UNIVERSITÄT TÜBINGEN

UNIVERSITÀ DEGLI STUDI DI TRENTO

INSTITUT LAUE LANGEVIN

Studiengang Physik Diplom

Laurea specialistica in Fisica

Diplomarbeit

Elaborato finale

Static and Dynamic Light Scattering from Aqueous Solutions of Bovine Serum Albumin and Salts

Gutachter/Relatori:

Prof. Dr. Frank Schreiber
Prof. Marina Scarpa

Diplomand/Laureando:

Fabio Zanini

Wintersemester 2009 - 2010

Anno Accademico 2009 - 2010

Hiermit wird bestätigt, dass die Diplomarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt wurden.

Tübingen, den .....

Unterschrift:

.....

Abstract

For several decades, light scattering has been a standard experimental technique for collecting information on structure, thermodynamics, and kinetics of macromolecular solutions. In this thesis, we show that light scattering plays a pivotal role on the stage of protein solutions, because of its simple setup, its sound theoretical basis, its versatility, and its relative inexpensiveness. In the following chapters, we present the outcome of a series of static and dynamic light scattering experiments on aqueous solutions of the globular protein bovine serum albumin (BSA) and various light salts. After introductory chapters presenting the minimal physicochemical background, we proceed with a description and a thorough discussion on the experimental findings. First, we discuss the solutions with sodium and calcium chloride, which are physiologically relevant [1] and well-suited for theoretical colloidal modelling [2]. Then, we discuss those with yttrium chloride, which involve a variety of recently discovered phenomena including re-entrant condensation [3] and liquid-liquid phase separation (LLPS) [4].

First, we find that the solutions without added salt, with NaCl, and with $CaCl_2$ can be described satisfactorily, on a semi-quantitative level, on the basis of two simple concepts, electrostatic screening and steric repulsion. The screened Coulomb repulsion between charged BSA molecules dominates at low salt concentration, causing low osmotic compressibility, high second virial coefficient, and fast collective diffusion. The samples at high salt concentration behave almost as suspensions of hard spheres instead.

Second, the measurements on yttrium chloride yield exciting results, which add useful information to the BSA-YCl_3 phase diagram. Our findings relate both to re-entrant condensation itself, suggesting a widespread presence of small oligomers in the third regime, and to the aggregation processes taking place in the condensed regime. Here, we observe a transition from transparent solutions of slightly attracting monomers towards turbid ones, via a few intermediates that are optically transparent but populated by large protein clusters ($\sim 10^4$ monomers). The latter possess a typical, strongly Q-dependent scattering profile, which resembles the oft-debated cluster peak observed in lysozyme solutions [5].

Abstract

Seit vielen Jahrzehnten ist die Lichtstreuung eine übliche experimentelle Methode zur Untersuchung von Struktur, Thermodynamik und Kinetik von makromolekularen Lösungen. In dieser Diplomarbeit zeigen wir, dass die Lichtstreuung, dank ihres einfachen und bezahlbaren Aufbaus, ihrer soliden theoretischen Basis, ihrer Flexibilität eine wichtige Rolle in der Forschung über Proteinlösungen spielt. In den folgenden Kapiteln stellen wir die Ergebnisse unserer Experimente mit statischer und dynamischer Lichtstreuung an wässerigen Lösungen verschiedener Salze und des globularen Proteins Rinderserumalbumin (BSA) vor. Nach den Einführungskapiteln, in denen wir den minimalen physikalisch-chemischen Hintergrund vorstellen, diskutieren wir die experimentellen Entdeckungen. Erst behandeln wir die Lösungen mit Natrium- und Kalziumchlorid, die physiologisch relevant [1] und geeignet für Kolloidmodelle [2] sind. Dann beschreiben wir die Lösungen mit Yttriumchlorid; dieses Salz verursacht eine Vielzahl von neu entdeckten Phänomenen einschließlich "Re-entrant Condensation" [3] und Flüssig-Flüssig-Phasenseparation (LLPS) [4].

Zuerst finden wir, dass die Lösungen ohne zusätzliches Salz und mit Natrium- oder Kalziumchlorid auf einem semiquantitativen Niveau anhand zwei einfacher Begriffe interpretiert werden können: elektrostatische Abschirmung und sterische Abstoßung. Die abgeschirmte Coulomb'sche Abstoßung dominiert zwischen geladenen BSA-Molekülen bei kleinen Salzkonzentrationen und ist Grund der kleinen osmotischen Kompressibilität, des großen zweiten Virialkoeffizientes und der schnellen kollektiven Diffusion. Die Proben bei großer Salzkonzentration verhalten sich hingegen ähnlich zu Suspensionen von harten Kugeln.

Zweitens liefern die Messungen mit Yttriumchlorid interessante Resultate, die sehr nützliche Informationen über das BSA-YCl_3 Phasendiagramm beinhalten. Unsere Ergebnisse betreffen sowohl die "Re-entrant Condensation" selbst, indem sie auf eine große Zahl von kleinen Oligomeren hinweisen, als auch die Aggregationsprozesse im kondensierten Regime. In diesem Regime beobachten wir einen Übergang von durchsichtigen, aus sich leicht anziehenden Monomeren bestehenden, zu trüben Lösungen. Wir sehen einige Zwischenlösungen, die transparent sind, aber auch große Proteinaggregate ($\sim 10^4$ Monomere) enthalten und ein typisches, stark Q-abhängiges Streumuster zeigen; das erinnert an den umstrittene, in Lysoyzmlösungen gemessenen "Cluster Peak" [5].

Abstract

Da molti anni, la diffusione di luce (light scattering) è una tecnica sperimentale ampiamente utilizzata per raccogliere informazioni sulla struttura, sulla termodinamica e sulla cinetica di soluzioni macromolecolari. In questa tesi, mostriamo che la diffusione di luce ricopre un ruolo di primo piano nel campo delle soluzioni proteiche, grazie alla sua semplicità strumentale, ad una base teorica solida, alla sua versatilità ed al costo di utilizzo ridotto. Nei vari capitoli della tesi, presentiamo il risultato di una serie di esperimenti di diffusione statica e dinamica di luce, condotti su soluzioni acquose di vari sali e di albumina del siero bovino (BSA), una proteina globulare. Dopo alcuni capitoli introduttivi, nei quali presentiamo una breve introduzione fisicochimica, procediamo con una descrizione ed una esauriente discussione dei risultati sperimentali. Trattiamo separatamente le soluzioni con cloruro di sodio e cloruro di calcio, che sono fisiologicamente rilevanti [1] ed appropriate per una modellizzazione con particelle colloidali [2], da quelle con cloruro di yttrio, che coinvolgono vari fenomeni di recente scoperta, tra cui la condensazione rientrante proteica [3] e la separazione di fase liquido-liquido (LLPS) [4].

Innanzitutto, troviamo che le soluzioni senza sale aggiunto o con NaCl o $CaCl_2$ possono essere descritte, ad un livello semiquantitativo, sulla base di due semplici concetti, la schermatura elettrostatica e la repulsione sterica. Mentre la repulsione di Coulomb schermata tra molecole cariche di BSA domina a basse concentrazioni saline, dove è causa di una bassa compressibilità osmotica, di un alto secondo coefficiente di viriale e di una rapida diffusione collettiva, i campioni ad alta concentrazione salina si comportano pressoché come sospensioni di sfere rigide.

Inoltre, le misure sul cloruro di yttrio rivelano scoperte molto interessanti, che incrementano sensibilmente le conoscenze sul diagramma di fase BSA-YCl_3. Questi risultati riguardano sia la condensazione rientrante in sé, suggerendo ina presenza diffusa di piccoli oligomeri nel terzo regime, sia i processi di aggregazione che hanno luogo nella regime condensato. In questo regime centrale, osserviamo una transizione da soluzioni trasparenti di monomeri, che si attraggono debolmente, verso soluzioni torbide, passando per degli stadi intermedi che sono otticamente trasparenti ma popolati da grandi aggregati proteici ($\sim 10^4$ monomeri). Questi ultimi campioni posseggono un profilo di scattering tipico e fortemente dipendente da Q, che richiama il picco di aggregazione osservato, non senza alcune critiche, in soluzioni di lisozima [5].

Preface

> It may be here remarked that most animals and plants keep to their proper homes, and do not needlessly wander about; we see this even with migratory birds, which almost always return to the same spot.
>
> Charles Darwin [6]

In November 2009, I started an internship program at the Institut Laue Langevin (ILL) in Grenoble, France. In the four months I worked there as *stagiaire*, supervised by kind and tireless Dr. Tilo Seydel and promptly helped by Mr. Marcus Hennig, I performed a series of light scattering experiments on aqueous solutions of a blood globular protein, Bovine Serum Albumin (BSA). On the one hand, I wanted to collect precise and systematic experimental data on samples with physiologically relevant salts, such as NaCl and $CaCl_2$, in order to investigate the dynamic behaviour of proteins in a controlled but native-like environment. On the other hand, I was looking forward to exploring the recently discovered phenomenon of protein re-entrant condensation [3].

Of course, even the detailed description provided in this thesis does not represent but a small fragment of the mosaic of situations experienced in Prof. Schreiber's research group. Among them, I mention one normal and anomalous small-angle X-ray scattering beamtime at the European Synchrotron Radiation Faciliy (ESRF), one neutron spin-echo, one time-of-flight, and several quasi-elastic neutron scattering beamtimes at the ILL, a few research trips to the University of Tübingen and the Jülich Center for Neutron Science (JCNS), and one intercontinental research trip, funded by the JCNS, to the BASIS backscattering spectrometer at the Oak Ridge National Laboratory (ORNL) in Tennesee, USA.

As a grammatical side note, I anticipate that the first person plural is in use throughout the thesis, except for this preface and the acknowledgements section, which I have decided to keep in singular form for their biographical character.

Acknowledgements

This thesis would not have been possible without the help of several people and institutions. First of all, I would like to thank my principal supervisor, Prof. Frank Schreiber at the Universität Tübingen (UT), for his scientific, logistic, financial, and psychological support during this year, as well as for the stimulating discussion on the physical reality of the Schrödinger equation. I am also indebted with Prof. Marina Scarpa at the Università degli Studi di Trento, for her willingness to act as secondary supervisor, her reassuring calmness, and her constant enthusiasm towards this work.

It is a pleasure to thank, not only as co-supervisors, Dr. Tilo Seydel at the Institut Laue Langevin (ILL) and Dr. Fajun Zhang at the UT, for their steady support, in terms of scientific guidance, cultural enrichment, and coordination, especially during my stay in Grenoble and in Oak Ridge. It is a honor to thank Dr. Ralf Schweins at the ILL as well, for his manifold and tireless help on the light scattering instrument, the sample preparation, and the scientific literature.

I would like to show my gratitude to the other group members at the UT for their wonderful seminar talks and their friendship. In particular, I am grateful to Felix Roosen-Runge for his scientific and philosophical advice, to Elena Jordan for her kindness and vitality, to Marcell Wolf for the time spent with me in the laboratory and his resilience to my pessimistic opinions, to Moritz Schollbach for his good organization, and to Andrea Sauter for her corrections to this thesis. I owe my gratitude to Marcus Hennig for his scientific, linguistic, and technological competence.

I acknowledge financial and logistic support of the Jülich Center for Neutron Science and the ILL, in particular the Time of Flight/High Resolution group for the *stagiaire* position, the neutron and light scattering beamtime, and the Large-Scale Structures group for help in the chemistry and soft-matter laboratories.

Finally, I would like to thank my relatives and friends for their trust, and Giulia for her patience.

Contents

Abstract **5**

Preface **9**

Acknowledgements **11**

Table of Contents **13**

List of Tables **17**

List of Figures **19**

List of Symbols **21**

1 Introduction **23**
1.1 Motivation 23
1.2 Literature Review 24

2 Materials **29**
2.1 Water 29
2.2 Bovine Serum Albumin 29
2.2.1 Molar Mass 30
2.2.2 Shape and Dimensions 30
2.2.3 Volume and Effective Radius Estimates 31
2.2.4 Specific Volume 32
2.2.5 Electrostatic Charge 33
2.2.6 Purity and Self-Association 33
2.2.7 Biological Function 34
2.3 Salts 35
2.3.1 Differences among NaCl, $CaCl_2$, and YCl_3 35
2.3.2 Meaning of the “salt-free” solutions 36

3 Theory **39**
3.1 Static Light Scattering . . . 39
3.1.1 Scattered Field of an Infinitesimal Element . . . 41
3.1.2 Total Scattered Field . . . 42
3.1.3 Detected Intensity and Rayleigh Ratio . . . 43
3.1.4 Thermodynamics of SLS . . . 45
3.1.5 Solvent Subtraction . . . 46
3.1.6 Classic SLS Equations . . . 47
3.1.7 Remarks on Multi-Component SLS . . . 51
3.2 Dynamic Light Scattering . . . 53
3.2.1 Working Principle . . . 54
3.2.2 Collective Diffusion Kernel and Constant . . . 59
3.2.3 Length and Time Scales for BSA Diffusion . . . 60
3.2.4 Fit Strategies for the Autocorrelation . . . 61
3.2.5 Diffusion in Dilute and Crowded Systems . . . 67

4 Experimental Methods **73**
4.1 Sample Preparation . . . 73
4.2 Light Scattering . . . 75
4.2.1 Experimental Setup . . . 75
4.2.2 Absolute Calibration of SLS . . . 78
4.3 UV Absorption . . . 79
4.4 Numerical Data Analysis with Matlab® . . . 80

5 Results and Discussion **83**
5.1 Overall View . . . 83
5.1.1 BSA Molar Mass . . . 85
5.1.2 Dilute-Limit D_0 and Hydrodynamic Radius . . . 87
5.2 Samples with Sodium and Calcium Chloride . . . 88
5.2.1 Static Light Scattering . . . 89
5.2.2 Dynamic Light Scattering . . . 94
5.2.3 Hydrodynamics and Sedimentation . . . 99
5.2.4 Theoretical Interpretation . . . 101
5.2.5 Comparison with Existing Literature . . . 106
5.3 BSA-YCl_3 Phase Diagram . . . 108
5.3.1 Red-Blue Transition: Regime II → Regime I . . . 112
5.3.2 Red-Green Transition: Regime II → LLPS . . . 113
5.3.3 Yellow Region: Regime III . . . 116

6 Summary and Outlook **119**
6.1 Summary 119
6.2 Outlook 121

Bibliography **125**

List of Tables

2.1 Radii of the ions used in this work. 35

3.1 Qualitative summary of the possible values of b_2 and the corresponding interpretations. 50

5.1 Molar mass estimates from SLS. 87
5.2 Estimates for D_0 and r_h at low c_p, from DLS. 88
5.3 B_2 estimates from SLS at low c_p. 90
5.4 Estimates for the slope coefficient k_d from DLS at low c_p. . . . 96
5.5 Estimates for the second hydrodynamic coefficient k_h from DLS/SLS at low c_p. 100

List of Figures

2.1 Crystal structure of HSA in schematic view. 31
2.2 Crystal structure of HSA showing surface accessible to water. 31

3.1 Simple representation of the scattering geometry. 40
3.2 Scattering from a single volume element 42
3.3 Schematic illustration of the scattering volume. 43
3.4 Typical time-resolved scattered intensity. 55
3.5 Typical intensity autocorrelation for a protein solution. 56
3.6 Schematic of the relaxation of a concentration fluctuation. . . 57
3.7 Schematic summary of the DLS technique. 58
3.8 Example of a cumulant fit. 63
3.9 Example of an inverse Laplace transform of a field autocorrelation. 64
3.10 Example of an ad-hoc double exponential fit. 66
3.11 Schematic illustration of the hydrodynamic interaction. 68

4.1 Schematic illustration of sample preparation. 74
4.2 Schematic illustration of sample filtration. 74
4.3 Schematic illustration of the light scattering instrument. . . . 76
4.4 Real picture of the ALV light scattering instrument at the ILL. 76
4.5 Schematic lateral view of the index-matching bath. 77
4.6 Typical near-UV absorbance spectrum of a BSA solution. . . . 79

5.1 Overview of the SLS/DLS results: Kc/R and D as a function of protein concentration for various salt species and concentrations. 84
5.2 SLS results for solutions without added salt, with NaCl, and with $CaCl_2$. 92
5.3 Normalized osmotic compressibility for solutions without added salt, with NaCl, and with $CaCl_2$. 93

5.4 Diffusion constant for solutions without added salt, with NaCl, and with $CaCl_2$, obtained by the ad-hoc double exponential decay fit routine. 97
5.5 Diffusion constant for solutions without added salt, with NaCl, and with $CaCl_2$, obtained by the first-order cumulant fit routine. 98
5.6 Normalized sedimentation coefficient for solutions without added salt, with NaCl, and with $CaCl_2$. 102
5.7 Normalized osmotic compressibility as a function of volume fraction, for the dataset of BSA in high-c_s solutions, compared with the theoretical prediction for hard spheres in the Percus-Yevick closure. 106
5.8 Phase diagram of BSA-YCl_3. 109
5.9 SLS results for samples with YCl_3. 110
5.10 DLS results for samples with YCl_3. 111
5.11 Diffusion constant for samples at equal ionic strength, and different salt species. 112
5.12 Q-dependent SLS results for red-green samples, with 5 mM YCl_3. 114
5.13 Speculative scenarios for the increase in $I(Q)/c_p$ at higher Q values than the maximum of light scattering. 115

List of Symbols

SLS	Static light scattering
DLS	Dynamic light scattering
BSA	Bovine serum albumin
c, c_p	Protein concentration
c_s	Salt concentration
k	Boltzmann constant, norm of the light wave vector
T	Temperature
Q	Scattering vector, protein charge
Kc/R, $\frac{Kc_p}{R}$	Normalized inverse Rayleigh ratio from SLS
D	Diffusion constant
D_0	Dilute-limit diffusion constant
$B_2(\mathbf{SLS})$	Second virial coefficient in experimental units
$B_2(\mathbf{t.u.})$	Second virial coefficient in theoretical units
b_2	Normalized second virial coefficient
k_d	Second diffusion coefficient
k_h	Second hydrodynamic coefficient
μ	Chemical potential
μ_0	Chemical potential of the ideal solution
Π	Osmotic pressure
Π_0	Osmotic pressure of the ideal solution
χ_T	Osmotic compressibility
χ_{T0}	Osmotic compressibility of the ideal solution
κ	Total compressibility, Debye screening parameter
n	Refractive index of the solution
n_w	Refractive index of the solvent
Δn	$n - n_w$

δn	Thermodynamic fluctuations of the refractive index
R, ΔR	(Background-subtracted) Rayleigh ratio
I_0	Incident light intensity
I	Scattered light intensity
I/c, $\frac{I}{c_p}$	Concentration normalized scattered intensity
ν	Frequency of the light
t	Time
LLPS	Liquid-liquid phase separation
HSA	Human serum albumin
UV	Ultraviolet (radiation)
SAXS	Small-angle X-ray scattering
SANS	Small-angle neutron scattering
3D-DLS	Three-dimensional dynamic light scattering
ASAXS	Anomalous small-angle X-ray scattering
USAXS	Ultrasmall-angle X-ray scattering
USANS	Ultrasmall-angle neutron scattering
DWS	Diffuse wave spectroscopy
XPCS	X-ray photon correlation spectroscopy
QENS	Quasi-elastic neutron scattering
NSE	Neutron spin-echo
GFC	Gel-filtration chromatography
SIC	Self-interaction chromatography

Chapter 1

Introduction

In this chapter, we introduce the subject of the thesis, the static and dynamic statistical properties of globular proteins in aqueous solution, studied by light scattering. In the first section, we list the reasons that make this topic so interesting. In the second one, we quickly review the past research in this field.

1.1 Motivation

Physicochemical studies on proteins are in general a fascinating research field. They require and stimulate an interdisciplinary competence between physics, chemistry, material science, computer science, and biology, which is difficult to find elsewhere. The results from basic research are readily applied, both practically and intellectually, in all those fields, and also in medicine. In this context, the rationale behind this thesis is based on four reasons.

The first and main reason is the fascination generated by the exciting discovery of the *re-entrant condensation* in protein systems, due to Zhang *et al.* [3]. In short, when a trivalent salt such as yttrium chloride is added to an aqueous solution of globular proteins, the mixture behaves very differently from the expectations based on monovalent and divalent salts. Whereas "normal" salts affect the proteins mainly by screening their mutual electrostatic repulsion, YCl_3 binds them strongly, changing the surface charge dramatically [7]. If enough YCl_3 is added, a protein is thought to undergo *charge inversion*, which changes its physicochemical properties. Furthermore, at intermediate yttrium concentration, the surface charge is almost neutralized, and a plethora of new thermodynamic phenomena arises. Among them, here we only mention metastable oligomerization and protein crystallization. We contribute to the general understanding of re-entrant condensation by per-

forming a systematic light scattering study on a typical re-entrant protein, Bovine Serum Albumin (BSA).

The second reason is the need to understand the basic mechanisms that regulate the thermodynamics and diffusion of proteins in native-like liquid environments. The protein used in our experiments, BSA, is naturally found in the blood, the physiological liquid environment par excellence. Although the blood is made of many different components, including whole cells and other proteins, one important effect on its generic physicochemical properties is due to small free ions, because their concentration controls the strength of the electrostatic repulsion among charged molecules. In this work, we study this phenomenon by performing static and dynamic measurements with two dissociable physiological salts, sodium and calcium chloride, in a wide range of protein and salt concentrations. We collect a rich variety of precise experimental data on this topic; we then explain them quantitatively or semi-quantitatively using microscopic concepts from the physics of colloids [2, 8].

The third reason is the intuition that the fundamental law of molecular biology, the *structure-function hypothesis* [9], needs to be extended. The hypothesis can be stated as follows: "The geometrical structure of biologically active molecules implies necessarily its biological function.". This assumption underestimates the importance of dynamic features, naively overlooking that most biomolecules accomplish their tasks essentially by *moving*. In this thesis, we see experimental results not only on protein structure, but also on protein dynamics. The dynamic properties turn out to be connected to the static ones, but not entirely determined by them. Therefore, we demonstrate that static experiments are not sufficient to predict the biological function of a macromolecule.

The last reason, not primarily scientific, is of course the thrilling sensation of geographic delocalization enjoyed all along the development of this work. After decades of sedentary life in Trento, I have taken part eagerly – but not without problems, nor without criticism – to the typically post-modern, social experiment that Bauman calls *liquid life* [10, 11].

1.2 Literature Review

The interest in proteins-salt systems arises very early in the physical chemistry literature. Hofmeister publishes key studies already in the nineteenth century [12]. Afterwards, the scattering phenomena in liquids and solutions gains the attention of the community thanks to the publications of Smoluchowski [13] and Einstein [14]. However, a complete theory of light scattering from solutions of macromolecules is developed only in the nineteen forties,

by Zimm [15], Debye [16], Stockmayer [17], and Kirkwood and Goldberg [18]. The application of the static light scattering technique on protein solutions, in terms of Rayleigh ratio or turbidity, is due to Edsall [19], Timasheff [20], and others. At the same time, the main chemicophysical properties of BSA are being theorized and measured, among others, by Scatchard [21, 22].

The early papers explore a variety of chemical and physical properties of BSA. The theories are expressed in terms of activity coefficients, Donnan equilibria, association and dissociation reactions, conformation changes, and denaturation, while the experimental parameters are the pH, the protein charge, and its binding state [23–28]. In general, the thermodynamics of globular proteins, including BSA, turns out to depend strongly on all those parameters. For neutral or slightly acidic pH and common salts such as NaCl, it is also found that BSA is negatively charged, so that the electrostatic repulsion is an important ingredient for the stability of the solution.

In the theories of McMillan and Mayer [29] and the Ornstein-Zernike integral scheme [30–33], the interparticle interaction $u(r)$ is treated explicitly, and the influence of the various contributions – electrostatics, excluded-volume, van der Waals – can be studied quantitatively, for instance by the Debye-Hückel [34] or the Derjaguin-Landau-Verwey-Overbeek (DLVO) [35,36] models. This approach is applicable on a wide range of scattering vectors, and is thus developed further after the discovery of small-angle X-ray and neutron scattering methods [37–39]. The experiments explore the space of parameters related to the protein-protein interactions, that are ionic strength, temperature, pressure, protein concentration. Of course, the direct use of $u(r)$ in the explanation of experiments on proteins is an ambitious task. As a matter of fact, proteins are very complicated objects, and it is impossible to determine the exact $u(r)$ starting from experimental data. Therefore, several authors try to develop semi-phenomenological models of protein solutions, which assume some drastic simplifications but are hoped to retain the key physical concepts. These models, inspired by the success of colloid physics, attribute to the proteins so-called *effective properties*, such as an effective charge, radius, volume. In turn, the latter differ from the real, bare, or *net properties*; they are semi-quantitative values that fill the gap between the complexity of biology and the accuracy of physics. This is the position chosen by most later scientists, until the present day [40–45].

The older studies on protein diffusion make use of various experimental methods such as diaphragm diffusion cells [46–48]. Dynamic light scattering is applied to protein solutions in the nineteen seventies, with studies on BSA similar to this one [49, 50]. Simultaneous is the development of the first colloidal theories of diffusion in concentrated solutions, first in terms of hard spheres and as power expansions in the protein concentration [51–55].

Later, the analysis of the direct and hydrodynamic interactions improves; the derivation of the generalized Svedberg equation is an important step [56], followed by more and more precise analytical theories that try to build a true many-body description, not limited to pairwise interactions [57–59]. The instrumental technology progresses as well, especially that of auto- and cross-correlators, enabling more reliable and extended studies [60–62]. The principal outcome of those investigations, as far as this thesis is concerned, is that the collective diffusion of BSA in concentrated solutions is substantially higher when the protein is natively folded and charged, i.e. at $6 \lesssim \mathrm{pH} \lesssim 7.5$, and the electrostatic repulsion is only weakly screened, which means at low salt concentration. The diffusion is slower at the isoelectric point instead, of whenever BSA unfolds as a consequence of denaturating agents in solutions, such as urea, or strong acids or bases.

In the latest years, thanks to the advances in the theoretical framework and the increased availability of instruments, the focus of a part of the research community shifts towards a unified vision of self- and collective diffusion, interactions, sedimentation, and viscosity. Several studies are performed in this direction [63–67].

In this thesis, we pursue two goals. On the one hand, we collect a complete and precise set of light scattering results on static and dynamic properties of BSA solutions. On the other hand, we try to connect the thermodynamic data from SLS with the DLS data on diffusion, using standard theoretical methods from the theory of colloids, which are not specific to BSA itself.

The literature on re-entrant condensation of proteins deserves a special paragraph. The very discovery of this phenomenon is recent [3], and much is still to be explained. Prof. Schreiber's research group is very active in this field, with a lot of successful experimental work in the last two years [4,7,68]. The main experimental finding are the BSA-XCl_3 phase diagrams, where X is (the neutral form of) one among a few studied trivalent cations, including notably Y^{3+} (see chapter 5). The conceptual explanation for the phase diagrams is still not polished, but an approximate picture is established. According to this view, BSA is negatively charged in yttrium-free solutions, but acquires a more and more positive charge upon increasing salt concentration, via strong association with Y^{3+} cations. At intermediate YCl_3 concentrations, the total protein charge is neutralized by the cations. Under these conditions, the stability of the solution is compromised, but not towards an irreversible aggregative instability, but rather towards a reversible, long-lived metastability; the BSA solutions look turbid. At even higher yttrium concentrations, the surface charge is reversed by the high number of Y^{3+} ions bound; the protein samples turn transparent again, or *re-enter*. In order to quantify the experimental results, the interpretative framework of Gros-

berg is tentatively invoked, albeit its restrictive assumptions on the surface charge distribution [69]. In this thesis, we continue the experimental work on this topic, in the same holistic spirit that characterizes the yttrium-free investigations.

Chapter 2

Materials

In this chapter, we describe the substances used in the samples. Basically, these consist of aqueous solutions of a protein and a salt. More specifically, the solutions have the following components. The first component is liquid water. The second component is the globular protein Bovine Serum Albumin (BSA). These substances are present, in different proportions, in all samples. Some samples also include a third component, a light chloride salt. We describe three different salts, which are sodium chloride (NaCl), calcium chloride ($CaCl_2$), and yttrium chloride (YCl_3). In this thesis, two different salts are never used together in the same solution.

2.1 Water

We prepare all solutions described in this thesis using deionized light water (H_2O) of conductivity 100 μS/cm, purified through a Millipore SIMPAK0D2 purification system. The residual ions in the water are not relevant for the results. In fact, the ions released by BSA itself are much more than those residual in the water, as we show below. For example, at a rather low BSA concentration of 1 mg/ml, about 0.5 mM of ions are released by the proteins in solution, an amount much higher than any possible residual ion molarity coming from the water itself. Therefore, for all practical purposes, we can consider this water as essentially pure.

2.2 Bovine Serum Albumin

Bovine Serum Albumin (BSA) is a compact, small and relatively light protein, found in cow blood. It is a standard protein in the physicochemical literature, because of its stability against denaturation, its easy purification,

and because it is comparatively inexpensive. Many aspects of BSA chemistry and biology are well-known; for a review, see e.g. Ref. [1]. Presently, many commercial preparations of BSA are available on the market; we purchase all BSA used for this thesis as a lyophilized powder from Sigma, cat. A3059, batch # 108K1295.

In the next sections, we briefly overview the most relevant physicochemical properties of BSA, in order to install our experimental results in the appropriate context. We also touch on the biological function of serum albumin.

2.2.1 Molar Mass

The BSA molar mass M_{BSA}, as calculated from the amino acid sequence, is 66.3 kDa [1]. BSA is thus a medium-light protein; M_{BSA} is much higher than the mass of light inorganic substances such as carbon dioxide ($\sim$ 40 Da), but still much lower than the mass of other proteins, such as ATP synthase ($\sim$ 400 kDa [70]). The molar mass of BSA in solution is not exactly equal to the presented value, for several reasons. First of all, when BSA is dissolved in water, ionic association/dissociation takes place on the protein surface. In other words, the external protein residues can release or bind a certain number of water or salt ions ($\sim$ 10-100), thus altering the protein mass. Second, we have to consider the anionic association of medium-weight substances, such as fatty acids, a very common phenomenon for serum albumins. In total, these ligands can cause an increase of approximately 2 kDa. Thus, we have to distinguish the *dry mass* in vacuum from the slightly larger *wet mass* in solution.

In chapter 5, among our experimental results, we show some estimates of the wet mass, measured by static light scattering. To distinguish these experimental values from the tabulated one (wet or dry), we use the symbol M instead of M_{BSA}. In some datasets, M is slightly higher than the expected wet mass. This is a normal technical issue, due to a non-excellent sample purity, which nonetheless creates some difficulties in the calculation of other experimental results, such as the second virial coefficient. See chapter 5 for a thorough discussion of this point.

2.2.2 Shape and Dimensions

The exact three-dimensional atomic structure of BSA is not known for lack of high quality crystals. Nevertheless, we can estimate it in several ways. First of all, using low-resolution techniques such as Small Angle X-ray Scattering (SAXS), an approximate form can be fitted. We use the best shape, an oblate

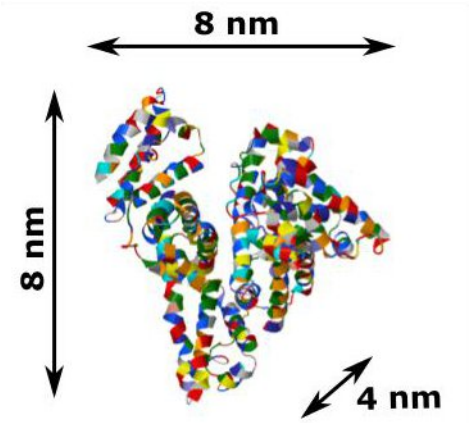

Fig. 2.1: Crystal structure of HSA in schematic view. From [71].

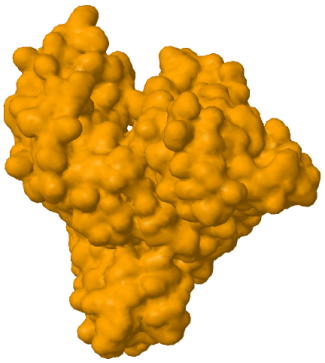

Fig. 2.2: Crystal structure of HSA showing surface accessible to water. From [71].

ellipsoid of semiaxes $a = 1.7$ nm and $b = 4.2$ nm [44]. Alternatively, we look at the fully resolved 3D structure of similar proteins, that have been crystallized. The most notable among them is Human Serum Albumin (HSA), the analogous molecule found in human blood. In Fig. 2.1 and Fig. 2.2, we see scaled representations of HSA, produced loading the atomic coordinates from the Protein DataBase [71] into the rendering program *Jmol* [72]. Fig. 2.1 is a schematic representation of HSA, with colors indicating the different residues. In Fig. 2.2, we see the effective surface accessible to water when HSA is dissolved in solution. As concluded by other researchers [73], this shape is approximately heart-like, with linear dimensions in agreement with the SAXS ellipsoidal estimate.

2.2.3 Volume and Effective Radius Estimates

The volume of BSA is an important property for the comparison of our experimental results with theoretical models. For instance, for the abscissae of most plots in Chapter 5, we use the BSA concentration c_p, expressed in milligrams of protein over milliliter of solution, whereas the colloidal theory prefers the BSA volume fraction ϕ, i.e. the volume occupied by BSA molecules over the total solution volume. We can switch between c_p and ϕ through a simple formula, in which the molecular volume of BSA, v, plays a key role,

$$\phi \quad = \quad v \cdot \frac{N_a}{\mathrm{M_{BSA}}} \cdot c_p \ ,$$

where N_a is the Avogadro constant.

Unfortunately, when BSA is dissolved in solution, the molecular volume becomes an ill-defined quantity. At first, we take the plain SAXS estimates for the protein dimensions and calculate the geometric volume of the ellipsoid, $v_{\text{geo}} = 126$ nm^3. We use this quantity to calculate the radius of a sphere with the same volume, $r_{\text{geo}} = (ab^2)^{1/3} = 3.11$ nm. Finally, it is possible to check the consistency of such an *effective radius* with a number of geometry-related experimental and theoretical techniques [8, 74]: ultraviolet absorption spectroscopy, dynamic light scattering, intrinsic viscosity, and small-angle x-ray scattering, together with Perrin factors [75], Simha factors [76, 77], and Hard-Spheres second virial coefficient calculations [78–80]. From all these methods, we find that r_{geo} and v_{geo} are clearly underestimations. The phenomenological effective radius is $r_{\text{phen}} \approx 3.33$ nm instead, and the molecular volume $v_{\text{phen}} \approx 155$ nm^3. The reasons for this discrepancy are still rather obscure, but an important role is surely played by the first water layer surrounding the protein, and maybe by the micromotions of some residues. Anyway, we use v_{phen} in the calculation of ϕ instead of v_{geo}, with a correction towards higher volume fractions of $\sim 25\%$.

2.2.4 Specific Volume

The specific volume quantifies the increase of volume of a solution upon addition of a certain mass of BSA. This property depends also on the aqueous environment, not only on the mass and volume of the dry protein. The difference is remarkable, because the solution could shrink or expand as a consequence of various physical and chemical processes involving the protein. As a prominent example, the first layer of water molecules around any protein, the so-called *hydration layer*, has a different structure and density from the bulk water. In principle, we should define

$$v_{\text{spec}} \quad := \quad \left(\frac{\partial V}{\partial m_p} \right) \quad ,$$

where V is the solution volume, and m_p the total weight of BSA in solution. This quantity should depend on the physicochemical parameters of the solution, such as protein concentration, salt concentration, temperature, and pressure.

