Frank Lüdel

Self aggregation of IgG fragments from different perspectives

PhD Thesis - Bielefeld University - 2016

Dissertation – submitted in partial fulfillment of the requirements for the degree of Dr. rer. nat. in the Faculty of Chemistry at Bielefeld University.

This dissertation was carried out at the Institute of Physical Chemistry I at the University of Bayreuth from February 2010 to September 2010 and continued at the Institute of Physical Chemistry III at the University of Bielefeld from September 2010 to October 2016, supervised by Prof. Dr. Thomas Hellweg.

I certify that this thesis does not incorporate, without acknowledgement, any material submitted for a degree or diploma at any university, and that, to the best of my knowledge, it does not contain any material previously published or written by another person except where due reference is made in the text. The research and writing embodied in this thesis are my own, except for the contributions made by others as described in the acknowledgements.

Submitted by: Frank Lüdel (*26.04.1982, Nürnberg)

Approved by: Prof. Dr. Thomas Hellweg (Advisor)
Prof. Dr. Hartmut Niemann

Copyright © 2016 Frank Lüdel

No part of this thesis may be reproduced or distributed without the written permission of the author.

First printing, November 2016

Contents

Contents 3

Summary 7

Acknowledgements 9

1 Introduction 11

2 Basics of the Immune System 15

2.1 Basic principles . . . 15

2.2 The immunoglobulins . . . 16

2.3 Applications of antibodies and resulting requirements . . . 17

3 The Aim of this Thesis 21

4 Numerical Methods and Software 23

4.1 CRYSOL . . . 23

4.2 SASREF . . . 25

4.3 HYDROPRO . . . 27

5 Material and Methods 29

5.1 Small angle x-ray scattering . . . 29

5.1.1 Basic scattering theory . . . 29

5.1.2 Laboratory SAXS beamline - XEUSS . . . 31

5.1.3 SAXS - Hands on . . . 32

5.1.4 Initial data treatment . . . 32

5.1.5 Further data treatment . . . 33

5.1.6 Data evaluation . . . 34

5.1.7 Molecular weight determination with SAXS . . . 35

5.2 Light scattering . . . 36

5.2.1 Photon correlation spectroscopy - Basic theory . . . 36

5.2.2 PCS - Hands on . . . 40

5.2.3 PCS - Initial data treatment . . . 40

5.2.4 Static light scattering - Basic theory . . . 40
5.2.5 Molecular weight determination with SLS - The Zimm-plot . . . 41
5.2.6 Static light scattering - Hands on . . . 42
5.3 Size exclusion chromatography . . . 43
5.4 Size exclusion chromatography - Hands on . . . 44
5.5 Gel electrophoresis - SDS-PAGE . . . 44
5.6 Gel electrophoresis - Hands on . . . 44
5.7 Further parameters . . . 45
5.8 Calculations . . . 45

6 Results and Discussion **47**
6.1 Initial cleaning and polishing of the 3H3 diabody . . . 47
6.2 Light scattering . . . 48
6.2.1 Photon correlation spectroscopy . . . 48
6.2.2 Static light scattering - Zimm-analysis . . . 52
6.3 Small angle x-ray scattering . . . 53
6.3.1 General data preparation and analysis . . . 53
6.3.2 Molecular weight determination . . . 57
6.4 Calculations based on homologous protein structures . . . 61
6.4.1 Calculation of SAXS data with CRYSOL . . . 61
6.4.2 Outlook: Possible structure optimisation using SASREF . . . 62
6.4.3 Preliminary data treatment steps . . . 65
6.5 Calculation of the hydrodynamic data using HYDROPRO . . . 66

7 Conclusion and Outlook **69**

8 List of Publications **73**

9 Bibliography **75**

A Appendix **91**
A.1 Used software . . . 92
A.2 Chemical crosslinking . . . 93
A.3 Sequence and generated sequence parts of 1LMK . . . 94
A.4 Calculation of the hydroynamic properties of the protein after structure optimisation . . . 95

Dedicated to my parents.

Summary

Within this thesis the experimental investigation of the solution structure of the anti-MET [1] antibody 3H3 using scattering techniques applied to solutions is presented.

Depending on the linker length between the V_H and the V_L domain, single-chain Fv (scFv) antibody fragments form monomers, dimers (diabodies) or higher oligomers. The research group of Prof. Niemann aimed at generating a diabody of the anti-MET antibody 3H3 in order to use it as a crystallisation chaperone to promote crystallisation of the MET ectodomain, via introducing a pre-formed twofold axis of symmetry. Size exclusion chromatography, however, suggested the protein to be monomeric.

Non-invasive light and x-ray scattering techniques were used to further investigate the oligomerisation state of 3H3 in solution. Via the combination of these experiments with model calculations a profound result was able to be obtained.

The small angle x-ray scattering (SAXS) curve measured for the protein nicely fits to the scattering curve calculated from the known crystal structure of a similar diabody, but does not correlate well to the calculated scattering curves of a monomeric scFv or a triabody. Concentration-dependent photon correlation spectroscopy (PCS) measurements revealed a hydrodynamic radius of 3.4 nm at infinite dilution and a negative interaction parameter k_D, indicating attractive interactions that are beneficial for crystallisation.

Both SAXS and light scattering (static and dynamic) measurements combined with model calculations, clearly show that the 3H3 antibody fragment in fact predominantly forms diabodies in solution.

Furthermore, based on the small angle x-ray scattering data, a structural model of 3H3 was developed via the structural optimisation of a homologous protein.

[1] The MET receptor is also known as hepatocyte growth factor receptor (HGFR).

Acknowledgements

I would like to convey my warm thanks to all the people who supported me with direct help, all kinds of assistance, fruitful discussions and suggestions, thus contributing to the successful development of this thesis.

First and foremost, I would like to express my deep gratitude to my advisor Prof. Dr. Thomas Hellweg, for his patient guidance and useful critiques as well as his great support during the time of my research for this thesis. Thomas not only always gave me the freedom to pursue the research areas in which I was interested, he also supported me by giving valuable insights into different fields of scientific research and management. Thank you Thomas for this opportunity!

My gratitude also goes toward both Dr. Sabine Rosenfeldt and Dr. Katja Henzler who introduced me to the "world of small angle scattering" (which can be very confusing at the beginning). Furthermore, I would like to acknowledge the open and honest collaboration within the "Patchy worm" project with Prof. Dr. Axel Müller, Dr. Holger Schmalz and Dr. Joachim Schmelz.

I would especially like to thank Dr. Sebastian Höhn for the great teamwork not only in the SAXS laboratory, as well as Franz Fischer and Karl-Heinz Lauterbach for the foresighted support concerning technical challenges particularly regarding the maintenance of the good old x-ray generators I was working with during my time in Bayreuth. In this context the team of the mechanical workshop was also always an absolutely reliable partner.

The financial support which was provided by the Deutsche Forschungsgemeinschaft (DFG) within the project SFB840 is gratefully acknowledged. The European Synchrotron (ESRF: Dr. Peter Boesecke) as well as the the Jülich Center for Neutron Science (JCNS: Dr. Aurel Radulescu, Dr. Vitaliy Pipich) and the Helmholtz-Zentrum Berlin (BER II: Dr. Miriam Siebenbürger, Dr. Karsten

Vogtt) gave me both invaluable access to outstanding x-ray and neutron scattering experiments as well as excellent support during and beyond the experiments given by the local contacts, and for this I am very grateful.

Thank you for the great support, assistance and the openness to discussions goes to the members of the collaborating research group of Prof. Dr. Hartmut Niemann, especially to Willem Bleymüller and Sandra Bufe. Sandra Bufe was also responsible for the purification of the protein samples used for the experiments presented in this thesis and for the chemical crosslinking experiments. Moreover, special thanks go to Sandra Bufe, Hauke Voß and Anna Barbara Dommes for near endless patience when explaining the world of "the green stuff" to me. I also greatly appreciate the time and effort given by Oliver Wrede, Sandra Bufe, Hauke Voß and Anna Barbara Dommes for the proofreading of the manuscript of this thesis, as well as the dedication shown by Daniela Kohl-Czertzick and Jenny Pilhofer for the linguistic proofreading.

I would like express my thanks to my colleagues of the Institutes of Physical Chemistry in Bayreuth and in Bielefeld for the great atmosphere and their support in every respect. In partiular, my grateful thanks are extended to Jana Dulle, Stefan Hauschild, Bastian Wedel, Yvonne Hertle, Oliver Wrede, Johannes Bookhold, Saskia Bannister, Lea Schröder, Jessica Klocke and Elena Herman.

Thank you Elisabeth Düngfelder and Sandra Gericke for the support regarding administrative questions.

Warm thanks go to thank Anna Barbara for her support and encouragement.

Last, but by no means least, I am deeply grateful to my parents for their everlasting patience and support as they accompany me on my academic adventure. Thank you!

1 Introduction

Protein crystallography represents a well established method for the structure determination of proteins in their crystalline state. To be able to study the protein structure in solution is an important option, particularly in the cases where no crystal structure is available or if it is suspected that the crystal structure might deviate from the native structure in solution.

Studying biopolymers in their native state offers the possibility of determining, for example, their molecular weight, their diffusion behaviour and their radius. This approach also allows the investigation of the reaction of these biopolymers under the influence of outer stimuli, such as changes in the pH-value or varying salt concentration, or the change of the oligomerisation state with time.

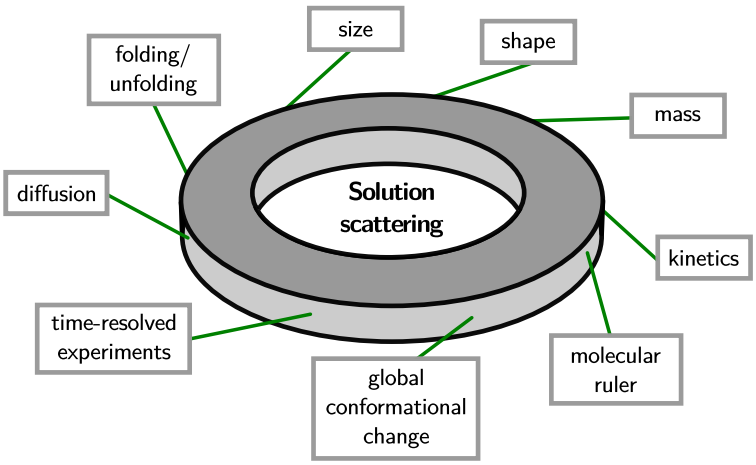

Figure 1: Possible applications of solution scattering techniques for the characterisation of biological samples ([1], modified).

Very powerful methods for the analysis of biological macromolecules in solution are small angle x-ray [2–16] [1] and neutron scattering [12, 17–22] (SAXS and SANS). During the last decades, these methods have evolved more and more into standard methods due to the developments in instrumentation and computational performances.

[1] The citations given in this chapter represent only a small selection of the literature available on this topic.

Small angle x-ray scattering and small angle neutron scattering offer a variety of applications for studying the structure of biological macromolecules in solution. These methods allow the investigation of the structural organisation of biomolecules regarding the structure of individual macromolecules [14, 23–25] and complexes [26], as well as the studying of the equilibrium composition of mixtures [27].

Furthermore, it is possible to perform time resolved experiments, using such methods as the stopped flow technique [28–30], which allow to study the kinetics of biological processes.

Besides focusing on single biological molecules, small angle scattering also allows the investigation of possible interactions between dissolved biomolecules [20, 31] which is of interest in biological, medical and biophysical research.

As well as the use of x-rays or neutrons as probe in solution scattering experiments, coherent laser light is also widely utilised. These static and dynamic light scattering techniques allow the determination of, for example, the molecular weight, the diffusion behaviour and the hydrodynamic radius of biomolecules in solution. Moreover, this kind of technique can easily be used in combination with purification methods such as size exclusion chromatography, where light scattering techniques are widely used for the analysis of the mass or the state of aggregation of the eluting species [32–34]. Furthermore, the high throughput analysis of reactions carried out in microplates [35] is possible using these techniques. Light scattering methods are highly sensitive to changes in size, thus they allow the examination of e.g. the aggregation [36] or dissociation [37] behaviour of biological molecules.

In addition to the described scattering techniques, which can only provide a low resolution structure compared to x-ray crystallography, model calculations can be done. They allow a refinement of the obtained structural models.

The present thesis combines SAXS and static as well as dynamic light scattering techniques for the first time to give a consistent description of the solution structure of an antibody fragment which is of great interest from medical point of view.

The present thesis combines SAXS and static as well as dynamic light scattering techniques for the first time. The combination of

these techniques has enabled the obtaining of a consistent description of the solution structure of an antibody fragment. This is of great interest from the medical point of view.

2 Basics of the Immune System

2.1 | Basic principles

The immune system protects the organism from intruding pathogens and from possible damage which they may provoke. To trigger an immune response, first an intruder has to be recognised by, for example, macrophages or dendritic cells. This leads to the activation of different defence mechanisms. Thereby, an important step is the presentation of antigens (distinct surface structures of the pathogens) to cells of the adaptive immune response, the T and B lymphocytes, resulting in a specific and long-lasting immune response. The lymphocytes fight the pathogens either by direct interaction with the infected host cells, as it is done by killer T cells (cellular immune response), or by producing antibodies - also called immunoglobulins, as it is done by the B cells (humoral immune response) [38, 39]. Antibodies bind the antigen at distinct chemical structures - the epitopes - on the surface of the antigen. Each B cell produces its unique antibodies with a particular binding specificity for one of these epitopes. Antibodies produced by identical immune cells are called monoclonal antibodies. It is possible that several different antibodies bind one antigen at different epitopes. These antibodies are produced by different immune cells and called polyclonal antibodies.

monoclonal antibody - produced by identical immune cells
polyclonal antibody - produced by several different immune cells

All processes involved in the immune response are highly regulated to protect host cells from immune response mediated damage. Failures of this regulation are responsible for allergy and for autoimmune diseases [38, 39].

2.2 | *The immunoglobulins*

In total five different classes of immunoglobulins - denoted as IgA, IgD, IgE, IgG and IgM - are known to be involved at different stages of the immune response.

In the course of an infection IgM, IgD and IgG are produced. After the activation of the B cells the levels of IgM and IgD rise. In a later stage of the infection IgG is produced. The two remaining types of immunoglobulins are IgE and IgA. IgE is mainly involved in the defence against parasites but is also responsible for allergic reactions by inducing the secretion of histamine, whereas IgA protects the mucose membranes [38, 39].

