Nature* **256** 705–8
[5] Richards F M 1974 The interpretation of protein structures: total volume, group volume distributions and packing density *J. Mol. Biol.* **82** 1–14

[6] Kauzmann W 1959 Some factors in the interpretation of protein denaturation *Adv. Protein Chem.* **14** 1–63
[7] Janin J and Chothia C 1990 The structure of protein–protein recognition sites *J. Biol. Chem.* **265** 16027–30
[8] Gittes F, Mickey B, Nettleton J and Howard J 1993 Flexural rigidity of microtubules and actin filaments measured from thermal fluctuations in shape *J. Cell. Biol.* **120** 923–34
[9] Howard J 2008 Molecular Mechanics of Cells and Tissues *Cell. Mol. Bioeng.* **1** 24–32

Chapter 4

[1] Amit A G, Mariuzza R A, Phillips S E and Poljak R J 1986 Three-dimensional structure of an antigen–antibody complex at 2.8 A resolution *Science* **233** 747–53
[2] Chothia C and Janin J 1975 Principles of protein–protein recognition *Nature* **256** 705–8
[3] Bhat T N, Bentley G A, Fischmann T O, Boulot G and Poljak R J 1990 Small rearrangements in structures of Fv and Fab fragments of antibody D1.3 on antigen binding *Nature* **347** 483–85
[4] Fischmann T O, Bentley G A, Bhat T N, Boulot G, Mariuzza R A, Phillips S E, Tello D and Poljak R J 1991 Crystallographic refinement of the three-dimensional structure of the FabD1.3-lysozyme complex at 2.5-Å resolution *J. Biol. Chem.* **266** 12915–20
[5] Stein A, Rueda M, Panjkovich A, Orozco M and Aloy P 2011 A systematic study of the energetics involved in structural changes upon association and connectivity in protein interaction networks *Structure* **19** 881–89

Chapter 5

[1] DeVos A M, Ultsch M and Kossiakoff A A 1992 Human growth hormone and extracellular domain of its receptor: crystal structure of the complex *Science* **255** 306–12
[2] Cunningham B C, Ultsch M, DeVos A M, Mulkerrin M G, Clauser K R and Wells J A 1991 Dimerization of the extracellular domain of the human growth hormone receptor by a single hormone molecule *Science* **254** 821–5
[3] Brooks A J, Dai W, O'Mara M L, Abankwa D, Chhabra Y, Pelekanos R A and Waters M J 2014 Mechanism of activation of protein kinase JAK2 by the growth hormone receptor *Science* **344** 1249783
[4] DeLano W L *et al* 2000 Convergent solutions to binding at a protein–protein interface *Science* **287** 1279–83
[5] Corper A L, Sohi M K, Bonagura V R, Steinitz M, Jefferis R, Feinstein A, Beale D, Taussig M J and Sutton B J 1997 Structure of human IgM rheumatoid factor Fab bound to its autoantigen IgG Fc reveals a novel topology of antibody–antigen interaction *Nat. Struct. Biol.* **4** 375–81
[6] Deisenhofer J 1981 Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from *Staphylococcus aureus* at 2.9- and 2.8-Å resolution *Biochemistry* **20** 2361–70

Chapter 6

[1] Clackson T and Wells J A 1995 A hot spot of binding energy in a hormone–receptor interface *Science* **267** 383–6
[2] Cunningham B C and Wells J A 1993 Comparison of a structural and a functional epitope *J. Mol. Biol.* **234** 554–63

[3] Amit A G, Mariuzza R A, Phillips S E and Poljak R J 1986 Three-dimensional structure of an antigen–antibody complex at 2.8 A resolution *Science* **233** 747–53
[4] Atwell S, Ultsch M, De Vos A M and Wells J A 1997 Structural plasticity in a remodeled protein–protein interface *Science* **278** 1126–8
[5] Guerois R, Nielsen J E and Serrano L 2002 Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations *J. Mol. Biol.* **320** 369–87
[6] Kortemme T and Baker D 2002 A simple physical model for binding energy hot spots in protein–protein complexes *Proc. Natl. Acad. Sci. USA* **99** 14116–21
[7] Reichmann D, Rahat O, Cohen M, Neuvirth H and Schreiber G 2007 The molecular architecture of protein–protein binding sites *Curr. Opin. Struct. Biol.* **17** 67–76
[8] Akiba H and Tsumoto K 2015 Thermodynamics of antibody–antigen interaction revealed by mutation analysis of antibody variable regions *J. Biochem.* **158** 1–13
[9] Moretti R *et al* 2013 Community-wide evaluation of methods for predicting the effect of mutations on protein–protein interactions *Proteins* **81** 1980–7
[10] Lensink M F, Velankar S and Wodak S J 2017 Modeling protein–protein and protein–peptide complexes: CAPRI 6th edition *Proteins* **85** 359–77