However, this definition is not practical, for two reasons. First of all, the experiments in this thesis alone do not enable to calculate this quantity directly. Second, since the maximal protein concentration used is not very high (~ 150 mg/ml), the total amount of water in the hydration layer is small,

and v_{spec} is essentially constant. Therefore, we prefer a simpler definition,

$$v_{\mathrm{spec}} \quad := \quad \frac{v_{\mathrm{phen}}}{\mathrm{M_{BSA}}} \quad = \quad 1.40 \text{ ml/g}\,.$$

This choice is intuitively consistent, because we determine v_{phen} on the basis of experiments on aqueous solutions, and $\mathrm{M_{BSA}}$ is independent of protein and salt concentration, which are the two free parameters of this study.

2.2.5 Electrostatic Charge

Some residues of BSA, mostly the ones on the surface, become charged in solution by binding or releasing water or salt ions. The thermodynamic properties of BSA are mainly influenced by the total charge Q, the algebraic sum of its charged amino acids and the ionic ligands,

$$Q = \left\{ -N_{\mathrm{H^+}} + N_{\mathrm{OH^-}} - \sum_i Z_i\, N_i \right\} e \quad . \tag{2.1}$$

In this expression, $N_{\mathrm{H^+/OH^-}}$ is the number of hydrons/hydroxides released by the protein, N_i ($-N_i$) the number of other ions released (bound), and Z_i their charge in elementary charge units e. The ionic species, i, include both the protein counterions and the added salt ions. In physiological solutions, the net charge of BSA is $Q \approx -10e$ [1]. More generally, the value of Q in a certain solution depends on the chemical equilibria of all ionizable species (residues + water + salt). Experimentally, Q can be tuned by varying one or more of the following parameters: protein concentration, salt concentration, pH, and temperature. Theoretically, on the one hand, Q is not easily predictable from *ab initio* quantum chemistry calculations, while, on the other hand, it is a central requirement of any colloidal model.

We cannot determine the protein charge only by the results of this thesis. This quantity can be either assessed by complementary experimental methods, such as electrophoretic light scattering, or fitted by theoretical modelling.

2.2.6 Purity and Self-Association

BSA is purified from cow blood or, occasionally, cow placenta, after filtering out blood cells. The BSA powder used for this thesis has been prepared as follows: the cow plasma is heated at temperatures around 65 °C, at which globulins and other proteins denaturate and are filtered out (*heat-shock fractionation*); then, BSA is lyophilized. It is well known that any

BSA powder prepared in this way includes a $\sim 10\%$ amount of dimers and higher oligomers. Moreover, the BSA molecules are most probably bound to Na^+, Ca^{2+}, and Cl^- ions, since these are abundant in the physiological environment. These ions are probably transferred to the solution when BSA is dissolved in water. Finally, we expect about 1-2 fatty acids bound to every BSA molecule, causing a slight micro-heterogeneity, and the small increase in molar mass [1].

The fatty acids do not represent a real problem for our experiments, because of their small amount and size. The presence of dimers and heavier oligomers is a bigger issue though, because they interfere measurably with the light scattering of monomers. We reach this conclusion from the three following experimental evidences. First of all, from the static light scattering experiments, we obtain various independent estimates on the mean solute molar mass, which often turn out to be higher than the molar mass of a BSA monomer, M_{BSA}. Second, in the dynamic light scattering correlograms (see chapter 5), we often observe, in addition to the faster decay D_1 of BSA monomers, a slower decay, D_2. In particular, D_2 is sometimes half as fast as D_1, sometimes five or even ten times slower. However, as the Rayleigh ratio of the same samples is still independent of the scattering vector, Q, we guess that this slower decay be caused by middle-sized BSA oligomers (~ 10 molecules). Third, in a series of BSA solutions with YCl_3 inside the so-called Regime II, we explicitly see a strongly Q-dependent Rayleigh ratio and diffusion constant. This fact indicates that middle-sized BSA self-association is a feasible physical process. It is not clear whether those BSA self-associates are just an unwanted byproduct of the lyophilization, too small to be filtered away in the sample preparation, or they are (meta)stable in solution. Only in the first scenario would a better sample preparation solve the problem. Actually, some authors in the BSA literature, who mention explicitly this point, seem to be able to obtain completely monomeric solutions by adding a gel filtration chromatography (GFC) step in the preparation [81, 82]. Furthermore, the disturbances seen in our experiments seem to be partially random, which would be strange if the oligomers were genuinely (meta)stable. Therefore, it is probable that the disturbing medium-sized complexes are residual waste of the BSA purification procedure. On the basis of this conclusion, we emphasize that GFC would represent an important improvement in future studies.

2.2.7 Biological Function

BSA is the most abundant protein of cow blood plasma, and is highly soluble in polar solvents. Analogous protein are found in many other mammal

Table 2.1: Radii of the ions used in this work ("effective radii" in the original paper) [84]. Error bars are to be considered about 1 on the last digit, since the precise ionic radius depends on the environmental conditions.

Species	Ionic radius [Å]
Na^+	0.10
Ca^{2+}	0.10
Y^{3+}	0.09
Cl^-	0.18

species such as rats, dogs, and humans. In general, Serum Albumins have at least four important biological functions. The first one is to regulate the pH of the plasma, binding and releasing both H^+ and OH^-. The second one is to maintain the osmotic pressure of blood vessels. This is essential to regulate the amount of salts and nutrients taken up by the cells. Then, Serum Albumins are well known for their transport properties for a variety of substances that cannot flow freely in the blood. For instance, most fatty acids and about one half of calcium ions Ca^{2+} in the blood are bound to Serum Albumin [1]. Finally, Serum Albumin is an important antioxidant. It is thought to capture most of the reactive oxygen species (ROS) in the plasma, as well as potential ROS-generating species such as Cu^{2+} and Fe^{2+} ions [83].

2.3 Salts

In these light scattering experiments, we use three different salts, that are NaCl, $CaCl_2$, and YCl_3. We use three commercial salt preparations from Sigma, delivered as anhydrous powders: NaCl, cat. S9625 (purity $\geq 99.5\%$); $CaCl_2$, cat. C4901 (purity $\geq 96\%$); and YCl_3, cat. 451363 (purity $\geq 99.99\%$). We do not mix two salts in any solution, but only use one *or* the other. The maximum salt concentrations are 500 mM for NaCl, 167 mM for $CaCl_2$, and 83 mM for YCl_3. These values are below the respective solubility limits, thus we see no precipitates.

2.3.1 Differences among NaCl, $CaCl_2$, and YCl_3

The main difference among the salt species is the charge of the cation, respectively +1, +2, and +3. Geometrically, the ionic radii of the cations are comparable; the Cl^- anion is much larger (see Table 2.1). From a functional point of view, both NaCl and $CaCl_2$ are physiological salts, whereas YCl_3 is

not found in the blood.

Most relevantly, from a physicochemical point of view, the first two salt species, NaCl and $CaCl_2$, seem to produce similar effects in BSA solutions. We confirm this statement by light scattering. Basically, anticipating the discussion presented in chapter 5, these salts bind BSA only weakly, and their main effect is to increase the ionic strength of the solution at a mean-field level. Yttrium chloride behaves in a totally different way. Intensive studies are being carried out at present in this field, but it seems probable that BSA has at least a few very reactive binding sites for the Y^{3+} cation. This strong binding cannot be explained in mean-field schemes, and the meaning and numerical value of the ionic strength in solutions with YCl_3 remains questionable. Zhang *et al.* observe the phenomenon of protein re-entrant condensation only with this and similar trivalent-cationic salts, such as $AlCl_3$ and $LaCl_3$ [3]. As explained in chapter 5, light scattering confirms this special behaviour of YCl_3. We find results in agreement with Ref. [3] and subsequent studies. Moreover, the results fit generally well, at a semi-quantitative level, in the current conceptual framework underlying the re-entrant condensation phenomenon, albeit with a few surprises.

2.3.2 Meaning of the "salt-free" solutions

Finally, we discuss briefly, in the context of the salts used in our experiments, the meaning of the so-called "salt-free" solutions. Basically, with this expression we indicate certain samples, which we prepare as mixtures of pure water (in the sense explained above) and BSA powder only, without any additional salt powder. As mentioned above in section 2.2.6, however, the BSA powder includes a certain amount of physiological ions, because it is purified from cow blood plasma without thorough dialysis against pure water. Those ions are partially released again when the protein powder is dissolved in water, giving rise to the total protein charge discussed in section 2.2.5. Therefore, even the solutions without *added* salt are not really salt-free. Moreover, the concentration of released ions most probably increases with increasing BSA concentration, if not linearly, at least monotonically. Since both cations (e.g. Na^+, Ca^{2+}) and anions (e.g. Cl^-) are released, their total concentration cannot be linked directly to the total protein charge. Nevertheless, we can use equation (2.1) as a rough estimate for this property, assuming that the protein releases as many ions as its total charge in elementary units:

$$c_{\mathrm{TOT}} \quad \sim \quad \frac{|Q|}{e}\,\frac{c_p}{\mathrm{M_{BSA}}} \quad \sim \quad 0.15\; c_p \;\mathrm{mM},$$

where c_p is the BSA concentration measured in mass over solution volume, for instance mg/ml, and c_{TOT} is measured in moles of ions over solution volume. Albeit only a rule of thumb, this estimate is useful; for instance, a solution at $c_p = 10$ mg/ml possesses an intrinsic salt concentration of about 1 mM.

Chapter 3

Theory

Light Scattering is a powerful tool for studying liquid solutions of macromolecules. However, the experimental data are not as directly interpreted as in other methods, for instance optical microscopy. The reason for this state of affairs is threefold. First of all, scattering phenomena are naturally described in reciprocal space, rather than in real space. Second, coherent scattering is intrinsically an interference effect. Third, concentrated solutions are many-body systems with complex interparticle forces. Unfortunately, our human intuition for Fourier analysis, wave dynamics, and collective effects is rather poor. A solid theory is thus necessary for guiding our intuition in the interpretation of the experimental results. We provide an essential theoretical introduction to light scattering in this chapter. In the first part, we describe static light scattering (SLS). This is the easier scattering technique, because we only measure the average intensity of the scattered light. The other technique, dynamic light scattering (DLS), needs a slightly more involved explanation, which we outline in the second part of the chapter. Basically, DLS probes the temporal intensity fluctuations of the scattered light, tracking their origin in the diffusive dynamics of proteins.

We emphasize that this chapter does not replace any theory book. For more comprehensive analyses, the reader can refer e.g. to Refs. [85–89]. In particular, Ref. [88] is a modern and simple introduction for a novice reader. Ref. [87] could be useful for more expert researchers instead.

3.1 Static Light Scattering

Simple scattering is a physical process, in which one or more particles are illuminated by a weak radiation field and react accordingly, spreading the radiation also in directions different from the incoming one(s). It is a univer-

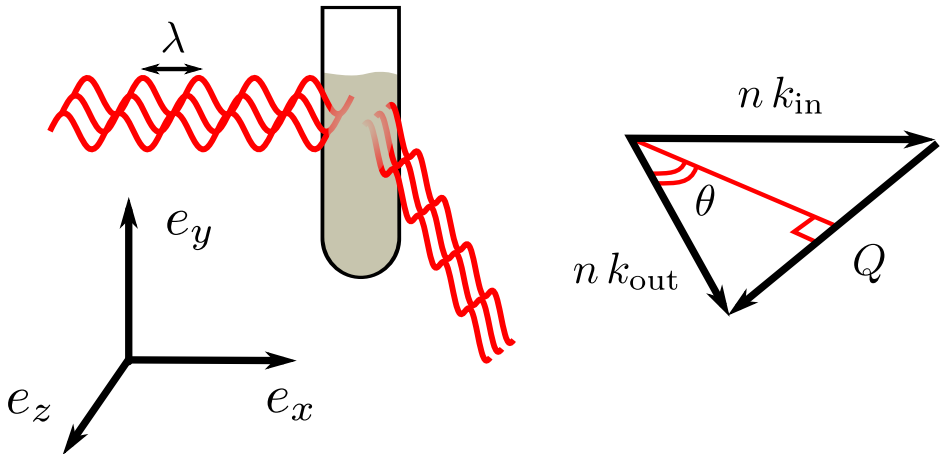

Fig. 3.1: Simple representation of the scattering geometry. Legend: e_x, e_y, e_z, right-hand cartesian coordinate system; k_{in}, incoming wave vector; k_{out}, outgoing wave vector; θ, half scattering angle; Q, scattering vector.

sal phenomenon; the scattering particles must be generically coupled to the radiation, but need not possess any wavelength-specific property, such as absorption or emission edges, fluorescence, or phosphorescence. In our actual experimental setup, the incoming radiation is a plane and monochromatic (red) light wave, which falls on a liquid solution of BSA (protein) molecules. The matter-field coupling is provided by the valence electrons of all atoms that build up the proteins. These electrons react to the weak light field, in first approximation, as classical electric dipoles, oscillating forcedly, in synchrony with the impinging light [90, 91]. In turn, the collective motion of charged particles produces a wave of scattered light, of toroidal form. Actually, we detect the scattered light in a two-dimensional plane instead of the whole space, so we must project the toroid onto the detection plane. This projection depends on the light polarization; in the simplest case, the resulting scattered field is circularly symmetric. We sketch the geometry in Fig. 3.1.

We start our overview of SLS by defining the relevant quantities. Let E_{in} be the electric field of the incoming wave, of the simplest form,

$$E_{\text{in}} \;:=\; E_0 \cdot \exp\Big[2\pi i\nu t - ikx\Big]\, e_y \quad \text{, with} \quad k_{\text{in}} := k\, e_x \,.$$

The plain wave propagates along the unit vector e_x and is polarized along e_y. We call I_0 its intensity, λ its vacuum wavelength, and $k_{\text{in}} =: k\, e_x$ its vacuum wave vector. We put the detector for the scattered light in the xz-plane, at an angle 2θ from e_x. The electric field of the outgoing light wave be E_{out}, its intensity I_{out} or I, and its vacuum wave vector k_{out}. The scattering is *elastic*,

i.e. there is no observable photon energy shift. The *scattering vector* is defined by $Q := n(k_{\text{out}} - k_{\text{in}})$, where n is the refractive index of the solution[1]. For elastic scattering, it has modulus:

$$|Q| \quad = \quad \frac{4\pi\, n}{\lambda} \cdot \sin\theta \quad ,$$

as we see in Fig. 3.1. Finally, let the solution have temperature T, protein concentration c_p, and salt concentration c_s. We discuss the units of protein concentration later on. The unit for salt concentration is always

$$\text{mM} \quad := \quad \frac{\text{millimoles of salt}}{\text{volume of } \textit{water} \text{ in liters}} \, .$$

The theory of SLS can be conceptually divided in two steps. In the first step, we derive the scattered field and intensity as a function of the incoming field. Then, we calculate the effect of the experimental parameters T, c_p, and c_s, in the long wavelength limit. In the following sections, we start from the first step. First of all, we split the illuminated volume V in infinitesimal elements, and calculate the scattered field of a single element. Second, assuming full coherence of the light field, we add up the scattered electric field of all elements, finding the total scattered field E_{out}. Third, we calculate the detected intensity by norm-squaring E_{out}. Then, we proceed to the second step, providing a thermodynamic description of our samples, and the effect of thermodynamics on light scattering. For approaches similar to mine, the reader should refer to Refs. [15–17, 92].

3.1.1 Scattered Field of an Infinitesimal Element

The scattered field of a single, infinitesimal volume element of the protein solution can be studied by means of classical electromagnetism. In first approximation, the reaction of a neutral element to a nonresonant, incoming light wave is the induced oscillating electric dipole [91]. In our simple geometry – polarization along e_y, detection in the orthogonal space – the scattered field generated by the oscillating dipole is a circular wave [15, 16],

$$E^{\text{el}}_{\text{out}}(r) = \frac{2\pi n\, E_0}{\lambda^2} \cdot \Delta n \cdot \frac{e^{iQr}}{|r|} \quad , \tag{3.1}$$

[1] Actually, in the light scattering literature, different definitions of Q are used. In particular, it is not clear whether the refractive index of the solution, n, or that (smaller) of the pure solvent (mostly water), n_w, is to be used. This is no practical problem for our experimental results, for the BSA concentration is low enough that $n \approx n_w$ in all relevant equations.

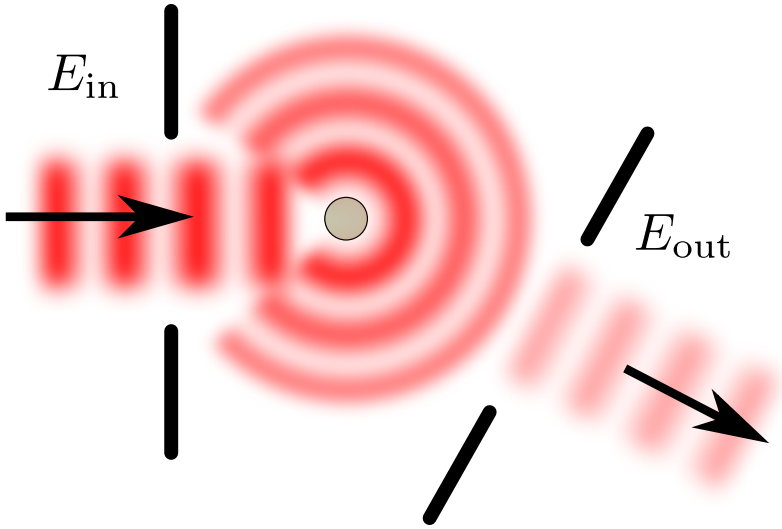

Fig. 3.2: Schematic illustration of the simple scattering from a single volume element, in electric dipole approximation. See equation (3.1).

where r is the vector connecting the scattering element to the detector. $\Delta n := n - n_w$, where n_w is the refractive index of the solvent, water. For the sake of brevity, we omit the time-dependent factor $e^{2\pi i\nu t}$ in the calculations. Strictly speaking, (3.1) only holds if $n \approx n_w$, but this condition is always verified for our samples. Moreover, a trivial geometrical correction to (3.1) would be required in the real experimental setup, related to the polarization of the incoming light. However, this correction is automatically included in the calibration of the SLS instrument, thus we will omit it for the sake of clarity.

We illustrate schematically the geometry of (3.1) in Fig. 3.2. The reader should remember that the circular wave, at fixed detector angle and large distances, can be safely approximated by a plane wave.

3.1.2 Total Scattered Field

Equation (3.1) expresses the scattered field of a single volume element. In order to calculate the total scattered field, we have to sum up the contributions of all volume elements. Since our light source is a spatially very coherent laser, we can simply sum all $E^{\text{el}}_{\text{out}}(r)$. The total scattered electric field becomes

$$E_{\text{out}} = \int_V E^{\text{el}}_{\text{out}}(r)\,\mathrm{d}r = \frac{2\pi n\, E_0}{\lambda^2 |r|} \cdot \Delta n \cdot \int_V \rho(r) e^{iQr}\,\mathrm{d}r \quad , \tag{3.2}$$

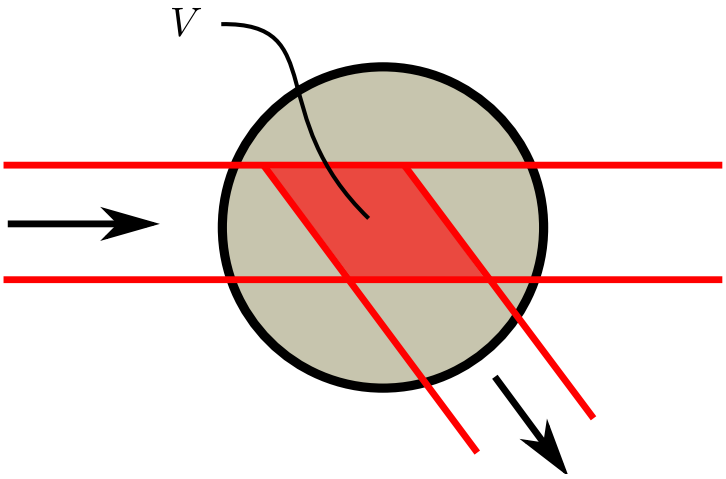

Fig. 3.3: Schematic illustration of the scattering volume.

where $\rho(r)$ is the density of valence electrons at position r. In this formula, we already approximated the circular wave by a plain wave, by taking $|r|$ out of the integral. The integral is performed over the so-called *scattering volume*, V, which is the volume of protein solution that sees contemporaneously the light source and the detector. V is much smaller than the total solution volume and, as we show in Fig. 3.3, also angle-dependent.

From (3.2), the reader can see why light scattering is so easily managed in reciprocal space: the experimental setup (fixed angle, coherent fields) corresponds to a spatial Fourier transformation.

3.1.3 Detected Intensity and Rayleigh Ratio

Finally, we have to norm-square the total scattered field to find the scattered intensity I_{out}, which is the quantity detected by the photodiodes. Actually, we normalize I_{out} by some trivial factors, by defining the so-called *Rayleigh ratio*,

$$R := I_{\text{out}} \cdot \frac{|r|^2}{I_0 V} \quad .$$

The advantage of R over I_{out} is that it does not depend on geometrical details of the experimental setup, such as the power of the light source. It also counterbalances the angle dependence of the scattering volume (see Fig. 3.3). We can easily calculate the Rayleigh ratio using (3.2),

$$R \quad = \quad \frac{4\pi^2 n^2}{\lambda^4} V \, (\Delta n)^2 \quad \cdot \quad \left| \frac{1}{V} \int_V \rho(r) e^{iQr} \, \mathrm{d}r \right|^2 \quad .$$

The first part of this expression represents the coupling strength between field and matter, i.e. the optical contrast. The second part describes the interference among scattering elements instead. We can rewrite it as follows:

$$\frac{1}{V^2} \iint \rho(r_1)\rho(r_2)\, e^{iQ(r_1-r_2)}\, \mathrm{d}r_1\, \mathrm{d}r_2 \quad ,$$

which makes the interference effect intuitive. In fact, if two elements are close to each other, $r_1 - r_2$ is small and the scattered fields interfere. If two elements are far apart, the complex exponential oscillates rapidly, and the interference is smeared out to zero. Of course, the length scale for the interference is set by the inverse scattering vector, Q^{-1}. In our case, the protein molecules (~ 3 nm) are much smaller than $Q^{-1} \sim 100$ nm. Therefore, two volume elements belonging to the same protein always interfere constructively, because $e^{iQ(r_1-r_2)} \to 1$. However, the interference pattern coming from different proteins is still complicated. We can write it as follows:

$$\frac{NV_P^2}{V^2} \quad \cdot \quad \left[\frac{1}{N} \sum_{i,j=1}^{N} e^{iQ(\bar{r_i}-\bar{r_j})}\right] \quad ,$$

where V_P is the volume of one protein, $\bar{r_i}$ is the mean position of the i-th molecule, and N is the number of molecules. The quantity in square brackets is called *static structure factor*, $S(Q)$, and is connected to the probability distribution of distances between two proteins.

Finally, we notice that the factor $\rho(r_1)\rho(r_2)$ in the interference probes the joint probability of finding a scattering element at the position r_1, and one at r_2. Since the solution is homogeneous, this is equivalent to the probability of finding two scattering elements at a distance $r := |r_1 - r_2|$. In turn, this probability depends on the local fluctuations of element numbers at the typical length scale r, that is $(\delta N)^2(r)$. After the Fourier transform, this is in turn proportional to the fluctuations in the particle number at an inverse length scale $Q = 2\pi/r$,

$$\frac{1}{V^2} \iint \rho(r_1)\rho(r_2)\, e^{iQ(r_1-r_2)}\, \mathrm{d}r_1\, \mathrm{d}r_2 \quad \propto \quad (\delta N)^2\,(Q) \quad .$$

The scattering measured at low Q correspond to the particle fluctuations at macroscopic length scales, which are described by the usual thermodynamics. For a rigorous treatment of this connection between microscopic and macroscopic properties, we refer to the derivation of the compressibility relation in Ref. [33].

3.1.4 Thermodynamics of SLS

We observe, in most SLS experiments for this thesis, that the Rayleigh ratio is independent of Q. The reason is that the light wavelength is much larger than the typical length scale at which the distance distribution changes significantly. Thus, as mentioned in the last section, the interference factor is proportional to the macroscopic number fluctuations in the sample:

$$\frac{1}{V^2}\iint \rho(r_1)\rho(r_2)\, e^{iQ(r_1-r_2)}\,\mathrm{d}r_1\,\mathrm{d}r_2 \quad \longrightarrow \quad (\delta N)^2\,(Q\to 0) \quad .$$

In turn, the scattering elements number fluctuations causes analogous fluctuations of the refractive index n, so that we find, after the appropriate normalization,

$$\left|\frac{1}{V}\int_V \rho(r)e^{iQr}\,\mathrm{d}r\right|^2 \quad \longrightarrow \quad \left\langle\left(\frac{\delta n}{\Delta n}\right)^2\right\rangle \quad . \tag{3.3}$$

In this expression, found originally by Einstein by another route [14], (δn) represents the thermal fluctuations of the refractive index, in the long wavelength limit. Equation (3.3) states that their variance is directly proportional to the scattering intensity at vanishingly small Q vectors. We can apply the classical theory of fluctuations to this result (see e.g. Ref. [93]). First of all, we select the free enthalpy G as a thermodynamic potential. The refractive index is a function of the natural variables of G, that are temperature T, pressure P, and particle numbers N_i of all species in solution. In principle, these would include water molecules (N_w), proteins (N_p), counterions (N_c), and, if present, salt ions (N_+, N_-). However, the theory for multi-solute mixtures would be, predictably, quite complex, and a comparison with experiments would require too large a number of tunable parameters. For this reason, we restrict ourselves to a single-solute mixture. We consider the proteins as the solute, and group the other species in the other component, the solvent. We outline the multi-component treatment only briefly, in section 3.1.7.

The refractive index is an intensive quantity. Therefore, it does not depend explicitly on both N_w and N_p, but only on their ratio. Equivalently, we can fix N_w and implicitly consider all other extensive quantities, such as N_p and V, as ratios to N_w. We can expand the fluctuations of n by the chain rule,

$$\mathrm{d}n = \left(\frac{\partial n}{\partial T}\right)\mathrm{d}T + \left(\frac{\partial n}{\partial P}\right)\mathrm{d}P + \left(\frac{\partial n}{\partial N_p}\right)\mathrm{d}N_p \quad .$$

At this point, we assume that n depends on T and P only through the mean density ρ. Without repeating the whole derivation given e.g. in Ref. [17], we

only state the final result,

$$\left\langle (\delta n)^2 \right\rangle = kT \left[\frac{\kappa}{V} \left(\frac{\partial n}{\partial \ln \rho} \right)^2 + \left(\frac{\partial n}{\partial N_p} \right)^2 \left(\frac{\partial N_p}{\partial \mu_p} \right) \right] \quad , \tag{3.4}$$

where κ is the total compressibility of the solution and μ_p the chemical potential of a protein molecule. The right-hand side of (3.4) consists of two parts. The first part refers to density fluctuations, and is present also in pure liquids, the so-called Einstein-Smoluchowski formula [13,14]. The second part is only present in liquid mixtures instead. It is related to fluctuations of the protein concentration, and describes the interesting thermodynamical features of the system. Combining (3.3) and (3.4), the Rayleigh ratio of a protein solution can be written as follows:

$$R = \frac{4\pi^2 n^2}{\lambda^4} \, kT \cdot \left[\kappa \left(\frac{\partial n}{\partial \ln \rho} \right)^2 + V \left(\frac{\partial n}{\partial N_p} \right)^2 \left(\frac{\partial N_p}{\partial \mu_p} \right) \right] \quad . \tag{3.5}$$

3.1.5 Solvent Subtraction

As mentioned in the last paragraph, equation (3.5) includes a contribution that is also present in pure liquids. In order to focus on the interesting physics of proteins, we try to subtract this contribution. For this purpose, we measure the scattering intensity of the solvent alone, independently. The solvent Rayleigh ratio is

$$R_w = \frac{4\pi^2 n_w^2}{\lambda^4} \, kT \cdot \kappa_w \left(\frac{\partial n_w}{\partial \ln \rho_w} \right)^2 \quad ,$$

where κ_w is the solvent compressibility and ρ_w its density.

At this point, we have to introduce a key approximation. There is no guarantee that R_w equals exactly the density-related part of R. Anyway, for not too small protein concentrations, 0.1 mg/ml $\lesssim c_p$, the density-related part of the scattering is much smaller that the concentration-related part. Besides, at not too high protein concentration, $c_p \lesssim 300$ mg/ml, the compressibility, the refractive index, and its increment in the solution are not very different from those in the solvent,

$$\left(\frac{\partial n}{\partial \ln \rho} \right) \approx \left(\frac{\partial n_w}{\partial \ln \rho_w} \right) \text{ , } n \approx n_w \text{ , and } \kappa \approx \kappa_w \text{ .}$$

Thus, the density-related scattering subtraction is only a minor correction, and we can indeed approximate it with R_w, that is

$$n^2 \kappa \left(\frac{\partial n}{\partial \ln \rho} \right)^2 \quad \approx \quad n_w^2 \kappa_w \left(\frac{\partial n_w}{\partial \ln \rho_w} \right)^2 \quad .$$

Therefore, we obtain the net Rayleigh ratio,

$$R - R_w = \frac{4\pi^2 n^2}{\lambda^4} \, kTV \, \left(\frac{\partial n}{\partial N_p}\right)^2 \left(\frac{\partial N_p}{\partial \mu_p}\right) \quad . \tag{3.6}$$

In the rest of the thesis, with a slight notational abuse, we will call this quantity simply R.

3.1.6 Classic SLS Equations

Formula (3.6) describes the interesting part of the Rayleigh ratio, due to protein concentration fluctuations, as a function of the thermodynamic and optical properties of the solution. A similar expression is used typically in the theory of Small-Angle X-ray or Neutron Scattering (SAXS/SANS). In this thesis, as common in the SLS literature, we prefer to recast (3.6) in a more comfortable form, by introducing the protein concentration c_p in lieu of the protein number N_p. Using the concentration units discussed in Chapter 4, i.e. milligrams of BSA over milliliter of solution, we can write

$$R = \left[\frac{4\pi^2 n^2}{\lambda^4 N_a} \left(\frac{\partial n}{\partial c_p}\right)^2\right] \cdot M_p \cdot kT \cdot \left(\frac{\partial c_p}{\partial \mu_p}\right) \quad ,$$

where N_a is the Avogadro constant and M_p the protein molar mass. The quantity in square brackets is an optical constant of the solution, and we can express it more concisely,

$$K \quad := \quad \frac{4\pi^2 n^2}{\lambda^4 N_a} \left(\frac{\partial n}{\partial c_p}\right)^2 \quad .$$

Moreover, the derivative $\partial_{\mu_p} c_p$ is a rather unusual quantity in chemical physics, in comparison with its reciprocal $\partial_{c_p} \mu_p$, which is easily found from the equation of state (EOS). In the ideal solution, $\partial_{c_p} \mu_p = kT/c_p$. Therefore, we calculate the reciprocal of R, and normalize it with respect to the ideal solution:

$$\frac{Kc_p}{R} = \frac{1}{M_p} \cdot \frac{\partial_{c_p} \mu_p}{(\partial_{c_p} \mu_p)_{\text{ideal}}} \quad . \tag{3.7}$$

This formula is the central result of our overview of the SLS theory. It is sometimes referred to as the *Debye equation* or the *Zimm equation*, but the latter term should only be used when the Q-dependence of Kc/R is considered [15, 16, 94].

SLS in the Dilute Limit

We use equation (3.7) as a link between theory and experiments, because it is directly comparable with measured quantities (R, c_p, n, etc.), but also of simple interpretation. If a solution is dilute enough, it follows the ideal EOS, and static light scattering yields the molar mass of the protein molecules,

$$\lim_{c_p \to 0} \frac{K c_p}{R} = \frac{1}{M_p} \quad . \tag{3.8}$$

In this case, the total scattering intensity is just the sum of the scattering intensities of the single solute (i.e. protein) molecules. Since no interference between different molecules is left, we call this situation *single-body SLS*. Of course, the field scattered by single electric dipoles inside a molecule still interfere constructively.

In principle, formula (3.8) is only valid for infinitely dilute solutions, but this criterion is of no practical use. In real experiments, we always work with finite solute concentrations. The actual range of validity of this equation is thus limited by the following effects. At too low c_p, the signal-to-noise ratio becomes too unfavourable; at too high c_p, the scattering ratio changes from M_p^{-1}, because of the interparticle interactions (see next section). In the experiments performed for this thesis, the lower boundary is located at $c_p \sim 10^{-1}$ mg/ml and the upper one around $c_p \sim 1$ mg/ml. The only exception is the dataset without added salt, in which the interparticle interactions are visible even at the lowest concentration, $c_p \approx 0.1$ mg/ml; in this case, formula (3.8) is only valid as an asymptotic limit.

In most biophysical studies in the literature, the molar mass of a protein is actually determined by other methods, such as mass spectroscopy. This point notwithstanding, we must emphasize that every SLS study on protein solutions should include some measurements at small protein concentrations, in order to check whether the measured molar mass agrees with previous estimations. In other words, equation (3.8) can be used as a self-consistency check, to be done before trying to interpret any experiment on more concentrated solutions. A discrepancy between the expected M_p and the measured one is a clear warning sign for the good experimentalist, for the very applicability of the theoretical scheme may be in danger because of technical problems. Such problems are quite common indeed and include low purity of the solution, self-association of proteins, and presence of proteases. We discuss this difficulty, for our solutions, in Chapter 5.

Interaction Effects in SLS: Virial Expansion

If a solution is too concentrated for the single-body SLS formula to hold, we must start to take into account the effect of interparticle forces. From a theoretical point of view, we can expand the solution EOS in powers of c_p, using the so-called *osmotic virial expansion* (see e.g. Refs. [30, 32, 33, 93]). The first-order corrected formula is

$$\frac{Kc_p}{R} = \frac{1}{M_p} + 2\,B_2\,c_p + \dots \quad , \tag{3.9}$$

where B_2 is called *osmotic second-virial coefficient*. B_2 is an indicator of the dominant effective interaction between two proteins. Interactions among three and more particles are neglected in this approximation; we thus call this case *two-body SLS*. A positive B_2 hints at an overall repulsion, a negative one at an attraction. In some situations, generally called θ-condition, B_2 vanishes, and the solution behaves similarly to an ideal one.

Equation (3.9) is valid in a wider range of solute concentrations than its single-body counterpart, equation (3.8), because interactions are, if only roughly, taken into account. From a practical point of view, we can locate the upper c_p boundary in a series of SLS measurements at increasing protein concentration, where the measured $Kc/R\,(c_p)$ curve starts to differ from an affine function, or to show a nonzero curvature. Among the SLS results presented in this thesis, we observe this change already at $c_p \approx 4$ mg/ml in the solutions without added salts; at much higher protein concentration, if at all, in all other datasets.