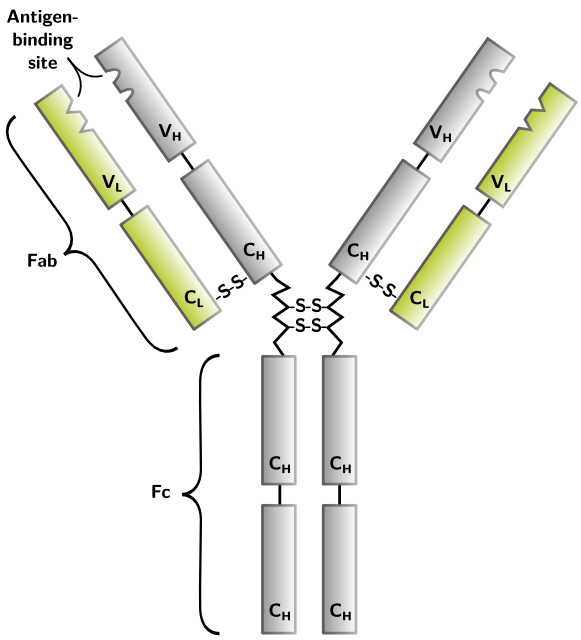

Figure 2: Schematic drawing of an immunoglobulin G, showing the constant (C) and variable (V) domains, the subdivision in Fab and Fc fragment which are connected by the flexible hinge region and the heavy (Index: H) and light (Index: L) chain. The antigen-binding sites are located at the end of the variable domain. The heavy and light chains are connected through disulphide bonds in the region of the Fab domain.

IgG is the predominant antibody class occuring in blood serum. Besides its role in the immune response it is also the mostly used antibody for therapeutical purposes. The overall molecular structure of this antibody class exhibits a Y shape (Fig. 2) with the two branches being formed by the antigen-binding fragments (Fab) and the base by the basal fragment or fragment crystallisable (Fc) [40–42]. The immunoglobulin consists of four polypeptide chains: two heavy chains (Index: H) and two light chains (Index: L). In the Fc part, the heavy chains interact with each other, whereas in the Fab part an interaction of the heavy and the light chains takes

place. The intersection between both parts is called the hinge region, which provides structural flexibility to the Fab fragments and simplifies the binding of the epitopes.

All chains forming the antibody exhibit constant (C) and variable domains (V). The antigen binding site is located at the end of each branch of the Fab part - the N-terminus - in the variable domain. The Fc part with the C terminus stabilises the antibody and binds to the Fc receptor during the phagocytosis in a later step of the immune response.

In 1972 the Nobel Prize in Physiology or Medicine was awarded to Gerald Edelman [43] and Rodney Porter [44] for their investigations of the chemical structure of antibodies.

2.3 | *Applications of antibodies and resulting requirements*

Antibodies feature highly specific binding to a wide range of antigens. For this reason antibodies are increasingly being used for therapeutical purposes such as passive immunisation [45, 46], immunomodulation (e.g. as therapy for multiple sclerosis) [47, 48], cancer therapy (e.g. Trastuzumab [49–51]) or for medical imaging (e.g. immunoscintigraphy [52]). When applying substances *in vivo* as pharmaceutical reagents, whether intended for therapeutic or diagnostic purposes, ideally they should act solely at their desired target and without adverse effects. Especially regarding cancer therapeutics, optimised pharmacokinetics and biodistribution are indispensable. On the one hand the tumour tissue has to be penetrated very fast combined with a rapid blood clearance, and on the other hand a rapid elimination, for example through the kidneys, without affecting the tumour, must be prevented [53, 54].

To meet or even to exceed these demands, the application of antibody fragments (Fig. 3), as opposed to the use of whole antibodies, came into the focus of research during recent years [53, 54]. Antibody fragments provide several further advantages and can be produced easily by expression in *E. coli* and *P. pastoris* with high yields [55, 56]. These fragments can be easily modified [57, 58] to control valency, affinity and avidity [59] and they can be used as building blocks in biochemical engineering.

Regarding the structure of an IgG various fragments can be derived thereof. These fragments are typically smaller than the

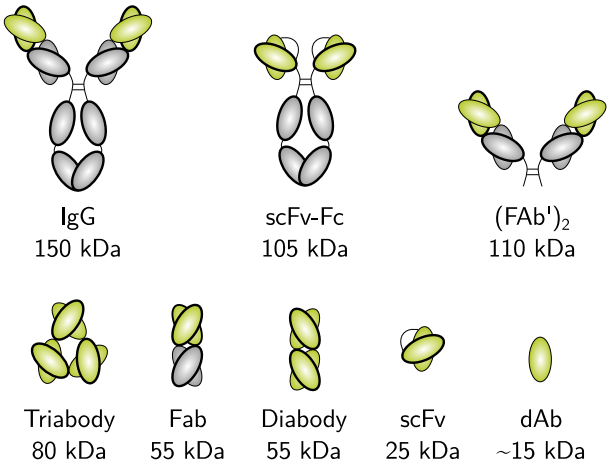

Figure 3: Schematic drawing of an IgG and different types of immunoglobulin fragments together with their estimated molecular weight (modified, [60, 61]). The constant domains are coloured in grey, the variable domains are represented in green. Light and bold outlines denote the light and heavy chains respectively.

corresponding antibodies and do not feature an Fc region. Several classes of antibody fragments are shown in figure 3, the smallest possible fragment represents the single-domain antibody fragment (dAb) with a molar mass of about 15 kDa. The smallest engineered antibody fragment containing the parental specificity is the variable fragment (Fv). This is composed of the variable region of the heavy (V_H) and the light chains (V_L) [62, 63]. Connecting both of these chains by a linker of a few amino acids results in a single-chain variable fragment (scFv). The biotherapeutic agent pexelizumab (Alexion) belongs to this class of antibody fragments [64]. The antigen-binding fragment (Fab) and the diabody (Db) are two classes of larger fragments with approximately the same molecular weight of about 55 kDa. The important difference between both of these classes is their number of binding sites. Fabs have only one binding site whereas Dbs are equipped with two binding sites, which can be engineered with equal or different binding specifities, as is already being done with antibodies for example in the therapeutic Removab (R) (catumaxomab, Fresenius/Trion) [64]. Other classes of antibody fragments are F(ab)2 fragments, consisting of two Fabs, and triabodies, consisting of three scFvs. All the fragments mentioned previously can be chemically modified for example by "PEGylation" to meet individual requirements.

To explain which structural changes lead to the formation of diabodies or higher oligomers of scFv, a closer look at the structure of the scFv is necessary. This fragment consists - as already mentioned above - of the variable region of the heavy (V_H) and the light (V_L) chains. These chains are connected by a linker of a few amino acids,

mostly Serin and Glycin, which provides flexibility and makes it resistant against proteolysis [65]. A variation in the number of amino acids in this linker leads to the formation of different oligomers instead of an scFv. ScFvs are typically formed at a linker length of more than 12 amino acids. Here it is possible that the variable regions of a chain can assemble in their natural conformation [66–68]. Reducing the amount of amino acids prohibits the binding between the V_H and V_L domains of one chain and thus forces the assembly of V_H and V_L from different chains, resulting in the formation of a dimer (diabody). A linker length of less then four amino acids finally leads to tri- and tetramers (tria- and tetrabodies) [69–71].

Usually there is a strict separation between the formation of different oligomers. Varying the linker length by only one amino acid may shift the equilibrium completely to another oligomeric state with the lowest complexity being most favourable [71]. As far as my last search [1] in the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) [72] showed, there are only six crystal structures of diabodies available so far [2] [66, 74–76]. These show that the two antigen binding sites are located at opposite sides of the protein. Due to this structural arrangement, diabodies have the ability to bind two antigens or even to bind two different epitopes of the same antigen. In addition, this means they are bivalent molecules featuring a higher avidity than monomeric antibody fragments.

[1] 07/2016

[2] A further structure is listed but with the status "unreleased" [73].

Furthermore, these two binding sites introduce a twofold symmetry to the molecule which can also be beneficial for crystallisation [77].

The diabody (3H3) described in this thesis could be used for this purpose. It binds to the MET receptor [3], a receptor tyrosine kinase often involved in the formation of tumours [78].

[3] The MET receptor is also known as hepatocyte growth factor receptor (HGFR).

3 *The Aim of this Thesis*

The antibody fragment 3H3 examined in this thesis is assumed to form a diabody. This diabody may be able to act as a crystallisation chaperone [79] for the extracellular domain of MET, which has not yet been crystallised in its entirety [80]. Size exclusion chromatography was used for the determination of the molecular weight of 3H3, which enabled the identification of the degree of oligomerisation. This method was also used for the purification of the protein solution necessary for crystallisation. These steps were performed by the group of Prof. Niemann and led to contradictory results about the size of 3H3.

3H3: The 3H3 fragment is genetically engineered based on an antibody found in sheeps and can be produced in *pichia pastoris* in the laboratory.

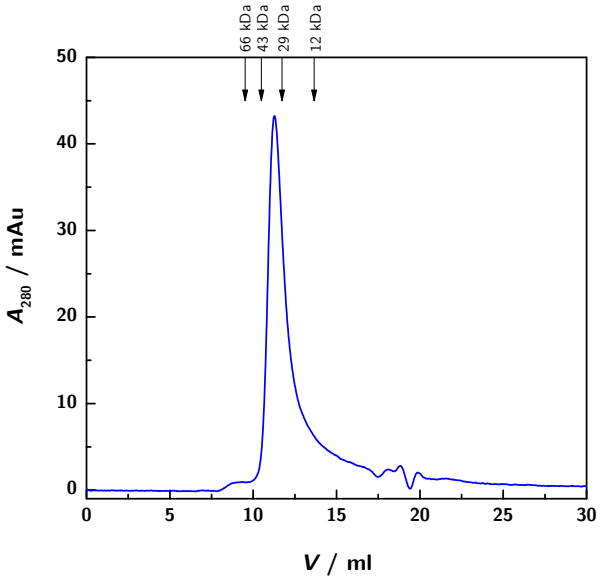

Figure 4: Analytical size exclusion chromatography of 3H3. 3H3 elutes from a Superdex 75 10/300 column with an elution volume of 11 ml to 11.3 ml corresponding to a molecular weight of approximately 36 kDa to 32 kDa. The elution volumes of the calibrants with the corresponding molecular weight are indicated by the arrows.

The elution profile of an analytical size exclusion chromatography run as presented in figure 4 shows a single peak with tailing towards higher elution volume, as well as a shoulder at lower elution volume. This shoulder is probably caused by aggregates. The position of the main peak at around 11.2 ml corresponds to a molecular weight of about 36 kDa to 32 kDa. A variation of the buffer

composition, the salt concentration as well as the pH value, which was also researched by the group of Prof. Niemann, had no influence on the peak position. The obtained molecular weight indicates the presence of a non-globular monomeric 3H3 scFv fragment with an estimated molecular weight of 27 kDa. Based on this SEC result one would conclude that there is no diabody formed but rather an scFv, not suitable for generating the desired twofold symmetry. However, crosslinking experiments indicate that the 3H3 is a dimeric protein (Fig. 5 and Appendix).

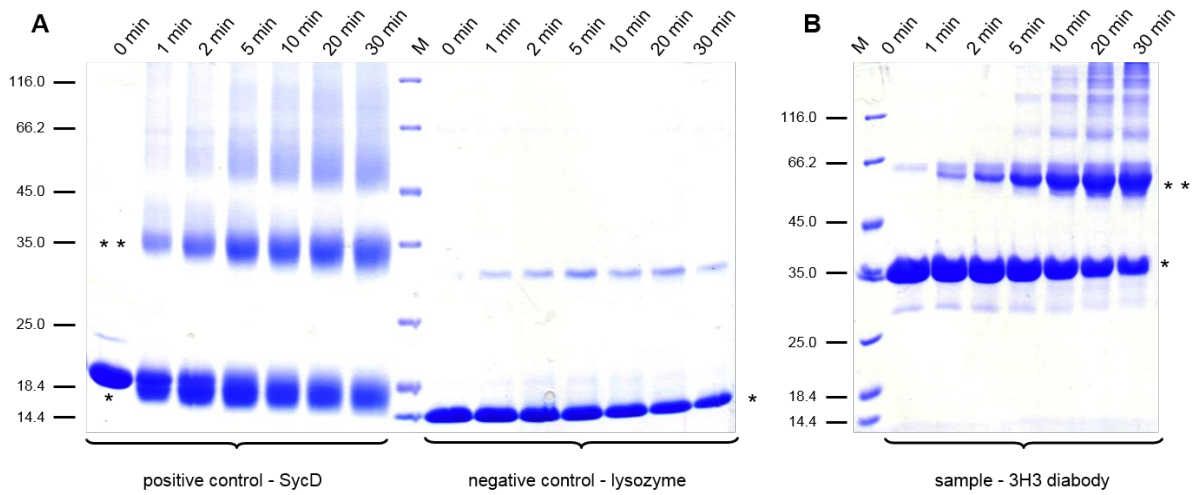

Figure 5: Analysis of chemical crosslinking with glutaraldehyde by SDS-PAGE [81]. (A) positive control SycD21-163 (previously shown to form dimers, [82]), negative control lysozyme. The positive control shows a band in the region of the dimer (**) increasing with time while that of the monomer decreases in intensity. The band in the region of the dimer appearing in the negative control is negligible. (B) 3H3 diabody purified by SEC. Due to the crosslinking a band at the size of the dimer can be seen comparable to the positive control.

Based on these results, the aim of this thesis is to investigate the 3H3 fragment with the help of scattering methods applied to solutions of this protein. This approach allows the investigation of this antibody fragment in its native state. The combination of photon correlation spectroscopy and small angle x-ray scattering provides complementary information on the overall size, molecular weight, the hydrodynamic properties as well as the low resolution structure of 3H3. The resulting overall picture emerging from the data related to all of these aspects may lead to being able to form a clear conclusion as to the existence of the 3H3 fragment as an scFv or as a diabody.

4 Numerical Methods and Software

4.1 | CRYSOL

The preparation of new or the modification of already known protein structures sooner or later requires a detailed structure determination. In the best case it is possible to obtain protein crystals which allow the determination of the exact atomic structure of the molecule. It is not always possible to obtain these crystals in a reasonable period of time or it might even be impossible to obtain usable crystals. In this case solution scattering methods, especially small angle scattering, offer the chance to get at least a low resolution model of the protein structure and some basic information such as the molecular weight, the radius of gyration or the overall shape. Using this method also requires modelling because all data obtained are in the reciprocal space.

To obtain an initial idea about the shape of a molecule, power laws are applicable. Another approach is the comparison of the experimental scattering data with computed scattering intensities based on structurally homologous atomic models. In the latter case the taking into account of the scattering intensities arising from the solvent (e.g. excluded volume, hydration shell) is important. The software used for these comparisons was CRYSOL [83], which allows the calculation of small angle scattering data based on structural data obtained from the PDB in consideration of the solvent properties. Furthermore a direct comparison of the calculated scattering data with experimental data is possible.