Chapter 7

[1] Erickson H P 1989 Cooperativity in protein–protein association: the structure and stability of the actin filament *J. Mol. Biol.* **206** 465–74
[2] Chothia C and Janin J 1975 Principles of protein–protein recognition *Nature* **256** 705–8
[3] Horton N and Lewis M 1992 Calculation of the free energy of association for protein complexes *Prot. Sci.* **1** 169–81
[4] Crothers D M and Metzger H 1972 The influence of polyvalency on the binding properties of antibodies *Immunochemistry* **9** 341–57
[5] de Vos A M, Ultsch M and Kossiakoff A A 1992 Human growth hormone and extracellular domain of its receptor: crystal structure of the complex *Science* **255** 307–12
[6] Clackson T, Ultsch M H, Wells J A and De Vos A M 1998 Structural and functional analysis of the 1:1 growth hormone:receptor complex reveals the molecular basis for receptor affinity *J. Mol. Biol.* **277** 1111–18

Chapter 8

[1] Koren R and Hammes G G 1976 A kinetic study of protein–protein interactions *Biochemistry* **15** 1165–71
[2] Smoluchowski M v 1916 Drei Vorträge über diffusion, Brownsche Molekularbewegung und Koagulation von Kolloidteilchen *Physik. Zeitschr.* **17** 585–99
[3] Northrup S H and Erickson H P 1992 Kinetics of protein–protein association explained by Brownian dynamics computer simulation *Proc. Natl. Acad. Sci. USA* **89** 3338–42
[4] Pang X, Qin S and Zhou H X 2011 Rationalizing 5000-fold differences in receptor-binding rate constants of four cytokines *Biophys. J.* **101** 1175–83
[5] Cunningham B C and Wells J A 1993 Comparison of a structural and a functional epitope *J. Mol. Biol.* **234** 554–63
[6] Pollard T D and De La Cruz E M 2013 Take advantage of time in your experiments: a guide to simple, informative kinetics assays *Mol. Biol. Cell* **24** 1103–10

Chapter 9

[1] Tangemann K and Engel J 1995 Demonstration of non-linear detection in ELISA resulting in up to 1000-fold too high affinities of fibrinogen binding to integrin aIIbb3 *FEBS Lett.* **358** 179–81

[2] Friguet B, Chaffotte A F, Djavadi-Ohaniance L and Goldberg M E 1985 Measurements of the true affinity constant in solution of antigen–antibody complexes by enzyme-linked immunosorbent assay *J. Immunol. Meth.* **77** 305–19

[3] Fields S and Song O 1989 A novel genetic system to detect protein–protein interactions *Nature* **340** 245–46

[4] Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M and Seraphin B 1999 A generic protein purification method for protein complex characterization and proteome exploration *Nat. Biotechnol.* **17** 1030–32

[5] Roux K J, Kim D I, Raida M and Burke B 2012 A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells *J. Cell. Biol.* **196** 801–10

[6] Trinkle-Mulcahy L 2019 Recent advances in proximity-based labeling methods for interactome mapping. F1000Res 8.