The second virial coefficient is a central quantity for microscopic models of protein solutions, for it is the easiest and most natural way to classify nonideal systems. From a statistical point of view, the first-order virial expansion corresponds to taking into account two-body effects, and neglecting higher-order correlations. Of course, theoreticians do not use mass-molar concentration units, but rather protein number over volume; B_2 is defined by:

$$B_2 = -\frac{1}{2}\int_V \left[e^{-\beta u(r)} - 1\right] \mathrm{d}r \quad , \tag{3.10}$$

where $\beta := (kT)^{-1}$, u is the energy potential between two particles, and r the vector between their positions. We can convert the B_2 measured by SLS to the theoretical units, used in the definition (3.10), by an apparently trivial normalization,

$$B_2(\textbf{t.u.})\ [\mathrm{m}^3] \ = \frac{M_p^2}{N_a} \cdot B_2(\textbf{SLS})\ [\mathrm{mol\ ml/g}^2] \quad , \tag{3.11}$$

Table 3.1: Qualitative summary of the possible values of b_2 and the corresponding interpretations.

b_2	Interpretation
$b_2 > 1$	Repulsive forces dominate.
$0 < b_2 < 1$	Attractive forces dominate, except for the steric repulsion (excluded volume).
$b_2 < 0$	Attractive forces dominate.

where **t.u.** stays for "theoretical units". We stress that the theoretical definition of B_2, equation (3.10), includes the effective intermolecular potential u. Unfortunately, our experimental data are in too narrow a range of scattering vectors to recover the full functional form of u. However, when we present the experimental results, we list the B_2 in theoretical units, so the interested theoretician can compare her predictions with our data.

We cannot omit a last remark on the second virial coefficient of hard spheres solutions (HS). The latter are used by theoreticians as a null-model for complicated objects, and consist in a single-solute solution of imaginary, uncharged spherical particles that repel each other by elastic contact collisions. The second virial coefficient of HS is simply

$$B_2^{\mathbf{HS}}(\mathbf{t.u.}) = +4\, v\,, \tag{3.12}$$

where v is the molecular volume. In the colloid physics community, It is also common to express the virial coefficient as ratio to the $B_2^{\mathbf{HS}}$ of the best hard sphere that models the colloid. In our case, since the effective volume of BSA is $v_{\text{phen}} = 155\ \text{nm}^3$, this ratio can be easily computed as follows:

$$b_2 := \frac{B_2(\mathbf{t.u.})}{4\, v_{\text{phen}}} \quad . \tag{3.13}$$

As a rough estimate, a so-normalized second virial coefficient smaller than one indicates the action of attractive interactions, for instance van der Waals forces. Of course, these are less important than the repulsive one as long as $b_2 > 0$ [95]. We summarize the latter conclusions in Table 3.1.

Further Effects of Interactions on SLS: Osmotic Compressibility

In concentrated solutions, many-body effects arise. The limit of validity of the first-order virial expansion is set, in SLS experiments, by the emergence of a non-affine dependence in $c_p \mapsto Kc/R$. Equation (3.7) requires the full

chemical potential of the solution. However, we can recast it in a slightly different form, which makes only use of the osmotic EOS, i.e. the protein osmotic pressure Π as a function of c_p and T. We write this result without explicit proof:

$$\begin{aligned} \frac{Kc_p}{R} &= \frac{1}{N_a kT c_p \, \chi_T} \\ &= \frac{1}{N_a kT} \cdot \left(\frac{\partial \Pi}{\partial c_p} \right)_T \quad . \end{aligned} \tag{3.14}$$

We introduce here the osmotic isothermal compressibility χ_T, defined as follows:

$$\chi_T := \frac{1}{c_p} \left(\frac{\partial c_p}{\partial \Pi} \right)_T = \left[c_p \left(\frac{\partial \Pi}{\partial c_p} \right)_T \right]^{-1}$$

The osmotic compressibility of the ideal solution is simply given by

$$\begin{aligned} \chi_{T0} &= (\Pi_0)^{-1} \\ &= \frac{M_p}{N_a \, c_p \, kT} \quad , \end{aligned} \tag{3.15}$$

where Π_0 is the osmotic pressure of the ideal solution. Combining (3.14) with (3.15), we can quickly calculate the normalized compressibility from the SLS measurements,

$$\frac{\chi_T}{\chi_{T0}} = \left[M_p \, \frac{Kc_p}{R} \right]^{-1} \quad , \tag{3.16}$$

and vice versa,

$$\frac{Kc_p}{R} = \frac{1}{M_p} \, \frac{\chi_{T0}}{\chi_T} \quad . \tag{3.17}$$

The osmotic compressibility can be defined theoretically in a fashion similar to the expression (3.10) for the second virial coefficient, including the interparticle potential u. However, we will not pursue that goal in this thesis. We will provide a solid set of experimental data for χ_T instead, at various c_p and c_s. Those results, which we present in Chapter 5, can be easily compared by the interested theoretician with her predictions.

3.1.7 Remarks on Multi-Component SLS

In a previous section, in the derivation of the expression (3.4), we mention the possibility of an SLS theory of multi-solute solutions. Such a theory would be very useful for our results indeed, for the samples do not only include

BSA and water, but also counterions, salt cations, and salt anions. Even worse, it is highly probable that, in some of the samples, self-association of BSA takes place, so that the solute species include also BSA dimers, trimers, and heavier oligomers. Since most of our results can be described satisfactorily by a single-solute solution model, we do not use any multi-solute SLS theory operatively. Nevertheless, in this section, we outline briefly the conceptual idea behind such a theory, so that it can be exploited with more ease in further research studies. The backbone of our treatment is taken from Stockmayer [17].

The starting point of our short overview is equation (3.4). As already mentioned, only the second part of this expression is an interesting property of the solution, the first part being present, in comparable magnitude, also in the pure solvent. The key ingredient is thus the quantity

$$\mathcal{X}_{pp} \quad := \quad \left(\frac{\partial n}{\partial N_p}\right)^2 \quad \cdot \quad \left(\frac{\partial N_p}{\partial \mu_p}\right) \quad ,$$

which is made up of an optical part, on the left, and a purely thermodynamic one, on the right. We introduce a more convenient notation, in which the protein component is denoted by the number 1,

$$\Psi_1 \; := \; \left(\frac{\partial n}{\partial N_p}\right) \; , \quad a_{11} \; := \; \left(\frac{\partial N_p}{\partial \mu_p}\right) \; , \text{and} \quad \mathcal{X}_{11} := \mathcal{X}_{pp} \; .$$

If the solution has more than one solute, say N solutes, we can simply replace Ψ_1 by an N vector, and a_{11} by a symmetric $N \times N$ matrix, or second-rank tensor,

$$\Psi_i \; := \; \left(\frac{\partial n}{\partial N_i}\right) \; , \quad a_{ij} \; := \; \left(\frac{\partial N_i}{\partial \mu_j}\right) \quad \text{with} \quad i, j = 1..N \; .$$

In the partial derivatives of Ψ_i, we keep all particle numbers except N_i fixed. Analogously, in a_{ij}, we keep fixed all chemical potentials except μ_j. We recall that the solvent particle number is always fixed. The SLS property, $\mathcal{X}_{11}$, becomes a trace:

$$\mathcal{X}_{11} \quad \longrightarrow \quad \mathcal{X} \; := \; \Psi_i \, a_{ij} \, \Psi_j \; ,$$

where we use the Einstein convention on repeated index summation. The substitution $\mathcal{X}_{11} \to \mathcal{X}$ is the only relevant change between the single-solute and the multi-solute SLS theories. This result might sound a bit too optimistic, but $\mathcal{X}$ really describes a variety of experimental situations. For instance, we sketch briefly two cases in the next paragraphs, protein dimerization and protein-salt ion complexation.

Protein Dimerization

Both a protein and its dimer are present in solution, with particle numbers N_1 and N_2, respectively. The SLS trace is given by

$$\begin{aligned}\mathcal{X} &= \Psi_1^2 a_{11} + 2\Psi_1\Psi_2 a_{12} + \Psi_2^2 a_{22} \\ &= \Psi_1^2 \ [a_{11} + 4a_{12} + 4a_{22}] \quad ,\end{aligned}$$

where we use in the last passage the tentative, approximate equivalence $\Psi_2 = 2\Psi_1$. If the protein tend to dimerize, a_{11} is large; a big increase in N_1 is needed in order to vary the chemical potential μ_1 significantly (bunching). At the same time, if only dimers, but no trimers and higher oligomers are present, then a dimer has to repel both a monomer and another dimer. This means that a_{12} and a_{22} are quite small. Thus, a careful balance between these two trends controls the dimerization. If the number of monomers and dimers can be monitored by another technique, SLS can yield a good amount of information on the thermodynamics of this phenomenon.

Protein-Salt Ion Complexation

The solution is made of a protein, in number N_p, and a small salt ion, e.g. a cation in number N_+. The SLS trace becomes

$$\begin{aligned}\mathcal{X} &= \Psi_p^2 a_{pp} + 2\Psi_p\Psi_+ a_{p+} + \Psi_+^2 a_{++} \\ &\approx \Psi_p^2 \ \left[a_{pp} + 2\,\frac{\Psi_+}{\Psi_p}\,a_{p+}\right] \quad .\end{aligned}$$

We neglected the last summand, because $\Psi_+ \ll \Psi_p$. However, we cannot apply the same line of reasoning to the second summand. In fact, as the protein and the salt ion tend to complexate, i.e. attract one another, the thermodynamic coefficient a_{p+} is large, and could compensate the smallness of the ratio Ψ_+/Ψ_p. Also in this case, if N_p and N_+ can be determined by an independent measurement, SLS can yield valuable thermodynamic information. In particular, this phenomenon should be studied, if possible, at a wavelength resonant with one of the components. A small change in λ would alter the relative weight of the two summands in $\mathcal{X}$, enabling the separate determination of a_{pp} and a_{p+}. The same principle is used in Anomalous Small-Angle X-ray Scattering (ASAXS).

3.2 Dynamic Light Scattering

Dynamic light scattering is the natural complementary technique of static light scattering. It focuses on the temporal properties of the scattered inten-

sity, which is not simply averaged anymore, but detected time-resolved by one or more very fast photon counters (APDs, see Chapter 4). In the following sections, we outline the theory that connects the time-resolved intensity to its physical origin, the diffusive dynamics of the proteins in solution.

Our explanation proceeds as follows. First of all, we present the working principle of DLS and the conceptual chain that links the scattered intensity to the BSA diffusion constant. Second, we touch on the theory of translational, macromolecular diffusion in liquids, including the meaning of the diffusion kernel. We also explain why the kernel simplifies to a diffusion constant. Third, we discuss the fit strategies used in the DLS data analysis, commenting on their advantages and drawbacks. Finally, we give a short overview on the Generalized Stokes-Einstein and Svedberg equations, which we exploit later for the interpretation of the DLS results. As a general orientation, we follow the treatment given by Nägele [89].

3.2.1 Working Principle

DLS is based on the time-resolved measurement of the scattered intensity, $I(t)$. We see an example of this function, taken from a real measurement, in Fig. 3.4. Evidently, $I(t)$ fluctuates around its mean value, $\langle I \rangle$, which we already measure by SLS. The fluctuations of $I(t)$ are caused by three simultaneous physical processes, classical laser fluctuations, photon statistics, and protein diffusive dynamics. First, as an example of the classical fluctuations, we plot in Fig. 3.4 a low-pass filtered fit of $I(t)$, the red line. We are not interested in this contribution; in the data analysis, we can eliminate it by means of a so-called *monitor diode*. In essence, we normalize out these fluctuations by splitting the laser beam in two parts and measuring the incident intensity with a slow-response detector (see Chapter 4). Second, the photon noise contribution is only important in two limiting cases, when the intensity of the scattered light is very low or very high [96]. We avoid both situations by means of a variable intensity attenuator, as explained in Chapter 4. Therefore, our DLS signal can be considered as entirely due to the third contribution, the protein diffusion. In the following, we focus on this contribution. First of all, we will define the key concept, the autocorrelation. Then, we explain the connection between the protein diffusion and the intensity fluctuations.

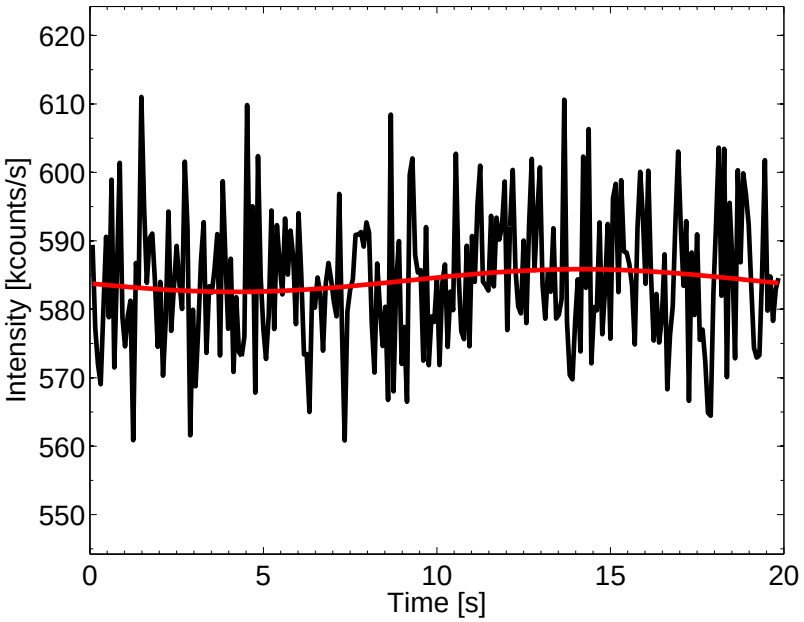

Fig. 3.4: Typical time-resolved scattered intensity. The ordinate axis units are the thousands of photons detected in one second. The red line is a low-frequency fit of the smoothed data, and shows the slow, classical intensity fluctuations of the laser power. The black line is the fully fluctuating scattering intensity instead.

Intensity Autocorrelation

The most natural and common way to study random fluctuations in a time-resolved manner is the *intensity autocorrelation function*,

$$g_I(t) \quad := \quad \frac{\langle I(t+t_0) \cdot I(t_0) \rangle}{\langle I \rangle^2} \; . \tag{3.18}$$

We plot an example of intensity autocorrelation in Fig. 3.5, showing the decay due to the protein diffusion. In the next sections, we call such a curve also *correlogram*. From an experimental point of view, the correlograms are calculated from the time-dependent intensity by a hardware electronic box, called autocorrelator (see Chapter 4). Roughly speaking, g_I quantifies the degree of causal correlation between the scattered intensity at a time, say t_0, and the scattered intensity at a later time $t_0 + t$. For instance, if nothing happens in the protein solution in the meanwhile, the scattered intensity

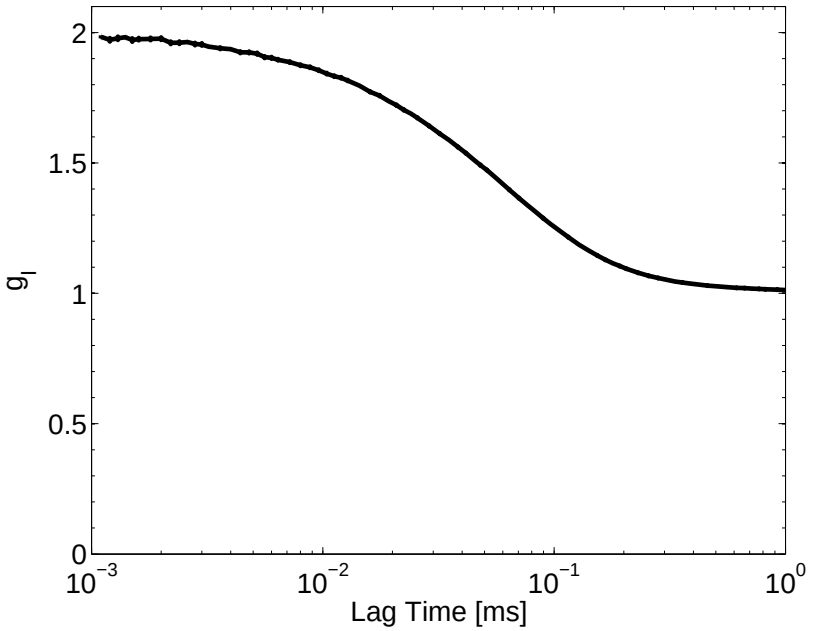

Fig. 3.5: Typical intensity autocorrelation for a protein solution. The reader can verify that the limits $g(0) = 2$ and $g(+\infty) = 1$ are fulfilled.

should remain constant,

$$\text{nothing happens:} \quad g_I(t) \longrightarrow \frac{\langle I^2 \rangle}{\langle I \rangle^2} \; = \; 2 \; ,$$

where the last equality is a basic result of Quantum Optics [97]. In this case, the intensity is completely correlated. Of course, for very short lag times t, the intensity is always completely correlated, whereas for very long lag times it is completely uncorrelated,

$$\lim_{t \to 0} g_I(t) = \frac{\langle I^2 \rangle}{\langle I \rangle^2} \; = \; 2$$

$$\lim_{t \to +\infty} g_I(t) = \frac{\langle I \rangle^2}{\langle I \rangle^2} \; = \; 1 \; .$$

The theory of DLS deals with the connection between a correlogram and its underlying diffusive dynamics.

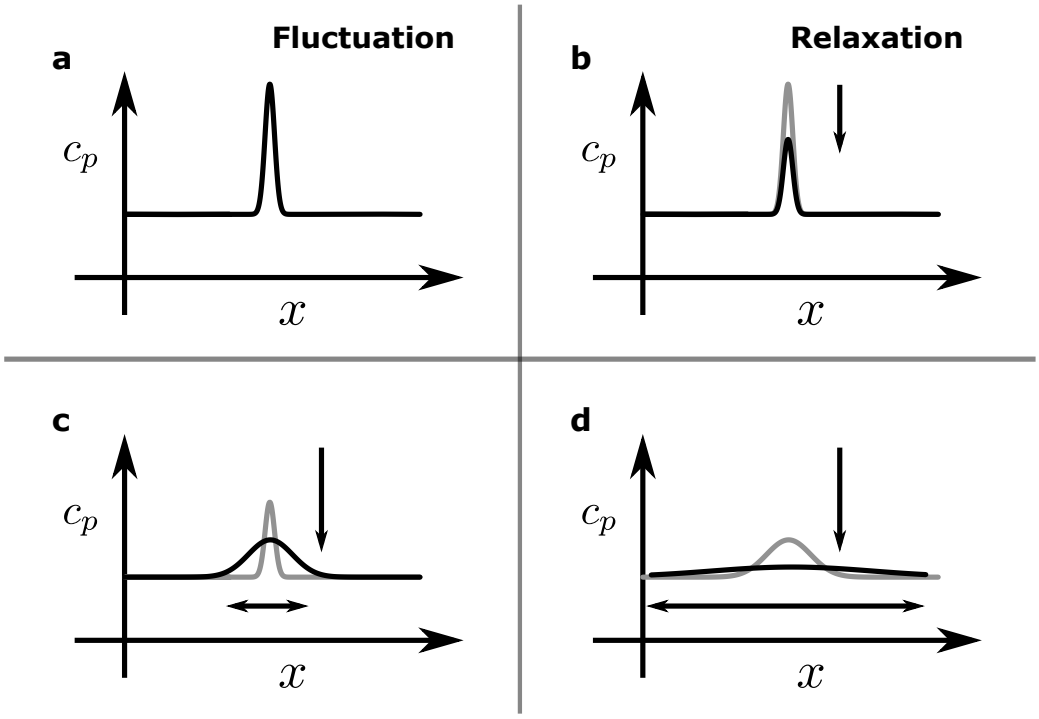

Fig. 3.6: Schematic the relaxation of a concentration fluctuation. At the beginning, the fluctuation, which arises randomly, is localized at a certain position x and relatively big (a). Relaxing towards equilibrium, it becomes smaller (b). It becomes smaller and smaller, and more widespread (c), until it vanishes (d).

Concentration Fluctuations and Their Relaxation

The physical origin of the detected intensity fluctuations (Fig. 3.4) and their autocorrelation (Fig. 3.5) lies in the local thermal fluctuations of protein concentration in the studied solution. These concentration fluctuations arise randomly. However, once a fluctuation is present, it relaxes back to equilibrium according to deterministic laws. We show a schematic of such a relaxation in Fig. 3.6. In the first image, **a**, the fluctuation has just arisen, and the local concentration around the peak is higher than the mean protein concentration of the solution. In the following images **b**, **c**, and **d**, the concentration peak becomes smaller and broader, till it finally disappears. This *relaxation dynamics* is obviously linked to the diffusive protein dynamics; the damping of a concentration peak takes place exactly because the proteins slowly diffuse away from the peak location, swimming towards the

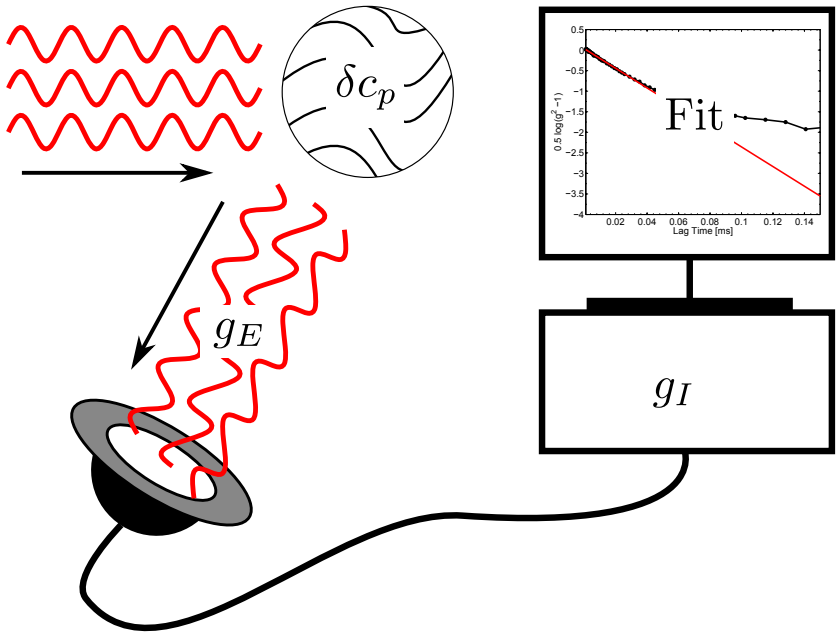

Fig. 3.7: Schematic summary of the DLS technique. Legend: δc_p, concentration fluctuations in the sample; g_E, field autocorrelation in the scattered beam; g_I, intensity autocorrelation computed by the autocorrelator; Fit, data analysis in the computer.

surrounding regions. We skip a deeper discussion in this paragraph; the central result of liquid relaxation theory is the relaxation equation,

$$\frac{\partial c_p}{\partial t}(r,t) = -D \cdot \nabla^2 c_p(r,t) \quad . \tag{3.19}$$

Here, c_p is the protein concentration, ∇^2 is the Laplace operator, and D is an intrinsic property of the protein solution, called *diffusion constant*. The key ingredient of this expression is the diffusion constant.

Connection between Relaxation and Autocorrelation

For a thermally fluctuating solution, light scattering is like a camera that captures the fluctuations in reciprocal space. In other words, when some proteins diffuse away from a concentration peak in a relaxation process, they continue to scatter light all along, and their swimming movement leaves a trace in the scattering interference pattern. The bridge between the relax-

ation and the intensity autocorrelation is the electric field autocorrelation,

$$g_E(Q,t) \quad := \quad \frac{\langle E(Q,t+t_0)\cdot E^*(Q,t_0)\rangle}{\langle E(Q)\cdot E^*(Q)\rangle} \ . \tag{3.20}$$

This quantity describes the degree of correlation in the temporal dependence of the scattered electric field. On the one side, g_E is connected to the relaxation (or diffusive) dynamics because it satisfies the same evolution equation (3.19), but in reciprocal space,

$$\frac{\partial g_E}{\partial t}(Q,t) = -D\cdot Q^2\, g_E(Q,t) \quad . \tag{3.21}$$

On the other side, g_E is directly linked to the intensity autocorrelation, which is the quantity actually detected, via the so-called *Siegert relation* [98],

$$g_I(Q,t) = 1+\beta\ |g_E(Q,t)|^2 \quad , \tag{3.22}$$

where β is a nonessential geometrical factor connected to the surface of the detector. We can easily solve equation (3.21) for the field and intensity autocorrelation, by using the boundary conditions of complete correlation (decorrelation) for zero (infinite) time lag. We find

$$g_I(Q,t) = 1+\beta\cdot\exp\left[-2D\cdot Q^2\cdot t\right] \quad , \tag{3.23}$$

which describes the correlogram in Fig. 3.5. This equation is the main result of our theoretical introduction to DLS. It connects the apparently meaningless, noisy raw data, $I(t)$, which we show in Fig. 3.4, all the way up the causal chain to the diffusion constant of BSA, D. We see a scheme of this argumentation in Fig. 3.7. Formula (3.23) also concludes our brief overview of the working principle of DLS. In the next sections, we discuss some aspects in deeper detail.

3.2.2 Collective Diffusion Kernel and Constant

We have already mentioned the equation that governs the diffusion of proteins and the relaxation of fluctuations, (3.19). We have also recalled that the field autocorrelation, g_E, obeys the same equation, in reciprocal space. In the context of macromolecular dynamics, this kind of behaviour is called *collective diffusion*, to distinguish it from a different phenomenon, *self-diffusion*, in which the motion of single proteins is observed using incoherent radiation (e.g. in neutron scattering from samples containing ^{1}H, with its large incoherent scattering cross-section). Actually, even in the restricted context of

collective diffusion, equation (3.19) is not valid in the general case, which requires the following expressions instead [89]:

$$\begin{aligned} \frac{\partial c_p}{\partial t}(r,t) &= -\nabla \cdot j(r,t) \\ j(r,t) &= -\int_V \mathrm{d}r_0 \int_0^t \mathrm{d}t_0 \; D_{\mathrm{ker}}(r-r_0, t-t_0) \cdot \nabla_0 c_p(r_0,t_0) \quad . \end{aligned} \tag{3.24}$$

The first line is simply the continuity equation, which states that, when a concentration peak decreases, the proteins move in all directions. The second line is a result of linear response theory. If the concentration peak is small, as normally the case for thermal fluctuations, the flux of proteins is linearly related to the gradient of concentration. In the simplest case of equation (3.19), the linearity reduces to a constant, D. In the general case (3.24), it is a kernel operator instead, the *collective diffusion kernel*, $D_{\mathrm{ker}}(r-r_0, t-t_0)$.

Of course, the experimental results potentially described by the general equations (3.24) are much more complicated than we have outlined in the previous section. For instance, equations (3.24) are able to predict non-exponential decay of the intensity correlations, in contrast with the simple result in formula (3.23). However, it turns out that, in our experimental situation, we need not extend the result (3.23) in this sense. The reason is that the scattering vectors of DLS are so small that we can approximate the collective diffusion kernel as follows [89]:

$$D_{\mathrm{ker}}(r-r_0, t-t_0) \;\longrightarrow\; D\,\delta(r-r_0)\delta(t-t_0) \quad . \tag{3.25}$$

This is also called *hydrodynamic regime*, and the resulting D is called (*zero-Q*) *long-time, collective diffusion constant*. Of course, the very concept of "smallness" of Q requires a typical length scale for comparison. We discuss this comparison briefly in the next paragraph. For a deeper discussion of macromolecular diffusion, see Ref. [89] and references therein.

3.2.3 Length and Time Scales for BSA Diffusion

Actually, not one but two parameters control whether a certain experiment falls in the hydrodynamic regime, allowing the simplification (3.25). These are the magnitude of the scattering vector, Q, and the time scale of the correlation function g_I. The scattering vector in our DLS experiments ranges from 6×10^{-4} Å^{-1} to 2.5×10^{-3} Å^{-1} (see Chapter 4). We have to compare this Q window with the typical correlation length of our protein samples, which is given by the position Q_c of the maximum of the main peak of the structure

factor. In turn, Q_c is related to the most probable distance between two BSA molecules, r_c, by the approximate formula $Q_c \sim 2\pi/r_c$ [89]. We cannot measure Q_c by light scattering, but recent estimates from SAXS experiments yield $Q_c \sim 5 \times 10^{-2}$ Å^{-1} [8]. This comparison shows that all scattering vectors probed by DLS are much smaller than this correlation peak. This conclusion is further confirmed by the experimental observation that, apart from a few special cases which we discuss separately, the diffusion constant of BSA is always independent of Q within the range of our instrument.

The time scale of the autocorrelation depends on the scattering vector. As a rough estimate, in our DLS experiments, the correlograms decay on a time scale τ of about 10^4 ns. According to colloid theory, we must compare our τ with the typical time it takes a BSA molecule to diffuse for a distance approximately equal to its radius [89]. We call the latter τ_c. Although BSA is not spherical, we can estimate its linear dimension by setting its phenomenological molecular volume, v_{phen}, equal to that of an hypothetical sphere. This sphere would have a radius $r \sim 3$ nm. Then, using the dilute-limit diffusion constant $D_0 \approx 6$ Å^2/ns, we can calculate the time τ_c by the approximate formula

$$\tau_c \sim \frac{r^2}{D_0} \sim 200 \text{ ns}.$$

This means that the correlograms measured by our DLS measurements describe time scales much larger than τ_c.

The conditions of the hydrodynamic regime, i.e. that the experimental scattering vector is much smaller than Q_c and that the correlograms decay much more slowly than τ_c, are thus verified. Therefore, the simplification in equation (3.25) is justified.

3.2.4 Fit Strategies for the Autocorrelation

In the last section, we have outlined the theory of diffusion underlying our DLS experiments, concluding that we can simplify the diffusion kernel by a constant, the collective, long-time diffusion constant D. In order to extract D from the correlograms, we proceed schematically as follows:

$$g_I(t) \quad \xrightarrow{\text{fit}} \quad D(Q) \quad \xrightarrow{average} \quad \langle D \rangle_Q \ .$$

The Q-average, performed in the second step, is easily done, because the measured $D(Q)$ is practically independent of Q. The fit of g_I is much less trivial. In principle, the direct formula (3.23) for the decay of the intensity autocorrelation should hold. However, we always find systematic discrepancies with the experimental correlograms. The reason is not completely clear,

but probably resides in the sample preparation. Roughly speaking, the BSA solutions are not pure enough for the ideal function (3.23) to work flawlessly, and we need a more sophisticated data treatment. Actually, we fit $g_I(t)$ using various methods that extend the range of applicability of equation (3.23) to slightly impure samples. In the DLS literature, there are several alternative extensions, which we overview in the next sections. None of them provides a definitive solution, but the ad-hoc double exponential decay routine is expected to be, for protein solutions, the most trustful and less biased strategy.

Cumulant Analysis

A first option, preferred in the past when autocorrelators were not so good as they are nowadays, consists in calculating the logarithm of the experimental $\beta\,|g_E|$ through the Siegert relation, and fitting it to a polynomial of degree 1 to 4, that is

$$Q^{-2} \cdot \log\left[(g_2(Q,t)-1)^{1/2}\right] \quad \xrightarrow{\text{fit}} \quad p_0 - p_1 \cdot t + p_2 \cdot t^2 - \dots$$

This is known as the *Cumulant expansion*. The initial fit slope, p_1, yields a mean diffusion constant of the solution, while the second parameter p_2 gives an estimate of the sample polydispersity, i.e. the amount of diffusing species different from the desired one. The reason why cumulants were so popular is twofold. On the one hand, it focuses on a narrow time window around $t = 0$, enabling good fits also for autocorrelators of limited dynamic range. On the other hand, it is perfectly suited for studies of artificially synthesized polymers and colloids, made e.g. of latex, rubbers, silicones, etc.. In fact, such samples have a broad size distribution, and a mean D is often the only quantity one hopes to extract from them.

In our case, the cumulant analysis suffers from a serious drawback. Our solutions include a certain fraction of non-monomeric BSA, which diffuse more slowly than single proteins. Since the cumulant analysis yields an average D by definition, this diffusion constant is a weird numerical mixture of many different protein species, and is systematically smaller than the "real" diffusion constant of BSA monomers. Nevertheless, we perform first-order cumulant fits on our samples, in order to compare our results with those already published. For these fits, we restrict the lag-time window to middle-short times, which can be reasonably fitted linearly. The very short lag-time region shows large electronic noise, and the long lag-time region is outside the region of applicability of the cumulant expansion. An example fit of our autocorrelations is shown in Fig. 3.8.

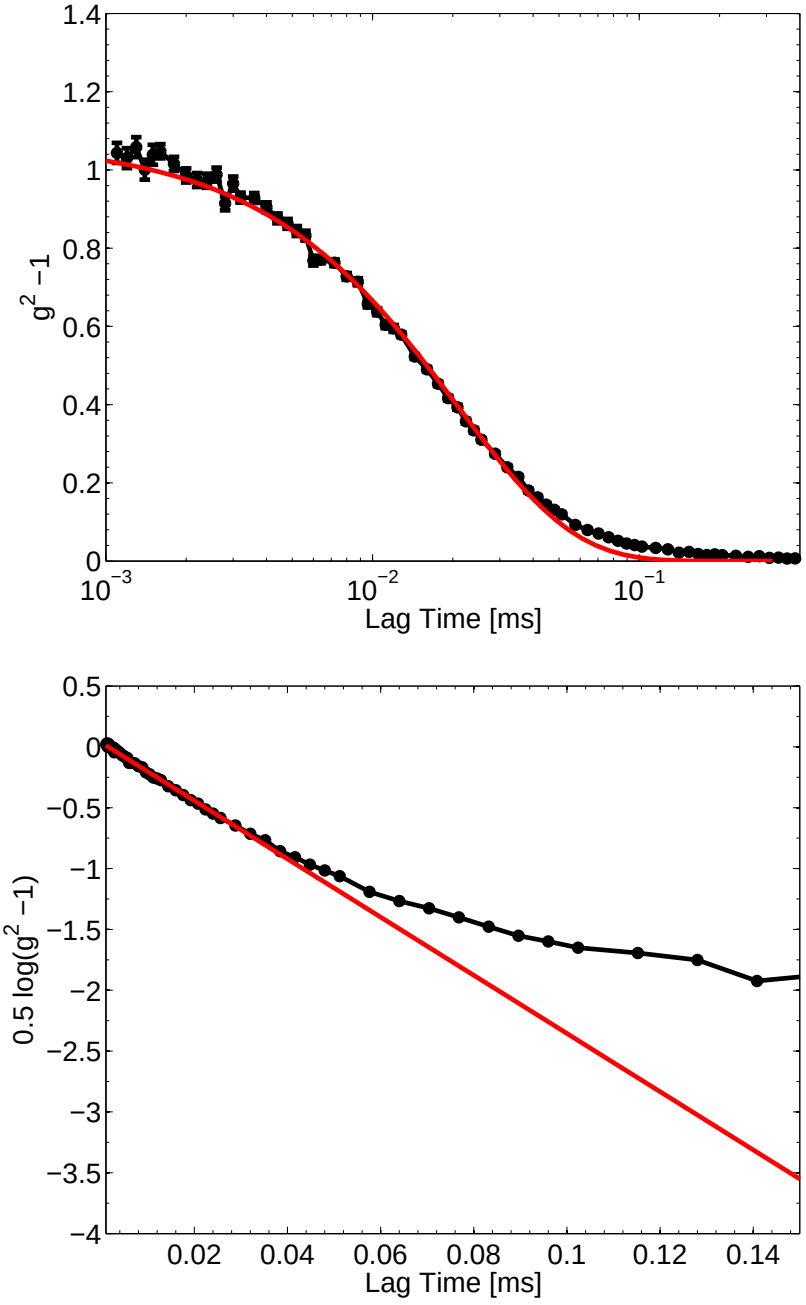

Fig. 3.8: Example of a cumulant fit of a raw intensity autocorrelation. The log-lin view (top) emphasizes the high dynamic range of the autocorrelator. The lin-log view (bottom) is the best way to observe non-exponential decays instead. The black lines are guides to the eye.