In general the scattering of a particle *in vacuo* can be calculated by [83]:

$$I(q) = \langle |A(q)|^2 \rangle_\Omega \tag{1}$$

$I(q)$	Scattering intensity
q	Scattering magnitude
$A(q)$	Atomic scattering
$\langle\ \rangle_\Omega$	average over all particle orientations
Ω	Solid angle in reciprocal space

When regarding a particle in solution this term has to be expanded by at least two terms. According to the schematic drawing of a protein in solution as shown in figure 6, the excluded volume ($\rho_0 A_c(q)$) and the hydration shell ($\delta\rho A_b(q)$, with $\delta\rho = \rho_b - \rho_0$) must be taken into account. This leads to the following formula [83]:

$$I(q) = \langle |A_a(q) - \rho_0 A_c(q) + \delta\rho A_b(q)|^2 \rangle_\Omega \tag{2}$$

Scattering intensities:

$A_a(q)$	of the particle *in vacuo*
$A_b(q)$	of the border layer
$A_c(q)$	of the excluded volume

Scattering densities:

ρ_0	of the solvent
ρ_a	of the particle
ρ_b	of the border layer
$\delta\rho$	$= \rho_b - \rho_0$

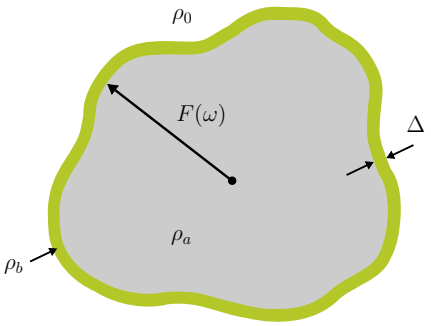

Figure 6: Schematic drawing of a protein in solution (after [83]), $F(\omega)$ is the envelope function, Δ is the effective thickness of the borderlayer, ρ stands for the scattering densities.

with

$$\rho(r) = \begin{cases} 1 & 0 \leq r \leq F(\omega) - \text{inside the protein} \\ 0 & r \geq F(\omega) - \text{outside the protein} \end{cases} \tag{3}$$

Equation 2 claiming the averaging over all particle orientations, $\langle\ \rangle_\Omega$, can be solved analytically by the multipole expansion approach according to Stuhrmann [84, 85] and Lattmann [86] using spherical harmonics. As a result eq. 2 can be writen as [83]:

$$I(q) = \sum_{l=0}^{L} \sum_{m=-l}^{l} |A_{lm}(q) - \rho_0 C_{lm}(q) + \delta\rho B_{lm}(q)|^2 \tag{4}$$

using the multipole expansion with atomic coordinates with the scattering amplitude of the particle with N atoms *in vacuo*:

$$A_{lm}(q) = 4\pi i^l \sum_{j=1}^{N} f_j(q)\ j_l(qr_j)\ Y^*_{lm}(\omega_j) \tag{5}$$

the scattering amplitude of the boarder layer:

$$B_{lm}(q) = i^l (2/\pi)^{1/2} \int_\omega Y^*_{lm}(\omega)\ \mathrm{d}\omega \int_{F(\omega)}^{F(\omega)+\Delta} j_l(qr) r^2\ \mathrm{d}r \tag{6}$$

N	Number of atoms
$\mathbf{r}$	$= (r, \omega) = (r, \theta, \varphi)$ Atomic coordinates
$f_j(q)$	Atomic form factor of the particle
$j_l(qr)$	Spherical bessel functions
$Y^*_{lm}(\omega_j)$	Spherical harmonics
$F(\omega)$	Envelope function describing the border layer
Δ	Effective thickness of the border layer
$g_j(q)$	Atomic form factor of the excluded volume
L	Resolution of the representation of the particle

and the scattering amplitudes of the excluded volume:

$$C_{lm}(q) = 4\pi i^l \sum_{j=1}^{N} g_j(q)\ j_l(qr_j)\ Y^*_{lm}(\omega_j) \tag{7}$$

During a CRYSOL run, several output files containing different sets of information are generated, the three most important files being the ALM, FIT and LOG files. The ALM file contains the scattering amplitudes as calculated according to equations 5 to 7. This file is required for a later rigid body modelling with SASREF. Another important file is the FIT file, here the experimental data and the result of the fitting is saved. An overview of the input parameters and the screen output is written in the LOG file. A description of the content of the FLM, INT and SAV files can be found in the CRYSOL manual available on the ATSAS homepage [1].

[1] https://www.embl-hamburg.de/biosaxs/software.html

4.2 | SASREF

The structure determination of (bio-)macromolecules remains a challenge even to the present day. Despite sophisticated techniques like x-ray crystallography, using high performance x-ray sources and detectors, and multidimensional nuclear magnetic resonance spectroscopy together with powerful computer systems, there remains still a huge effort of manpower, necessary to obtain a high resolution structure. In the case of x-ray crystallography, moreover, crystals of the molecules are needed and this crystalline state can differ from the appearance in solution of the sample [87].

Another possibility for obtaining structural information albeit not on the atomic level, is the combination of small angle scattering techniques with *ab inito* or rigid body modelling approaches. For the latter, homologuos protein structures are commonly used as a starting point. The software collection ATSAS developed by D. Svergun et al. [88] provides several modules for the different modelling approaches. In this thesis the program SASREF [89] was used. It uses rigid body modelling, based on the simulated annealing approach and provides the possibility to account for known intramolecular distances. This modelling is performed against experimental small angle x-ray scattering data.

The rigid body modelling is performed with a number of K subunits whose structures are known. The sum of these subunits represents the protein. For each subunit the scattering amplitude

$C^{(k)}(q)$ is determined, then the subunits are rearranged stepwise by shifting and rotating. The resulting scattering amplitude of the generated structure is calculated and compared with the experimental data [89]. The scattering intensity of the whole protein is given by [90]:

$$I(q) = \left\langle \left| \sum_{k=1}^{K} A^{(k)}(q) \right|^2 \right\rangle_{\Omega} \tag{8}$$

with

$$A^{(k)}(q) = \exp(iq\mathbf{r}_k)\Pi(\alpha_k\beta_k\gamma_k)\left[C^{(k)}(q)\right] \tag{9}$$

In equation 9 $\Pi(\alpha_k\beta_k\gamma_k)$ is the rotational operator with the Euler rotation angles $\alpha_k, \beta_k, \gamma_k$ and $\mathbf{r}$ is the vector of the shift.

Using a multipole expansion on spherical harmonics leads to the analytical term of the scattering intensity:

$$I(q) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-1}^{1} \left| \sum_{k=1}^{K} A_{lm}^{(k)}(q) \right|^2 \tag{10}$$

After each modelling step the goodness-of-fit of the actual arrangement of the subunits to the experimental data $I_{\exp}(q)$ is measured by [89]:

$$\chi^2 = \frac{1}{N-1} \sum_j \left[\frac{I_{\exp}(q_j) - kI(q_j)}{\sigma(q_j)} \right]^2 \tag{11}$$

N	Number of atoms
k	Scaling factor
$\sigma(q_j)$	Experimental error at q_j

In addition, it is necessary to account for a reasonable arrangement of the subunits, avoiding sterical clashes (cross) and loose contacts (cont) or incorporating information from other methods or known restraints (dis). This is realised by employing a target function E (eq. 12) using the penalty terms $\alpha_i P_i$.

During the fitting the target function E is minimised [89]:

$$E = \sum \left(\chi^2\right)_i + \alpha_{\text{dis}} P_{\text{dis}} + \alpha_{\text{cross}} P_{\text{cross}} + \alpha_{\text{cont}} P_{\text{cont}} \tag{12}$$

As an initial point, SASREF needs a file with the scattering amplitudes as provided by CRYSOL for each subunit (ALM-file), a file with the experimental scattering data and an optional file containing the contact conditions (e.g. restraints in distance between the subunits).

4.3 | *HYDROPRO*

It is possible to predict the hydrodynamic properties of rigid macromolecules as well as proteins on the basis of their crystal structure using for example a bead or a shell modelling approach. As a starting point the atomic-level structure (Fig 7) or the residues can be used. During modelling the structure is approximated by replacing either each non-hydrogen atom or each residue by identical beads with a defined radius (Fig 8, 9). For the determination of the radius of gyration this bead-model (also called primary model) is used.

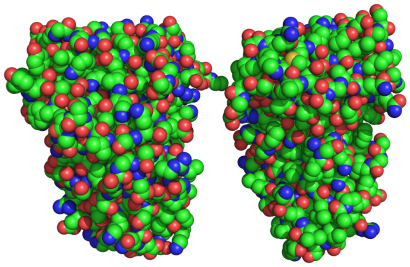

Figure 7: Atomic-level structure of 3H3 predicted from the structure of 1LMK and the sequence of 3H3 with the program 3D-JIGSAW [91, 92]. The figure was generated from the obtained atomic coordinates using PyMOL [93].

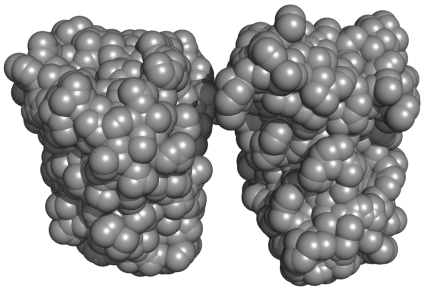

Figure 8: Bead-model of 3H3 composed of beads with a radius of 0.29 nm. The model is based on the atomic coordinates predicted by 3D-JIGSAW.

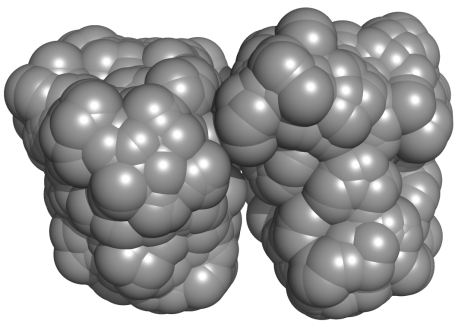

Figure 9: Bead-model of 3H3 composed of beads with the radius of 0.59 nm, taking a hydration layer into account. The model is based on the atomic coordinates predicted by 3D-JIGSAW.

For the determination of the translation diffusion coefficient only the beads in contact with the solvent are used to generate a shell-model consisting of a number N_s of small, nonoverlapping

minibeads with the radius σ. During the bead-model calculation the radius of the minibeads is decreased while the number of beads is increased in a given range [94]. At the end an extrapolation to $\sigma \to 0$ is done. This shell-model approach was first introduced by V. Bloomfield et. al. [95, 96].

The calculations described above can be done with the help of the program HYDROPRO, developed by J. Garcia de la Torre et. al [94, 97]. HYDROPRO uses a variable number of beads in the range of ∼200 to a (recommended) current maximum of ∼2000. The limitation of the maximum of beads is set by the fact that the computational costs are proportional to N_s^3. Despite this limit, fluid-dynamics calculations [98] and experiments with well-defined nanoparticles [99] showed sufficient compliance with the results obtained by HYDROPRO [94]. All calculations of the hydrodynamic properties of the 3H3 were performed with a bead radius of 0.59 nm for the primary model. This is an enlargement of 0.3 nm [2] regarding the recommended bead radius of 0.29 nm by J. Garcia de la Torre in order to take the hydration layer into account [100].

[2] 0.3 nm is a good approximation for the diameter of a water molecule.

5 Material and Methods

5.1 | Small angle x-ray scattering

5.1.1 Basic scattering theory

During a scattering experiment a collimated beam of radiation is incident on a sample. The main part of the radiation passes through the sample without any interaction (transmission), a very small amount is absorbed and a certain part is scattered elastically. This scattered part contains information about the sample structure. Hence, in a scattering experiment a comparison of the intensity of the incident beam I_0 with the scattered intensity I_{sc} at a certain angle 2θ at a distance L is made using a detector with $A_{det} = \Delta\Omega L^2$ [101].

I_0	Intensity of the primary beam
I_{sc}	Scattered intensity
2θ	Scattering angle
L	Sample to detector distance
A_{det}	Area of the detector
$\Delta\Omega$	Solid angle collected by each detector element
$d\sigma/d\Omega$	Differential scattering cross-section per unit solid angle

The ratio of the two intensities is defined as the differential scattering cross-section per unit solid angle $d\sigma/d\Omega$ and has a squared length as unit:

$$\frac{d\sigma}{d\Omega} = \frac{I_{sc}}{I_0} L^2 \tag{13}$$

The scattering cross section σ is defined as [102]:

$$\sigma = \frac{\text{scattered energy}}{\text{incident energy per unit area}} \tag{14}$$

$d\Sigma/d\Omega$	Differential scattering cross-section per unit sample volume

Regarding the sample volume V in a further step leads to the differential scattering cross-section per unit sample volume:

$$I = \frac{d\Sigma}{d\Omega} = \frac{1}{V}\frac{d\sigma}{d\Omega} \tag{15}$$

In general, for small angle scattering analysis the magnitude of the scattering vector q [1] is given by:

$$q = |\overrightarrow{q}| = \frac{4\pi n}{\lambda} \sin\left(\frac{2\theta}{2}\right) \tag{16}$$

[1] Caution! Instead of q sometimes h or s is used; s can also be $s = \frac{q}{2\pi}$

with the refractive index of the medium n [2], the wavelength of the incident beam λ and the scattering angle 2θ. As shown in the vector diagram (Fig. 10) the scattering vector $\vec{q}$ is the vector between the incident beam $\vec{k_i}$ and the scattered beam $\vec{k_s}$.

[2] In the case of x-ray and neutron scattering $n = 1$

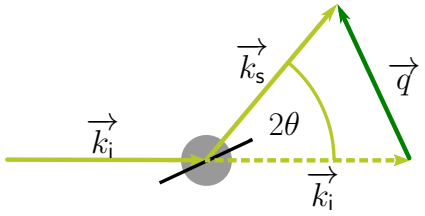

Figure 10: Vector diagram for elastic scattering. The scattering vector $\vec{q}$ is defined as the vector between the incident beam $\vec{k_i}$ and the scattered beam $\vec{k_s}$. 2θ is the scattering angle.

The definition of the scattering angle as 2θ results from the consideration made by Bragg as shown in figure 11. This also leads to the well-known Bragg equation (eq. 17) and to equation 18, correlating the scattering vector q with the intermolecular distance d.