[7] Hung V, Udeshi N D, Lam S S, Loh K H, Cox K J, Pedram K, Carr S A and Ting A Y 2016 Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2 *Nat. Protoc.* **11** 456–75

[8] Chen Y and Erickson H P 2011 Conformational changes of FtsZ reported by tryptophan mutants *Biochemistry* **50** 4675–84

[9] Cooper J A, Walker S B and Pollard T D 1983 Pyrene actin: documentation of the validity of a sensitive assay for actin polymerization *J. Muscle Res. Cell Motil.* **4** 253–62

[10] Siarheyeva A and Sharom F J 2009 The ABC transporter MsbA interacts with lipid A and amphipathic drugs at different sites *Biochem. J.* **419** 317–28

[11] Pollard T D 2010 A guide to simple and informative binding assays *Mol. Biol. Cell* **21** 4061–67

[12] Schuck P and Minton A P 1996 Analysis of mass transport-limited binding kinetics in evanescent wave biosensors *Anal. Biochem.* **240** 262–72

[13] Seidel S A *et al* 2013 Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions *Methods* **59** 301–15

[14] Engvall E and Perlmann P 1972 Enzyme-linked immunosorbentassay, ELISA *J. Immunol.* **109** 129–35

[15] Peleg-Shulman T, Roisman L C, Zupkovitz G and Schreiber G 2004 Optimizing the binding affinity of a carrier protein: a case study on the interaction between soluble ifnar2 and interferon beta *J. Biol. Chem.* **279** 18046–53

Chapter 10

[1] Pierschbacher M D and Ruoslahti E 1984 Cell attachment activity of fibronectin can be duplicated by small synthetic fragments of the molecule *Nature* **309** 30–3

[2] Leahy D J, Aukhil I and Erickson H P 1996 2.0 Å crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region *Cell* **84** 155–64

[3] Van Agthoven J F, Xiong J P, Alonso J L, Rui X, Adair B D, Goodman S L and Arnaout M A 2014 Structural basis for pure antagonism of integrin alphaVbeta3 by a high-affinity form of fibronectin *Nat. Struct. Mol. Biol.* **21** 383–88

[4] Amit A G, Mariuzza R A, Phillips S E and Poljak R J 1986 Three-dimensional structure of an antigen–antibody complex at 2.8 A resolution *Science* **233** 747–53
[5] Frank R 2002 The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports—principles and applications *J. Immunol. Methods* **267** 13–26
[6] Pardon E, Laeremans T, Triest S, Rasmussen S G, Wohlkonig A, Ruf A, Muyldermans S, Hol W J G, Kobilka B K and Steyaert J 2014 A general protocol for the generation of nanobodies for structural biology *Nat. Protoc.* **9** 674–93
[7] Koide A, Bailey C W, Huang X and Koide S 1998 The fibronectin type III domain as a scaffold for novel binding proteins *J. Mol. Biol.* **284** 1141–51
[8] Smith G P 1985 Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface *Science* **228** 1315–17
[9] Bloom L and Calabro V 2009 FN3: a new protein scaffold reaches the clinic *Drug Discovery Today*
[10] Jost C and Pluckthun A 2014 Engineered proteins with desired specificity: DARPins, other alternative scaffolds and bispecific IgGs *Curr. Opin. Struct. Biol.* **27** 102–12
[11] Gilbreth R N and Koide S 2012 Structural insights for engineering binding proteins based on non-antibody scaffolds *Curr. Opin. Struct. Biol.* **22** 413–20

Chapter 11

[1] Oldfield C J, Meng J, Yang J Y, Yang M Q, Uversky V N and Dunker A K 2008 Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners *BMC Genom.* **9** Suppl 1:S1
[2] Amit A G, Mariuzza R A, Phillips S E and Poljak R J 1986 Three-dimensional structure of an antigen–antibody complex at 2.8 A resolution *Science* **233** 747–53
[3] Cunningham B C, Ultsch M, DeVos A M, Mulkerrin M G, Clauser K R and Wells J A 1991 Dimerization of the extracellular domain of the human growth hormone receptor by a single hormone molecule *Science* **254** 821–5
[4] Kamada R, Toguchi Y, Nomura T, Imagawa T and Sakaguchi K 2016 Tetramer formation of tumor suppressor protein p53: structure, function, and applications *Biopolymers* **106** 598–612
[5] Borgia A *et al* 2018 Extreme disorder in an ultrahigh-affinity protein complex *Nature* **555** 61–6
[6] Ditlev J A, Case L B and Rosen M K 2018 Who's in and who's out-compositional control of biomolecular condensates *J. Mol. Biol.* **430** 4666–84
[7] Wright P E and Dyson H J 2015 Intrinsically disordered proteins in cellular signalling and regulation: *Nat. Rev. Mol. Cell Biol.* **16** 18–29