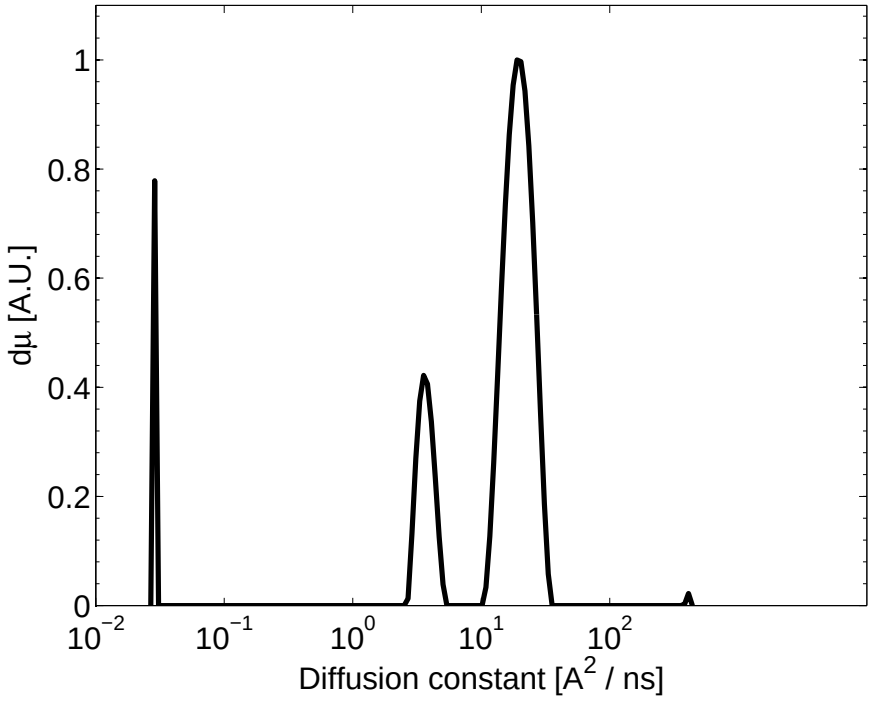

Fig. 3.9: Example of an inverse Laplace transform of a field autocorrelation. Only the two central peaks are physically meaningful. The regularizator is proportional, in this case, to the squared numerical first derivative of $\mathrm{d}\mu$.

Inverse Laplace Transform Schemes

A second option is offered by Inverse Laplace Transform-based schemes. Their most famous implementation is the CONTIN program of Provencher [99, 100]. The basic idea is that, if a single exponential decay is not sufficient to fit the data, a whole distribution of decays will be so. In other words, instead of a single diffusion constant D, we look for a distribution function $\mathrm{d}\mu(D)$, such that

$$(g_2(Q,t)-1)^{1/2} = \beta\ |g_1| \quad \longleftarrow \quad \int_0^\infty \mathrm{d}\mu(D) \exp\left[-D\cdot Q^2\cdot t\right] \quad .$$

This transformation from $\beta\ |g_1|$ to $\mathrm{d}\mu$ is called *Inverse Laplace Transform* (ILT). In case of a single, perfect, exponential decay, with diffusion constant

D_{perfect}, the distribution reduces to a delta peak,

$$\mathrm{d}\mu(D) \quad \longrightarrow \quad \delta(D - D_{\text{perfect}})\,\mathrm{d}D \quad .$$

Of course, in nonideal situations, $\mathrm{d}\mu(D)$ is more complicated. We present an example of Laplace transform of a field autocorrelation g_E in Fig. 3.9. Evidently, $\mathrm{d}\mu(D)$ is not a single delta peak. Apart from the fit artifacts at the boundaries, we obtain two peaks. The main, faster peak refers to BSA monomers. The smaller, slower peak could be due to several reasons, e.g. BSA monomers and oligomers.

The main issue concerning this approach is the serious risk of overfitting. Compared to the cumulant fit, which has 1 to 4 free parameters, here the full $\mathrm{d}\mu(D)$ function must be reconstructed, with a far higher (conceptually infinite) number of free parameters. Moreover, the inverse Laplace transformation of experimental data is an infamous ill-conditioned problem. In other words, a small change in the experimental points may cause a huge change in $\mathrm{d}\mu(D)$. This makes the physical interpretation of the distribution very delicate. An improvement of this scheme exploits the so-called regularizators. These are abstract functions, without any physical significance, which are introduced in the fitting routine to restrict the field of candidates for $\mathrm{d}\mu(D)$. The main effect of a regularizator is to force $\mathrm{d}\mu(D)$ to be smooth enough, *de facto* excluding delta peaks and favouring broad, bell-shaped distributions. Actually, in Fig. 3.9, we show a regularized fit, in which the regularizator is proportional to the squared numerical first derivative of the distribution. Note that the peaks of $\mathrm{d}\mu(D)$, as a consequence of the regularization, are recovered with a fairly low accuracy (the abscissa is in logarithmic scale).

In conclusion, Laplace transforms represent a nice opportunity if the experimental scientist has absolutely no idea about the possible diffusion modes in the sample, or if broad distributions are expected. They also present several drawbacks, because they are either unstable or very imprecise, and their physical interpretation remains often dubious. For this thesis, we use ILTs mainly as a qualitative tool for building an intuition about the kind of diffusive modes present in our samples. However, we do not analyze the results of the ILTs quantitatively.

Ad-hoc Fit Routines

Finally, a third route remains open, that is the use of *ad hoc* fit models for g_I. This is a feasible option if the physics of the sample is already partially known, so that the choice of the fit function can be justified, at least heuristically. This is the case of BSA solutions, for the following reasons. It is known that all commercially available BSA powders, including the ones used for

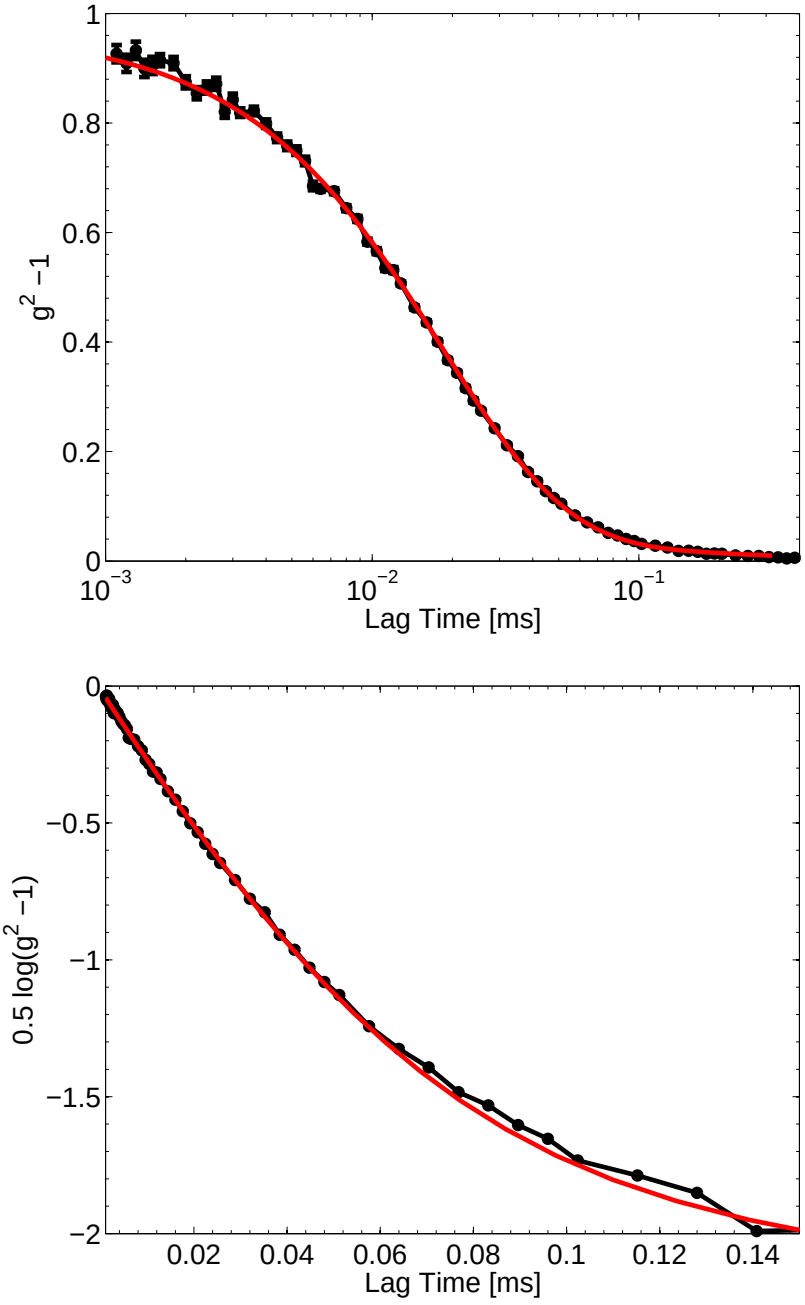

Fig. 3.10: Example of an ad-hoc double exponential fit of a raw intensity autocorrelation. The log-lin view (top) emphasizes the high dynamic range of the autocorrelator. The lin-log view (bottom) is the best way to observe non-exponential decays instead. The black lines are guides to the eye. Note that the slope of the bottom plot is *not* simply the first decay, $\Gamma_1 = D_1\, Q^2$.

this thesis, contain a certain fraction of protein complexes, dimers or higher oligomers (see Chapter 2). These species are too small for the filters used in the sample preparation, but still large enough for their diffusion behaviour to differ quantitatively from that of a single BSA molecule. If the fraction of monomers is large, we can assume that they move independently of heavier species. Thus, the field autocorrelation has two independent contributions,

$$g_E(Q,t) = \sum_{i=1}^{2} A_i \cdot \exp\left[-D_i \cdot Q^2 \cdot t\right] \quad , \tag{3.26}$$

where the index 1 refers to BSA monomers, and the species 2 works as a garbage collector for all other species. According to our intuitive picture, we choose $D_1 > D_2$, for smaller particles diffuse faster than larger ones. The function (3.26) is squared and fitted directly to the measured intensity autocorrelation, g_I. This fit routine is nonlinear, thus much more slowly than alternative strategies such as cumulants. However, the computational power of a modern laptop is high enough to perform such a fit in less than one second.

Because of the ability of (3.26) to decouple, at least partially, the effect of disturbing species on the fitted diffusion constant, it is the main fit strategy adopted for this thesis. An example of such a fit is shown in Fig. 3.10.

3.2.5 Diffusion in Dilute and Crowded Systems

Thanks to the long-time small-Q condition, the diffusion kernel D_{ker} reduces to the collective diffusion coefficient D. Of course, the value of D still depends on the physical parameters of the system, such as the BSA concentration c_p, the salt concentration c_s, the temperature, and the pressure. Actually, the main theoretical challenge is the prediction of this function, $D(c_p, c_s, T, P)$, from first principles. In general, this is a difficult task, because the theory has to describe direct interactions as well as hydrodynamic ones. The former are the usual molecular forces, for instance the electrostatic and the van der Waals interactions. Hydrodynamic interactions are mediated by the solvent instead, and are depicted in Fig. 3.11. Schematically, they work as follows. One particle interacts with its surrounding solvent, exchanging momentum and energy. The inhomogeneity just created is then spread by the solvent particles throughout the whole volume, by very fast travelling relaxation processes. On the time scale of protein diffusion, the flow disturbance created by the motion of a particle by its solvent friction is transmitted quasi-instantaneously to the place of a second one. This leads to change in

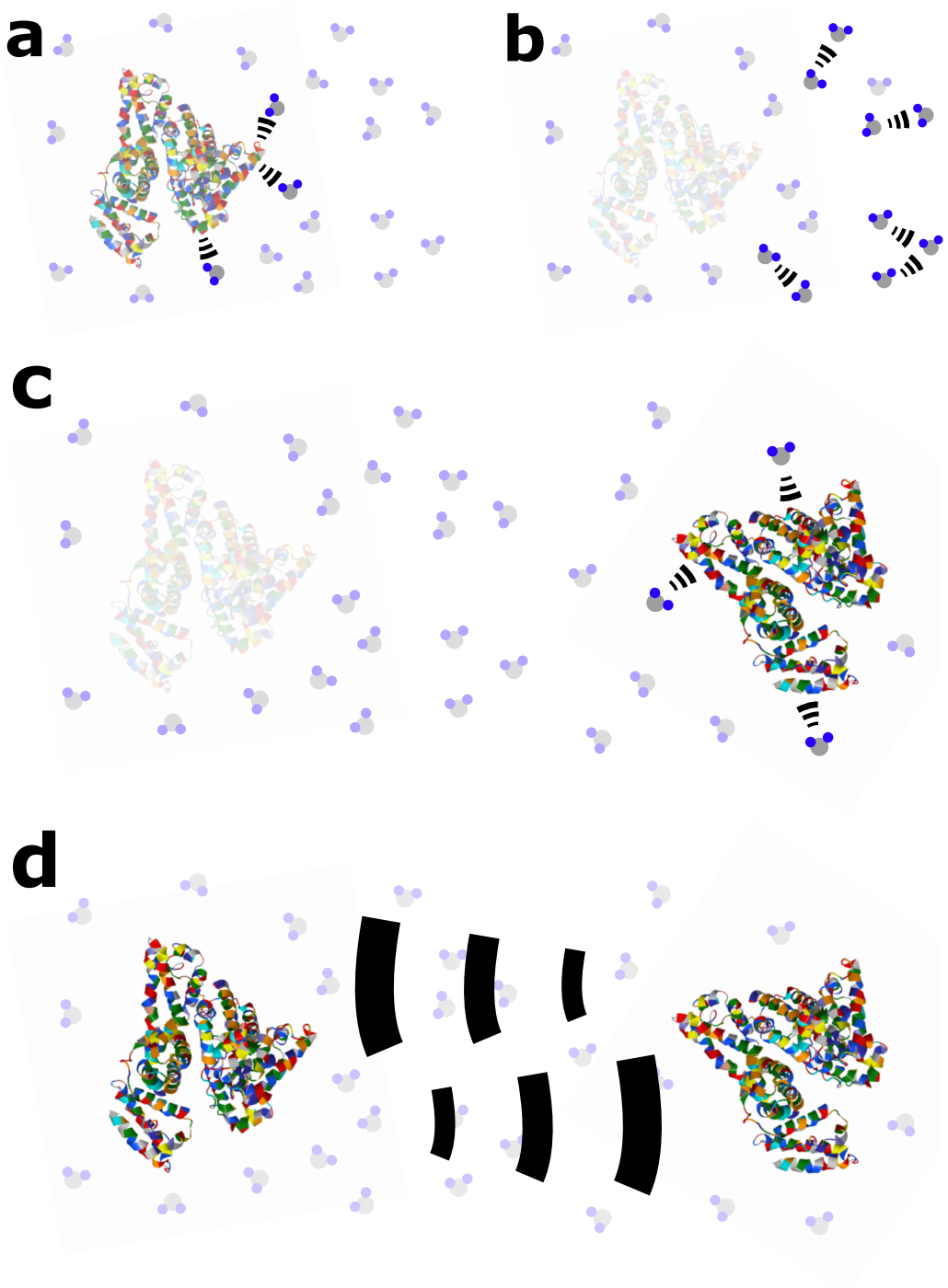

Fig. 3.11: Schematic illustration of the hydrodynamic interaction. A protein exchanges energy and momentum with the surrounding water molecules (a). The excitation propagates across the solution (b). A second protein receives a small part of the energy/momentum, and reacts accordingly (c). The overall effect is an interaction between proteins (d).

the velocity of the second particle causing, in turn, an additional flow perturbation which reacts back on the first particle. In principle, there is an infinite series of increasingly weakening back reflections of flow perturbations.

Of course, the farther away two particles are, the less important the hydrodynamic interactions, because the solvent relaxation waves are damped at large distance. In the limiting case of an infinitely dilute solution, the hydrodynamic interactions, just like the direct ones, vanish. D becomes, in first approximation, independent of all force-related parameters, such as c_s. We call the dilute-limit diffusion constant D_0,

$$\lim_{c_p \to 0} D \quad =: \quad D_0 \quad .$$

We analyze this quantity in detail in the following paragraph. Here, we only mention that, at low but nonzero protein concentration, we can expand D in powers of c_p in a fashion that recalls the virial expansion for the osmotic pressure (or compressibility),

$$\frac{D}{D_0} = 1 + k_d \, c_p + \ldots \quad .$$

We call the expansion parameter, k_d, *second diffusion coefficient*. Analogously to B_2, the second diffusion coefficient encodes the first correction to the diffusion constant due to the interparticle interactions, both direct and hydrodynamic. However, the power expansion of D is actually more delicate, because the second diffusion coefficient tends to infinity for vanishing salt concentration [67]. We observe a very high second diffusion coefficient in our solutions without added salt indeed, as explained in Chapter 5. Furthermore, since the theory measures the concentration in other units, protein number over solution volume, we can express the second diffusion coefficient in theoretical units by the following conversion:

$$k_d(\textbf{t.u.})\ [\mathrm{m}^3] \ = 10^{-3} \cdot \frac{M_p}{N_a} \cdot k_d(\textbf{DLS})\ [\mathrm{ml/mg}] \quad .$$

The Stokes-Einstein Equation

In very dilute solutions, the proteins are separated by very large distances, and all interactions among them are practically absent. This makes the theory much easier, for it only has to take into account the dynamics of a single protein in an external solvent flow. The calculation of the dilute-limit diffusion constant, D_0, has been solved a century ago by Einstein and Smoluchowski for a spherical particle [101, 102]. The resulting formula is

called *Stokes-Einstein equation* (SE):

$$D_0 = \frac{kT}{6\pi\,\eta_0\,r_h} \quad , \tag{3.27}$$

where η_0 is the solvent viscosity, and r_h is the radius of the sphere. The SE can be applied also for nonspherical particles, even if they posses no "radius". In this case, the parameter r_h relates then to some typical linear dimension of the particle, and is called *hydrodynamic radius*. For instance, the SE has been proven for ellipsoids of revolution, using the so-called *Perrin factors* to calculate r_h from the semiaxes [75]. For particles of complex geometry, such as proteins, the applicability of (3.27) is more questionable. The SE equation is derived solving the Navier-Stokes equations under the assumptions of laminar flow and stick boundary conditions on the sphere surface. For nonspherical molecules with a complex, vine-like-covered surface, such as proteins, both hypothesis loose appeal. However, the SE can be safely used for rough estimates. Usually, instead of starting from r_h and verifying the SE equation, the SE itself is assumed valid, the diffusion constant D_0 is measured in dilute protein solutions, and the hydrodynamic radius is extracted, as an order-of-magnitude estimate of the protein size. In Chapter 5, we will follow this approach.

Generalized Stokes-Einstein Equations

Formula (3.27) is very useful, but only valid for very dilute solutions, in which the interactions are virtually absent. In the literature, there are many efforts to extend it to concentrated solutions of particles, which interact via direct as well as hydrodynamic forces. One famous approach seeks a relation of the following kind:

$$\frac{D}{D_0} = f(\eta, \lambda_i) \quad ,$$

where η is the viscosity of the concentrated solution, and λ_i is a small set of parameters describing the interactions, e.g. the osmotic compressibility. η and the λ_i can depend, in turn, on the basic thermodynamic variables c_p, c_s, T, and P. Relations of this form are collectively called *Generalized Stokes-Einstein* (GSE) equations. Examples are Refs. [82, 103]. It turns out that these generalizations of the SE equations are quite difficult to carry out in practice. Indeed, the most recent progresses in the field of macromolecular diffusion come from another approach, the Svedberg equation.

Svedberg Equation and Generalizations

The SE equation relates viscosity and diffusion in dilute systems. A similar formula, relating diffusion with sedimentation, is the Svedberg Equation (SV) [104–106],

$$\frac{D_0}{s_0} = \frac{N_a}{M_p\ (1 - v_p/v_w)}\ kT \quad , \tag{3.28}$$

where v_p and v_w are the specific volumes of protein and water, respectively. s_0 is called *sedimentation coefficient*, and represents the stationary velocity at which proteins settle down in a centrifugation cell. Please note that the viscosity plays no direct role in the Svedberg equation; its effect is hidden in the coefficient s_0. In other words, $s_0 = s_0(\eta_0)$ holds, and vice versa, $\eta_0 = \eta_0(s_0)$.

The SV is interesting by itself, but even more useful are its generalizations to concentrated solutions. In fact, under some reasonable assumptions on the specific volumes, the *Generalized Svedberg Equation* (GSV) holds [56,89,106],

$$\frac{D}{D_0} = \frac{s}{s_0} \cdot \frac{\chi_{T0}}{\chi_T} \quad , \tag{3.29}$$

where s, D, and χ_T are the sedimentation coefficient, the diffusion constant, and the osmotic compressibility of the concentrated solution, and s_0, D_0, and χ_{T0} their dilute-limit counterparts. Here, the main point is that, differently from the GSE equations, which require viscosity measurements to be applied, the GSV suits very well our experimental data collection. In fact, we measure D, D_0, and χ_T at the same time. This enables a calculation of the sedimentation coefficient ratio.

Equation (3.29) has been proven, using the generalized Smoluchowski equation, not only in the hydrodynamic regime, but also for Q-dependent short-time diffusive dynamics. In that framework, the GSV is usually written as follows [89]:

$$\frac{D(Q)}{D_0} = H(Q) \cdot \frac{1}{S(Q)} \quad ,$$

where S is the static structure factor, and H is called *hydrodynamic function*. At small Q, it can be shown that $H(Q \to 0) \longrightarrow s/s_0$ indeed [89]. Thus, our simultaneous determination of χ_T and D by SLS and DLS yield the low-Q hydrodynamic function. We show the results of this analysis in Chapter 5.

Second Hydrodynamic Coefficient

In absence of hydrodynamic interactions, it results $H(Q) \to 1$, that is $s \to s_0$. This result holds experimentally in the dilute limit, just like the other limiting

equations $\chi_T \to \chi_{T0}$ and $D \to D_0$. In solutions at low protein concentration, the hydrodynamic interactions have a small but non-negligible effect. We can describe their influence with a one-number property, just like the second virial coefficient for the osmotic pressure. For this purpose, we expand the normalized sedimentation coefficient in powers of c_p and keep only the first-order correction [89],

$$\frac{s}{s_0} = 1 + k_h\, c_p + \dots \quad .$$

We call the coefficient k_h *second hydrodynamic coefficient*. It plays a role similar to that of B_2 or k_d, because it is the principal correction to the sedimentation due to protein-protein interactions. Moreover, the theory predicts that, in salt-free solutions, k_h tends to $-\infty$, because s/s_0 is not analytic at $c_p = 0$ anymore [67]. Indeed, in our solutions without added salt, we fit a very large, negative second hydrodynamic coefficient (see Chapter 5).

The GSV itself can be expanded in powers of c_p, exploiting the expansions of χ_T, D/D_0, and s/s_0,

$$\frac{D}{D_0} = [1 + k_h\, c_p + \dots] \cdot [1 + 2M_p B_2\, c_p + \dots] \quad ,$$

finding an expression for the expansion coefficients,

$$k_h \quad = \quad k_d - 2M_p B_2 \quad . \tag{3.30}$$

Using this formula, we can combine our SLS and DLS experimental data at low protein concentration to investigate the second hydrodynamic coefficient. Actually, we can study k_h from two different points of view. First, we could calculate k_d from DLS by expanding D, then B_2 from SLS by expanding χ_T, and finally perform the subtraction to obtain k_h. We do not follow this path, because the error bars on k_d and B_2 would cause a huge uncertainty on k_h. We choose the second alternative instead, which consists in calculating numerically s/s_0 for all concentrations, and fitting this quantity linearly afterwards.

We can convert the second hydrodynamic coefficient to theoretical units according to the following expression:

$$k_h(\mathbf{t.u.})\ [\mathrm{m}^3] \ = 10^{-3} \cdot \frac{M_p}{N_a} \cdot k_h(\mathbf{exp})\ [\mathrm{ml/mg}] \quad .$$

This is the same kind of conversion already explained for B_2 and for k_d.

Chapter 4

Experimental Methods

In this chapter, we describe the methods applied in the experiments and data analysis. The first part relates to the sample preparation. In the second part comes a detailed description of the light scattering experiments. The third part is an introduction to the technique of UV absorption spectroscopy, used for measuring the protein concentration. The chapter ends with a brief description of the main tool used for the data analysis, the Matlab® code[1]. We will not show any Matlab® code itself in this thesis.

4.1 Sample Preparation

The preparation of samples for scattering experiments proceeds in two steps, mixing and filtration. We see the first step schematically in Fig. 4.1. We take a clean glass veil of about 4 ml volume. We put BSA powder into the veil, weighing the amount with an analytical balance (1). We add water, measuring the volume with calibrated pipettes (2). We put the resulting mixture on a rolling bed for about thirty minutes, till it becomes homogeneous (3). Then, we add a pre-mixed solution of water and salt, using the pipettes (4), and finally put the glass veil on the rolling bed for slow mixing, for about thirty minutes (5). Obviously, we skip the passages 4 and 5 when we prepare salt-free samples. The balance used for weighing is a Mettler Toledo Excellence XS204 DeltaRange Analytical Balance, with readability 0.1 mg. The pipettes used for dilution are Eppendorf Research 20-200 μl, cat. 3111000157 and 0.1-1 ml, cat. 3111000165, with readability 0.2 μl and 1 μl, respectively.

The second step of the preparation is filtration; we illustrate this schematically in Fig. 4.2. We extract approximately 1.5 ml of solution, using a dis-

[1] ©1994-2010 The MathWorks, Inc.

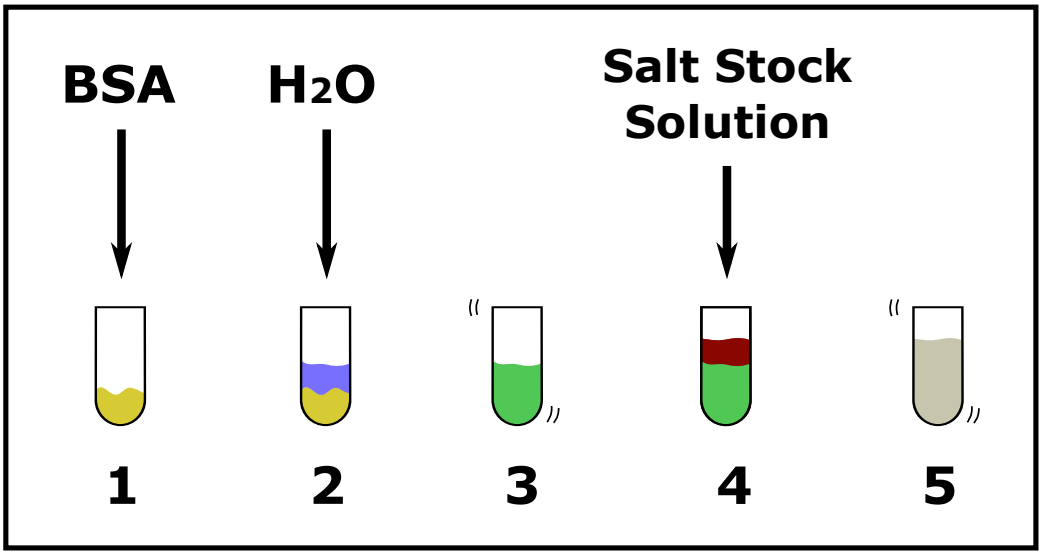

Fig. 4.1: Schematic illustration of sample preparation.

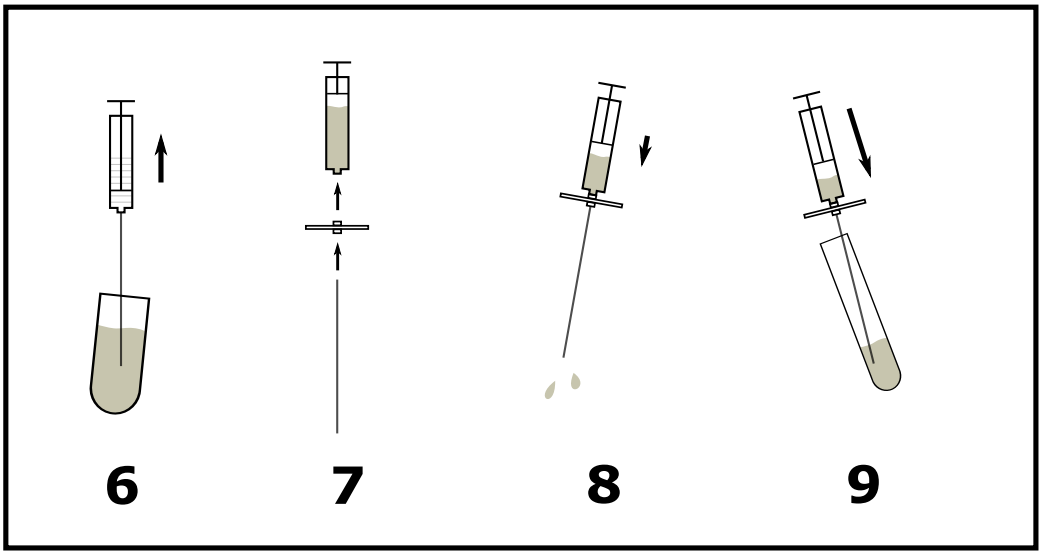

Fig. 4.2: Schematic illustration of sample filtration.

posable plastic syringe with steel needle (6). Then, we remove the old needle. We connect a disposable syringe filter with a new needle (7). We pour out some droplets of solution to wet the filter (8), and pour the rest of the sample into a borosilicate glass test-tube (9), which we seal immediately. During filtration, the pressure exerted by the finger is low, in order to stay below the bubble point of the filter. For the same reason, we avoid too narrow syringes (1 ml). We use Whatman Puradisc 13 nylon syringe filters of pore size 0.1μm cat. 6789-1301. Once in a while, we test a random filter against pure water for cleanness, checking the latter by the absence of time correlations in the scattered light.

4.2 Light Scattering

The main laboratory apparatus used to perform the experiments is a multiangle static/dynamic light scattering instrument. We give a short introduction to the theory of light scattering from solutions in chapter 3; here, we only illustrate briefly the physical principle. The basic idea is to illuminate a solution with a laser beam and watch the light reflected by the sample in various directions. The simplest measurable quantity is the mean light intensity as a function of the angle of observation. This is probed by Static Light Scattering (SLS), and relates to time-independent, equilibrium thermodynamic properties of the sample. Additionally, the near-equilibrium fluctuation dynamics can be investigated by tracking the time statistics of the scattered light. This latter technique is called Dynamic Light Scattering (DLS).

4.2.1 Experimental Setup

We sketch the experimental setup of the light scattering instrument in Fig. 4.3. Fig. 4.4 is a picture of the real machine. The light is generated in a Helium-Neon gas laser operating at a vacuum wavelength of 632.8 nm (the standard red lines), with an output power of 22 mW. The power is low enough to neglect any heating of the sample. Moreover, the laser frequency is far away from any resonance, so that the optical absorption is practically zero.

Multiangle measurements are available thanks to a high-precision goniometer controlled by a step motor (ALV CGS3). The total angle range is 25-150°, but the smallest angle used for this thesis is 30°. A typical multiangle measurement includes from five to thirteen equidistant angles. The magnitude of the scattering vector depends on the refractive index of the solution, but is approximately in the range from 7 to 25 $\times 10^{-4}$ Å^{-1}. This translates into a rule of thumb for the instrument static spatial resolution:

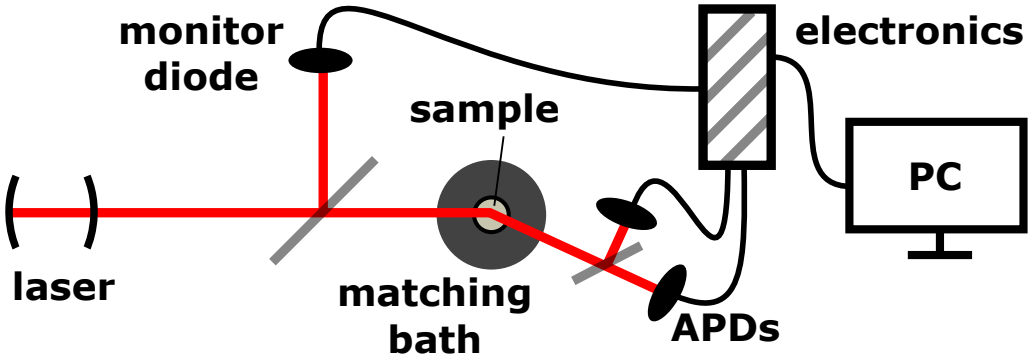

Fig. 4.3: Schematic illustration of the light scattering instrument. The light comes from the laser and is split towards the sample and a monitor diode, which calibrates the intensity. The sample, surrounded by a refractive-index matching toluene bath in order to reduce the reflections, and by a water thermal bath (not shown) to control the temperature, scatters the light. This is split again, in order to enhance the dynamic range at fast response times, and detected in terms of photon counts by the APDs. In turn, these send the count information to the autocorrelator, and finally to the computer.

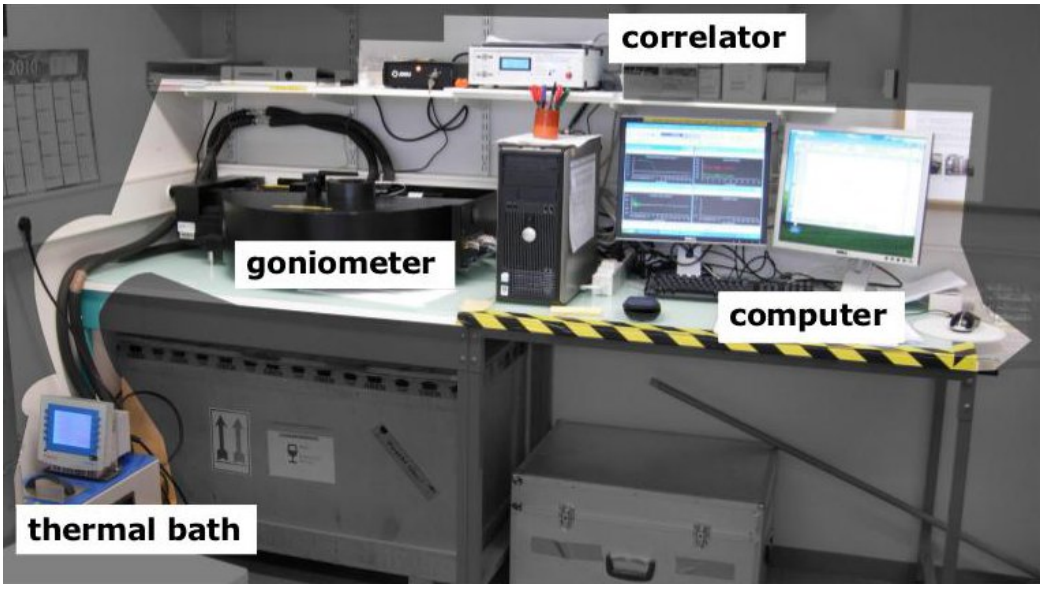

Fig. 4.4: Real picture of the ALV light scattering instrument at the ILL.