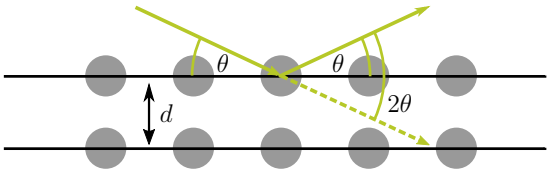

Figure 11: Bragg reflection from crystal planes with distance d.

$$n\lambda = 2d \sin\theta \tag{17}$$

$$d = \frac{2\pi}{q} \tag{18}$$

The scattering intensity of the sample $I(q)$ obtained in a scattering experiment is influenced by several factors (eq. 19), including different parameters about the sample as described below:

$$I(q) = \frac{N}{V}\,\Delta b^2\,\mathrm{P}(q)\,\mathrm{S}(q) = K\,c\,M\,\mathrm{P}(q)\,\mathrm{S}(q) \tag{19}$$

Here N/V is the number density of scatterers,

$$\Delta b = b_{\text{sample}} - b_{\text{buffer}} \tag{20}$$

the excess scattering length of the sample, $\mathrm{P}(q)$ the form factor representing intramolecular interactions and $\mathrm{S}(q)$ the structure factor representing intermolecular interactions. It is assumed that the

N/V	Number density of scatterers
b_{sample}	Scattering length of the sample
b_{buffer}	Scattering length of the buffer
Δb	Excess scattering length
$\mathrm{P}(q)$	Form factor
$\mathrm{S}(q)$	Structure factor
K	Optical constant
c	Concentration of the solute
M	Molecular weight of the solute

samples are diluted in such a way that no intermolecular interactions occur and S(q) = 1. The scattering intensity is furthermore dependent on the concentration c of the sample, the molecular weight M and on the sample specific optical constant K. In the case of small angle x-ray scattering (SAXS) K is defined as:

$$K_{\mathrm{SAXS}} = \frac{\mathrm{N_A}\Delta b^2}{M^2} = \frac{(\overline{\nu}\Delta\rho)^2}{\mathrm{N_A}} \tag{21}$$

$\mathrm{N_A}$	Avogadro number
$\overline{\nu}$	Specific volume
$\Delta\rho$	Excess scattering length density
V_{sample}	Volume of the sample

with

$$\overline{\nu} = \frac{\mathrm{N_A} V_{\mathrm{sample}}}{M} \tag{22}$$

Thus, the optical constant K_{SAXS} contains information about the specific volume $\overline{\nu}$ of the sample and the excess scattering length density:

$$\Delta\rho = \rho_{\mathrm{sample}} - \rho_{\mathrm{buffer}} \tag{23}$$

5.1.2 Laboratory SAXS beamline - XEUSS

SAXS measurements were performed using a XEUSS system (XENOCS, Sassenage, France). This provides a monochromatic Cu K_α [3] radiation and a variable sample-to-detector distance, covering a range between 50 cm and 277.5 cm. The setup consists of an evacuated collimation path with two variable scatterless slits, an evacuated X-ray scattering path (flight tube) and a hybrid-pixel area detector (Pilatus 300k, Dectris, Switzerland). The slits consist of four independently movable blades equipped with a single crystal on the edge pointing towards the beam, providing a beam of squared cross section [4].

[3] $\lambda = 0.154\,11$ nm

There are several sample environments available which are interchangeable. Flow-through capillaries made of quartz glass or Kapton with an outer diameter of 1 mm are used for the measurement of dissolved samples with poor contrast.

[4] Modes:
high resolution: 0.4 mm x 0.4 mm,
high flux: 0.8 mm x 0.8 mm

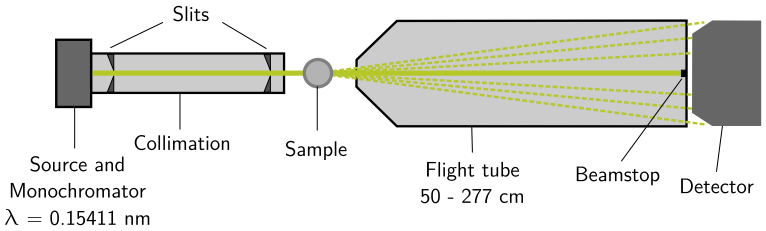

Figure 12: Schematic drawing of the XEUSS SAXS system.

5.1.3 SAXS - Hands on

The SAXS measurements were performed using the XEUSS-beamline operated in the high-resolution mode, equipped with the flow-through setup and a 1 mm quartz glass capillary. To cover the maximum q-range, measurements were made at three different sample-to-detector distances (50 cm, 133 cm and 278 cm). The samples were prepared directly before the measurements, filtered through a 0.2 µm syringe filter (Cellulose Acetate, VWR International GmbH, Darmstadt, Germany) and centrifuged for 15 minutes at 21,000 x g at 4 °C. All measurements were performed at room temperature.

5.1.4 Initial data treatment

To be able to perform a complete and satisfying data treatment, in each scattering experiment a complete set of data has to be collected.

This comprises knowing the intensity of the incoming beam I_0, the transmission of the sample T, the sample-to-detector distance, the dark count rate of all detectors I_d used during the measurement, and the scattering pattern of an absolute calibration standard, e. g. glassy carbon [103, 104]. It is also vital to have a sufficiently long period of time of measurement of the sample itself to guarantee good statistics.

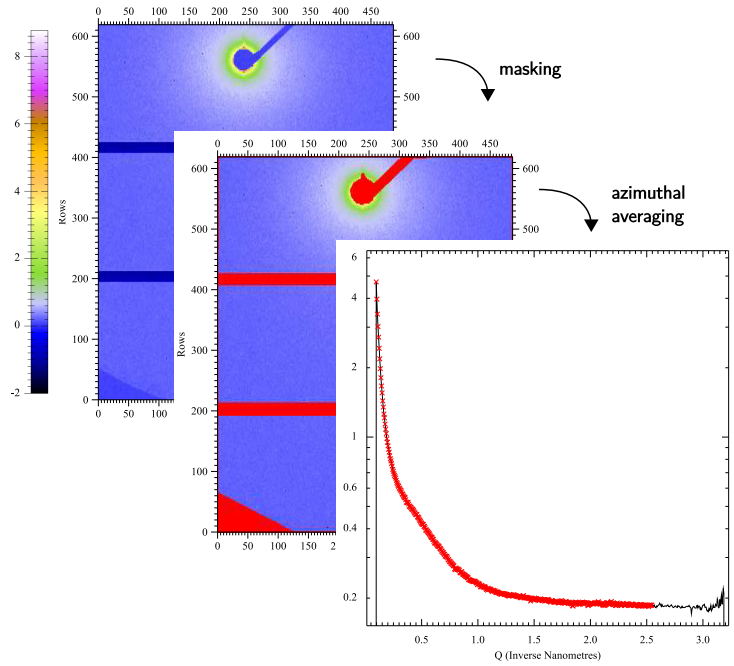

Figure 13: Schematic overview of initial SAXS data treatment steps from the 2D detector image to the 1D plot of the scattering intensities.

Masking

The first step to be taken for the data treatment after finishing the measurement is to mask all the regions on the 2D scattering pattern which do not contain any information about the sample (beam stop, dead pixels, dead areas of the detector).

Azimuthal averaging

Then azimuthal averaging is performed to obtain the 1D scattering data (figure 13), this procedure is only applicable in the case of isotropic scattering patterns. Anisotropic scattering patterns have to be analysed in 2D or particular areas have to be selected for partial azimuthal averaging.

Averaging of exposures

To enhance the signal to noise ratio, single short time measurements are averaged now. This can also be performed before the azimuthal averaging.

Normalisation

In the following normalisation step the scattering data has to be normalised with the transmission T and the acquisition time t according to the equations 24 and 25:

$$I_{\mathrm{norm}}(q) = \frac{I_{\mathrm{sc}}(q)}{T\,t} \tag{24}$$

with the transmission

$$T = \frac{I_{\mathrm{t}} - I_{\mathrm{d}}}{I_0 - I_{\mathrm{d}}} \tag{25}$$

$I_{\mathrm{norm}}(q)$	Scattered intensity, normalised
$I_{\mathrm{sc}}(q)$	Scattered intensity without corrections
T	Transmission
t	Acquisition time
I_{t}	Intensity of the transmitted beam
I_{d}	Darkcountrate of the detector
I_0	Intensity of the incoming beam

All the initial data treatment steps described until now have to be applied to the sample as well as to the background and to standard measurements.

5.1.5 *Further data treatment*

The use of an inhouse SAXS beamline with a limited flow compared to synchrotron beamlines necessitates longer acquisition times. For the purposes of this thesis each sample was measured for 24 h, representing a good compromise for obtaining acceptable statistics without having instabilities in the sample. Especially biological samples may degrade or form aggregates during the measurement time. To account for this, the measurements were subdivided in frames of 10 min. After the azimuthal averaging, the data of the first and last 10 frames were averaged and compared in order to exclude any changes of the sample.

Damage control

Background substraction

In a further data treatment step a careful background substraction is performed to eliminate contributions from the solvent (e.g. buffer) and from the setup. Especially when examining particles,

the scattering intensities of the buffer $I_{sol}(q)$ and the empty capillary $I_{cap}(q)$ have to be subtracted volume fraction ϕ weighted [5]. In the case of fuzzy systems such as proteins, the volume fraction is very small so it is common to subtract the buffer without any weighting [6]. After the buffer subtraction, the scattering intensities of the sample without any influences of the sample environment $I_{bs}(q)$ are obtained.

[5] $I_{bs}(q) = I_{sample}(q) - (1-\phi)I_{sol}(q) - \phi I_{cap}(q)$

[6] $I_{bs}(q) = I_{sample}(q) - I_{sol}(q)$

Absolute calibration

The scattering intensities obtained so far are on a relative scale. For some further evaluation steps, e.g. determining the molecular weight, it is necessary to bring the scattering intensities on an absolute scale. In order to do this, the measurement of a known calibration standard is used, and a shifting factor is determined and applied to the scattering intensities of the sample.

5.1.6 Data evaluation

The scattering intensities obtained during SAXS experiments enable various conclusions to be drawn about the sample. As shown in equation 19, the scattering intensities depend on the structure factor, the form factor, the molecular weight and the concentration of the solute. Another parameter, which can be extracted from the scattering intensities, is the radius of gyration R_g which can be obtained by a Guinier analysis using the following equation [105, 106]:

Guinier analysis

$$I(q) = I(0)\exp\left(-\frac{1}{3}R_g^2 q^2\right) \tag{26}$$

To perform the Guinier analysis the logarithmic intensity is plotted against the squared scattering vector and a linear fit in the lower q-region is performed. The valid interval for the fit is limited by $q_{max}R_g \leq 1.3$. From the slope of the fit the radius of gyration can then be calculated according to equation 26. The intercept of the y-axis gives $I(0)$ which can be used for molecular weight determination [7] as described in the next chapter.

[7] This requires the intensity to be on absolute scale and normalised with the concentration!

In contrast to other methods of analysis, the data obtained by scattering experiments are in the reciprocal space, this means a Fourier transformation is done in the experiment. With the growing potential of the computers available during the last few years, several attempts have been made to provide a user-friendly software for an inverse Fourier transformation, resulting in software packages such as PCG [107] by O. Glatter or the ATSAS-package [88] by

Inverse Fourier transformation

D. Svergun. As a result of the inverse Fourier transformation the pair distance distribution function $p(r)$ is obtained:

$$p(r) = \frac{r}{2\pi^2} \int_0^\infty qI(q) \sin qr \, \mathrm{d}q \tag{27}$$

In 1955 Guinier and Fournet [106] showed that half the normalised second moment of the $p(r)$ function is the squared radius of gyration:

$$R^2_{\mathrm{g,real}} = \frac{\int_0^\infty p(r) r^2 dr}{2 \int_0^\infty p(r) dr} \tag{28}$$

The ATSAS-package provides the program SHANUM [108]. This tool combines the inverse Fourier transformation with the Shannon sampling approach [109–111] and allows the determination of the useful data range of an experimental scattering data set.

5.1.7 Molecular weight determination with SAXS

Small angle x-ray scattering as a static scattering technique allows the determination of several parameters of substances diluted in a solvent. One of these parameters is the molecular weight M, which gives a clear indication as to whether the substance exists in a monomeric or oligomeric state. There are several different ways described in literature allowing the determination of molecular weight out of scattering data [112]. One possibility described by A. Guinier and O. Glatter emanates from the scattering data on absolute scale and uses the extrapolation to $I(0)$. The molecular weight can then be calculated using the following expression [113]:

$$M_{\mathrm{Prot}} = \frac{I(0)_{\mathrm{abs}} \mathrm{N_A}}{c_{\mathrm{Prot}} (\Delta\rho\overline{\nu})^2} \tag{29}$$

$I(0)_{\mathrm{abs}}$	Scattering intensity on absolute scale at $q = 0$
c_{Prot}	Sample concentration
$\Delta\rho$	Scattering contrast
$\overline{\nu}$	Partial specific volume

Additionally the scattering contrast and the partial specific volume of the sample have to be known.

An alternative way of determining the molecular weight is the use of a secondary scattering standard e.g. lysozyme [114] or BSA [115]. In this case the sample as well as a standard have to be measured under the same conditions and the scattering intensities have

to be normalised according to their concentrations. The molecular weight can then be determined by:

$$M_{\text{Prot}} = \frac{I(0)_{\text{Prot}}}{c_{\text{Prot}}} \cdot \frac{M_{\text{Std}}}{I(0)_{\text{Std}}/c_{\text{Std}}} \tag{30}$$

Fischer *et. al.* [116, 117] describe a completely different approach for the determination of the molecular weight from SAXS measurements. They use the Kratky function calculated from the scattering intensities on relative scale and perform an integration of the area under the curve to determine the volume of the sample (e.g. protein) and from this the molecular weight.

$$Q = \int_0^\infty I(q)q^2 \mathrm{d}q = 2\pi^2(\Delta\rho)^2 V^2 \tag{31}$$

$$Q' = \int_0^{q_{\max}} I(q)q^2 \mathrm{d}q \tag{32}$$

$$V' = 2\pi^2 \frac{I(q)}{Q'} \tag{33}$$

Due to the fact that the scattering intensities are only available in a limited q-range, this method yields an apparent volume which has to be corrected using a linear (A) and an angular coefficient (B):

$$V = A + BV' \tag{34}$$

From the corrected volume the molecular weight can be calculated by:

ρ_m Density of the molecule

$$M = V\rho_m \tag{35}$$

This evaluation method is implemented in the online tool SAXS MoW [8] [117]. The tool needs as input the pair distance distribution function calculated using the program GNOM [118] included in the ATSAS package [88].

[8] http://www.if.sc.usp.br/~saxs

5.2 | *Light scattering*

5.2.1 Photon correlation spectroscopy - Basic theory

In a classical photon correlation spectroscopy (PCS) experiment the intensity fluctuations of the scattered light (Fig. 15) (Tyndall effect)

are recorded by a photosensitive detector using the setup shown in figure 14 [119, 120].