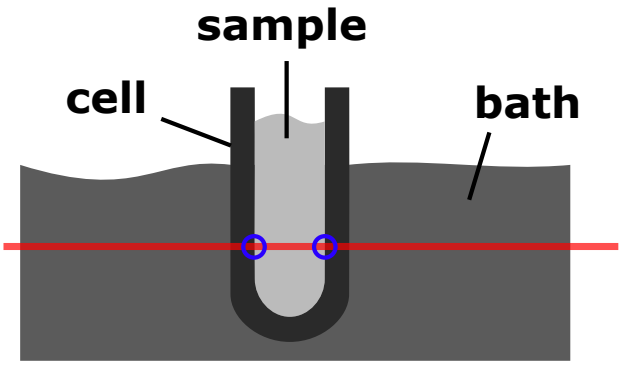

Fig. 4.5: Schematic lateral view of the index-matching bath. Blue points indicate possible reflections.

particles smaller than $\sim$ 20 nm, including BSA molecules, show the same scattering pattern at all angles.

The scattered light is filtered by a slit, split into two beams by a 50/50 cube, and collected at the chosen angle by a pair of avalanche photodiodes (APDs). These detectors, which measure photon counts digitally, offer a fast response and a short dead time; these features are essential for DLS. The electric signal coming from the diodes is fed into an ultrafast, digital, multi-tau hardware cross-correlator with a minimal correlation time of 2.3 ns (ALV 7004/FAST). The electronic system calculates simultaneously the mean scattering intensity over a chosen integration time, used for SLS, and the intensity time cross-correlation between the two APDs, used for DLS. The results are then transmitted to the computer for further analysis. The two separate APDs are preferred over a single detector in order to suppress afterpulsing problems at very short lag times (*pseudo-cross-correlation*) [107].

APDs have different efficiencies depending on the mean detected intensity. In order to optimize the data quality, in particular to avoid large photon noise at very low count rates [96], a servo-controlled variable attenuator is set in the light path just after the laser tube. Before every run is started, the best attenuator is automatically selected depending on the scattering power of the sample. Moreover, a further 50/50 beam splitter sends part of the incoming light to a monitor optical diode, allowing absolute calibration of the laser intensity.

The light scattering cell is made of borosilicate glass, with a refractive index $n_{\text{cell}} \approx 1.51$ relative to vacuum. It is immersed in a relatively large bath of liquid toluene, of refractive index $n_{\text{tol}} \approx 1.50$ (see Fig. 4.5). The rationale of this so-called *index-matching bath* is the following: because the cell and the surrounding medium have similar refractive indices, there is

almost no reflection at the outer cell walls. Such reflections would, if present, complicate the interpretation of the scattering signal.

Furthermore, the sample is kept at constant temperature by a circulating water Haake bath, shown in Fig. 4.4 in the bottom-left corner. We set the temperature manually to 22 °C. Before any light scattering measurement, we wait for approximately 15 minutes for the temperature to stabilize.

The polarizations of the incoming and scattered light are not selected specifically, but this point does not represent a problem. In SLS, the absolute intensity is calibrated experimentally using a standard liquid, pure toluene, *de facto* factoring out the geometrical factors coming from polarization. We describe this point further in the next section. DLS data are not affected either, since the dynamic depolarizing effects of BSA are very fast, hardly visible in the time window of the detection system.

4.2.2 Absolute Calibration of SLS

SLS at small scattering vectors yields valuable thermodynamic information but, as outlined in chapter 3, the absolute scattering intensity is required. In principle, we could calculate this quantity starting from the nominal power of the laser and the beam size. However, that method would be quite inaccurate, for it ignores the many defects of a real experimental setup, such as fluctuations in the laser power, imperfect alignment of the detector, and partial reflection at the air-glass interfaces. We prefer a more convenient method, which is actually the standard approach. First of all, we measure the scattering power of ultrapure toluene W_{tol} in any units, e.g. photocounts per second. Second, the scattering power of a protein solution W_{sol} is measured in the same units. Finally, the ratio between them is converted into a real intensity via the Rayleigh ratio of toluene R_{tol}, which is a tabulated quantity,

$$R_{\mathrm{sol}} = \frac{W_{\mathrm{sol}}}{W_{\mathrm{tol}}} \cdot R_{\mathrm{tol}} \quad . \tag{4.1}$$

We use the same measurement of W_{tol} also for the calibration of the solvent scattering intensity, needed for the background subtraction.

A last important effect is the apparent shrinking of the scattering volume, as seen by the APDs, because of refraction at the solution-glass and bath-air optical interfaces. It has been shown that the correction factor is equal to the squared refractive index of the sample [108]. Therefore, the observed Rayleigh ratio of the solution in (4.1) becomes

$$R_{\mathrm{sol}} \qquad \longrightarrow \qquad R_{\mathrm{sol}} \cdot \frac{n_{\mathrm{tol}}^2}{n^2} \quad ,$$

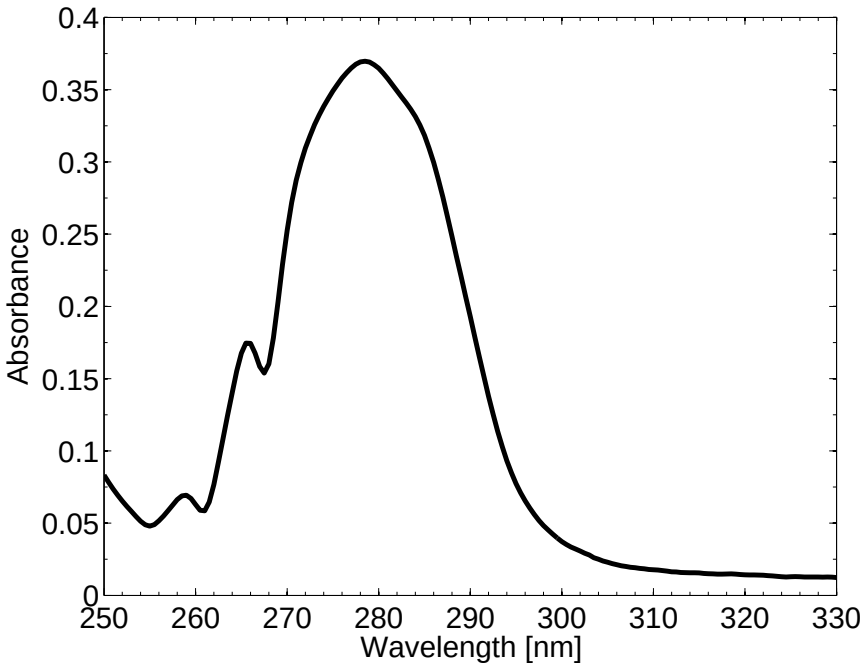

Fig. 4.6: Typical near-UV absorbance spectrum of a BSA solution of protein concentration around 0.5 mg/ml.

where n_{tol} is the refractive index of toluene, and n that of the protein solution. We perform an analogue correction for the solvent.

4.3 UV Absorption

In the experiments carried out for this thesis, we use UV absorption spectroscopy for measuring the BSA concentration of the samples. We prepare the solutions by mixing a known weight of BSA with a known volume of solvent, either pure water or a salt-water solution. Therefore, after the preparation, we know the so-called *nominal* BSA concentration, c_{nom}. However, after the mixing, we filter the samples, in order to remove dust and other impurities. As good as the nylon filters may be, they nonetheless retain a non-negligible amount of protein, so that the actual concentration is

$$c_{\mathrm{act}} < c_{\mathrm{nom}} \quad .$$

UV absorption spectroscopy thus is the standard way to circumvent this

problem. This method is based on the Beer–Lambert law. According to this well known relation, the light intensity transmitted through a semitransparent sample decays exponentially with the length L of the optical path inside the sample,

$$\begin{aligned} \frac{I_{\text{out}}}{I_{\text{in}}} &= \exp(-A) \\ &= \exp(-\epsilon_{\text{BSA},\,\lambda}\; c_{\text{p}}\; L) \quad . \end{aligned} \tag{4.2}$$

In this equation, A is the so-called absorbance, the quantity actually measured in a practical experimental situation. Of course, A depends on the solute species and concentration, and on the wavelength of light λ. We present a typical absorption spectrum of BSA in Fig. 4.6. The second version of the law, equation (4.2), connects the measured absorbance with the protein concentration. In this formula, L is the length of the absorption cell, i.e. of the optical path, in our case $L = 1$ cm. c_p is the protein concentration, in the following units:

$$c_p \quad := \quad \frac{\text{mass of BSA}}{\text{volume of } \textit{solution}} \quad .$$

Finally, $\epsilon_{\text{BSA},\,\lambda}$ is called *extinction coefficient* and is typical of the protein species; it depends on λ too. It is customary to measure the absorption at its maximum, at $\lambda = 279$ nm; the extinction coefficient for BSA takes the value $\epsilon_{\text{BSA},\,279\ \text{nm}} = 6.67\ \text{l g}^{-1}\ \text{cm}^{-1}$. From a physicochemical point of view, the absorption of UV light in proteins is a localized phenomenon that takes place in some specific amino acids. For instance, UV absorbing amino acids are those with aromatic rings, such as phenylalanine and tyrosine.

4.4 Numerical Data Analysis with Matlab®

The numerical data analysis is a crucial and often overlooked step between an experiment and its physical interpretation. As far as this thesis is concerned, it is also the step that takes most of the time. In fact, the amount of raw data used for this thesis is quite large, and its interpretation far from trivial. In order to manage this situation, we need a flexible, stable, and automation-friendly data analysis software. Among various valid options, Matlab® turns out to be the best choice, mostly because of compatibility with colleagues and their previous experience with it. Starting from an absolute ignorance of the Matlab® syntax, the data analysis has required about 8500 lines of code.

It is crucial to respect some good programming practices in the code. Before using a new algorithm, for instance a fit routine, we try it on some test data. We check explicitly the plausibility of all physically meaningful results, their orders of magnitude, and their units. Moreover, we tend not to save unnecessary intermediate versions of the numerical data, to avoid confusion. In order to reduce the intrinsic arbitrariness of human decisions, we try to automatize the data analysis as much as possible. Before averaging the SLS/DLS results, we filter them automatically and/or manually, excluding self-evident outliers that could invalidate any important conclusions. In some cases, we calculate the same quantity in two different ways, and check for mutual consistency. For example, in all cases when the diffusion constant is independent of Q, we calculate it both via direct weighed averaging, and from the slope of a linear fit of $\Gamma(Q)$. Furthermore, we use fit and normalization routines with care, only when a theoretically sound basis exists; we avoid them otherwise. For instance, we do not fit B_2 in the samples with YCl_3; in these solutions, the SLS results appear too complicated, to be naively reduced to a single number.

Even with all these precautions, there is no guarantee that the code is always correct, or best suited for a certain purpose. As a matter of fact, some algorithms used are really unstable, such as the inverse Laplace transform, or theoretically slightly doubtful, such as the ad-hoc fit strategy. In these situations, we try to repeat the analysis with little modifications, or different technical implementations. The results finally presented in this thesis are reasonably safe, physically founded, and numerically stable. The reader is free to contact Prof. Schreiber's group to obtain a more detailed description of the numerical data analysis, or to view the code itself. We release the code under the GNU General Public License, version 3.

Chapter 5

Results and Discussion

In this chapter, we present and discuss the experimental results. First, we give a general overview, and perform the consistency checks on the dilute-limit properties, the molar mass and the diffusion coefficient D_0. Then, we examine the results from samples without added salt, with NaCl, and with $CaCl_2$. We discuss their interpretation in terms of second virial and second hydrodynamic coefficients, osmotic compressibility, crowding effect, and sedimentation coefficient. We show that all results from NaCl and $CaCl_2$ solutions found on two simple physical concepts, that are electrostatic screening and excluded volume. Finally, we deal with the YCl_3 samples. We show that the previous interpretation is not able to explain the light scattering results. We discuss them using the BSA-YCl_3 phase diagram instead. We demonstrate that a three-phase diagram cannot capture the richness of phenomena observed by light scattering, and introduce a four-phase diagram in which the old regime II is split in two very different subregimes.

5.1 Overall View

We perform static and dynamic light scattering on BSA-salt-water solutions, at various protein and salt concentrations. The salts are NaCl, $CaCl_2$, and YCl_3. The temperature is 22 °C, that is 295 K. As a general rule, the central quantity in SLS is the ratio Kc/R, which is proportional to the inverse osmotic compressibility. In DLS, it is the long-time, collective diffusion constant D.

In figure 5.1, we present Kc/R and D as a function of protein concentration c_p, for various salt species and concentrations c_s. Every line is measured at fixed salt concentration. The black right-pointing triangles correspond to samples without added salt. The coloured curves indicate samples with salt,

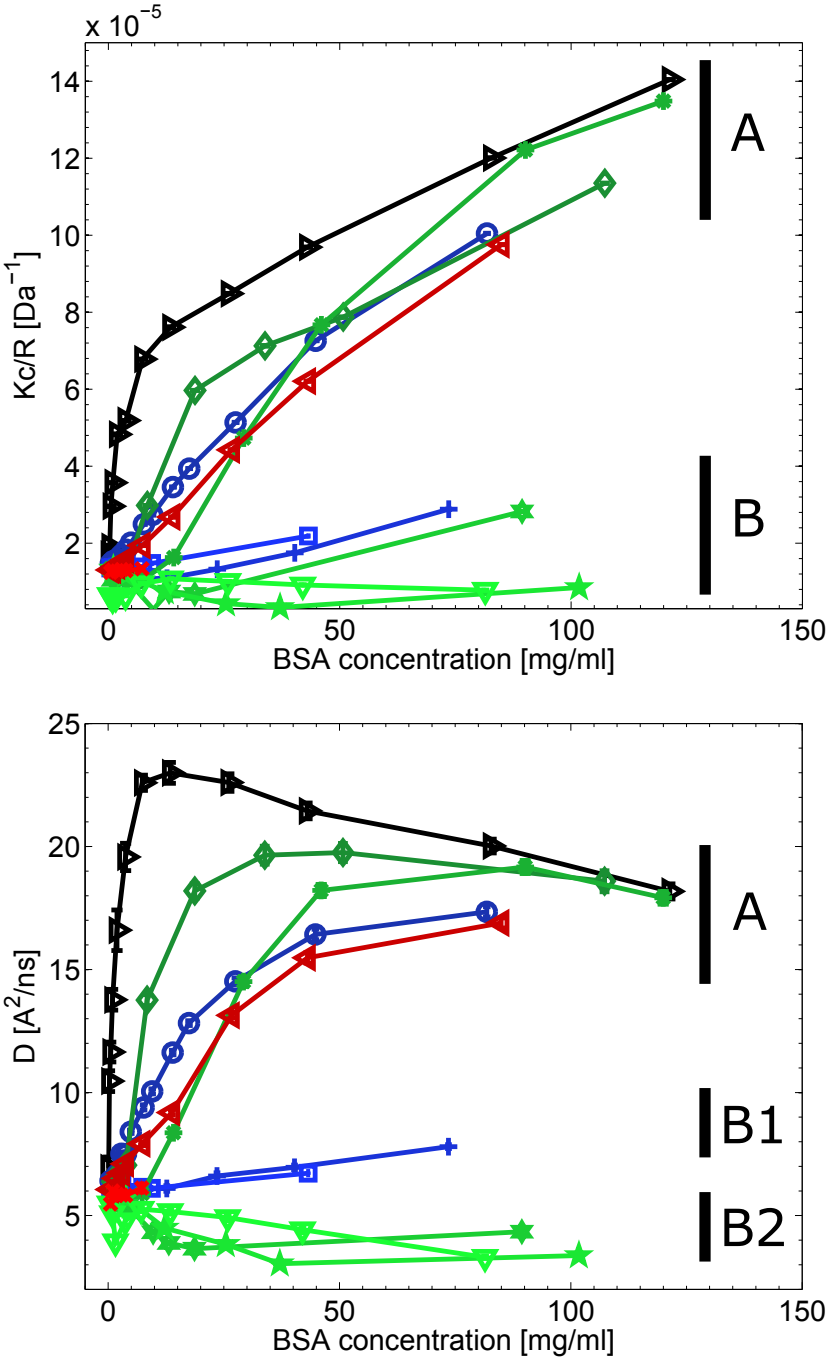

Fig. 5.1: Overview of the SLS/DLS results: Kc/R and D as a function of protein concentration for various salt species and concentrations. Legend: black right-pointing triangles, no added salt; blue circles (5 mM), plus signs (150 mM), and squares (500 mM), NaCl; red left-pointing triangles (1.67 mM) and crosses (167 mM), $CaCl_2$; green diamonds (0.2 mM), asterisks (0.83 mM), six-point stars (5 mM), five-point stars (8.3 mM), and downwards-pointing triangles (83 mM), YCl_3. Lighter tones (**B** areas) indicate higher salt concentrations. The lines are guides to the eye.

at various salt concentrations, as follows:

1. the blue curves with NaCl (circles 5 mM, plus signs 150 mM, and squares 500 mM),
2. the red curves with $CaCl_2$ (left-pointing triangles 1.67 mM and crosses 167 mM), and
3. the green curves with YCl_3 (asterisks 0.20 mM, diamonds 0.83 mM, six-point stars 5.0 mM, five-point stars 8.3 mM, and downwards-pointing triangles 83 mM).

In other words, the dark circles, left-pointing triangles, diamonds, and asterisks correspond to samples at low salt concentration. The light plus signs, squares, crosses, six- and five-point stars, and downwards-pointing triangles correspond to samples at intermediate or high salt concentration.

All solutions at low salt concentration show a similar behaviour, and form a first group, which we call **A**. Samples in this category have high inverse osmotic compressibility, i.e. high resistance to osmotic compression. They also show a fast collective diffusion. In this group, the green curves of samples with YCl_3 deserve a special remark, because Kc/R is a non-monotonic function of protein concentration. We are able to explain this peculiar behaviour on the basis of the BSA-YCl_3 phase diagram, as discussed in section 5.3.

A different physical picture describes the samples with a large amount of added salt, in group **B**. In these samples, the static light scattering is close to the ideal-solution value, $Kc/R = 1.5 \times 10^{-5}$ Da^{-1}. The diffusion constant shows a similar behaviour in solutions with NaCl and $CaCl_2$, at high c_s (group **B1**). However, the solutions with a high YCl_3 content possess a lower diffusion constant (group **B2**). The difference between the subgroups **B1** and **B2** is related to salt-specific effects of YCl_3, explained in section 5.3.

Evidently, there is a tight connection between the inverse osmotic compressibility and the diffusion constant. In fact, every sample showing a high Kc/R, i.e. a low compressibility, also possess a large D, and vice versa. Such a connection is expected from the GSV relation (3.29). We will discuss this point in section 5.2.3.

5.1.1 BSA Molar Mass

As explained in chapter 3, the ratio Kc/R should converge, at low protein concentration, to the inverse molar mass of BSA (see equation (3.8)). This is a necessary condition for the theoretical interpretation to be valid. Moreover, as seen in chapter 2, the molar mass of BSA in solution is not exactly

equal to its dry mass, the sum of the residues masses, because of the chemical association and dissociation of water ions, salt ions, and fatty acids. A reasonable value for the molar mass of BSA in solution is $M_{BSA} = 68$ kDa.

In order to verify the molar mass consistency of our SLS measurements, we take the experimental data points at low protein concentration for all datasets, and fit them polynominally as follows:

$$\frac{Kc_p}{R}\,(c_p) \quad \xrightarrow{\text{fit}} \quad k_0 + k_1\,c_p + k_2\,c_p^2 \quad ,$$

with order 1 or 2. We choose a first-order fit for solutions without added salt, with NaCl, or with $CaCl_2$; it corresponds to the virial expansion presented in section 3.1.6 (two-body SLS). For the solutions with YCl_3, we use either first- or second-order fits, as shown in Table 5.1. In this case, the virial expansion cannot be used, because the yttrium cations react chemically with BSA. However, the single-body SLS limit is still valid. Therefore, we can use the polynominal fit as a generic model function for the experimental Kc/R, and only consider the zero-order coefficient for the calculation of the measured molar mass. In fact, in all datasets, we extract the molar mass from the fit as $M = 1/k_0$, with the same relative uncertainty as k_0. In Table 5.1, we show the number of points N and the fit order O used for the various molar mass fits, the zero-order coefficients k_0, and the corresponding molar mass estimates M. As for all fit results that follow, we present the fit errors on the parameters, with 95% confidence, in parentheses, after the parameter value itself. These are errors on the last digit; for instance, the notation 74(13) is equivalent to 74 ± 13.

The found values for M are qualitatively in agreement with M_{BSA}, with the exception of the entry in bold font. However, most dataset yield a slightly higher molar mass than expected. We attribute the discrepancies to the presence of non-monomeric protein species, probably mostly dimers. This assumption is based on the observation of similar situations in the literature, e.g. in Ref. [109]. The authors of the latter and other articles, in order to obtain a higher fraction of monomers in solution, include Gel Permeation Chromatography in the sample preparation [62,82,109]. In any case, the presence of these unwanted species is not so relevant for Bovine Serum Albumin if compared to similar ones, such as Human Serum Albumin or Lysozyme [5, 110–112]. Of course, the numerical values for static properties from those samples must be taken *cum grano salis*, including the second virial coefficient. DLS and dynamic properties are not strongly affected, as explained in the next section.

Table 5.1: Molar mass estimates from SLS at low c_p. N is the number of points used for the fit, O the fit order, k_0 is the zero-order fit coefficient, and M its reciprocal, the molar mass.

	N	O	k_0 [10^{-5} Da^{-1}]		M [kDa]	
no salt	5	1	1.7	(7)	60	(24)
5 mM NaCl	5	1	1.43	(1)	70.6	(1)
150 mM NaCl	5	1	1.06	(3)	95	(2)
500 mM NaCl	5	1	1.33	(1)	75	(1)
1.67 mM $CaCl_2$	5	1	1.21	(2)	83	(1)
167 mM $CaCl_2$	5	1	1.25	(7)	80	(4)
0.20 mM YCl_3	3	1	1.2	(5)	80	(30)
0.83 mM YCl_3	5	2	1.40	(2)	71	(1)
5.0 mM YCl_3	4	1	1.2	(7)	85	(50)
8.3 mM YCl_3	5	2	1.3	(2)	74	(13)
83 mM YCl_3	5	2	**0.67**	**(1)**	**149**	**(3)**

5.1.2 Dilute-Limit D_0 and Hydrodynamic Radius

The theory of diffusion, outlined in chapter 3, predicts that the diffusion constant in the dilute limit depends only on the shape of a single BSA molecule, and not on the interactions. This statement is made quantitative by the SE equation (3.27). As long as the protein shape is unaffected by the presence of light salts, the extrapolation of D towards vanishing c_p should be almost the same for all datasets. We perform such an extrapolation by numerically fitting the diffusion constant, via the following fit function:

$$D\,(c_p) \quad \xrightarrow{\text{fit}} \quad D_0\,\left(1 + k_d\,c_p + k_{d2}\,c_p^2\right) \quad ,$$

where the fit parameters are D_0, k_d, and k_{d2}. The latter parameter is only used for few datasets, which cannot by fitted otherwise; for the other ones, we set $k_{d2} := 0$. We specify the number of points N and the fit order O, used for the different datasets, in Table 5.2. From the SE relation (3.27), it is possible to calculate the hydrodynamic radius r_h from D_0. Both properties are listed in Table 5.2 for our samples. These numbers agree with the literature value, $D_0 = 6.0$ Å^2 ns^{-1} [60, 61]. Strictly speaking, tiny differences among the datasets are expected, because the solvent viscosity is slightly changed by the different salts. However, these are negligible compared to the experimental error bars.

Comparing the DLS dilute-limit results in Table 5.2 with the SLS ones in Table 5.1, it seems that D is less sensitive than Kc/R to the presence of

Table 5.2: Estimates for D_0 and r_h at low c_p, from DLS. The number of points used for the fit is N, the fit order is O. The values of D_0 under different conditions may differ slightly, because of the change in solvent viscosity (see equation (3.27)).

	N	O	D_0 [Å^2 ns^{-1}]		r_h [nm]	
no salt	4	1	7	(5)	3	(2)
5 mM NaCl	5	1	6.1	(6)	3.7	(4)
150 mM NaCl	5	1	5.9	(2)	3.8	(2)
500 mM NaCl	5	1	5.9	(5)	3.8	(3)
1.67 mM $CaCl_2$	5	1	5.9	(2)	3.9	(1)
167 mM $CaCl_2$	5	1	5.7	(6)	4.0	(4)
0.20 mM YCl_3	3	1	6.1	(6)	3.7	(4)
0.83 mM YCl_3	5	2	6.3	(6)	3.6	(3)
5.0 mM YCl_3	4	1	6.1	(7)	3.7	(4)
8.3 mM YCl_3	5	2	6.0	(8)	3.7	(5)
83 mM YCl_3	5	2	**5**	**(3)**	**4**	**(2)**

BSA oligomers. We expected this result, because we calculate the diffusion constant via the double-exponential decay fit routine, which reduces strongly the disturbances of heavier solute species. The agreement of D_0 with the literature value is an ulterior confirmation of the validity of this fit strategy.

However, both the molar mass and the dilute-limit diffusion constant of the dataset with 83 mM YCl_3 seems to be still biased towards bigger protein complexes. We interpret this fact as a hint of real protein association. We refer the reader to section 5.3 for further details.

5.2 Samples with Sodium and Calcium Chloride

The BSA solutions with sodium and calcium chloride show relatively simple and intuitive light scattering results. We present and discuss them in the following sections. We start with static light scattering, including a calculation of the second virial coefficient and a more general discussion based on the osmotic compressibility. Then, we describe the results of dynamic light scattering, dealing with the dilute, as well as the concentrated solutions. We combine SLS and DLS in the third section, in which we derive the normalized sedimentation coefficient, and the second hydrodynamic coefficient. Finally, we present the theoretical ideas of electrostatic screening and short-range

interactions, in particular steric repulsion, proving them able to explain all light scattering results in a unified conceptual framework.

5.2.1 Static Light Scattering

In this section, we analyse the SLS results in detail. First of all, we present the findings, from dilute solutions, on the second virial coefficient. Second, we discuss the osmotic compressibility in crowded solutions.

Second Virial Coefficient

The second virial coefficient can be extracted from SLS measurements as explained in chapter 3. If the experimentally fitted molar mass agreed with the expected one for all dataset, we could simply fit Kc/R linearly, and interpret the first-order coefficient, i.e. the slope, as $B_2(\mathbf{SLS})$. In turn, this would yield, using the conversion factors in equation (3.11), the virial cofficient in theoretical units, $B_2(\mathbf{t.u.})$, to be compared with models. Unfortunately, the actual situation is not so good, because the values of M in Table 5.1 are not always equal to $\mathrm{M_{BSA}}$. For this reason, we must wonder whether to fix the intercept of the linear fit equal to $\mathrm{M_{BSA}}$. Actually, the discrepancy between M and $\mathrm{M_{BSA}}$ is probably due to oligomers in the solution. However, their number is not large, otherwise we would measure molar masses much higher than 100 kDa. They are also not big, because we see no Q-dependence of the scattering intensity. The influence of the few oligomers on M is pronounced, but that on B_2 is rather limited, because oligomers and monomers presumably generate similar force fields. Thus, we choose to let all fit parameters free, so that the bias in M is partially counterbalanced, without affecting much the second virial coefficient. As a consequence, these fits actually coincide with the fits used for the determination of the molar mass, presented in a previous section. We use five experimental points at different c_p for every dataset, as shown in Table 5.1.

It is less clear what molar mass is best suited for the subsequent conversions to theoretical units, needed by equations (3.11) and (3.13). The fitted one, M, would make our calculation self-consistent, but M is quite sensitive to the fit details, such as error bars, the number of points considered, and the degree of the polynomial. The literature value, $\mathrm{M_{BSA}}$, can be considered error-free instead, but its applicability to these slightly polydisperse solutions is doubtful. Since this question is subtle and cannot be settled easily, we calculate $B_2(\mathbf{t.u.})$ in both ways. We list the numerical estimates in Table 5.3. In the same table, we present the second virial coefficient normalized to the

Table 5.3: B_2 estimates from SLS at low c_p. The first column is the result of a linear or quadratic fit with all parameters free. The second column shows B_2 converted into theoretical units (**t.u.**) using the fitted values of M, given in Table 5.1. The third column shows B_2 calculating through M_{BSA}. The last column shows the second virial coefficient normalized to the hard-sphere one, b_2, calculated via M_{BSA}.

	B_2(**SLS**) [10^{-4} · mol ml g^{-2}]	B_2(**t.u.**) via M [10^{-24} m^3]	B_2(**t.u.**) via M_{BSA} [10^{-24} m^3]	b_2 via M_{BSA}
no salt	90 (30)	50 (20)	60 (20)	100 (30)
5 mM NaCl	4.9 (1)	4.0 (1)	3.7 (1)	6.0 (2)
150 mM NaCl	0.1 (2)	0.2 (1)	0.1 (1)	0.2 (1)
500 mM NaCl	1.5 (5)	1.3 (3)	1.0 (3)	1.6 (5)
1.67 mM $CaCl_2$	4.9 (5)	5.5 (5)	3.7 (4)	6.0 (6)
167 mM $CaCl_2$	1 (1)	1 (1)	1 (1)	2 (2)

hard-sphere value, b_2, as explained in equation (3.13), using M_{BSA} in the units conversion.

The results for the second virial coefficient look very interesting. The dataset without added salt has by far the largest, positive B_2. This is surely due to the electrostatic repulsion between like-charge BSA molecules, which makes the integrand in the definition (3.10) positive. Actually, the Kc/R curve from samples without added salt becomes nonlinear already at a fairly low protein concentration, around 4 mg/ml. For this reason, in order to collect enough data points in the linear region, we had to measure SLS in very dilute solutions, at c_p as low as 0.1 mg/ml. At $c_p \gtrsim 4$ mg/ml, many-body effects are important.

The opposite happens in dataset with high concentration of added monovalent and divalent salts, NaCl and $CaCl_2$. In these cases, B_2 is much smaller, and the whole Kc/R curve is linear. The most dilute measured solution in the datasets with salts has $c_p \approx 0.5$ mg/ml. The solutions with an intermediate amount of added salt present, consistently, an intermediate virial coefficient. Therefore, it seems that the addition of dissociable salt reduces the strength of the protein-protein repulsion, $u(r)$, but never so much to commute it into a net attraction. Furthermore, in Table 5.3, we observe that the decrease of B_2 at low salt concentration is very steep, so the protein-protein repulsion is neutralized with a fairly low amount of salt. At high salt concentration, the repulsion is completely switched off, because we observe no difference between the datasets with 150 mM and 500 mM NaCl.

In addition, we remark that the B_2 in the datasets at 5 mM NaCl and 1.67 mM $CaCl_2$ are equal, although the salt species and concentrations are different. The same is true, within the error bars, for the datasets at 500 mM NaCl and 167 mM $CaCl_2$.

All these phenomena can be explained using the concept of *electrostatic screening*. Basically, the free ions in the solution tend to screen the charge of a BSA molecule to the other ones. In other words, from an electrostatic point of view, the strong repulsion caused by the protein charge is weakened by the presence of the free ions. Moreover, the central ingredient of the protein-protein interactions, in dilute solutions, is not the salt concentration, nor the salt species, but the so-called *ionic strength*. We give a detailed explanation of this point in section 5.2.4.

As a side note, we notice that the smallest values for $B_2(\mathbf{SLS})$ and its uncertainty are of the order of magnitude of 10^{-4} mol ml g^{-2}, which is also the lowest limit of the ALV CGS3 instrument used for our experiments [113].

Osmotic Compressibility

At high protein concentration, the second-order virial expansion, which predicts a linear relation between c_p and Kc/R, fails. We report the SLS results from NaCl and $CaCl_2$ solutions in Fig. 5.2. The prominent feature of the curves Kc/R versus c_p shown is a monotonic, but sublinear behaviour. In order to present the data in terms of standard thermodynamic variable, we calculate the normalized osmotic compressibility, using the theoretical machinery introduced in chapter 3, in particular equation (3.16). Just as for the theoretical second virial coefficient, we need the BSA molar mass in this calculation. However, in this case we can only use the fitted M, because we want to assure the self-consistency in the dilute limit,

$$\lim_{c_p \to 0} \frac{\chi_T}{\chi_{T0}} \quad = \quad 1.$$

We show χ_T/χ_{T0} in Fig. 5.3. As a direct consequence of the positiveness of B_2 and the monotonicity of Kc/R, the osmotic compressibility of our protein samples is lower than that of an ideal solution. This effect is also caused by the protein-protein repulsion. In fact, the osmotic compressibility of salt-free solutions is much lower than that of high-c_s solutions with NaCl and $CaCl_2$. This general trend notwithstanding, the compressibility in more and more crowded solutions seems to become less and less dependent on the presence of salt, at least by comparing the black lines with the low-salt, dark ones. This phenomenon gives the hint that the electrostatic repulsion cannot be the only important interaction in crowded solutions. Different, generic

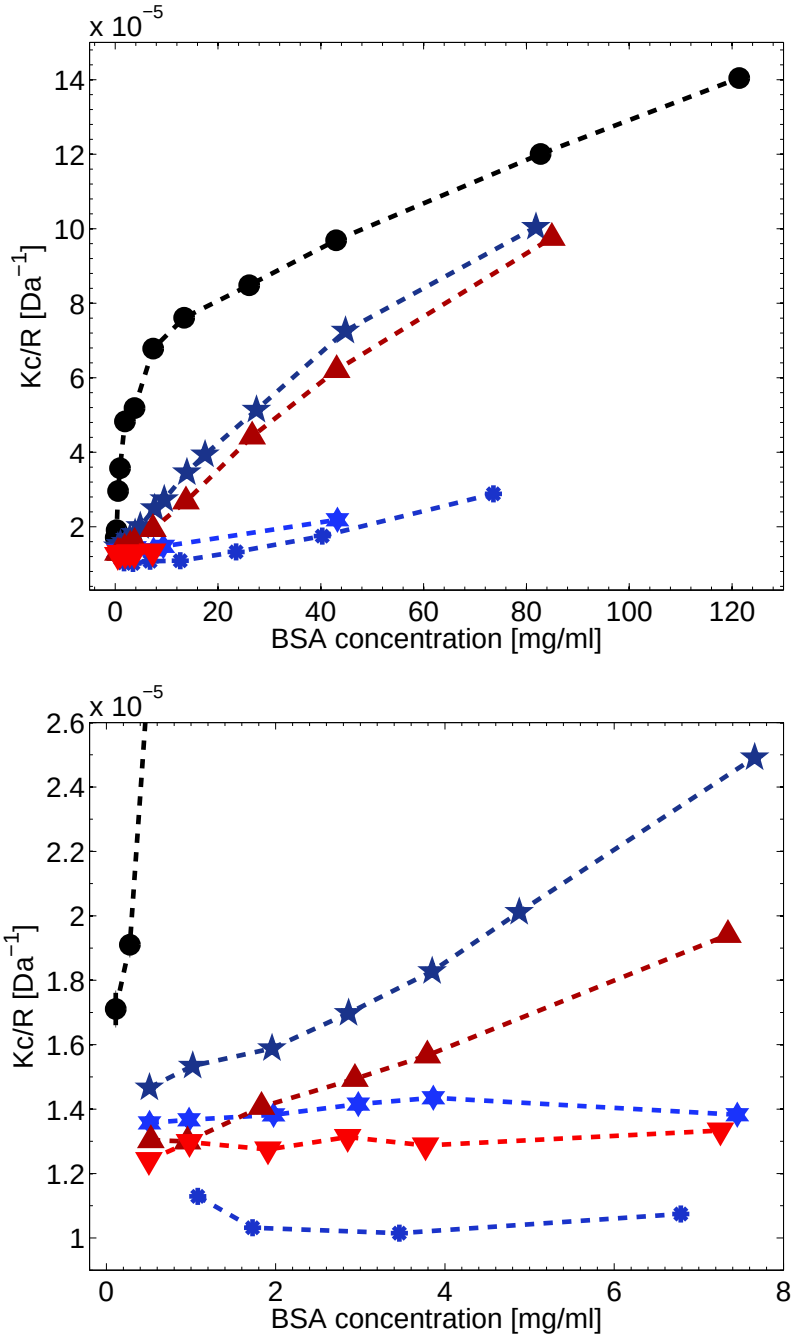

Fig. 5.2: SLS results for solutions without added salt, with NaCl, and with $CaCl_2$. The second plot is a zoom in the dilute region. Legend: no salt (black circles), 5 mM NaCl (dark blue five-point stars), 150 mM NaCl (medium blue six-point stars), 500 mM NaCl (light blue asterisks), 1.67 mM $CaCl_2$ (dark red upwards pointing triangles), and 167 mM $CaCl_2$ (light red downwards pointing triangles). The dotted lines are guides to the eye.

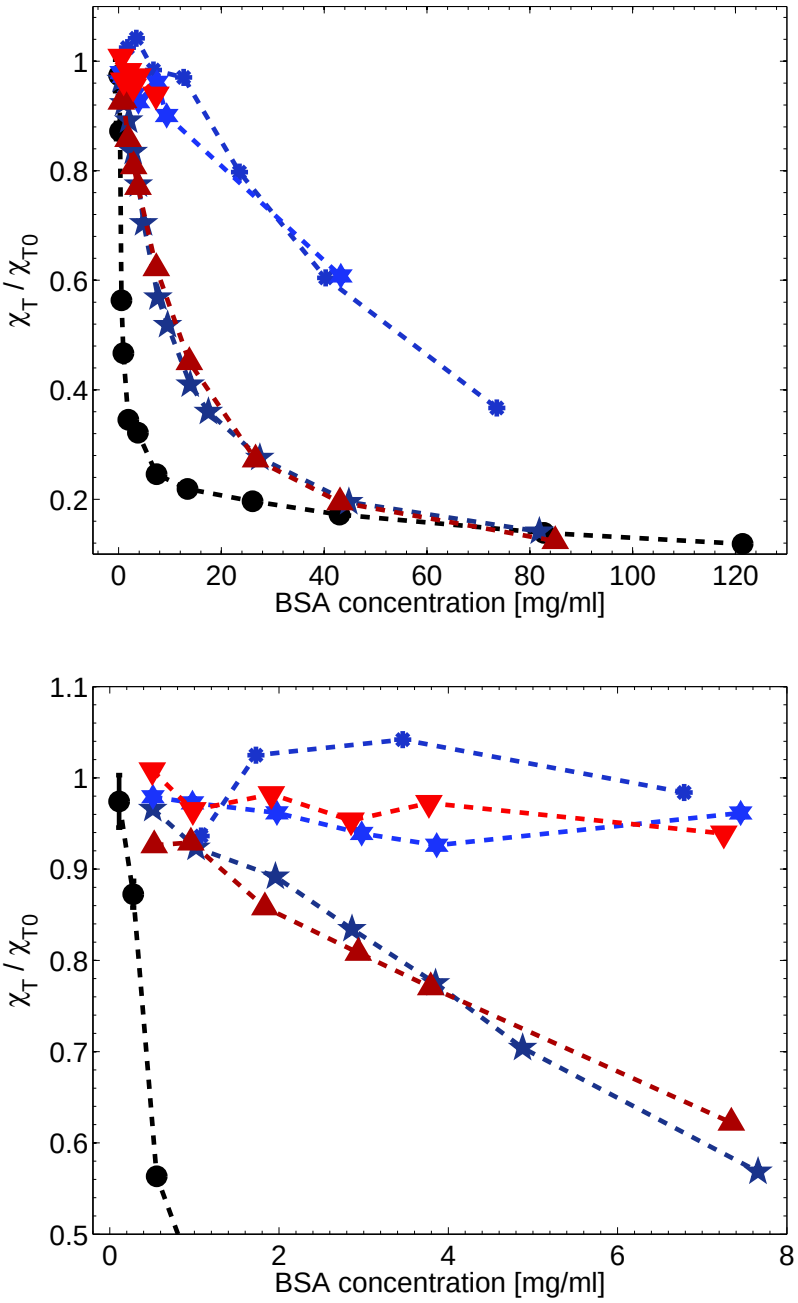

Fig. 5.3: Normalized osmotic compressibility for solutions without added salt, with NaCl, and with $CaCl_2$. The second plot is a zoom in the dilute region. Legend: no salt (black circles), 5 mM NaCl (dark blue five-point stars), 150 mM NaCl (medium blue six-point stars), 500 mM NaCl (light blue asterisks), 1.67 mM $CaCl_2$ (dark red upwards pointing triangles), and 167 mM $CaCl_2$ (light red downwards pointing triangles). The dotted lines are guides to the eye.

interactions have to be dominant in crowded solutions instead. Since the average protein-protein distance in increasingly crowded solutions becomes smaller, those additional forces are probably short-ranged. We will show below that the most important short-range interaction is the steric repulsion.

The osmotic compressibilities of solutions with 5 mM NaCl and with 1.67 mM $CaCl_2$ show a spectacular resemblance. This observation extends the likeliness of their second virial coefficients to the crowded region; those datasets really behave as one. This phenomenon confirms that the protein-protein interactions, in dilute as well as in crowded solutions, do not depend explicitly on the salt species nor on their concentration, but only on the ionic strength, as well as on the BSA concentration.

Summarizing, the SLS results seem to be in agreement with an explanation based on electrostatic screening, in the dilute region, and on generic short-range interaction, in the crowded region. This hypothesis is confirmed by the DLS results indeed, as we show in the next sections. See also section 5.2.4 for a detailed explanation.

In addition, we must emphasize that the *caveat* about the presence of BSA oligomers in the samples, already mentioned in the context of the second virial coefficient, holds invariably for these results on the osmotic compressibility. In other words, even if the normalization of χ_T through M reduces the bias in Kc/R due to oligomers, the presence of those disturbing species could slightly alter our plots. For crowded solutions, there is a further point of caution, because the fraction of oligomers in solution might in fact increase with increasing protein concentration. In fact, from an intuitive point of view, it seems more probable that two proteins touch and merge if the mean intermolecular distance is smaller, i.e. at high c_p. However, we confirm that, even in the highest concentrated solutions, we do not see any dependence of Kc/R on the scattering vector. We conclude that the heavier protein species are relatively small and in low number, and should not change our results but of a few percent.

5.2.2 Dynamic Light Scattering

The main result of DLS on protein solutions is the zero-Q, long-time, collective diffusion constant as a function of protein concentration, salt type, and salt concentration. Because of the probable presence of oligomers in solution, we use mainly the ad-hoc double decay fit strategy. We interpret the faster decay as the diffusion of interacting BSA monomers, and the slower decay as a disturbance coming from protein complexes. After several tentative data analyses, we conclude that the slower decay is too strongly affected by random aggregates to be studied systematically. Therefore, we do not present

any result on D_2, limiting the analysis to D_1 instead. We show results of cumulant analysis as well, for comparison with older literature.

This section is structured similar to the previous one. Firstly, we present the results at low protein concentrations, where we fit D as a function of BSA concentration. Then, we show our findings in the crowded regime.

Dilute Solutions: Linear Region

In solutions at low but nonzero protein concentration, the diffusion constant starts to deviate from its dilute limit value, because of the influence of interactions, both direct – electrostatic, van der Waals, steric – and hydrodynamic ones. In order to study this influence, we fit D affinely as a function of c_p, that is

$$D\ (c_p) \quad \xrightarrow{\text{fit}} \quad D_0\ (1 + k_d\, c_p) \quad ,$$

For the same reasons mentioned in the discussion of the second virial coefficient, we do not fix D_0 in the fits; both parameters, D_0 and k_d, are freely adjustable. We have already used the D_0 from those fits in Table 5.2, in which we also show the number of experimental points used for every fit. Just like B_2 for the osmotic compressibility (or pressure), k_d is the best one-number correction to the dilute limit diffusion coefficient in presence of interactions. Also in this case, k_d should represent a property of BSA monomers reasonably well, even if the solution is slightly polydisperse. We list the fit results for k_d in Table 5.4, both in experimental units and in theoretical ones. We use the fitted mass, M, for the unit conversion, but the choice is practically negligible in this case. We must remark that the reported error bars most likely overestimate the Bayesian uncertainty on k_d. For instance, a naked-eye observation of the curves of 5 mM NaCl and 1.67 mM $CaCl_2$ indicates that they have a very similar slope, just as their respective osmotic compressibilities. This is probably a consequence of an instability in the fitting procedure, but does not represent a problem, for two reasons. First, our results in the crowded region enable us to compare the diffusion of BSA solutions with different salt species directly, in a wide range of protein concentrations. Second, the values of the fit parameters, D_0 and k_d, are indeed in agreement with the experimental points, as we can check easily with the help of Fig. 5.4.

Concentrated Solutions: Crowding Effect

The results on diffusion in more concentrated protein solutions are biologically more relevant, because both the cellular environment and the extracellular one (e.g. the blood) are quite crowded, approximately between 35 and 50 mg/ml [1]. In Fig. 5.4, we plot the diffusion constant, as a function of

Table 5.4: Estimates for the slope coefficient k_d from DLS at low c_p. The values in the first column are expressed in experimental units. The second column shows k_h in theoretical units (**t.u.**), obtained using the values of M given in Table 5.1.

	k_d(**DLS**) [ml/g]		k_d(**t.u.**) via M [10^{-24} m^3]	
no salt	1000	(2000)	100	(200)
5 mM NaCl	60	(50)	7	(6)
150 mM NaCl	2	(6)	0.3	(9)
500 mM NaCl	0	(40)	0	(5)
1.67 mM $CaCl_2$	55	(15)	8	(2)
167 mM $CaCl_2$	9	(40)	1	(6)

protein concentration, from a double decay fit. Subsequently, we also report similar curves calculated by the first-order cumulant analysis, in Fig. 5.5. We must notice that, in the cumulant fits, the dynamic range of the correlator is restricted to a small, short-time window. These plots resemble, at first sight, the SLS plots of Kc/R, shown in Fig. 5.2. In other words, the following ordering is conserved: salt-free, low salt concentration, high salt concentration. This means that a high stiffness corresponds experimentally to a fast collective diffusion. Analogously, the solutions at high salt concentration, whose χ_T depends only weakly on c_p, show a correspondingly little variation of D. We analyze this important correspondence further in section 5.2.3, but we anticipate a short remark: the experimental observation that Kc/R and D possess similar dependences on c_p does not imply, *ipso facto*, any causal chain between the two quantities.

The DLS curves show a peculiarity, the maximum in the collective diffusion constant for samples without added salt. The samples at low salt concentration seem to possess a shallow maximum as well, at higher BSA concentration. This feature is best illustrated in the cumulant plots, in Fig. 5.5. The diffusion constant of the high-c_s dataset present no maximum instead, being mostly flat or slightly increasing. We are able to explain such maxima, at least qualitatively, using the same key concepts as for the SLS results, i.e. the ionic strength and generic short-range interactions. Here, we only elucidate the main effect, but we give a detailed interpretation in section 5.2.4. Basically, in crowded solution, the interprotein distance shrinks and the BSA molecules come more often in contact, with dramatic consequences on the dynamic properties of the protein. In fact, when two molecules touch each

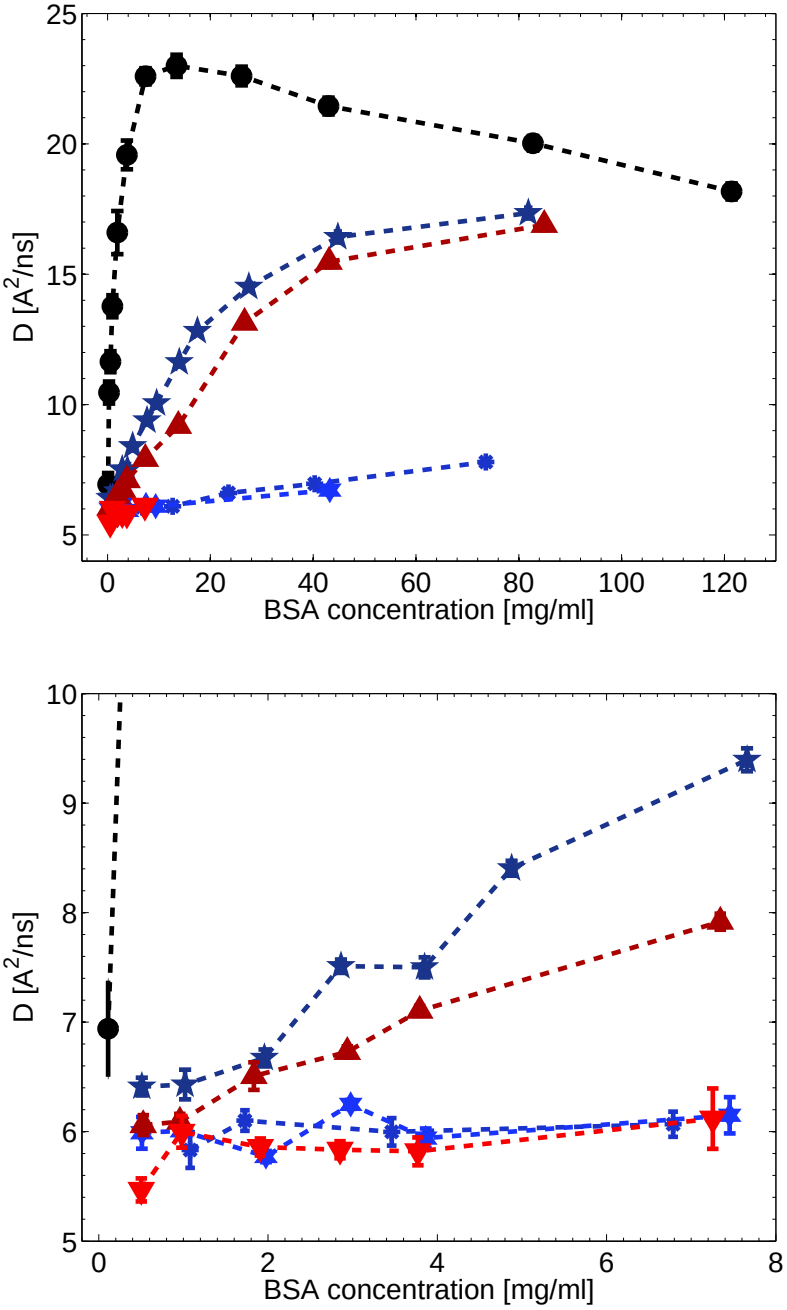

Fig. 5.4: Diffusion constant for solutions without added salt, with NaCl, and with $CaCl_2$, obtained by the ad-hoc double exponential decay fit routine. The second plot is a zoom in the dilute region. Legend: no salt (black circles), 5 mM NaCl (dark blue five-point stars), 150 mM NaCl (medium blue six-point stars), 500 mM NaCl (light blue asterisks), 1.67 mM $CaCl_2$ (dark red upwards pointing triangles), and 167 mM $CaCl_2$ (light red downwards pointing triangles). The dotted lines are guides to the eye.

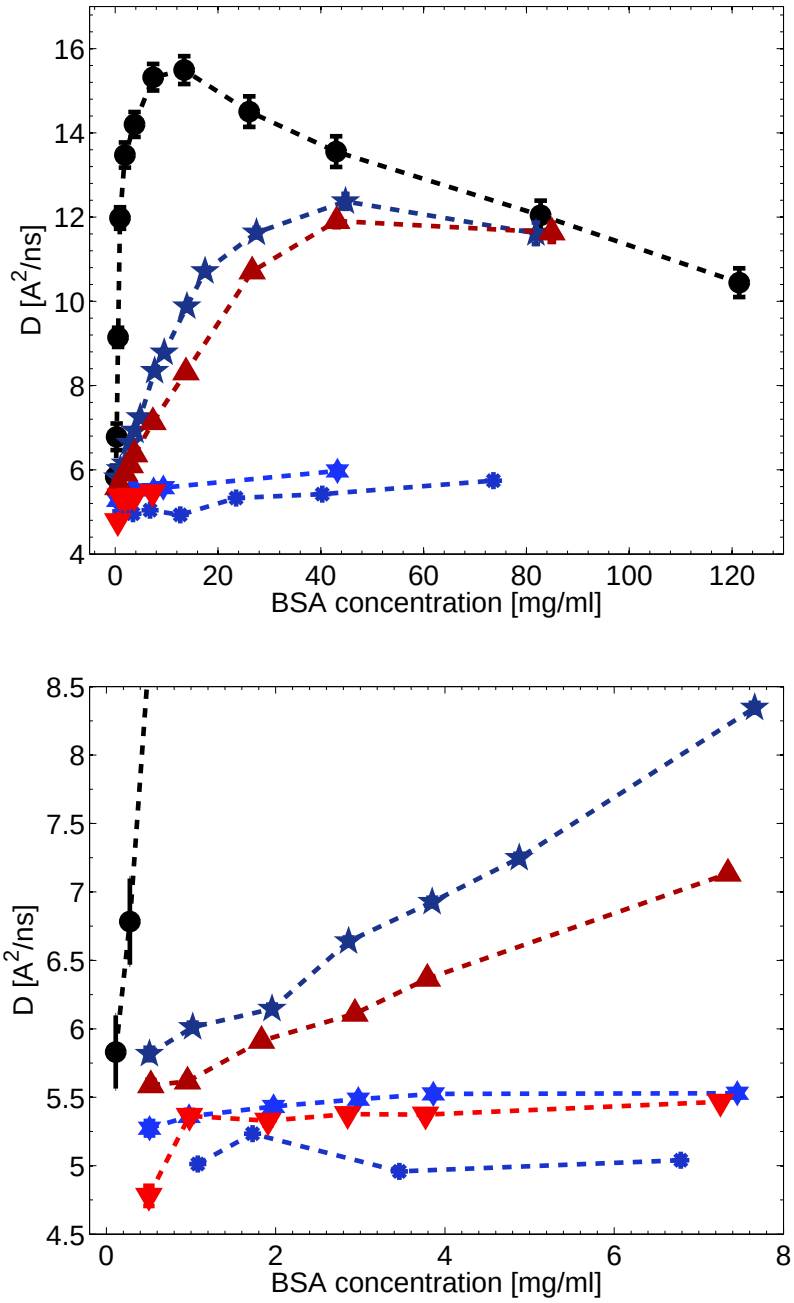

Fig. 5.5: Diffusion constant for solutions without added salt, with NaCl, and with $CaCl_2$, obtained by the first-order cumulant fit routine. The second plot is a zoom in the dilute region. Legend: no salt (black circles), 5 mM NaCl (dark blue five-point stars), 150 mM NaCl (medium blue six-point stars), 500 mM NaCl (light blue asterisks), 1.67 mM $CaCl_2$ (dark red upwards pointing triangles), and 167 mM $CaCl_2$ (light red downwards pointing triangles). The dotted lines are guides to the eye.

other, their surfaces cannot slip on one another, and a certain amount of kinetic energy is dissipated. More generally, at small interparticle distance, i.e. in crowded solutions, the solvent tends to dissipate more energy than in dilute solutions. The translational diffusion is thus vastly hindered by such situations of contact or proximity; although D tends to increase because of the stiffness-reactivity correspondence, it saturates at a certain BSA concentration, at which the contacts are frequent enough, and slows down afterwards.

Fig. 5.4 and 5.5 deserve a last comment. Surprisingly, in crowded solutions, D is never smaller than its dilute limit value D_0,

$$D(c_p, c_s) \geq D_0 \qquad \forall\, c_s,\, c_p.$$

In other words, concentrated protein solutions react to external disturbances faster than dilute ones. This result seem to contradict the physical intuition. At first sight, we would expect the high particle density to slow down relaxation dynamics, certainly not to *boost* it up! Actually, it this result can be made intuitive with a simple argument. The crucial observation is that BSA in solution is charged. If we imagine to increase continuously the protein concentration, we are forcing more and more like-charge molecules near to each other. Thus, the total potential energy increases steeply. Because of the high energy cost for two proteins to come close, they tend to quickly maximize their distance, making the relaxation energy landscape smoother. In conclusion, D is higher than D_0. We must stress, however, that this simplistic picture is probably invalid in very crowded solutions, where the proteins touch each other very often. Thus, we cannot exclude that $D < D_0$ at high enough c_p, say, above 300 mg/ml.

5.2.3 Hydrodynamics and Sedimentation

The simultaneous measurement of D and χ_T makes possible to exploit the generalized Svedberg equation (3.29) to calculate the normalized sedimentation coefficient, s/s_0. We emphasize that we have not performed any sedimentation measurements ourselves. We first compute a non-normalized sedimentation coefficient,

$$s_{\mathrm{nn}} := D \cdot \left[\frac{K c_p}{R}\right]^{-1} \quad ,$$

where D comes from DLS, and Kc/R from SLS. The subscript nn indicates that this quantity is not normalized to one in the dilute limit. In principle, the normalization constant should be simply $D_0\, \mathrm{M_{BSA}}$. However, the presence of

Table 5.5: Estimates for the second hydrodynamic coefficient k_h from DLS/SLS at low c_p. The first column shows the number of points, N, used for the fit. The second column shows k_h obtained from the fit procedure (see text). The third column shows the same quantity in theoretical units (**t.u.**), calculated using the molar mass fitted by SLS, given in Table 5.1.

	N	k_h(**exp**) [ml/g]	k_h(**t.u.**) via M [10^{-24} m^3]
no salt	5	-150 (300)	-5 (30)
5 mM NaCl	5	-8 (40)	-1 (4)
150 mM NaCl	5	0 (20)	0 (3)
500 mM NaCl	5	-10 (30)	-2 (4)
1.67 mM $CaCl_2$	6	-18 (6)	-2.5 (9)
167 mM $CaCl_2$	6	2 (10)	0 (1)

oligomers in solution makes the determination of this quantity very delicate. Therefore, we choose to normalize the data in another fashion. First, we fit the second hydrodynamic coefficient in the next section. Then, we use the normalization from those fits to calculate s/s_0. We plot this latter quantity in the second-next section.

Second Hydrodynamic Coefficient

The expansion of s/s_0 in powers of c_p yields the second hydrodynamic coefficient k_h. We obtain this quantity by fitting the sedimentation coefficient s_{nn}, output of the previous calculation, by the following affine function:

$$s_{\mathrm{nn}} \quad \xrightarrow{\text{fit}} \quad A\ (1 + k_h\ c_p) \quad .$$

The fit parameters are A and k_h. The second fit parameter, k_h, is the second hydrodynamic coefficient. We list the fit results in Table 5.5, both in experimental units and in theoretical ones. In the same table, we also show the number N of experimental points used for all fits. We only use the molar mass M fitted by SLS for the units conversion. The second hydrodynamic coefficient is always negative. This means that the protein dynamics is slowed down by the hydrodynamic interactions, which act essentially as an additional friction. Moreover, the ordering found for k_d is inversed, as a consequence of B_2 in equation (3.30). In other words, the sedimentation is much slower in salt-free solutions than in solutions at high salt concentration.

This phenomenon can be rationalized on the basis of the screening concept, which we describe further in section 5.2.4. If the repulsion among proteins is weak, as in the solutions at high salt concentration, the proteins have less energy to balance the gravity force, and tend to come near to each other towards the bottom of the experimental tube. In solutions without added salt, in which the electrostatic repulsion is strong, the proteins tend to maximize the mutual distance, even if this has a certain cost in terms of gravitational potential.

The counterintuitive hierarchy reversal demonstrates another relevant conclusion, that hydrodynamic interactions play a key role in determining the collective motions of proteins. In other words, a static knowledge of a motile protein is intrinsically insufficient for determining its biological function.

Many-Body Hydrodynamics

The fit of the second hydrodynamic coefficient yields, as a byproduct, the normalization constant for the sedimentation ratio, s/s_0, which we calculate as follows:

$$\frac{s}{s_0} = \frac{s_{\rm nn}}{A} \quad .$$

This quantity converges to one in the dilute limit, by construction. We plot the ratio as a function of protein concentration, in Fig. 5.6. We show no error bars for this plot, in order to keep the figure readable, but the reader can see from the point fluctuations that the data are quite noisy. Anyway, it is clear that the sedimentation coefficient of crowded solutions, for all salt concentrations, becomes much smaller than s_0. Furthermore, the virtual projections of all curves converge above 80 mg/ml, even for samples at high salt concentration. This convergence is more rapid, in the case of high-c_s solutions, than that of χ_T or D. We guess that the basic mechanism behind the convergence is the same, i.e. the generic short-range interactions overcoming electrostatics, but the Coulomb repulsion seems to be neutralized more effectively in s than in χ_T or D.

5.2.4 Theoretical Interpretation

Our discussion of our results from salt-free and normal-salt solutions is founded upon the concepts of *electrostatic screening* and *generic short-ranged interactions*, which we have already mentioned several time in the last sections. The main idea is the following. BSA in solution has two principal properties, its charge and its finite dimensions. The former causes a relatively long-ranged

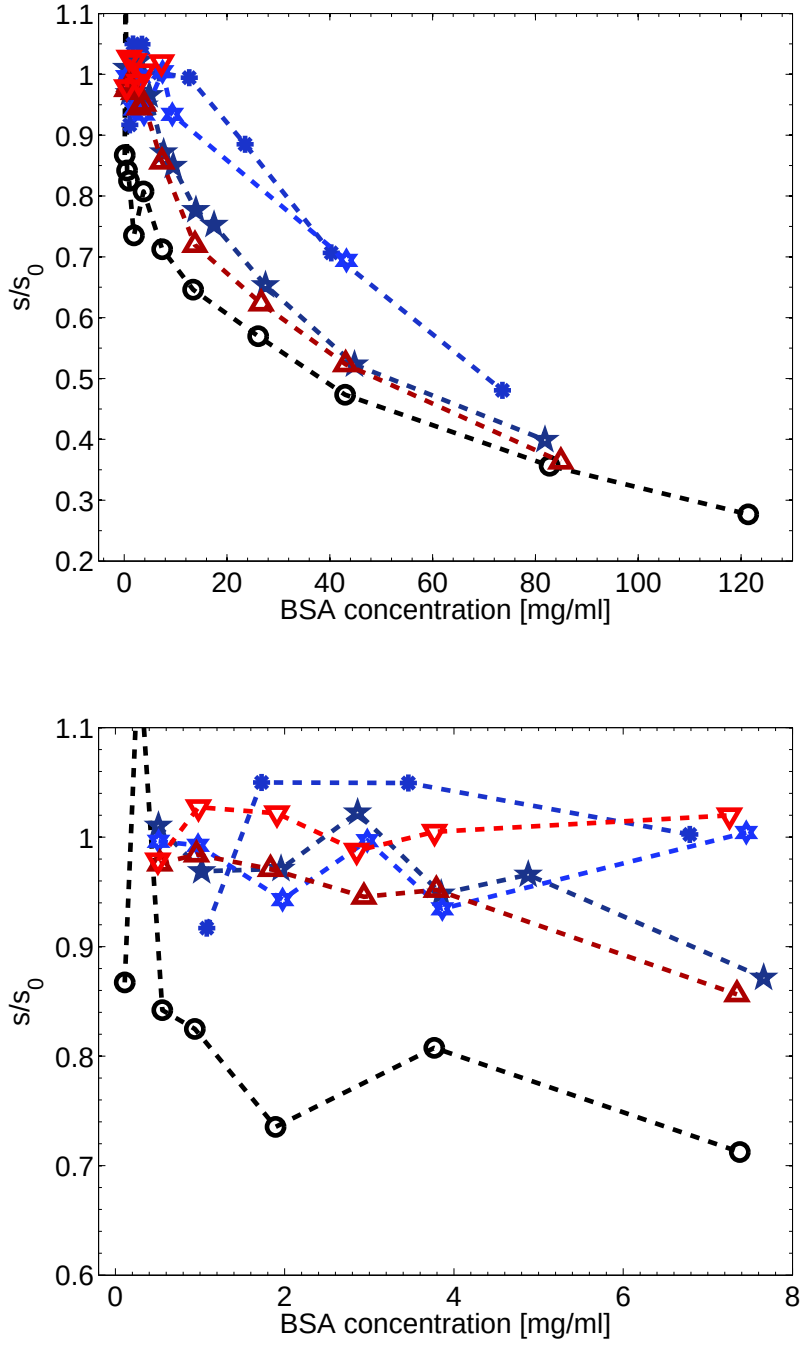

Fig. 5.6: Normalized sedimentation coefficient for solutions without added salt, with NaCl, and with $CaCl_2$. The second plot is a zoom in the dilute region. Legend: no salt (black circles), 5 mM NaCl (dark blue five-point stars), 150 mM NaCl (medium blue six-point stars), 500 mM NaCl (light blue asterisks), 1.67 mM $CaCl_2$ (dark red upwards pointing triangles), and 167 mM $CaCl_2$ (light red downwards pointing triangles). The dotted lines are guides to the eye.

repulsion of electrostatic origin and dominates in dilute solutions, because the mean interprotein distance is large. In more and more crowded solutions, the proteins are closer to one another, and short range interactions play an increasingly important role. Most of these short-range forces do not depend strongly on the protein charge nor on the salt concentration. Among them, we mention van der Waals attraction and steric repulsion.

In the following subsections, we present the two concepts in deeper detail, and discuss their relevance in the context of our experimental findings. First of all, we deal with electrostatic screening, then with the generic short-range interactions.

Electrostatic Screening

The concept of electrostatic screening is old, and several theories have been built on it, including notably those of Debye-Hückel [34], and of Derjaguin-Landau-Vervey-Overbeek (DLVO) [35]. In solution, BSA is negatively charged, because it releases a number of ions, called *counterions*, that prefer to associate with water than with the protein itself. Because of those counterions and, if present, of the ions of dissociated added salts, the protein environment is populated not only by polarizable water molecules, as in usual dielectrics, but also by mobile charges. We call i the generic ionic species in solution. Let its number concentration be N_i, and its charge Z_i, in elementary charge units. For instance, in solutions with NaCl, i includes Na^+ with $Z_{Na^+} = +1$, Cl^- with $Z_{Cl^-} = -1$, and any other ionic species released by BSA, plus the water ions, OH^- and H^+ – or its compounds, such as H_3O^+.

Of course, the protein-protein repulsion $u(r)$ depends on the position of all charges. Vice versa, the ions follow the electric field lines in order to minimize their energy. The simplest mean-field approximation neglects the ion-ion interactions, both of enthalpic, and of entropic nature. Under further limiting assumptions, among them the smallness of the protein and ion charges, and a not-too-high salt concentration, $u(r)$ is a power-exponential law

$$u(r) \quad \propto \quad \frac{e^{-\kappa r}}{r} \quad ,$$

where κ is the so-called *Debye screening parameter*, which describes the range of the repulsion. This parameter, in turn, depends on the amount of salt dissolved in solution,

$$\kappa \quad \propto \quad \sqrt{\frac{1}{2} \sum_i Z_i^2 \, N_i} \quad , \tag{5.1}$$

whereby the sum runs over all ionic species, including the ions dissociated from the added salt. The quantity under the square root in formula (5.1) is called *ionic strength*. We can explain some of our results based on this property alone. First of all, we explicate the hierarchy in B_2 and k_d,

$$\text{no salt} \quad > \quad \text{low-}c_s \quad > \quad \text{high-}c_s \quad .$$

In fact, the higher the salt concentration, the shorter and less effective the protein-protein repulsion; the mobile ions in solution are said to *screen* the interaction. At high salt concentration, the electrostatic repulsion is practically neutralized by the salt ions, which, a bit like a sponge, absorb most of the energy by conveniently rearranging themselves in space. In this case, the protein solution is expected to behave very similarly to a solution of hard ellipsoids. Second, we illustrate the reason for the extreme similarity in our results with 5 mM NaCl and 1.67 mM $CaCl_2$, or 500 mM NaCl and 167 mM $CaCl_2$. In fact, we have

$$\begin{aligned} c_{\mathrm{NaCl}} &= 3\, c_{\mathrm{CaCl_2}} \\ &\Downarrow \\ \left[Z^2_{\mathrm{Na^+}}\, N_{\mathrm{Na^+}} + Z^2_{\mathrm{Cl^-}}\, N_{\mathrm{Cl^-}}\right]_{\mathrm{NaCl}} &= \left[Z^2_{\mathrm{Ca^{2+}}}\, N_{\mathrm{Ca^{2+}}} + Z^2_{\mathrm{Cl^-}}\, N_{\mathrm{Cl^-}}\right]_{\mathrm{CaCl_2}} \\ &\Downarrow \\ \kappa_{\mathrm{NaCl}} &= \kappa_{\mathrm{CaCl_2}} \ , \end{aligned}$$

where c_{NaCl} and $c_{\mathrm{CaCl_2}}$ denote the salt concentrations. Assuming that the counterion concentration is constant, the NaCl and $CaCl_2$ solutions have the same ionic strength and Debye screening parameter. Therefore, the interprotein interaction $u(r)$ is also the same, as all parameters based on it, such as the second virial coefficient.

We must emphasize one further point. Although the ionic strengths of the solutions with NaCl and $CaCl_2$ mentioned above are equal, we cannot calculate this property exactly. The reason is that, in addition to the free ions released by the dissociation of the added salt, the ionic strength includes a contribution that comes from the ions released by the protein. As we explain in Chapter 2, this contribution basically depends on the purification process. This remark does not invalidate the main conclusion of this section; as long as the ionic strength is kept constant, the static and dynamic properties of BSA solutions do not depend on the valence of the added salt.

Generic Short-Range Interactions

As mentioned previously, we cannot justify all our light scattering results with NaCl or $CaCl_2$ by the electrostatic screening alone. In order to inter-

pret the results from crowded solutions, we have introduce other generic, short-range interactions. The most compelling evidence for this need is the observation that χ_T, D, and – even more rapidly – s/s_0 seem to converge at high BSA concentration to the same value independent of salt species and concentration. We show this behaviour in Fig. 5.3, Fig. 5.4, and Fig. 5.6.