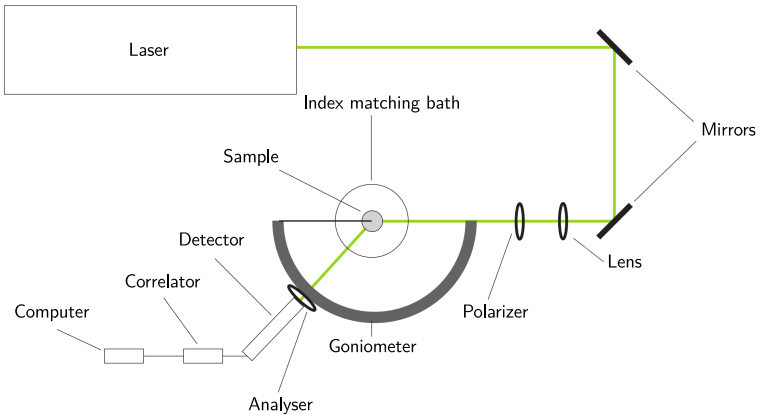

Figure 14: Schematic drawing of the light scattering setup for PCS and static measurements.

A PCS experiment recording directly the intensity fluctuations of the scattered light is called *homodyne* experiment. This homodyne experiment can be done at different scattering angles θ and hence, at different q-values using a goniometer setup.

These fluctuations are then analysed by a digital correlator which generates the intensity time auto correlation function $g^2(\tau)$ (Fig. 16) [119]:

$$g^2(\tau) = \frac{\langle E_s^*(t)E_s(t)E_s^*(t+\tau)E_s(t+\tau)\rangle}{\langle I\rangle^2} \tag{36}$$

$g^2(\tau)$ can be used to compute the normalised electrical field time correlation function:

$$g^1(\tau) = \frac{\langle E_s^*(t)E_s(t+\tau)\rangle}{\langle I\rangle} \tag{37}$$

by using the Siegert-relation:

$$g^2(\tau) = \left[1 + |g^1(\tau)|^2\right] B \tag{38}$$

Based on this information a conclusion about the translational motion of the scattering particles can be drawn. It is also possible to measure $g^1(\tau)$ directly in a so-called *heterodyne* experiment [120], but this is much more complicated [121].

In the case of monodisperse samples, $g^1(\tau)$ can be described by a single exponential function:

$$g^1(\tau) = \exp(-\Gamma\tau) \tag{39}$$

$g^2(\tau)$	Intensity time correlation function
τ	Delay or correlation time
$E_s(t)$	Scattered electric field
$E_s^*(t)$	Conjugate-complex of the scattered electric field
I	Mean scattered intensity

$g^1(\tau)$	Electrical field time correlation function
$E_s(t)$	Scattered electrical field
$E_s^*(t)$	Conjugate-complex of the scattered electrical field
I	Scattered intensity
B	Baseline parameter

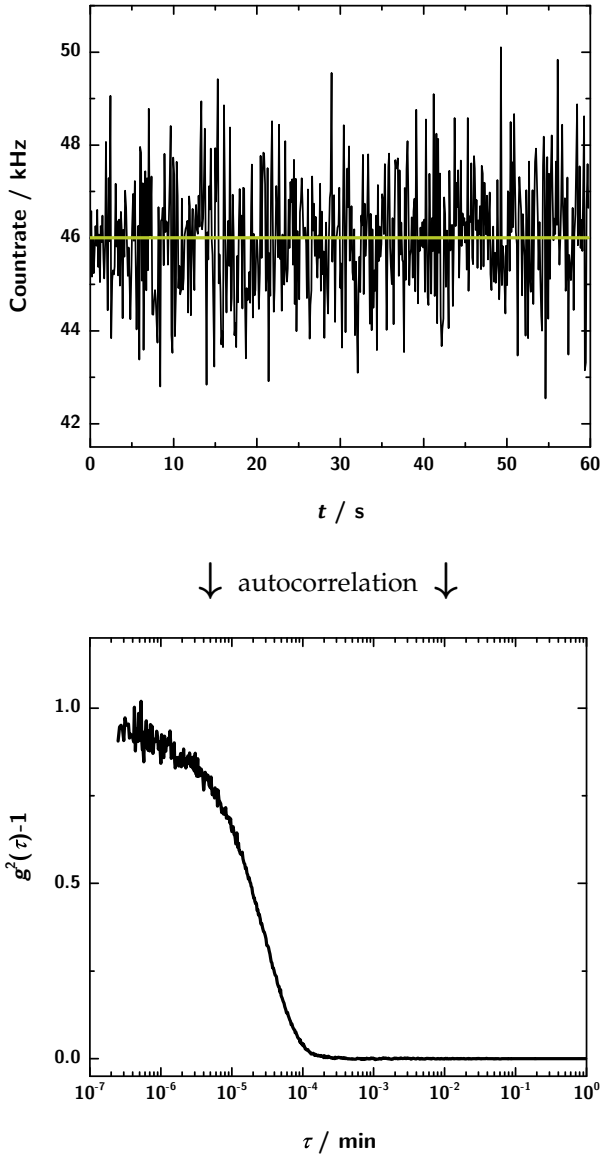

Figure 15: The intensity trace (countrate of the detector) versus time of a PCS measurement at a certain angle (black; in green: the mean value of the countrate).

Figure 16: The resulting intensity time autocorrelation function calculated from the measured intensity trace.

The relaxation rate Γ is given by:

$$\Gamma = D^{\text{exp}} q^2 \tag{40}$$

with the mutual translational diffusion coefficient D^{exp} [121].

However, real samples are polydisperse and might exhibit several relaxation modes stemming for example from aggregates of different sizes or from rotational and internal dynamics in the scattering biopolymer. In this case $g^1(\tau)$ is given by:

$$g^1(\tau) = \int_0^\infty G(\Gamma) \exp(-\Gamma\tau)\, \mathrm{d}\Gamma \tag{41}$$

Γ	Relaxation rate
D^{exp}	Mutual translational diffusion coefficient
$G(\Gamma)$	Relaxation rate distribution

with $G(\Gamma)$ being the relaxation rate distribution (Fig. 17). $G(\Gamma)$ can be computed by an inverse Laplace transformation [120].

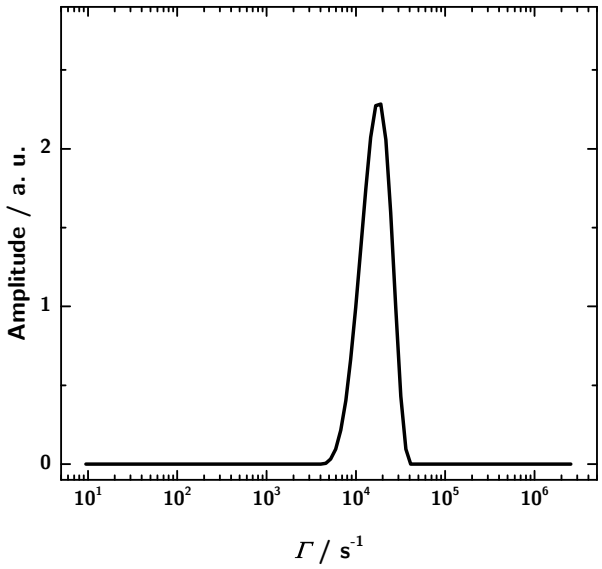

Figure 17: Relaxation rate distribution obtained from the intensity time autocorrelation function shown in Fig. 15.

For the determination of the z-averaged relaxation rate $\langle\Gamma\rangle$ the inverse Laplace transformation is used. This algorithm is implemented in the FORTRAN program CONTIN provided by S. Provencher [122]. If only a single relaxation is observed, $\langle\Gamma\rangle$ is related to the mutual diffusion coefficient $\langle D^{\exp}\rangle$ by:

$$\Gamma = \langle D^{\exp}\rangle q^2 \tag{42}$$

In this case $\langle D^{\exp}\rangle$ is related to the apparent hydrodynamic radius through the Stokes-Einstein equation [123]:

$$\langle D^{\exp}\rangle = \frac{k_B T}{6\pi\eta R_h} \tag{43}$$

with the temperature T, the Boltzmann constant k_B, the viscosity of the solvent η, and the radius of a hydrodynamically equivalent sphere R_h. In the case of an ideal diluted solution $\langle D^{\exp}\rangle$ corresponds to the self-diffusion coefficient D^0.

A further parameter which can be obtained from PCS experiments is the interaction parameter (dynamical virial coefficient) k_D:

$$\langle D^{\exp}\rangle = D^0\,(1 + k_D c) \tag{44}$$

k_D can easily be obtained from the slope of the averaged mutual translational diffusion coefficient $\langle D^{\exp}\rangle$ versus the concentration c.

k_B	Boltzmann constant
T	Temperature
η	Viscosity of the solvent
D^0	Viscosity of the solvent
k_D	Interaction parameter
c	Concentration of the sample

The self-diffusion coefficient D^0 is given by the y-intercept. A positive k_D points to repulsive intermolecular interactions, whereas a negative k_D indicates attractive interactions [124].

5.2.2 *PCS - Hands on*

Photon correlation spectroscopy measurements were carried out on a light scattering goniometer setup (LS instruments, Switzerland) equipped with a He-Ne Laser working at 632.8 nm (JDSU 1145P, USA). The samples were placed in a decaline filled vat, thermostated at 20 °C using a refrigerated heating circulator (Julabo F25ME, Julabo GmbH, Germany). The measurements were performed at two scattering angles, 60° and 135°, with nine different sample concentrations in the range of 2.5 mg ml^{-1} to 10 mg ml^{-1}. The samples with the highest and the lowest concentration were additionally measured angular dependent in steps of 10°. All samples were prepared directly before the measurements in a clean workbench, centrifuged for 15 min at 21,000 x g at 4 °C, filtered through a 0.2 µm syringe filter (Cellulose Acetate, VWR International GmbH, Darmstadt, Germany) directly into a cylindrical quartz cuvette (540.110-QS, Hellma GmbH & Co. KG, Germany) and finally centrifuged for 5 min at 2000 x g for the removal of dust and air bubbles. The cuvettes were previously rinsed with freshly distilled acetone.

5.2.3 *PCS - Initial data treatment*

The initial step of the data treatment is the evaluation of the autocorrelation function received from each measurement. This was performed using the bash program MSPLOT-C, developed in this workgroup [125]. The program performs a CONTIN-analysis [122] and allows a multiple data-treatment.

5.2.4 *Static light scattering - Basic theory*

The basic theory underlying static light scattering (SLS) experiments is the same as described for small angle x-ray scattering. Instead of x-ray irradiation, here the monochromatic, coherent light of a laser is used as probe. The scattered light from the sample is then detected angular dependent with the help of a similar setup as used for PCS (Fig. 14). This setup allows the measurement of

the scattered intensity in an angular range between 40° to 130°. In an SLS experiment the scattering contrast arises from the different refractive indices of the scattering objects and their environment. Thus it is correlated with the electron density of the atomic constitution of the individual material responsible for the scattering. To obtain the scattering information of the pure sample on an absolute scale and free of influences caused by the setup – called excess Rayleigh ratio R_{ex} – the raw intensity recorded by the detector I_{raw} has to be corrected angular dependent:

$$I_{scattered} = \left(\frac{I_{raw} \cdot \sin\theta}{I_{Laser}} \right) \tag{45}$$

and a normalisation with a standard, normally toluene, with a known Rayleigh ratio (scattering intensity on absolute scale) must be done as described in equation 46 [120, 126]:

$$R_{ex} = \left(\frac{I_s - I_b}{I_{ref}} \right) \left(\frac{n_b}{n_{ref}} \right)^2 R_{ref} \tag{46}$$

I_{raw}	Raw intensity, recorded by the detector
I_{Laser}	Intensity of the incoming beam
$I_{scattered}$	Corrected scattered intensity
I_s	Scattered intensity of sample
I_b	Scattered intensity of the buffer
I_{ref}	Scattered intensity of the reference
n_b	Refractive index of the solvent
n_{ref}	Refractive index of the reference
R_{ref}	Rayleigh ratio of the reference

5.2.5 *Molecular weight determination with SLS - The Zimm-plot*

The Zimm-equation 47 together with the resulting Zimm-plot (Fig. 18) provides the possibility of calculating several system parameters from the data obtained by a static light scattering experiment [127]. Zimm's equation relates the form factor $P(\theta)$ to the optical constant K, the mass concentration c, the molecular weight M and the second osmotic virial coefficient A_2.

$$\frac{Kc}{R_{ex}} = P(\theta)^{-1} \left(\frac{1}{M} + 2A_2 c \right) \tag{47}$$

When using the Guinier approximation, equation 47 can be written as:

$$\frac{Kc}{R_{ex}} = \left(\frac{1}{M} + 2A_2 c \right) \left(1 + \frac{q^2}{3} \left\langle R_g^2 \right\rangle \right) \tag{48}$$

or simplified

$$\frac{Kc}{R_{ex}} = \frac{1}{M} \left(1 + \frac{q^2}{3} \left\langle R_g^2 \right\rangle \right) + 2A_2 c \tag{49}$$

with

$$K = \frac{4\pi^2 \cdot n^2}{N_A \cdot \lambda_0^4} \cdot \left(\frac{dn}{dc} \right)^2 \tag{50}$$

K	Constant containing optical parameters, see eq. 50
c	Concentration of the sample
R_{ex}	Excess Rayleigh ratio
$P(\theta)$	Form factor
M	Molecular weight
A_2	Second virial coefficient
$\langle R_g \rangle$	Radius of gyration

N_A	Avogadro constant
λ_0	Wavelength of the light in vacuo
dn/dc	Refractive index increment

For the construction of the Zimm-diagram $\frac{Kc}{R_{ex}}$ is plotted against $q^2 + kc$ and an extrapolation to $q = 0$ and $c = 0$ is done. The scaling factor k only influences the appearance of the plot. Hence, the Zimm-equation and the plot consolidate all experimental and molecular parameters of the static light scattering experiment. The following special cases

- extrapolation to zero of the concentration and the scattering vector

$$\left(\frac{Kc}{R_{ex}}\right)_{c,q\to 0} = \frac{1}{M} \tag{51}$$

- extrapolation to zero of the scattering vector

$$\left(\frac{Kc}{R_{ex}}\right)_{q\to 0} = \frac{1}{M} + 2A_2c_2 \tag{52}$$

- extrapolation to zero concentration

$$\left(\frac{Kc}{R_{ex}}\right)_{c\to 0} = \frac{1}{M}\left(1 + \frac{q^2}{3}\left\langle R_g^2\right\rangle\right) \tag{53}$$

yield the molecular weight, the second virial coefficient, the radius of gyration and allow a cross-check of the experimental results.

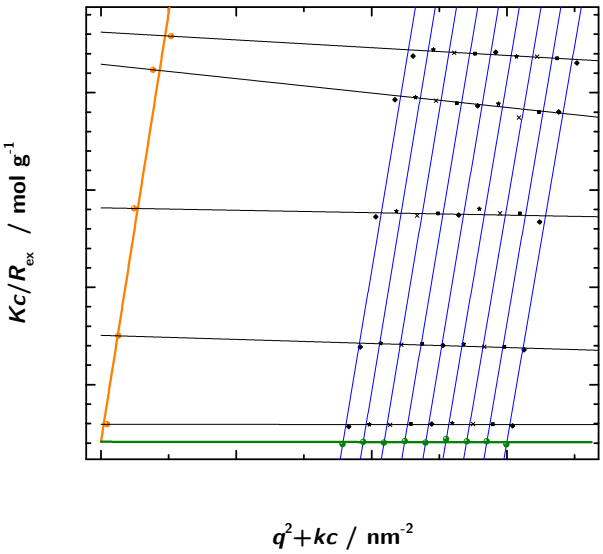

Figure 18: Example of a Zimm-diagram showing the measured data sets (rows of black symbols), the extrapolation to $q = 0$ for every dataset (orange dots), the extrapolation to $c = 0$ (green dots) and their corresponding linear fits (lines).