At the level of pure speculation, we see a number of candidate mechanisms that could play a role in this context. First of all, since BSA has a certain molecular volume and two proteins do not overlap, we cannot ignore the steric repulsion, which is very steep and very short-ranged. Second, we mention the van der Waals attraction, which is active between any kind of polarizable molecules, thus also between two proteins. Third, BSA has a static dipole moment that induces a further short-ranged attraction. Finally, we cannot exclude the eventuality of reversible self-association by BSA-BSA chemical bonding.

Actually, the influence of any of these effects on protein thermodynamics and diffusion is a topic of current research. These light scattering data are surely a piece needed for solving the puzzle, but cannot distinguish alone between one interaction and the other. However, standard colloidal models use the steric repulsion as a null model, on which other interactions can be built upon. Therefore, we try to check whether this null model is sufficient to interpret our light scattering results. For this purpose, we compare our SLS results on χ_T/χ_{T0} from the solutions at high salt concentration with the theoretical prediction for a suspension of hard spheres. For the latter, we use the analytical result in the Percus-Yevick closure [89, 114]. In the calculation of the BSA volume fraction, ϕ, we use the usual specific volume, given by $v_{\mathrm{phen}}/(\mathrm{M_{BSA}}/N_a) = 1.40$ ml/g. We plot this comparison in Fig. 5.7. Here, since the electrostatic repulsion is almost switched off the by salt ions, only the remaining short-range interactions should be important. The figure hints at a compatibility between experiment and theory. Of course, because of the low number of points, the predictive power of such a statement is not so high. Nevertheless, the agreement between the curves is far from obvious, and indicates that the steric repulsion, or excluded volume, is probably the most important short-range interaction indeed.

Predictive Power of the Interpretation

Our interpretation of the experimental results from both static and dynamic light scattering is still rather qualitative. However, we demonstrate that the basic physical mechanisms are quite universal, in any case not specific to Bovine Serum Albumin. This conclusion opens the fascinating perspective of building a full model on protein solutions based on the physics of colloidal

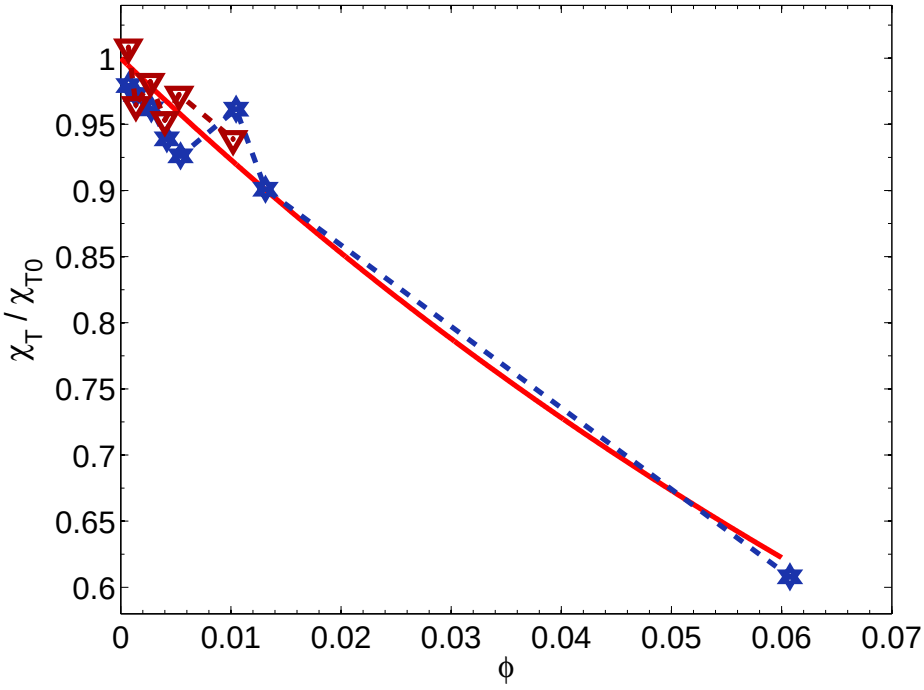

Fig. 5.7: Normalized osmotic compressibility as a function of volume fraction, for the dataset of BSA in high-c_s solutions, compared with the theoretical prediction for hard spheres in the Percus-Yevick closure. Legend: blue stars, 500 mM NaCl; dark red triangles, 167 mM $CaCl_2$; light red line, theory. The blue and red dashed lines are guides for the eye.

suspensions. As a matter of fact, this has been already tried in the literature, but the choice of the relevant interactions is always a delicate point. Through our data analysis we show that, for globular proteins in solutions of non-binding salts, the key forces are the electrostatic and the steric repulsions.

5.2.5 Comparison with Existing Literature

Our interpretation of the light scattering results from BSA solutions with NaCl or $CaCl_2$ is not particularly innovative. Several studies in the literature, both experimental and theoretical, deal with the thermodynamics and the diffusion of BSA in aqueous solutions. In this paragraph, we compare our findings with those previous investigations.

Tessier *et al.* measure the second virial coefficient of BSA in solutions with NaCl and neutral pH and $T = 25$ °C by SLS and self-interaction chromatography [115]. They have no data at very low ionic strength, but in high-c_s solutions, above ~ 100 mM, they measure $B_2 = 1 \times 10^{-4}$ mol ml/g^2. A similar value is found by Vilker *et al.* under the same conditions, $B_2 = 1.34 \times 10^{-4}$ mol ml/g^2, for BSA with 150 mM NaCl. Also the analysis of Minton, in Ref. [116], of the freezing-point depression osmometric data of Kanal, Ref. [117], performed in a supercooled solution at $T = -6.87$ °C, yield a value of $B_2 = 1.0 \times 10^{-4}$ mol ml/g^2. All these numbers are in quantitative agreement with our results from BSA solutions with 150 mM or 500 mM NaCl, or 167 mM $CaCl_2$.

Phillies *et al.* measure the collective diffusion constant of BSA in 150 mM and 22 °C [50]. Our results agree essentially with theirs. More recently, a similar study is carried out by Meechai *et al.* [61]. Their data on BSA solutions with 150 mM NaCl agree quantitatively with those presented here. Gaigalas *et al.* present experimental data on dilute salt-free solutions [60]. Their increase of D at low protein concentration is slightly less steep than ours, even when fitted by the cumulant expansion. However, those authors have very few experimental points at concentrations higher than 5 mg/ml. Moreover, those samples are probably not so pure and, since their cumulant routine computes an average diffusion constant, their D is probably affected by impurities. We point out, in any case, that this bias in D does not invalidate any of their qualitative statements, because all their samples are probably biased in a similar way. They also measure one single sample at $c_s = 25\ mM$, where D looks qualitatively comparable with our results, falling between our curves at $c_s = 0$ and $c_s = 150\ mM$. In Ref. [118], Bowen *et al.* study the diffusion of BSA in solutions with KCl. We expect these results to be similar to ours, because we show that the physical mechanisms are only dependent on the amount of dissociated electrolyte and not on its nature. These authors find, at all c_s, values of D higher than those of Gaigalas *et al.* but still slightly lower than ours. In any case, the hierarchy in D is found in their measurements as well. The theoretical paper of Yu *et al.* includes a figure (Fig. 4) on a model for the diffusion of BSA in 1 mM KCl [119]. That image refers to a lower temperature, but the dilute-limit D_0 is factored out, so it should be almost independent of T. Actually, we have no direct measurements at that ionic strength, but their solid curve is slightly below our results from salt-free samples, exactly as expected.

In conclusion, we see that our SLS and DLS experimental data on BSA solutions with NaCl and $CaCl_2$ agree with any published study that we have found, although not always perfectly.

5.3 BSA-YCl_3 Phase Diagram

The results of both SLS and DLS on solutions with yttrium chloride look qualitatively different from the previous ones. In fact, this trivalent salt has been shown experimentally to ignite a variety of latent physico-chemical phenomena, including induced protein isoelectricity, phase transitions, metastability, and re-entrant condensation [3]. The key concept for rationalizing all those effects is the BSA-YCl_3 phase diagram (PD), at fixed temperature and pressure. In Fig. 5.8, we present, together with the original phase diagram by Ianeselli [7], another version, in which we indicate the regions measured for this thesis by colored stripes. The colors (blue, red, green, and yellow) relate to qualitatively different categories of SLS/DLS results. They are completely phenomenological, and not related *a priori* with the original phase diagram. However, they overlap gracefully with it. This makes clear that our light scattering experiment are in full agreement with previous studies on re-entrant condensation, but also proceed beyond them. In particular, as we show by the color code, we need a four-phase diagram to explain our findings.

First of all, we call *blue* those solutions characterized by low osmotic compressibility and high diffusion constant, which are similar to salt-free solutions. The *red* samples possess low compressibility and low diffusion constant instead, which decrease with increasing BSA concentration. Crucially, the red samples are optically transparent, and we see no Q-dependence in Kc/R nor D. This implies that no large-scale protein aggregation takes place in the red samples. In the solutions that belong to the *green* region, we find a higher, Q-dependent scattering intensity. Some green samples are even too turbid to be measured. Therefore, the green solutions must possess some degree of macroscopic protein aggregation. Finally, the *yellow* region is characterized by low Kc/R and D, which increase with increasing concentration. This last feature excludes an explanation based on monomer-monomer attraction, suggesting a scenario of weakly repelling, stable oligomers instead.

We present the SLS results on these solutions in Fig. 5.9, the DLS results in Fig. 5.10. We also show the salt-free dataset for comparison purposes. The six curves correspond to the six colored stripes in the phase diagram, Fig. 5.8. Here, differently from Fig. 5.1, we plot the data in new color tones. The four regions of the phase diagram are not immediately distinguishable, thus we comment on them in the next paragraphs.

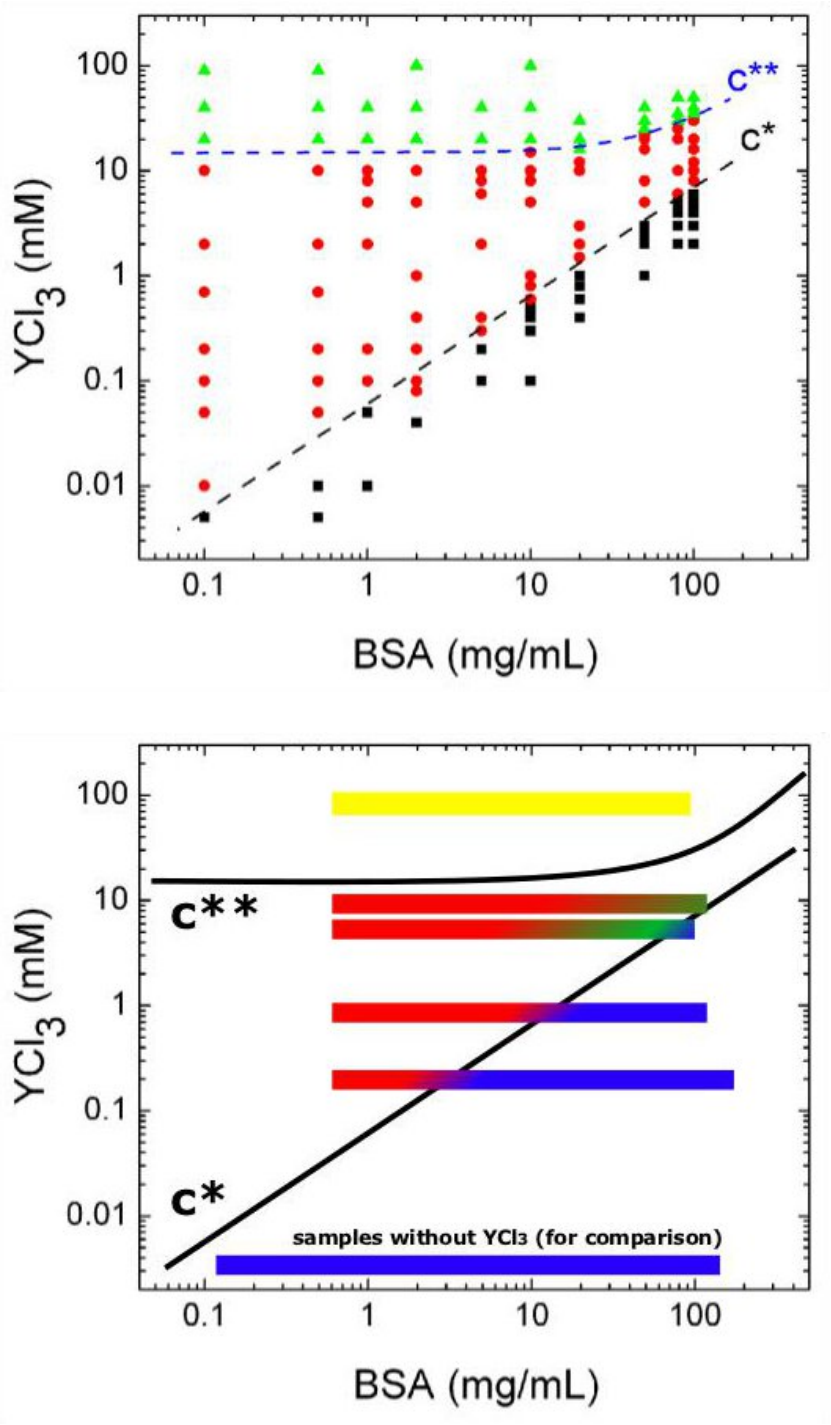

Fig. 5.8: Phase diagram of BSA-YCl_3. Original version from Ref. [7] (top) and modified version (bottom) emphasizing the regions studied in this thesis by light scattering (colored stripes). Different colors indicate a different light scattering behaviour of the samples, based on naked-eye observation, SLS, and DLS. Color legend: monomeric and strong repulsion (blue), monomeric and weak attraction (red), non-monomeric and turbid (green), partially monomeric and quasi-neutral (yellow). All samples were stable on the time scale of a few hours.

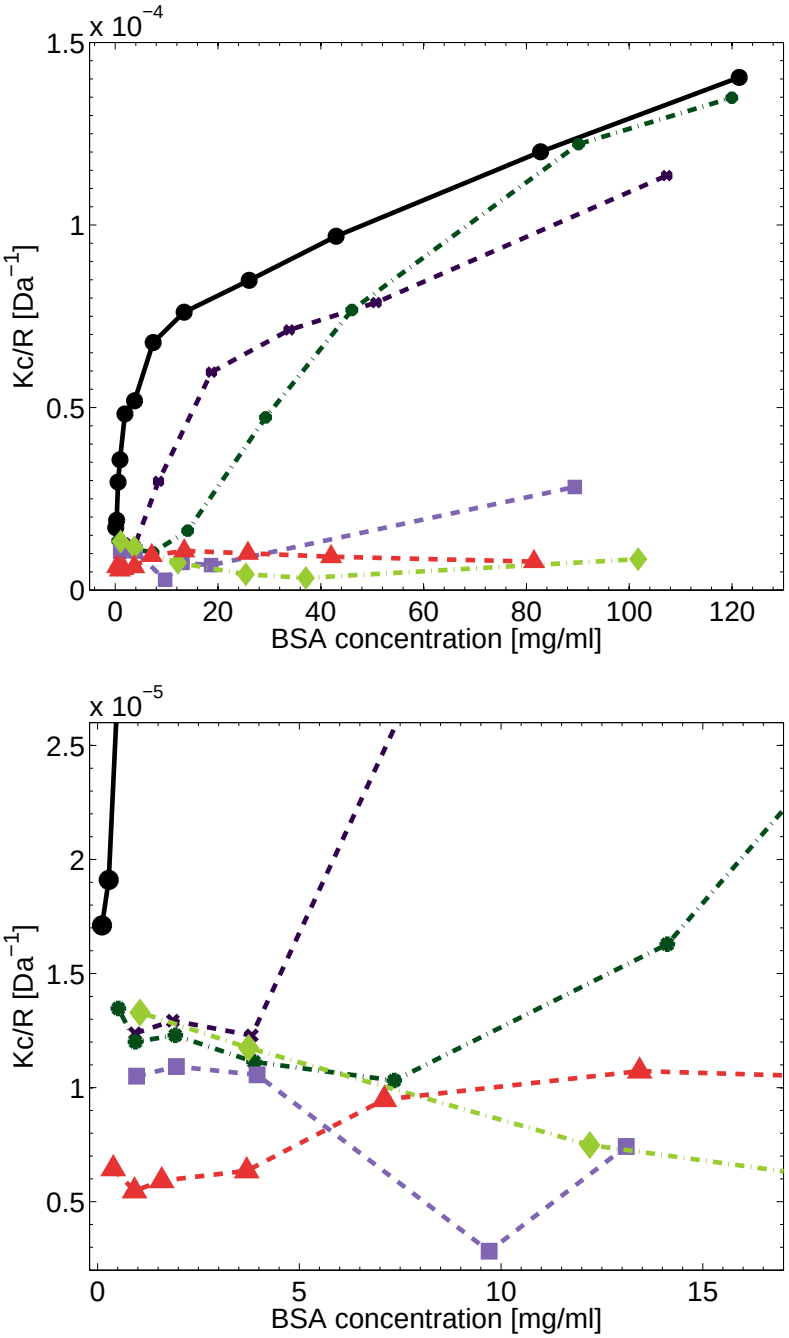

Fig. 5.9: SLS results for samples with YCl_3. The second image is a zoom of the first one. Legend: no salt (black circles), 0.20 mM YCl_3 (dark purple crosses), 0.83 mM YCl_3 (dark green asterisks), 5.0 mM YCl_3 (violet squares), 8.3 mM YCl_3 (light green diamonds), 83 mM YCl_3 (red triangles). The lines are guides to the eye.

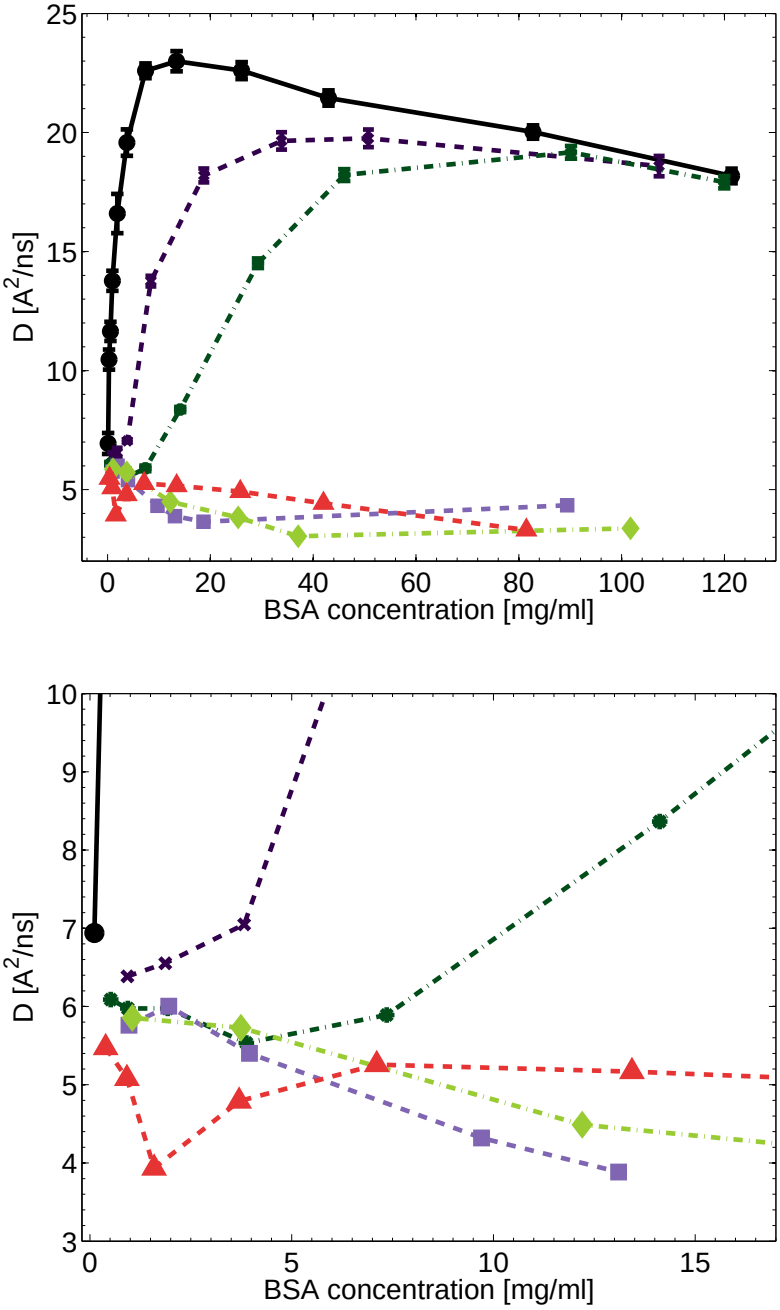

Fig. 5.10: DLS results for samples with YCl_3. The second image is a zoom of the first one. Legend: no salt (black circles), 0.20 mM YCl_3 (dark purple crosses), 0.83 mM YCl_3 (dark green asterisks), 5.0 mM YCl_3 (violet squares), 8.3 mM YCl_3 (light green diamonds), 83 mM YCl_3 (red triangles). The lines are guides to the eye.

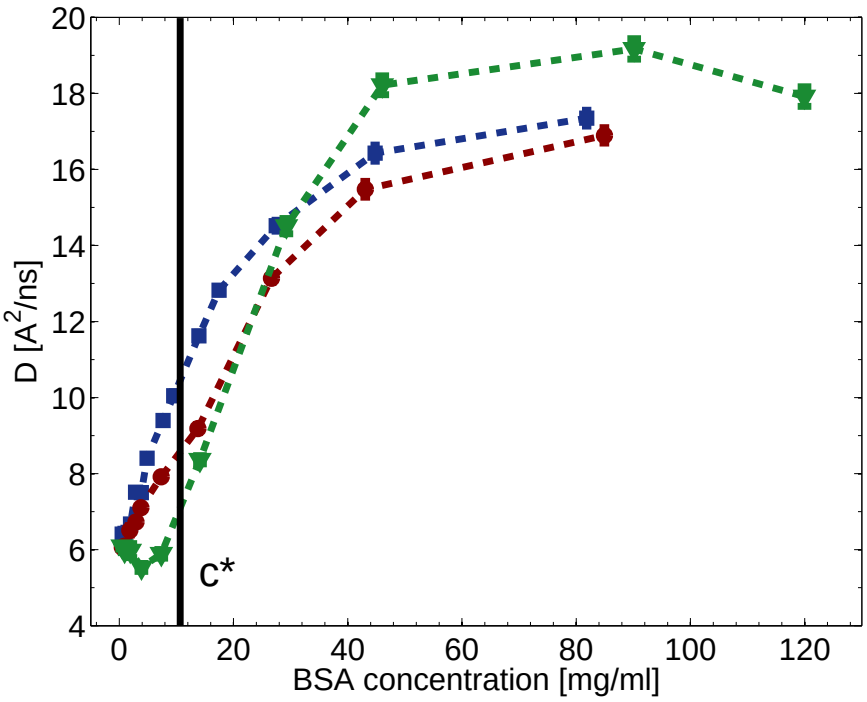

Fig. 5.11: Diffusion constant for samples at equal ionic strength and different salt species. Legend: 5 mM NaCl (blue squares), 1.67 mM $CaCl_2$ (red circles), and 0.83 mM YCl_3 (green triangles). The vertical line indicates the red-blue transition at c^*. Note the sudden change of slope of the green curve in correspondence with c^*. The dotted lines are guides to the eye.

5.3.1 Red-Blue Transition: Regime II → Regime I

The datasets at 0.2 mM and 0.8 mM YCl_3 start, at low c_p, with almost ideal χ_T and D, as typical of the red region. At moderate c_p, the SLS/DLS curves undergo a rather sharp transition towards low χ_T and high D. In other words, they turn blue. At even higher c_p, the curves are indistinguishable from the salt-free one. For both datasets at 0.2 mM and 0.8 mM YCl_3, the red-blue transition is located approximately on the c^* line of the phase diagram. This confirms quantitatively that the red-blue is nothing but the expected transition between Regime II and Regime I, observed also by SAXS and UV spectroscopy [3, 7].

The salt ionic strength of the dataset at 0.8 mM YCl_3 is the same of the

low-c_s datasets with NaCl and $CaCl_2$ presented in the previous sections. We compare the three curves for D in Fig. 5.11. We denote the YCl_3 samples by dark green, the NaCl ones by dark blue, and the $CaCl_2$ by dark red. Evidently, the diffusion curve for yttrium chloride is quite unique. At low protein concentrations, the samples with yttrium chloride show a much lower Kc/R and D than those with sodium and calcium chloride. Surprisingly, at higher c_p, after the red-blue transition, we observe the opposite phenomenon, the yttrium samples possessing higher D, almost like the salt-free ones. This exchange of roles confirms that the physical mechanism underlying the influence of YCl_3 on the interactions cannot be simple electrostatic screening, otherwise all three curves would look similar. As a matter of fact, at high ratios c_p/c_s, in the blue region, the screening efficiency of YCl_3 seems to be even lower than that of, say, NaCl.

5.3.2 Red-Green Transition: Regime II → LLPS

With the red-green transition, we want to describe a phenomenon taking place entirely in the so-called Regime II of the three-phase diagram, i.e. between c^* and c^{**}. We observe this transition particularly well in the dataset at 5 mM YCl_3. At low BSA concentration, the samples are optically transparent. Their SLS/DLS results, shown in Figs. 5.9 and 5.10 in light green, present slightly decreasing Kc/R and D. This is expected in solutions dominated by a net interprotein attraction. Moreover, both Kc/R and D are independent of Q, just like all normal samples with e.g. NaCl. These two features, the negative slope and the independence of Q, typify these red samples. This phenomenology changes at higher BSA concentrations. The samples start to become less transparent and, around 30-50 mg/ml, completely turbid. We do not plot these green solutions in Figs. 5.9 and 5.10, because they cannot be examined by the usual light scattering. Obviously, some aggregation process must take place. However, we see no precipitate on the bottom of the turbid solutions, thus the red-green transition cannot trigger a fluid-solid phase separation. Actually, recent experimental studies in solutions of BSA and similar proteins, such as HSA, seem to indicate that a liquid-liquid phase separation (LLPS) takes place in the green region [4]. Albeit a very interesting phenomenon, we cannot investigate the LLPS further in this thesis, mainly because of time constraints.

At intermediate BSA concentrations smaller than 30 mg/ml, we observe an interesting crossover, whose results we report in Fig. 5.12. The black line, shown for comparison, refers to the sample at 4 mg/ml BSA. At this concentration, the scattering is clearly independent of the scattering vector. The samples at 10 mg/ml, 13 mg/ml, and 19 mg/ml are still transparent, but

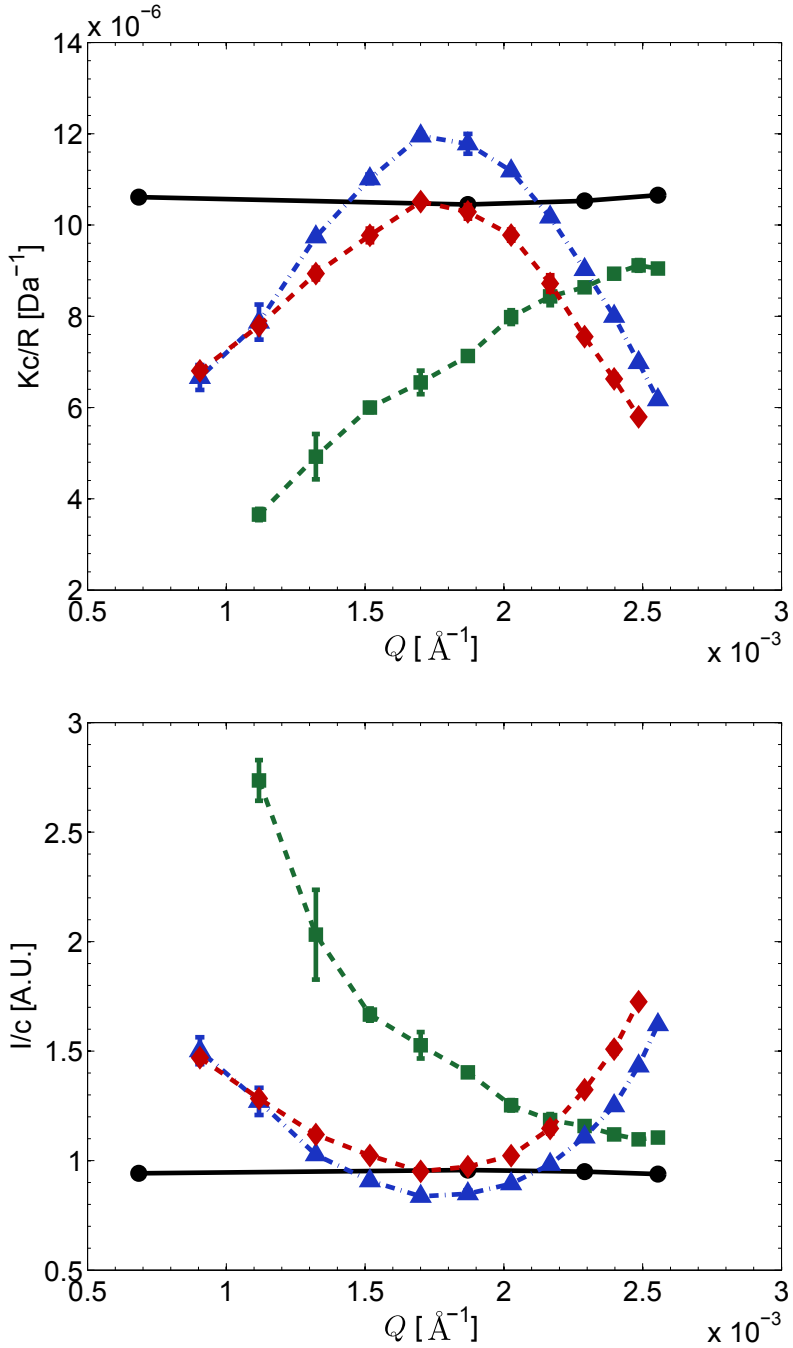

Fig. 5.12: Q-dependent SLS results for red-green samples, with 5 mM YCl_3. BSA concentrations are 10 mg/ml (green), 13 mg/ml (blue), and 19 mg/ml (red). The curve at $c_p \approx 4$ mg/ml (black), which is independent of Q, is shown for comparison. The ratio Kc/R (top) and its inverse, the scattering intensity over protein concentration, I/c (bottom), are shown. The latter quantity is shown in arbitrary units.

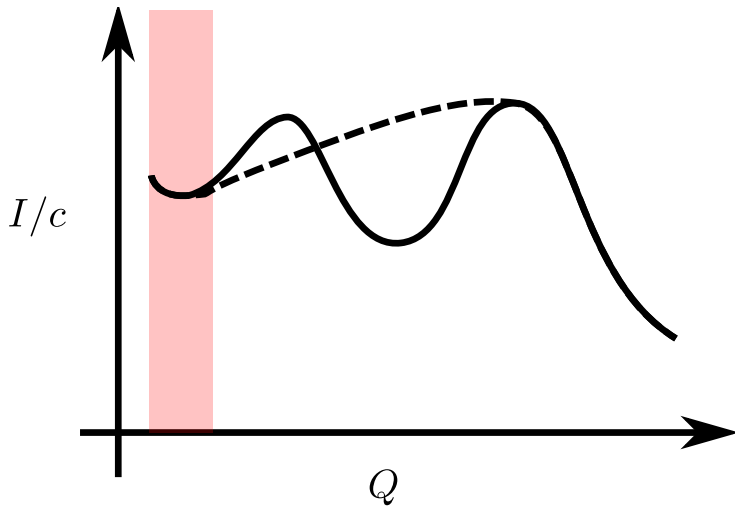

Fig. 5.13: Speculative scenarios for the increase in $I(Q)/c_p$ at higher Q values than the maximum of light scattering. The red region shows the Q window of light scattering. Legend: dashed line, slow increase prediction; continuous line, cluster peak prediction.

their scattering ratios are strongly Q-dependent. This phenomenon cannot be caused by BSA monomers, because they are far too small. Therefore, it must be due to oligomers. Since an SLS/DLS scan takes a couple of hours, the oligomers must be in a stable, or at least a long-lived metastable state. Moreover, since the measurement is repeated at least twice, we are sure that this effect is not due to kinetic phenomena. We can estimate roughly the size of the oligomers that seed the transition. We exploit, without proof, the following Zimm-Guinier-like formula:

$$\left[\frac{Kc_p}{R}\right]^{-1}(Q) \quad \underset{\sim}{\propto} \quad 1 - \frac{1}{3}\,R_g^2\,Q^2 \quad ,$$

where R_g is the raw estimate of the radius of gyration. We find $R_g \sim 100$ nm. This size scale means that the oligomers are formed, in average, by about 10^4 proteins. Moreover, because of this relatively large size, we intuitively guess that these oligomers effectively seed the liquid droplets, which develop fully only beyond the LLPS boundary.

The curves from the solutions at 13 mg/ml and 19 mg/ml, in the bottom plot of Fig. 5.12, posses a further interesting feature. Basically, we observe that their Q-dependence is not monotonic. Their scattering intensity, normalized by c_p, increases not only at vanishing Q values, which is a blueprint of aggregation in general, but also, in the other direction, at high Q. The

latter increase cannot be related to the main peak of the monomer-monomer structure factor, which is located at ten times a larger Q. So, it must be a typical feature of those oligomeric (meta)stable states. Moreover, we pose the natural question on how this increase evolves at even higher scattering vectors, outside the window of light scattering. We sketch the two alternative scenarios in Fig. 5.13. The first possibility is that the increase in $I(Q)/c_p$ saturates to a smaller slope, and finally merges with the main monomer-monomer peak (dashed line). Otherwise, that increase could be the first flank of an anomalous peak of the structure factor, which would describe the protein clusters (continuous line). In this discussion, the height of the peaks is unimportant. Obviously, we cannot take a decision merely on our experimental data. Nonetheless, we must mention that the second alternative picture would represent a definitive support for the long-debated results of Stradner on the so-called *cluster peak* [5, 111, 112, 120]. One direct way to study this phenomenon would be to complement light scattering, whose results we present here, with other scattering methods that probe the region of higher Q, such as small-angle X-ray or neutron scattering (SAXS/SANS). However, the typical experimental setup used in SAXS enables a smallest scattering vector of approximately 4×10^{-3} Å^{-1} [121], while the largest Q from SLS is 2.5×10^{-3} Å^{-1}, as explained in Chapter 4. Since the peak could be located in the region between those values, an ultrasmall-angle X-ray or neutron scattering (USAXS/USANS) measurement session would be even better.

Finally, we notice that, at even higher c_p, above 100 mg/ml, we can prepare a transparent sample again. This sample is the last light-green point in Figs. 5.9 and 5.10. It shows intermediate compressibility and diffusion constant. This agrees with the expectations based on the phase diagram, for this sample is already slightly beyond the c^* line, i.e. back in Regime I. For this reason, we color the edge of the stripe at 5 mM YCl_3 in Fig. 5.8 in blue.