5.2.6 Static light scattering - Hands on

The static light scattering measurements were performed with the same setup as used for the PCS. All measurements were done at 20 °C and the scattered intensities of the samples were recorded in

the scattering angle range between 40° and 130° in steps of 1°. For the measurements of the buffer solution and the toluene reference the same parameters were used. The samples were prepared directly before the measurements in a clean workbench, centrifuged for 15 min at 21,000 x g at 4 °C, filtered through a 0.2 µm syringe filter (Cellulose Acetate, VWR International GmbH, Darmstadt, Germany) directly into a quartz cuvette (540.110-QS, Hellma GmbH & Co. KG, Germany) and finally centrifuged for 5 minutes at 2000 x g for the removal of dust and air bubbles. The cuvettes were previously cleaned with freshly distilled acetone.

5.3 | *Size exclusion chromatography*

Size exclusion chromatography (SEC) is a method which can be used for analytical purposes, as well as for purification and polishing of substances [128]. For this a sample with several components flows through a column with a porous organic network structure as stationary phase. The single components are thereby retarded by their different penetration depth of the network of the stationary phase in a manner that small molecules are retarded more than bigger ones due to their higher penetration depth. Several different detectors (e. g. UV, refractive index, light scattering) can be used. A conclusion concerning the hydrodynamic radius can be drawn from the elution volume and it is possible to obtain the molecular weight of each component by a comparison with a standard. A typical size exclusion chromatography diagram is shown in figure 19.

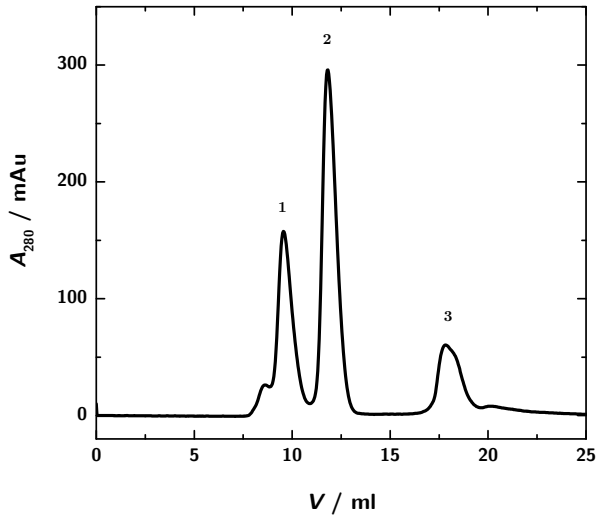

Figure 19: Example of an SEC elution diagram showing the separation of three proteins with different sizes (Albumin (1), Carbonic anhydrase (2), Aprotinin (3), calibration mixture).

5.4 | *Size exclusion chromatography - Hands on*

For the size exclusion chromatography a HPLC-system (ÄKTApurifier, GE Healthcare, Chalfont St Giles, UK) equipped with different columns for purification and analysis was used. The purification was done with a HiLoad 16/60 Superdex 75 PG column (GE Healthcare, Berlin, Germany) using a 20 mM MES, pH 6.5, 150 mM NaCl buffer. Whereas for the analytical size exclusion chromatography a Superdex 75 10/300 GL column (GE healthcare, Berlin, Germany) was used. This column was equilibrated with 1.3 times the volume of the column with 20 mM MES, 6.5, 500 mM NaCl buffer. After injecting the sample, the UV-absorption at 280 nm was measured for 1.3 column volumes.

MES: 2-(*N*-morpholino)ethanesulfonic acid

5.5 | *Gel electrophoresis - SDS-PAGE*

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) is a standard method for the analysis of proteins [129, 130]. It allows the determination of the molecular weight by comparison with a standard and the analysis of the purity of a sample. The basic principle of electrophoresis is the separation of charged molecules in an electrical field. In an initial step the samples are treated with the anionic detergent sodium dodecyl sulphate (SDS) causing a linearisation of the proteins and imparting a negative charge. Then they are loaded onto the gel and an electrical field is applied.

The gel, a polymer network consisting of acrylamide and bisacrylamide, restrains molecules depending on their size, e.g. larger molecules migrate slower than small molecules towards the anode. The first stage of the gel has a lower acrylamide content and is called stacking gel (Fig. 20). It is used to concentrate the samples and to obtain a defined starting point. To visualise the resulting bands the gel is stained with Coomassie Brilliant Blue.

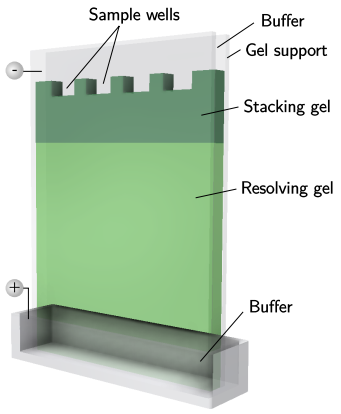

Figure 20: Schematic drawing of a gel electrophoresis chamber.

5.6 | *Gel electrophoresis - Hands on*

The SDS gel used for this thesis had an acrylamide content of 10 % (v/v). The detailed composition is shown in table 1. For the electrophoresis run 40 mA were applied for 35 min to 40 min. As molecular weight standard the unstained protein molecular weight

SDS sample buffer (6×):
300 mM Tris/HCl pH 6.8, 600 mM β-mercaptoethanol, 60 % (v/v) glycerol, 12 % (w/v) SDS, bromophenol blue, ultrapure water

SDS electrophoresis buffer (10×):
250 mM Tris, 1.92 M glycine, 1 % (w/v) SDS, ultrapure water

ladder (Fermentas, St. Leon-Rot, Germany) was used. All samples were treated with 6x SDS buffer and incubated for 5 min at 95 °C before being loaded on the gel. After the run the SDS gel was inserted in Coomassie staining solution, incubated for one minute in the microwave oven and then shaken for 20 min at room temperature. To highlight the bands, the gel was treated with destaining solution. For documentation purposes the gel was then incubated for 30 min in drying resolution and afterwards dried at room temperature.

Staining solution:
0.25 % (v/v) Coomassie Brillant Blue R-250, 30 % (v/v) Ethanol, 10 % (v/v) Acetic acid, Milli-q-water

Destaining solution:
40 % (v/v) ethanol, 10 % (v/v) acetic acid, ultrapure water

Drying solution:
50 % (v/v) methanol, 3.4 % (v/v) glycerol, ultrapure water

	Type of gel	
	stacking (4 %)	resolving (10 %)
Ultrapure water / ml	4.5	9.0
Acrylamide-Bisacrylamide (30 %, 37.5:1) / ml	0.9	7.5
1.5 M Tris-HCl pH 8.8 / ml	-	5.7
0.5 M Tris-HCl pH 6.8 / ml	1.5	-
SDS (10 %) / µl	75	200
APS (24 %) / µl	20	50
TEMED / µl	10	20

Table 1: Composition of the SDS-gels (APS: Ammonium persulfate, TEMED: *N,N,N′,N′*-Tetramethylethane-1,2-diamine; all quantities for four gels; the stacking gel has an acrylamide content of 4% (v/v) compared to the resolving gel with 10% (v/v))

All chemicals were purchased form AppliChem (Darmstadt, Germany), Fermentas (Leon-Rot, Germany), Fluka (Neu-Ulm, Germany), Merck (Darmstadt, Germany), Carl Roth (Darmstadt, Germany), Serva (Heidelberg, Germany), Sigma-Aldrich (Taufkirchen, Germany) or VWR (Leuven, Belgium) unless otherwise stated.

5.7 | Further parameters

For the evaluation of the light scattering data further parameters of the samples are needed: the density, the refractive index and the refractive index increment. The density and the refractive index measurements were performed on a DMA 4500 density meter and on an RXA 170 refractometer from Anton Paar (Graz, Austria). All measurements were repeated at least three times and the averaged values were taken for further calculations. The refractive index increment was determined using the α-Ref.-System (Version 2.1) from SLS-Systemtechnik (Denzlingen, Germany) with a quartz cuvette (410.45, Starna GmbH, Pfungstadt, Germany). All measurements were performed at 20 °C.

5.8 | Calculations

The static scattering experiments described in earlier sections lead to data sets describing the sample parameters in reciprocal space.

One possibility of obtaining information about the system in real space, is the use of the inverse Fourier transformation resulting in the $p(r)$-function [107]. Another approach is to calculate the scattering intensities emanating from a known structure or to calculate structure factors which are then fitted to the measured scattering data. When dealing with protein samples the program CRYSOL [83] from the ATSAS package [88] provides a sophisticated way for the calculation of scattering functions out of PDB files with the possibility of the adaption of the calculated and the measured data sets. If no details about the structure of a protein are known or if no homologous structure is available, it is necessary to use *ab-initio* calculations [27]. This method needs no further information than the scattering data itself and calculates a possible low resolution model of the shape of a molecule fitting the scattering amplitudes. Whereas if there are homologous structures available or the protein consists of known substructures, a rigid body modelling approach is useful [89]. Hereby a structural optimization by shifting and rotating fixed structural parts against each other, is performed within the limits of predefined restraints. Each optimization step is followed by a comparison of the calculated and the experimental scattering data. This kind of structure optimisation can be performed using the programs MASSAH [131, 132] or SASREF [89] also available from the ATSAS package. The latter program allows to define constraints, e.g. distances, that should not be exceeded. Besides the calculation or simulation of parameters arising from static scattering experiments, it is also possible to simulate hydrodynamic properties of known structures which then can be compared with photon correlation spectroscopy measurements. For this case the program HYDROPRO [94, 97, 133, 134] from J. Garcia de la Torre et. al. can be used [100, 135].

6 Results and Discussion

6.1 | Initial cleaning and polishing of the 3H3 diabody

The analysis of the overall structure and the molecular weight of the 3H3 diabody was conducted using scattering experiments. These types of experiments require highly pure samples. To fulfill this requirement the purified protein samples were polished by a preparative SEC run directly before each experiment. In figure 21 the elution profile of such an SEC run is shown. The main peak appearing at the elution volume of 65 ml denotes the 3H3 antibody fragment, whereas the small peaks at the elution volume of the 51 ml and 83 ml result as a consequence of impurities and aggregates.

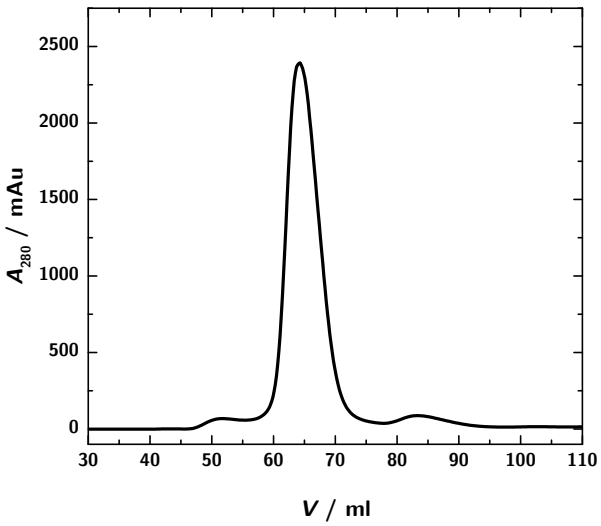

Figure 21: Elution profile of the preparative SEC run of the 3H3 diabody.

To distinguish between the fractions containing the pure polished protein and fractions containing impurities, the SEC fractions were analysed with SDS PAGE (Fig. 22). An SDS gel with an acrylamide content of 10 % was used for this analysis and 10 µl of each selected SEC fraction was deposited. Finally, the protein bands were visualised by staining with Coomassie blue solution.

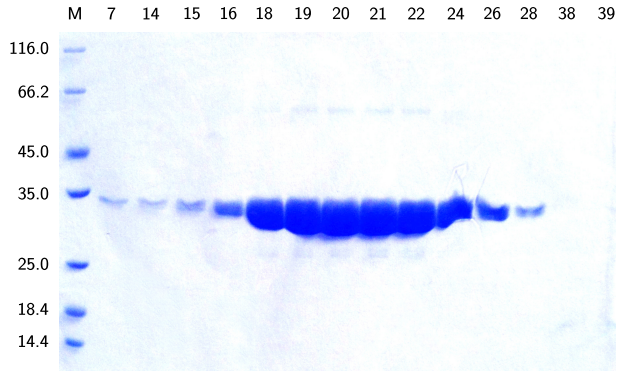

Figure 22: Analysis of the fractions of the SEC run with SDS PAGE. For the analysis a SDS gel with an acrylamide content of 10% was used and 10 µl of each fraction was deposited. Finally the gel was dyed with Coomassie blue solution.

The fractions 7 - 28 show the characteristic band at the expected molecular weight of the denaturated 3H3 diabody (27 kDa). Due to the small retention volume of fraction seven, it is assumed that this fraction contains aggregates which are separated by SEC. The main peak fractions (fraction 18-24) contain the maximum amount of 3H3 diabody. The very weak bands at approximately 25 kDa and at approximately 60 kDa are caused by impurities and can be neglected when seen in comparison to the total amount of protein in these fractions as shown in figure 22.

6.2 | *Light scattering*

6.2.1 Photon correlation spectroscopy

The determination of the hydrodynamic properties of the 3H3 antibody fragment was done by photon correlation spectroscopy (PCS) measurements. Different samples with a protein concentration between 2.5 mg ml^{-1} and 10 mg ml^{-1} were analysed at scattering angles of 60° and 135°. For the sample with the highest and the lowest protein concentration angular dependent measurements were performed in steps of 1°. Thus it is possible to analyse the system regarding changes in concentration and changes becoming visible in angular dependent measurements. Furthermore this method is highly sensitive to the particle size. Hence, it allows a clear distinction between scFv, diabody and triabody and can detect the coexistence of these structures. The resulting data sets were analysed using the software CONTIN [122] to obtain the relaxation rate Γ and the translational diffusion coefficient D^{exp}.

According to equation 40 the translational diffusion coefficient D^{exp} can also be obtained from a plot of the relaxation rate against

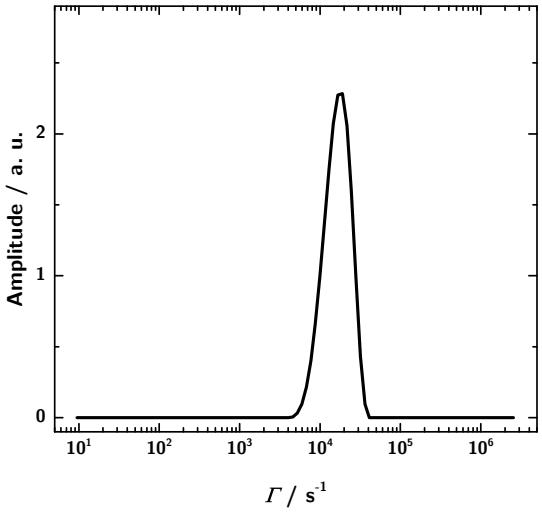

Figure 23: Relaxation rate distribution obtained by CONTIN of the 2.5 mg ml^{-1} sample recorded at an angle of 80°, showing only one relaxation rate.

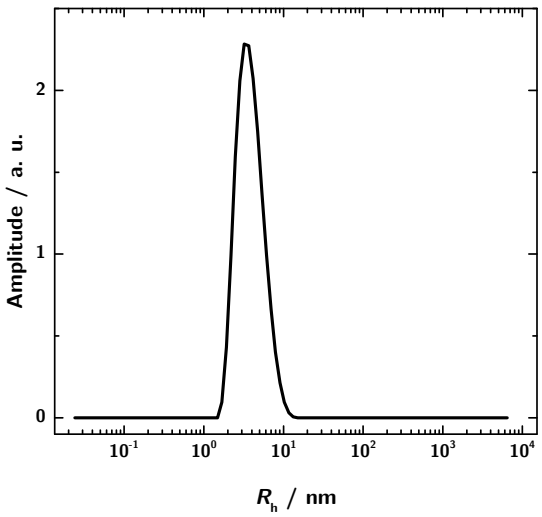

Figure 24: Distribution of the hydrodynamic radius calculated from the results obtained by CONTIN of the 2.5 mg ml^{-1} sample recorded at an angle of 80°, showing only one relaxation rate.

the squared scattering vector (Fig. 25) by performing a linear fit. In this approach the angular dependent measurements were evaluated, yielding a value for D^{exp} with a lower statistical error. All values of D^{exp} in this thesis originate from applying this approach.

Using the known translational diffusion coefficient, the hydrodynamic radius was calculated according to the Stokes-Einstein equation 43. The calculated values for the translational diffusion coefficient and the hydrodynamic radius obtained by CONTIN are summarised in table 2. The translational diffusion coefficient determined at 2.5 mg ml^{-1} was 5.95×10^{-11} m^2 s^{-1}. At 10 mg ml^{-1} a slower diffusion coefficient of 5.06×10^{-11} m^2 s^{-1} was found. This slower diffusion at a higher sample concentration can be explained

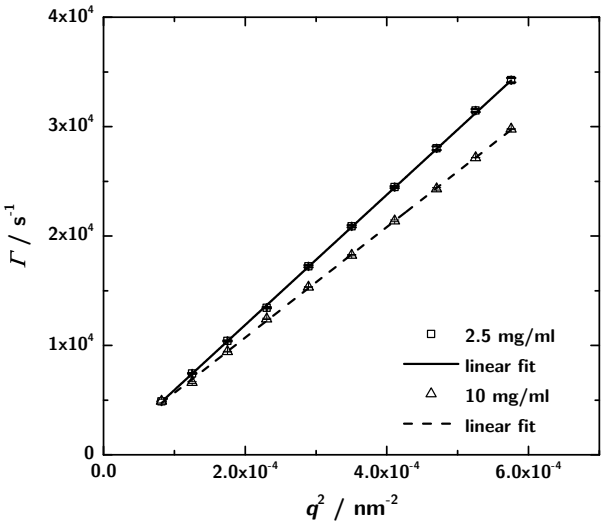

Figure 25: Determination of the translational diffusion coefficients for the $2.5\,\mathrm{mg\,ml^{-1}}$ and $10\,\mathrm{mg\,ml^{-1}}$ samples by a linear fit according to eq. 40.

by regarding the interaction parameter which is discussed in the next section.

Sample	Computed parameters	
c / $\mathrm{mg\,ml^{-1}}$	D / $\mathrm{m^2\,s^{-1}}$	R_h / nm
2.5	5.95×10^{-11}	3.55 ± 0.18
10	5.06×10^{-11}	4.18 ± 0.21

Table 2: Hydrodynamic parameters measured by PCS and evaluated using CONTIN and a linear fit approach combined with the Stokes-Einstein equation.

A further parameter that can be derived from the PCS measurements is the interaction parameter k_D, indicating whether there are attractive or non-attractive intermolecular interactions. It is represented by the slope of a concentration dependent plot of the translational diffusion coefficient (eq. 44, Fig. 26) and was determined to $-1.50 \times 10^{-2}\,\mathrm{ml\,mg^{-1}}$ for the measurements at an angle of 60° and to $-1.80 \times 10^{-2}\,\mathrm{ml\,mg^{-1}}$ at an angle of 135°. Negative values stand for attractive intermolecular interactions between protein molecules in the given solvent [124], which indicates a favourable precondition for crystallisation. The occurrence of attractive intermolecular interactions has a direct influence on the translational diffusion coefficient with increasing sample concentration. At higher concentration the diffusion of a single molecule is slowed down due to the attractive intermolecular interactions. Analysing the translational diffusion coefficient regarding the measurements at different sample concentrations, the extrapolation to zero concentration results

in the self-diffusion coefficient D^0. In table 3 the values calculated for D^0 and the connected $R_{h,o}$ for the concentration dependent measurements at 60° and 135° are shown. The self-diffusion coefficient describes the diffusion behaviour of a single scattering object.

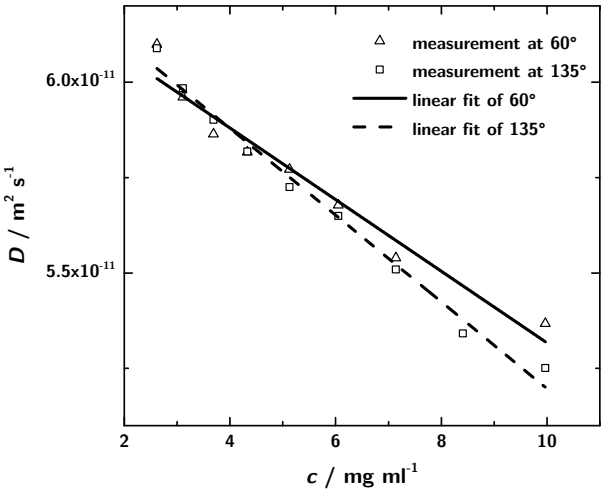

Figure 26: Determination of the interaction parameter (obtained form the slope of the linear fit) and the self-diffusion coefficient (the intersection of the linear fit with the y axis) at 60° and 135°.

Parameter	60°	135°
D^0 / $m^2\,s^{-1}$	6.26×10^{-11}	6.33×10^{-11}
$R_{h,o}$ / nm	3.4 ± 0.2	3.3 ± 0.2
k_D / $ml\,mg^{-1}$	-1.50×10^{-2}	-1.80×10^{-2}

Table 3: Hydrodynamical parameters of a single molecule and the interaction parameter calculated from concentration dependent PCS measurement data obtained at 60° and 135°.

At a glance the results determined by PCS show a monomodal relaxation rate distribution giving no evidence neither for the formation of aggregates nor for the coexistence of scFvs, diabodies and triabodies during the time of the measurement. The negative values obtained for the interaction parameter k_D indicate attractive intermolecular interactions which are a good precondition for crystallisation experiments. The experiments at 60° and 135° lead to identical results considering the assumed uncertainties. Finally, self diffusion coefficients in the range of $6.3 \times 10^{-11}\,m^2\,s^{-1}$ and a hydrodynamic radius of 3.4 nm were found.

6.2.2 *Static light scattering - Zimm-analysis*

In addition to the determination of hydrodynamic properties using photon correlation spectroscopy, light scattering provides the possibility of obtaining static parameters such as the molecular weight. For this reason static light scattering in combination with the Zimm-analysis was used. The basis for this analysis was the data obtained from angular dependent static light scattering measurements of samples with five different protein concentrations. A Zimm-analysis was then performed on the area of the data which had shown a linear behaviour. In figure 27 the results with a linear fit and the extrapolation according to the equations mentioned in chapter "Materials and Methods" are given. The intercept of both extrapolated lines with the y axis is at $2.0666 \times 10^{-5}\,\mathrm{mol\,g^{-1}}$. Thus, the resulting molecular weight obtained by static light scattering is 48.4 kDa. This is a good accordance with the calculated molecular weight of 53.9 kDa of the diabody.

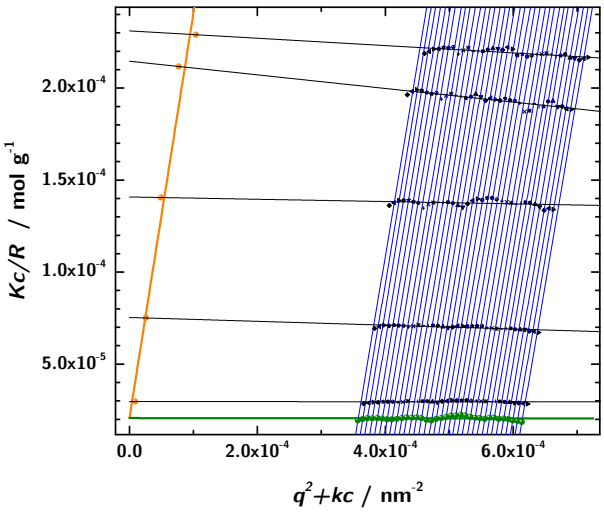

Figure 27: The Zimm-diagram obtained from the static light scattering measurements (black: measured data, orange: extrapolation to $q = 0$, green: extrapolation to $c = 0$, lines: linear fits).

Parameter	Value
c / $\mathrm{mg\,ml^{-1}}$	0.91, 2.58, 4.9, 7.71, 10.0
$R_{\text{ex, Toluene}}$ / $\mathrm{cm^{-1}}$	13.43×10^{-6}
$\mathrm{d}n/\mathrm{d}c$ / $\mathrm{cm^3\,g^{-1}}$	0.267
$n_{\text{b, MES-Buffer}}$	1.335
$n_{\text{ref, Toluene}}$	1.496
k / $\mathrm{nm\,g^{-1}}$	10
K / $\mathrm{mol(m/g)^2}$	$5.550\,38 \times 10^{-11}$

Table 4: Parameters and their values used for the Zimm-analysis

6.3 | *Small angle x-ray scattering*

6.3.1 *General data preparation and analysis*

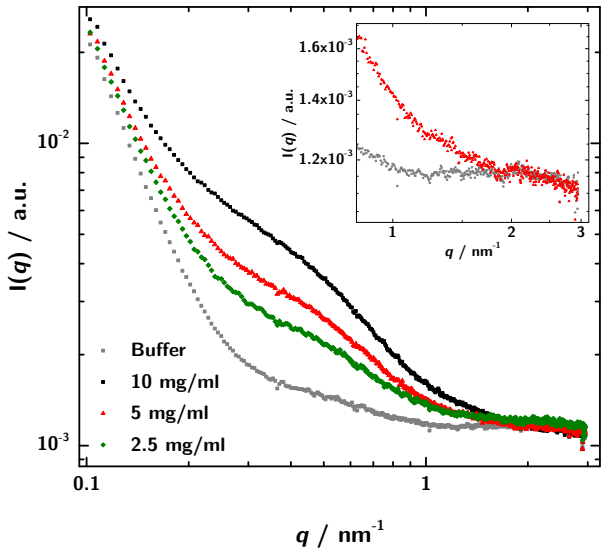

Figure 28: 1D plot of the scattering intensities of the different sample concentrations and the buffer without further treatment. The inset shows the q-range where buffer and sample data overlap.

Solution scattering methods provide an elegant way for the investigation of proteins in solution. They allow the determination of geometrical parameters as well as the molecular weight and can be used as the starting point for model calculations. The SAXS measurements were performed with sample concentrations of 2.5 mg ml^{-1}, 5 mg ml^{-1} and 10 mg ml^{-1} at three sample-to-detector distances 50 cm, 133 cm and 278 cm. These distances allow the coverage of a q-range from 0.04 nm^{-1} to 6.78 nm^{-1}. The potentially usable q-range after buffer subtraction and combining the results of the three distances ranges from 0.018 nm^{-1} to 1.67 nm^{-1} (Fig. 28, 30).

Glassy carbon was used as standard for the intensity calibration. The consequently determined calibration factors were: 594 at 278 cm, 100 at 133 cm and 75 at 50 cm.

To check for beam damage and possible instabilities of the setup, the measurements were subdivided in frames of 10 min, the data of the first and last 10 frames were averaged and compared. This is shown exemplarily in figure 29 for the measurement of the 5 mg ml^{-1} sample at a sample to detector distance of 133 cm.

The plot of the concentration normalised data (Fig. 30) shows only small differences in the scattering intensities of the three sample concentrations in the q-region below 1 nm^{-1}. Changes in the

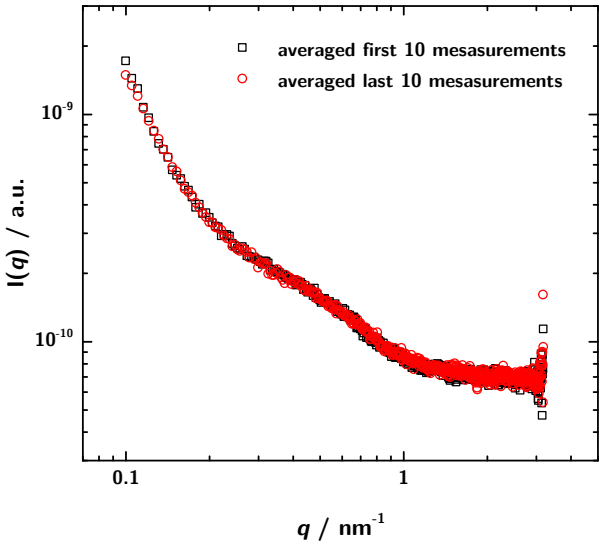

Figure 29: Comparison of the scattering data of the beginning of a measurement with those at the end (5 mg ml^{-1}, 133 cm sample to detector distance). To obtain a better statistic 10 frames were averaged respectively. Within the experimental precision no change in the scattering curves and therefore no sample degradation is observable.

scattering intensities with increasing sample concentration may point towards interparticle interaction and the formation of aggregates beside this, it can be caused by fluctuations of the primary beam intensity during the different measurements needed.