5.3.3 Yellow Region: Regime III

The yellow region in Fig. 5.8 represents a dataset in the so-called re-entrant regime, or regime III, of the phase diagram. We show the SLS and DLS curves of this dataset as the red curves in Fig. 5.9 and Fig. 5.10. In the dilute limit, neither Kc/R nor D converge to the expected limits, M^{-1} and D_0, respectively. They are both significantly lower instead, suggesting, also in this case, the presence of BSA self-associates. These must be rather small, because we observe no Q-dependence in Kc/R nor in D. Judging from the usual sensitivity of light scattering, they should have an average radius smaller than about 20 nm. Most probably, it is a mixture dimers and trimers.

At higher protein concentrations, the trends of the Kc/R and D curves appear less clear. However, we can safely state that no large-scale aggregation takes place in the yellow region, for both SLS and DLS results remain of the same order of magnitude as the ideal, monomeric solution. At the highest c_p, the samples are closer to the c^{**} line, but probably not close enough to cause a strong change in their scattering behaviour. At the same time, our measurements rule out the simplistic picture that the re-entrant samples are just charge-inversed, salt-free-like samples. Were this idea right, we would observe a strong increase in both Kc/R and D at low c_p; this is evidently not the case.

There are two important differences in the protein state between the normal regime I and the re-entrant regime III. First, although the charge of BSA is inversed because the Y^{3+} cations bind it tightly, the solution is still full of monovalent chloride ions. These act as new counterions for the charge-reversed protein. Therefore, the solution is transparent and stable, but the electrostatics is strongly screened by the high Cl^- ion concentration [122]. The second difference relates to the stable form of BSA in the third regime. The light scattering results indicate that, in the reentrant regime, BSA forms small associates. The driving force for this association is not clear, but a role of Y^{3+} as a chemical bridge is possible. Because of its extremely high charge density, the cation could catalize the dimerization reaction by drilling a localized hole in the average repulsive armor generated by the electrostatic monopole moment of BSA.

We conclude this section with a last remark on these oligomers – dimers, trimers – in the re-entrant regime. If the monomer is really unstable or metastable compared to those complexes, the BSA second virial coefficient must become strongly negative. This is a trivial statement; if the BSA monomers attract each other to the point that they merge, the potential $u(r)$ used in the definition (3.10) must be, for the biggest part, smaller than zero. Of course, this argumentation is not in contrast with the positive slope of the red Kc/R versus c_p curve in Fig. 5.9, because that slope, if tentatively interpreted as a virial coefficient, refers nonetheless to BSA oligomers, and not to the BSA monomer itself. In other words, although two monomers attract one another strongly, the positive slope is an indication that two BSA oligomers interact with each other by a weak repulsion, and with a positive *oligomeric* second virial coefficient.

Chapter 6

Summary and Outlook

In this last chapter, we summarize the contents of the thesis and draw the conclusions. Moreover, we give an outlook on the perspectives for future research studies.

6.1 Summary

In this thesis, we present the results of static and dynamic light scattering experiments (SLS/DLS) on aqueous solutions of bovine serum albumin (BSA) and three salts, which are sodium, calcium, and yttrium chloride.

We first discuss and explain the results from solutions without added salt, with NaCl and $CaCl_2$. From SLS, we show that the osmotic compressibility is lower than in the ideal solution, i.e. $\chi_T < \chi_{T0}$. Moreover, the samples without added salt show a sharp decrease of χ_T/χ_{T0} with protein concentration c_p. The samples at high salt concentration show $\chi_T \sim \chi_{T0}$ instead (see Fig. 5.3). Analogously, the second virial coefficient is always positive but decreases – by almost two orders of magnitude – from samples without added salt to samples at high salt concentration (see Table 5.3). We present evidence in support of the hypothesis that a BSA solution at high salt concentration interacts approximately as a hard-sphere suspension (see Fig. 5.7). From DLS, we show that the collective diffusion constant of BSA, D, is influenced by the direct protein-protein interactions as well as by the hydrodynamic ones. The overall picture matches the results found by SLS. In fact, the solutions at high salt concentration show only a slight dependence of D on the protein concentration, as expected in a suspension of hard spheres. The solutions without added salt or at low salt concentration show, at high dilution, a rapid increase of D with increasing c_p instead. We quantify this behaviour by the second diffusion coefficient (see Table 5.4). This increase is followed by a

decrease or a saturation in more crowded solutions (see Figs. 5.4 and 5.5). The diffusion constant is never smaller than the dilute-limit value D_0, but we expect this relation to fail in very crowded solutions. In addition, the simultaneous SLS and DLS measurements enable us to study indirectly the sedimentation coefficient s of BSA solutions. We show that s decreases with increasing protein concentration, particularly sharply in solutions without added salt (see Table 5.5). On this point, we conclude that hydrodynamic interactions cannot be ignored in any physical description of protein motion and, consequently, in any biological study on protein function. Furthermore, we notice that the sedimentation coefficient becomes independent of the salt concentration in crowded solutions (see Fig. 5.6).

Thanks to those experimental results, which represent a precise and systematic set of data, we are able to give a unified and self-consistent theoretical interpretation, based on the concepts of screened electrostatic repulsion and generic short-range interactions, especially steric repulsion. Electrostatics is more relevant at low and intermediate protein concentrations, and the protein-protein repulsion is strongest in the solutions without added salt. In addition, we find that the statics and dynamics of BSA in a salt solution only depend on the ionic strength, and not on the salt species, in agreement with mean-field descriptions such as the Debye-Hückel theory [34]. In crowded solutions, we find that short-rage interactions play a major role. In our range of protein and salt concentrations, the steric repulsion seems to be the main character among all short-range forces. We stress that no protein-specific interaction is required for the explanation of our light scattering results. Although we do not realize a fully quantitative model of BSA solutions in this thesis, we believe that such an attempt is feasible. We describe our efforts in this directions in section 6.2.

We devote a special section to the results from the solutions with yttrium chloride. We show that the outcome of the light scattering experiments fit well into and add valuable information to the phase diagram of BSA-YCl_3 (see Fig. 5.8). We measure the transition between regime II and regime I, at c^*, exactly where it was expected to be (see Fig. 5.11). We observe that an additional phase, at intermediate salt and protein concentrations, is required. We estimate the transition towards this new phase, at $c_s \sim 5$ mM,in the range $c_p \sim$ 30-50 mg/ml. Based on recent investigations by other group members, we interpret this phase as a metastable liquid-liquid phase separation region (LLPS) [4]. Surprisingly, we find a very peculiar hybrid region, well outside the LLPS, in which the samples are transparent but the light scattering is strongly Q-dependent (see Fig. 5.12). We tentatively relate this phenomenon to the formation of (meta)stable self-associates of $\sim 10^4$ proteins. We guess that these associates are precursors or nuclei of the liquid droplets of the

LLPS. Moreover, we present evidence that those complexes could induce a cluster peak in the structure factor of the solutions, which would be located in a Q region slightly above 2.5×10^{-3} Å^{-1} (see Fig. 5.13). This region is not covered by SLS, and might be outside the standard window of SAXS/SANS as well, but should definitely be accessible by ultrasmall-angle scattering methods. We discuss this idea in section 6.2. Finally, we analyze the light scattering results from solutions in the re-entrant regime. We confirm that the solutions are transparent and stable. However, our measurements suggest that BSA is not monomeric, but forms small, weakly repelling oligomers (see the red triangles in Figs. 5.9 and 5.10). Were this idea right, the second virial coefficient of BSA monomers would not re-enter to positive values, but become strongly negative instead. Nonetheless, the second virial coefficient of those oligomers would be larger than zero.

6.2 Outlook

In this thesis, we try to exploit the state-of-the-art instrumentation, theory, and numerical data analysis of light scattering to investigate the static and dynamic properties of Bovine Serum Albumin in solution. We exploit the most recent SLS and DLS technology in order to gather valuable information on globular proteins. On the one hand, we observe that, in most cases, the narrow window of scattering vectors only represents the $Q \to 0$ limit, and that even the fastest autocorrelator is restricted to the measurement of long-time properties. On the other, the intrinsically asymmetrical polydispersity of BSA, due to the presence of heavier-than-monomer protein species in solution, makes the theoretical framework quite fragile. Despite the conceptual difficulties caused by these limitations, we also show that light scattering is a powerful tool in the hands of the careful experimentalist.

On the topic of protein structure and diffusion in solution with common salts, the results of our thesis show that the basic ingredients for modelling protein solutions are electrostatic screening, short-range interactions including steric repulsion, and hydrodynamic interactions. Obviously, similar light scattering studies should be performed on other globular proteins as well, starting from similar ones, such as HSA, ovalbumin, and beta-lactoglobulin. Several group members are already accomplishing this task; one of them, Elena Jordan, is even using the same instrument used for this thesis. Furthermore, since no protein-specific force seems to be necessary, we can be confident that a quantitative model of globular proteins is feasible. This model should exploit the standard machinery of theoretical and computational colloidal physics, including the Ornstein-Zernike hierarchy with appro-

priate closure relations, analytical low-concentration expansions, and various dynamic simulations. Some recent research articles in this direction seem promising [2, 119].

At the same time, considered the complexity of protein systems, we envision a future research based on many mutually complementary experimental techniques. The spectrum of possibilities includes large-scale facilities methods, such as small-angle neutron scattering (SANS), quasi-elastic neutron scattering (QENS), neutron spin-echo (NSE), normal and anomalous small-angle X-ray scattering (SAXS and ASAXS), and X-ray photon correlation spectroscopy (XPCS), but also laboratory-based ones, such cross-correlation light scattering (e.g. 3D-DLS), diffuse wave spectroscopy (DWS), self-interaction chromatography (SIC), and near-field microscopy. Those methods should be combined with well-established analysis procedures such as sedimentation velocity and equilibrium measurements in the ultracentrifuge, viscosity measurements, electrophoretic light scattering, infrared, visible, ultraviolet, and Raman spectroscopy, circular dichroism, differential scanning calorimetry, cloud-point measurements, osmometry, pycnometry, and optical microscopy. All those methods explore different aspects of the same system. Since the list is long, we cite a few concrete potentially interesting perspectives. Small-angle scattering is a direct Q-range extension of SLS, is able to supply the complete structure factor of protein solutions, which is needed both for direct comparison with static colloidal models, and for the development of dynamic models of protein diffusion. We are currently participating to a collaboration between the Universität Tübingen, the JCNS, the Slovak Academy of Sciences, and the ILL, on a project, in which we try to link together SAXS, SLS, DLS, and viscosity measurements, on the basis of analytical and computational models [8]; the results presented in this thesis used for that study as well. The dynamic counterparts of SAXS/SANS are NSE, QENS, and XPCS. QENS is being heavily exploited by our colleagues Dr. Tilo Seydel and Marcus Hennig, with excellent results [123]. NSE or XPCS could explore the so-called *cage diffusion* and become, in this picture, the missing link between DLS, which probes zero-Q collective diffusion, and QENS, which probes self-diffusion. We have already performed the first NSE experiment on this subject, and verified the feasibility of this idea.

On the other topic, the re-entrant condensation of proteins in solution with trivalent salts, the research is progressing at a fast pace on various fronts. Our results confirm that the second regime does not only consist of attracting monomers or large, precipitating aggregates. We directly observe a situation in which BSA probably forms mid-sized self-associates that are nonetheless stable – or metastable – in suspension. Other evidences are brought forward in support of a liquid-liquid phase separation, which can hardly be studied by

light scattering because of the high solution turbidity [4]. The very existence of metastable protein self-complexes is particularly interesting, because they act, in many cases, as seeds for the nucleation of protein crystals. In turn, the relevance of protein crystallization for biophysics and medicine can be hardly overestimated, just as its notorious difficulty of explanation. Once again, the first step beyond our findings should consist in a similar study with other globular proteins that undergo re-entrant condensation. A specific point of discussion, in this context, is the observation and interpretation of a small-Q cluster peak in the structure factor. Our data seem to support at least the existence of the peak, but are limited not only by too narrow a range of scattering vectors, but also by the issue of multiple scattering in turbid solutions. Actually, both these limitations could be overcome by means of ultrasmall-angle X-ray/neutron scattering (USAXS/USANS), on the one hand, and 3D-DLS and DWS, on the other. At present, only 3D-DLS is being tested by another group member, Elena Jordan.

The physical meaning of the re-entrant regime itself is becoming less obscure as research proceeds, but still needs to be explored thoroughly. The experimental evidence brought forward in this thesis indicate that the re-entrant regime is populated by small protein oligomers. However, it is difficult to predict whether this feature is distinctive of BSA or common to other globular proteins. Therefore, once more, we stress the importance of future studies of similar nature but focussed on different protein species. Moreover, because of the high similarity between BSA monomers and dimers under many physical aspects, we think that the re-entrant regime should be investigated by the most sensitive techniques, such as SIC.

Bibliography

[1] T. Jr. Peters. Serum Albumin. *Advances in Protein Chemistry*, 37, 1985.

[2] J. Gapinski, A. Wilk, A. Patkowski, W. Häussler, A. J. Banchio, R. Pecora, and G. Nägele. Diffusion and microstructural properties of solutions of charged nanosized proteins: experiment versus theory. *The Journal of Chemical Physics*, 123(5):054708, August 2005.

[3] F. Zhang, M. W. A. Skoda, R. M. J. Jacobs, S. Zorn, R. A. Martin, C. M. Martin, G. F. Clark, S. Weggler, A. Hildebrandt, O. Kohlbacher, and F. Schreiber. Reentrant condensation of proteins in solution induced by multivalent counterions. *Physical Review Letters*, 101(14):148101, Sep 2008.

[4] M. Wolf. Liquid-liquid phase separation of human serum albumin with yttrium chloride. Diploma thesis, Universität Tübingen, 2010.

[5] A. Stradner, H. Sedgwick, F. Cardinaux, W. C. K. Poon, S. U. Egelhaaf, and P. Schurtenberger. Equilibrium cluster formation in concentrated protein solutions and colloids. *Nature*, 432(7016):492–5, November 2004.

[6] Charles Darwin. *On The Origin of Species by Means of Natural Selection*. 6th. edition, 1872.

[7] L. Ianeselli. Reentrant Condensation of Proteins Induced by Multivalent Ions in Solution. Diploma thesis, Universität Tübingen, 2008.

[8] F. Zanini, F. Roosen-Runge, M. Heinen, D. Fedunová, F. Zhang, M. Hennig, T. Seydel, R. Schweins, M. Antalík, G. Nägele, and F. Schreiber. Viscosity and Diffusion: Crowding and Charge Effects in Protein Solutions. Unpublished, 2010.

[9] C. Ghélis and J. Yon. *Protein folding*. Academic Press, New York, 1982.

[10] Z. Bauman. *Vita liquida*. Laterza, Roma-Bari, 2006.

[11] Z. Bauman. *Liquid life*. Polity Press, Cambridge, UK, 2005.

[12] F. Hofmeister. Zur Lehre von der Wirkung der Salze. *Naunyn-Schmiedeberg's Archives of Pharmacology*, 25(1):1–30, 1888.

[13] M. v. Smoluchowski. Molekular-kinetische Theorie der Opaleszenz von Gasen im kritischen Zustande, sowie einiger verwandter Erscheinungen. *Annalen der Physik*, 330(2):205–226, 1908.

[14] A. Einstein. Theorie der Opaleszenz von homogenen Flüssigkeiten und Flüssigkeitsgemischen in der Nähe des kritischen Zustandes. *Annalen der Physik*, 338(16):1275–1298, 1910.

[15] B. H. Zimm. Molecular theory of the scattering of light in fluids. *The Journal of Chemical Physics*, 13(4):141–145, 1945.

[16] P. Debye. Molecular-weight determination by light scattering. *The Journal of Physical and Colloid Chemistry*, 51(1):18–32, 1947.

[17] W. H. Stockmayer. Light scattering in multi-component systems. *The Journal of Chemical Physics*, 18(1):58–61, 1950.

[18] J.G. Kirkwood and R.J. Goldberg. Light Scattering Arising from Composition Fluctuations in Multi-Component Systems. *The Journal of Chemical Physics*, 18:54, 1950.

[19] J. T. Edsall, H. Edelhoch, R. Lontie, and P. R. Morrison. Light scattering in solutions of serum albumin: effects of charge and ionic strength. *Journal of the American Chemical Society*, 72(1948):4641–4656, 1950.

[20] S.N. Timasheff, H.M. Dintzis, J.G. Kirkwood, and B.D. Coleman. Light Scattering Investigation of Charge Fluctuations in Isoionic Serum Albumin Solutions. *Journal of the American Chemical Society*, 79(4):782–791, 1957.

[21] G. Scatchard. Physical Chemistry of Protein Solutions. I. Derivation of the Equations for the Osmotic Pressure. *Journal of the American Chemical Society*, 68(11):2315–2319, 1946.

[22] G. Scatchard, I.H. Scheinberg, and S.H. Armstrong Jr. Physical Chemistry of Protein Solutions. IV. The Combination of Human Serum Albumin with Chloride Ion1. *Journal of the American Chemical Society*, 72(1):535–540, 1950.

[23] G. A. Ballou, P. D. Boyer, and J. M. Luck. The Electrophoretic Mobility Of Human Serum Albumin As Affected By Lower Fatty Acid Salts. *Journal of Biological Chemistry*, 159:111, 1945.

[24] I.M. Klotz and J.M. Urquhart. The binding of organic ions by proteins. Comparison of native and of modified proteins. *Journal of the American Chemical Society*, 71(5):1597–1603, 1949.

[25] S. G. DeWitt. The Interaction of Human Serum Albumin with Long-chain Fatty Acid Anions. *Journal of the American Chememical Society*, 80(3), 1953.

[26] J.T. Yang and J.F. Foster. Changes in the Intrinsic Viscosity and Optical Rotation of Bovine Plasma Albumin Associated with Acid Binding1. *Journal of the American Chemical Society*, 1588(1588):1588–1595, 1954.

[27] C. Tanford and S.A. Swanson. Hydrogen Ion Equilibria of Bovine Serum Albumin. *Journal of the American*, I(1955):6414–6421, 1955.

[28] C. Tanford. The Viscosity of Aqueous Solutions of Bovine Serum Albumin between pH 4.3 and 10.5. *The Journal of Physical Chemistry*, 1956(1956), 1956.

[29] Jr. W. G. McMillan and J. E. Mayer. The statistical thermodynamics of multicomponent systems. *The Journal of Chemical Physics*, 13(7):276–305, 1945.

[30] J.E. Mayer and M.G. Mayer. *Statistical Mechanics*. John Wiley & Sons Inc, New York, 1940.

[31] H.L. Frisch, J.L. Lebowitz, and S.A. Rice. The equilibrium theory of classical fluids. *Physics Today*, 18:78, 1965.

[32] C.A. Croxton. *Liquid State Physics - A Statistical Mechanical Introduction*. Cambridge University Press, London, 1974.

[33] J.P. Hansen and I.R. McDonald. *Theory of Simple Liquids*. Academic Press, London, 1976.

[34] P. Debye and E. Hückel. Zur Theorie der Elektrolyte. I. Gefrierpunktserniedrigung und verwandte Erscheinungen. *Physikalische Zeitschrift*, 24(9):185–206, 1923.

[35] B. Derjaguin and L. Landau. A theory of the stability of strongly charged lyophobic sols and the coalescence of strongly charged particles in electrolytic solution. *Acta Phys.-Chim. USSR*, 14:633–662, 1941.

[36] E.J.W. Verwey and J. Overbeek. Theory of the stability of lyophobic colloids. *Elsevier Publ. Co. Inc., London*, 8(8):9, 1948.

[37] D.P. Riley and G. Oster. Some theoretical and experimental studies of X-ray and light scattering by colloidal and macromolecular systems. *Discussions of the Faraday Society*, 11:107–116, 1951.

[38] A. Vrij. Light scattering of a concentrated multicomponent system of hard spheres in the Percus-Yevick approximation. *The Journal of Chemical Physics*, 93(1):51–59, 1978.

[39] D. Bendedouch and S.H. Chen. Structure and interparticle interactions of bovine serum albumin in solution studied by small-angle neutron scattering. *The Journal of Physical Chemistry*, 87(9):1473–1477, October 1983.

[40] W.R. Bowen and P.M. Williams. The Osmotic Pressure of Electrostatically Stabilized Colloidal Dispersions. *Journal of Colloid and Interface Science*, 184(1):241–50, December 1996.

[41] P. Prinsen and T. Odijk. Optimized Baxter model of protein solutions: electrostatics versus adhesion. *The Journal of Chemical Physics*, 121(13):6525–37, 2004.

[42] L. Jin, Y. X. Yu, and G. H. Gao. A molecular-thermodynamic model for the interactions between globular proteins in aqueous solutions: applications to bovine serum albumin (BSA), lysozyme, alpha-chymotrypsin, and immuno-gamma-globulins (IgG) solutions. *Journal of Colloid and Interface Science*, 304(1):77–83, December 2006.

[43] S.H. Chen, M. Broccio, Y. Liu, E. Fratini, and P. Baglioni. The two-Yukawa model and its applications: the cases of charged proteins and copolymer micellar solutions. *Journal of Applied Crystallography*, 40:s321–s326, 2007.

[44] F. Zhang, M. W. A. Skoda, R. M. J. Jacobs, R. A. Martin, C. M. Martin, and F. Schreiber. Protein Interactions Studied by SAXS: Effect of Ionic Strength and Protein Concentration for BSA in Aqueous Solutions. *The Journal of Physical Chemistry B*, 111(1):251–259, 2007. PMID: 17201449.

[45] Y.K. Park and G. Choi. Effects of pH, salt type, and ionic strength on the second virial coefficients of aqueous bovine serum albumin solutions. *Korean Journal of Chemical Engineering*, 26(1):193–198, 2009.

[46] J.H. Northrop and M.L. Anson. A method for the determination of diffusion constants and the calculation of the radius and weight of the hemoglobin molecule. *The Journal of General Physiology*, 12(4):543, 1929.

[47] O. Lamm and A. Polson. The determination of diffusion constants of proteins by a refractometric method. *Biochemical Journal*, 30(3):528, January 1936.

[48] K.H. Keller, E.R. Canales, and S.I. Yum. Tracer and mutual diffusion coefficients of proteins. *The Journal of physical chemistry*, 75(3):379–387, 1971.

[49] P. Doherty and G. B. Benedek. The effect of electric charge on the diffusion of macromolecules. *The Journal of Chemical Physics*, 61(12):5426–5434, 1974.

[50] G. D. J. Phillies, G. B. Benedek, and N. A. Mazer. Diffusion in protein solutions at high concentrations: A study by quasielastic light scattering spectroscopy. *The Journal of Chemical Physics*, 65(5):1883–1892, 1976.

[51] G.D.J. Phillies. Effects of intermacromolecular interactions on diffusion. I. Two-component solutions. *The Journal of Chemical Physics*, 60(03):976, 1974.

[52] G.K. Batchelor. Brownian diffusion of particles with hydrodynamic interaction. *Journal of Fluid Mechanics*, 74(01):1–29, 1976.

[53] B.U. Felderhof. Hydrodynamic interaction between two spheres. *Physica A: Statistical and Theoretical Physics*, 89(2):373–384, 1977.

[54] B.U. Felderhof. Diffusion of interacting Brownian particles. *Journal of Physics A: Mathematical and General*, M(11), 1978.

[55] B.J. Ackerson. Correlations for interacting Brownian particles. II. *The Journal of Chemical Physics*, 69(74074):684, 1978.

[56] W. B. Russel and A. B. Glendinning. The effective diffusion coefficient detected by dynamic light scattering. *The Journal of Chemical Physics*, 74(2):948–952, 1981.

[57] C.W.J. Beenakker and P. Mazur. Self-diffusion of spheres in a concentrated suspension. *Physica A: Statistical and Theoretical Physics*, 120(3):388–410, 1983.

[58] U. Genz and R. Klein. Collective diffusion of charged spheres in the presence of hydrodynamic interaction. *Physica A: Statistical and Theoretical Physics*, 171(1):26–42, 1991.

[59] R. B. Jones and P. N. Pusey. Dynamics of Suspended Colloidal Spheres. *Annual Review of Physical Chemistry*, 42(1):137–169, October 1991.

[60] A. K. Gaigalas, J. B. Hubbard, M. Mccurley, and S. Woo. Diffusion of bovine serum albumin in aqueous solutions. *The Journal of Physical Chemistry*, 96:2355–2359, 1992.

[61] N. Meechai, A. M. Jamieson, and J. Blackwell. Translational Diffusion Coefficients of Bovine Serum Albumin in Aqueous Solution at High Ionic Strength. *Journal of Colloid and Interface Science*, 175:167–175, 1999.

[62] G. Chirico and M. Gardella. Photon cross-correlation spectroscopy to 10-ns resolution. *Applied Optics*, 38(10):284–288, 1999.

[63] W. Häussler, A. Wilk, J. Gapinski, and A. Patkowski. Interparticle correlations due to electrostatic interactions: A small angle x-ray and dynamic light scattering study. I. Apoferritin. *The Journal of Chemical Physics*, 117(1):413, 2002.

[64] T. Narayanan and X. Liu. Protein Interactions in Undersaturated and Supersaturated Solutions: A Study Using Light and X-Ray Scattering. *Biophysical Journal*, 84(1):523–532, 2003.

[65] S. Longeville. Myoglobin in crowded solutions: structure and diffusion. *Chemical Physics*, 292(2-3):413–424, 2003.

[66] A. J. Banchio and G. Nägele. Short-time transport properties in dense suspensions: from neutral to charge-stabilized colloidal spheres. *The Journal of Chemical Physics*, 128(10):104903, March 2008.

[67] M. Heinen, P. Holmqvist, A. J. Banchio, and G. Nägele. Short-time diffusion of charge-stabilized colloidal particles: generic features. *Journal of Applied Crystallography*, 43(5):970–980, August 2010.

[68] M. Ziller. Simulation of protein charge distributions under the influence of salt. Diploma thesis, Universität Tübingen, 2010.

[69] A.Y. Grosberg and T. T. Nguyen. Colloquium: The physics of charge inversion in chemical and biological systems. *Reviews of Modern Physics*, 74(2):329–345, 2002.

[70] M. Liu. Wolinella succinogenes nitrite reductase: Purification and properties. *FEMS Microbiology Letters*, 19(2-3):201–206, August 1983.

[71] S. Sugio, A. Kashima, S. Mochizuki, M. Noda, and K. Kobayashi. Crystal structure of human serum albumin at 2.5 A resolution. *Protein Engineering*, 12(6):439–446, 1999.

[72] Jmol: an open-source java viewer for chemical structures in 3d. http://www.jmol.org/.

[73] C. Leggio, L. Galantini, and N. V. Pavel. About the albumin structure in solution: cigar Expanded form versus heart Normal shape. *Physical Chemistry Chemical Physics : PCCP*, 10(45):6741–50, 2008.

[74] S. E. Harding and H. Colfen. Inversion formulas for ellipsoid of revolution macromolecular shape functions. *Analytical Biochemistry*, 228(1):131 – 142, 1995.

[75] F. Perrin. Mouvement Brownien d'un ellipsoide (II). Rotation libre et dépolarisation des fluorescences. Translation et diffusion de molécules ellipsoidales. *Sciences-New York*, pages 1–11, 1936.

[76] R. Simha. The influence of brownian movement on the viscosity of solutions. *The Journal of Physical Chemistry*, 44(1):25–34, 1940.

[77] J.W. Mehl, J.L. Oncley, and R. Simha. Viscosity and the shape of protein molecules. *Science*, 92:132–133, 1940.

[78] N. F. Carnahan and K. E. Starling. Equation of state for nonattracting rigid spheres. *The Journal of Chemical Physics*, 51(2):635–636, 1969.

[79] J. A. Barker and D. Henderson. What is "liquid"? understanding the states of matter. *Review of Modern Physics*, 48(4):587–671, Oct 1976.

[80] G. S. Singh and B. Kumar. Geometry of hard ellipsoidal fluids and their virial coefficients. *The Journal of Chemical Physics*, 105(6):2429–2435, 1996.

[81] M. Yamasaki, H. Yano, and K. Aoki. Differential scanning calorimetric studies on bovine serum albumin: II. Effects of neutral salts and urea. *International Journal of Biological Macromolecules*, 13(6):322–328, 1991.

[82] M. Placidi. A dynamic light scattering study on mutual diffusion coefficient of BSA in concentrated aqueous solutions. *Europhysics Letters*, 43(August):476–481, 1998.

[83] M. Roche, P. Rondeau, N. R. Singh, E. Tarnus, and E. Bourdon. The antioxidant properties of serum albumin. *FEBS letters*, 582(13):1783–7, 2008.

[84] R. D. Shannon. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallographica Section A*, 32(5):751–767, Sep 1976.

[85] B.J. Berne and R. Pecora. *Dynamic light scattering: with applications to chemistry, biology, and physics*. Dover Publications, New York, 2000.

[86] C.S. Johnson and D.A. Gabriel. *Laser light scattering*. Dover Publications, New York, 1994.

[87] W. Brown. *Dynamic light scattering: the method and some applications*. Oxford University Press, USA, New York, 1993.

[88] W. Schärtl. *Light scattering from polymer solutions and nanoparticle dispersions*. Springer Verlag, Berlin Heidelberg, 2007.

[89] G. Nägele. *The Physics of Colloidal Soft Matter*. Institute of Fundamental Technological Research, Warszawa, Poland, 2004.

[90] S. Focardi, I. Massa, and A. Uguzzoni. *Fisica Generale, II (elettromagnetismo)*. Casa Editrice Ambrosiana, Milano, 2003.

[91] J. D. Jackson. *Classical Electrodynamics*. John Wiley & Sons, New York, 3rd edition, 1998.

[92] D. Stigter. Interactions in Aqueous Solutions. III. On Statistical Thermodynamics of Colloidal Electrolytes. *The Journal of Physical Chemistry*, 64(7):838–842, 1960.

[93] L. D. Landau and E. M. Lifschitz. *Course of Theoretical Physics, Volume 5: Statistical Physics Part 1*. Butterworth-Heinemann, Oxford, 3rd edition, 1980.

[94] B. H. Zimm. The Scattering of Light and the Radial Distribution Function of High Polymer Solutions. *The Journal of Chemical Physics*, 16(12):1093–1099, 1948.

[95] F. Bonneté and D. Vivarès. Interest of the normalized second virial coefficient and interaction potentials for crystallizing large macromolecules. *Acta Crystallographica Section D*, 58(10 Part 1):1571–1575, Oct 2002.

[96] K. Schätzel. Correlation techniques in dynamic light scattering. *Applied Physics B: Lasers and Optics*, 42(4):193–213, 1987.

[97] C.C. Gerry and P.L. Knight. *Introductory Quantum Optics.* Cambridge University Press, Cambridge, UK, 2004.

[98] A.J.F. Siegert. Report. Technical report, MIT Radiation Lab, 1943.

[99] S.W. Provencher. CONTIN: a general purpose constrained regularization program for inverting noisy linear algebraic and integral equations. *Computer Physics Communications*, 27(3):229–242, 1982.

[100] S.W. Provencher. A constrained regularization method for inverting data represented by linear algebraic or integral equations. *Computer Physics Communications*, 27(3):213–227, 1982.

[101] A. Einstein. Zur Theorie der Brownschen Bewegung. *Annalen der Physik*, 324(2):371–381, 1906.

[102] M. von Smoluchowski. Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen. *Annalen der Physik*, 326(14):756–780, 1906.

[103] A.L. Kholodenko and J.F. Douglas. Generalized Stokes-Einstein equation for spherical particle suspensions. *Physical Review E*, 1995.

[104] T. Svedberg and K.O. Pedersen. *Die Ultrazentrifuge: Theorie, Konstruktion und Ergebnisse.* Verlag von Theodor Steinkopff, 1940.

[105] G.J. Hooyman, H. Holtan Jr, P. Mazur, and S.R. de Groot. Thermodynamics of irreversible processes in rotating systems. *Physica*, 19(1-12):1095–1108, 1953.

[106] A. Vrij and H. N. W. Lekkerkerker. On the relation between diffusion, sedimentation, and friction. *The Journal of Chemical Physics*, 78(March):2760–2763, 1983.

[107] ALV Laser Vertriebsgesellschaft m.b.H. *ALV-5000 MULTIPLE TAU DIGITAL CORRELATOR: Reference Manual*, 1993.

[108] J.J. Hermans and S. Levinson. Some Geometrical Factors in Light-Scattering Apparatus. *Journal of the Optical Society of America*, 41(7):460–465, 1951.

[109] D.G. Neal, D. Purich, and D.S. Cannell. Osmotic susceptibility and diffusion coefficient of charged bovine serum albumin. *The Journal of Chemical Physics*, 80:3469, 1984.

[110] R. Wetzel, M. Becker, J. Behlke, H. Billwitz, S. Böhm, B. Ebert, H. Hamann, J. Krumbiegel, and G. Lassmann. Temperature behaviour of human serum albumin. *European Journal of Biochemistry*, 104(2):469–478, 1980.

[111] A. Shukla, E. Mylonas, E. Di Cola, S. Finet, P. Timmins, T. Narayanan, and D. I. Svergun. Absence of equilibrium cluster phase in concentrated lysozyme solutions. *Proceedings of the National Academy of Sciences of the United States of America*, 105(13):5075–80, April 2008.

[112] L. Porcar, P. Falus, W. R. Chen, A. Faraone, E. Fratini, K. Hong, P. Baglioni, and Y. Liu. Formation of the Dynamic Clusters in Concentrated Lysozyme Protein Solutions. *The Journal of Physical Chemistry Letters*, 1(1):126–129, 2010.

[113] ALV-GmbH R. Peters. *Technical Data, ALV/CGS3*, 2002.

[114] M. S. Wertheim. Exact solution of the Percus-Yevick integral equation for hard spheres. *Physical Review Letters*, 1963.

[115] P. M. Tessier, S. D. Vandrey, B. W. Berger, R. Pazhianur, S. I. Sandler, and A. M. Lenhoff. Self-interaction chromatography: a novel screening method for rational protein crystallization. *Acta Crystallographica Section D Biological Crystallography*, 58(10):1531–1535, 2002.

[116] A. P. Minton. A molecular model for the dependence of the osmotic pressure of bovine serum albumin upon concentration and ph. *Biophysical Chemistry*, 57(1):65 – 70, 1995. Macromolecular Crowding.

[117] K.M. Kanal, G.D. Fullerton, and I.L. Cameron. A study of the molecular sources of nonideal osmotic pressure of bovine serum albumin solutions as a function of ph. *Biophysical Journal*, 66(1):153 – 160, 1994.

[118] W.R. Bowen, Y. Liang, and P.M. Williams. Gradient diffusion coefficients–theory and experiment. *Chemical Engineering Science*, 55, 2000.

[119] Y.X. Yu, A.W. Tian, and G.H. Gao. Prediction of collective diffusion coefficient of bovine serum albumin in aqueous electrolyte solution with hard-core two-Yukawa potential. *Physical Chemistry Chemical Physics*, 7(12):2423–2428, 2005.

[120] J. Trewhella. The different views from small angles. *Proceedings of the National Academy of Sciences*, 105(13):4967–4968, 2008.

[121] F. Zhang. Private communication, 2010.

[122] F. Roosen-Runge. Private communication, 2010.

[123] F. Roosen-Runge, M. Hennig, T. Seydel, F. Zhang, M.W.A. Skoda, S. Zorn, R.M.J. Jacobs, M. Maccarini, P. Fouquet, and F. Schreiber. Protein diffusion in crowded electrolyte solutions. *Biochimica et Biophysica Acta (BBA)-Proteins & Proteomics*, 1804(1):68–75, 2010.