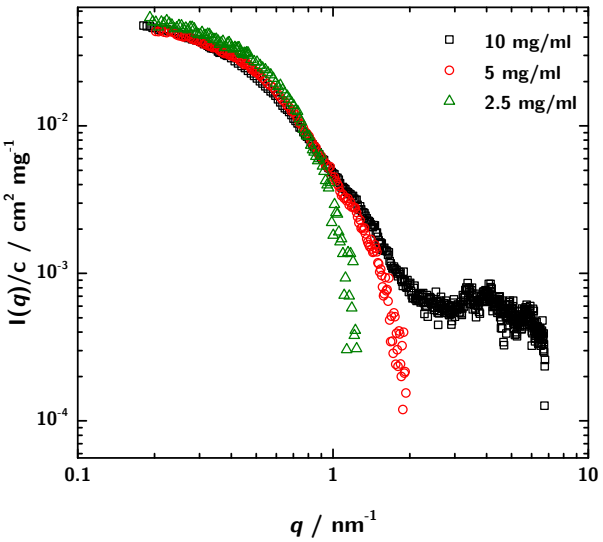

Figure 30: Experimental SAXS patterns on absolute scale with concentration correction for protein solutions at 10 mg ml^{-1} (square), 5 mg ml^{-1} (circle) and 2.5 mg ml^{-1} (triangle). Each curve was obtained from three measurements at different sample to detector distances.

As it is previously shown by PCS measurements, the formation of aggregates can be neglected, leading to the assumption that there is no significant concentration dependent structural change recogniseable. The slight decrease in intensity at low q-values with increasing sample concentration can be explained by the contribution of the non-avoidable structure factor.

The ascending noisiness of the scattering data with increasing q-values arises from the very small difference between the buffer

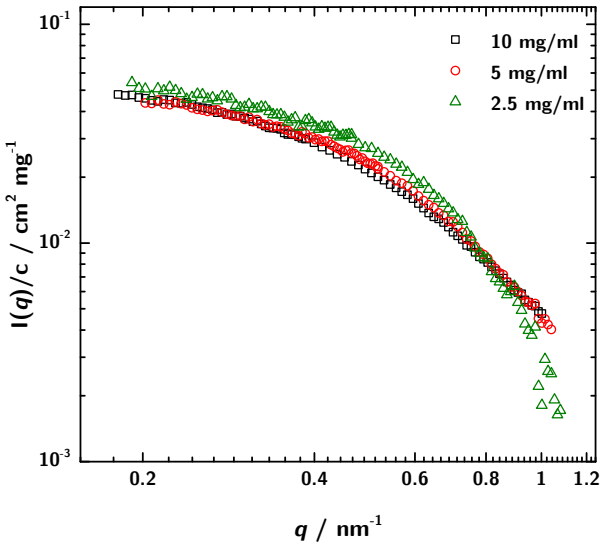

Figure 31: Experimental SAXS patterns as showen in Fig. 30 up to the maximum useful q-values calculated using SHANUM.

and the sample signal in this q- range (see also inset of Fig. 28). In general it can be stated that the set of data gathered from the $5\,\mathrm{mg\,ml^{-1}}$ sample is, in this case, the best possible compromise, taking into account both the statistics obtained at a measurement time of 24 hours as well as the tendency of aggregate formation at high levels of concentration. In 2015, P. Konarev and D. Svergun [108] described a method for determining the useful range of experimental data by employing the Shannon sampling algorithm to the scattering data. This method is implemented in the software SHANUM [108] which is part of the ATSAS package [88]. The results of the application of this approach to the obtained scattering data is shown in figure 31. It can be seen that the useful range is restricted to q-values around $1.1\,\mathrm{nm^{-1}}$.

The first evaluation step after this general initial data treatment was to employ the Guinier analysis (Fig. 32) followed by performing the indirect-transformation (GNOM).

The mean radius of gyration $R_{\mathrm{g,mG}}$ was estimated according to the Guinier approximation (eq. 26) to 3.41 nm (see table 5 for all details). The extrapolation of the Guinier fit to q = 0 resulted in the $I(0)$ values which allow in a further step the determination of the approximate molecular weight. The values obtained for R_{g} and $I(0)$ for the different sample concentrations are identical, considering the assumed uncertainties.

The GNOM calculation yields the pair distance distribution functions $p(r)$ of the SAXS data from the different concentrations

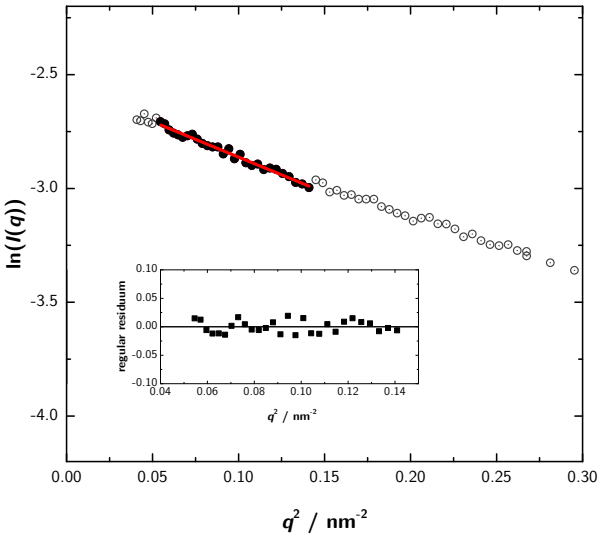

Figure 32: Guinier plot and fit of the scattering data obtained from the 5 mg ml^{-1} sample. The residua are shown in the inset.

	10 mg ml^{-1}	5 mg ml^{-1}	2.5 mg ml^{-1}
R_g / nm	3.66 ± 0.18	3.29 ± 0.16	3.27 ± 0.16
$I(0)$ / cm^2 mg^{-1}	0.056	0.052	0.059

Table 5: Structural parameters derived from Guinier analysis.

measured (Fig. 33). The data obtained provide the maximum dimension (D_{max}) denoted by the value of r where the $p(r)$-function reaches zero. A further parameter, the $R_{g,real}$, was obtained using this method (Tab. 6) according to eq. 28.

	10 mg ml^{-1}	5 mg ml^{-1}	2.5 mg ml^{-1}
$R_{g,real}$ / nm	3.13 ± 0.16	3.03 ± 0.15	2.89 ± 0.14
V_{Porod} / nm	84.39	76.07	102.77
D_{max} / nm	9.9	9.06	8.50

Table 6: Structural parameters derived from GNOM analysis.

Another basic step in the evaluation of scattering data is the analysis of the Kratky plot, showing the relation of the scattering intensity weighted by the square of the scattering vector, to the scattering vector (Fig. 35, 34). This plot gives evidence of the folding state of a protein [136, 137]. One single peak denotes a globular shape, a continuous increase of the values with higher q points to a natively unfolded protein, and several peaks can be identified if it is a multi-domain sample with flexible linkers. The Kratky plot in the q range specified by the SHANUM tool shows a single peak for the 2.5 mg ml^{-1} sample. The plots of the 5 mg ml^{-1} and 10 mg ml^{-1} sample show a small peak that does not return to the initial level at

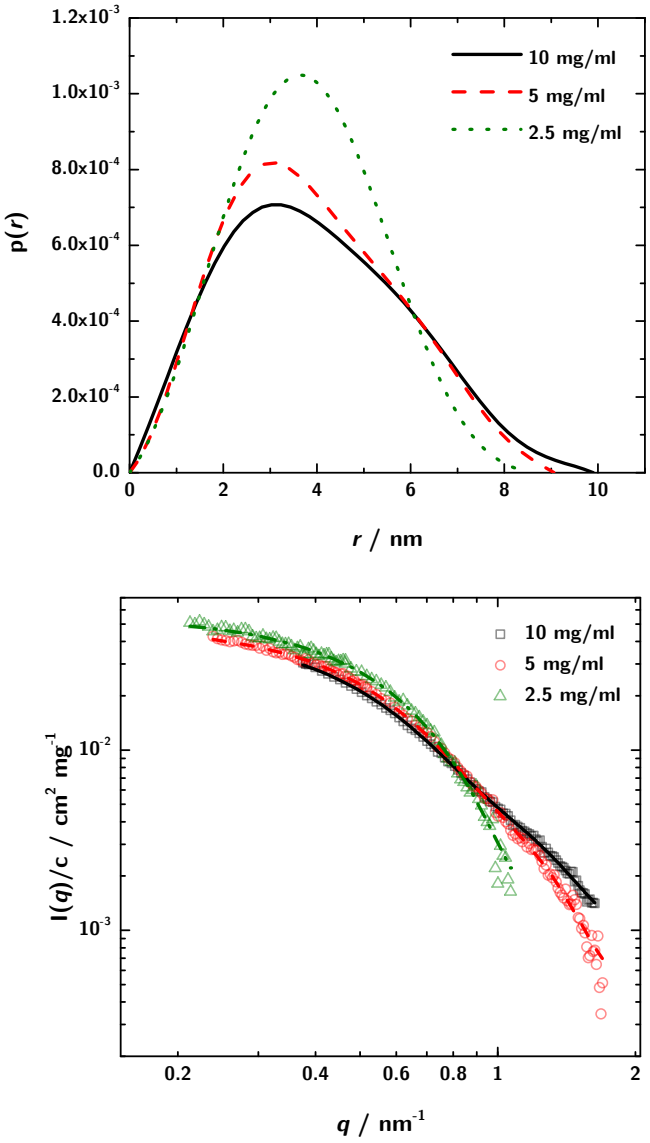

Figure 33: $p(r)$ - functions obtained by GNOM (upper graph) and the corresponding fit in the reciprocal space (lower graph).

higher q values. At this point it remains unclear if these curves run into a plateau or if further peaks will appear.

When analysing the q-range beyond the limits given by SHANUM, a second peak can be seen in the region of $q = 1.3$ at high sample concentrations. For the lowest sample concentration there is no data available in this region. Regarding the Kratky plot with the extended q range leads to the conclusion that the protein has multi-domains which are connected by flexible linkers.

6.3.2 Molecular weight determination

The main aim of this thesis is to study the aggregation state of the 3H3 antibody fragment which might primarily exist as monomer,

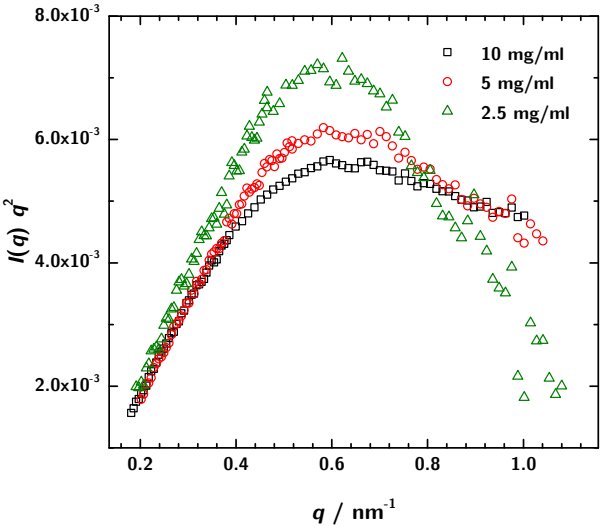

Figure 34: Kratky plot of the scattering data of three sample concentrations. The upper q-limit is chosen according to the values obtained by SHANUM. Regarding this q range, it can not be clearly concluded in which folding state 3H3 exists.

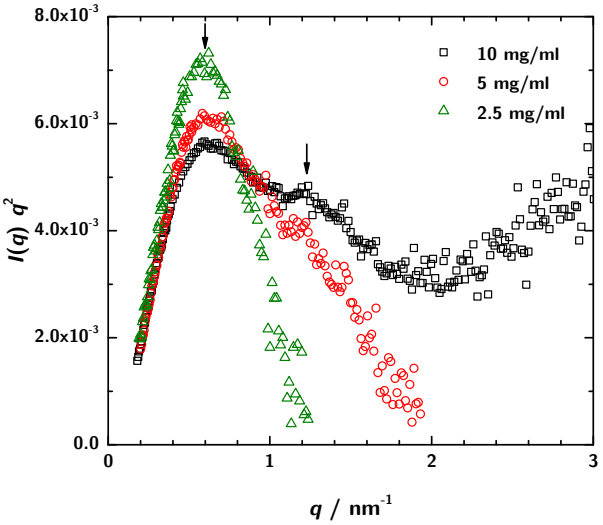

Figure 35: Kratky plot of the scattering data of three sample concentrations with an extended q-range. The data for 10 mg ml^{-1} and 5 mg ml^{-1} show two peaks (marked by arrows) denoting a protein with multi-domains connected by a flexible linker.

dimer or as oligomer. Beyond several structural parameters, the molecular weight is a suitable and easily accessible parameter to distinguish between the possible degrees of oligomerisation of a protein.

The molecular weight of a protein can be determined from the scattering data in several ways as described in the previous chapter. Using the $I(0)$ values obtained by the Guinier analysis of the absolute scattering intensities, a mean molecular weight of 70.8 kDa was obtained. The comparison with $I(0)$ of lysozyme as a secondary standard resulted in a mean molecular weight of 46.5 kDa, the analysis of the area under the Kratky plot with the online calculation tool "SAXS MoW" [117] yielded a mean value of 69.0 kDa (Tab. 7).

The molecular weight of the diabody calculated from the sequence is 53.9 kDa.

A comparison of the molecular weight determined from SAXS data using different methods, with the calculated theoretical molecular weight of the diabody, can be done. Using the average of the individual results of each of the methods a deviation in the range of −14 % to 31 % compared to the calculated value of 53.9 kDa was obtained.

In order to be able to verify the supposition that 3H3 possesses a diabody structure, a comparison with the molecular weights of the scFv and the triabody is also necessary.

In contrast, a further comparison based on the calculated molecular weight of the scFv, 26.9 kDa, yields a deviation of up to 81 % from this calculated value. This is a much higher deviation than that obtained by the comparison with the molecular weight of the diabody. The triabody molecular weight is shown to be much heavier than the experimental value obtained for the 3H3. In this case, the deviation features a range of −12 % to −57 % compared to the triabody molecular weight.

Hence, in despite of the spread in the molecular weight, the theoretical value for the diabody has the lowest over all deviation from the experimental findings.

Parameter	Experiment		
	10 mg ml^{-1}	5 mg ml^{-1}	2.5 mg ml^{-1}
$R_{g,Guinier}$ / nm	3.66 ± 0.18	3.29 ± 0.16	3.27 ± 0.16
$R_{g,real}$ / nm	3.13 ± 0.16	3.03 ± 0.15	2.89 ± 0.14
$I_{Guinier}(o)$ / cm^2 mg^{-1}	0.056	0.052	0.059
$I_{via\ Lys}(o)$ / cm^2 mg^{-1}	0.20	-	-
M_{abs} / kDa	70.0	66.1	76.2
$M_{via\ Lys}$ / kDa	46.8	43.4	49.3
$M_{via\ MoW}$ / kDa	68.6	69.4	–

Table 7: Structural parameters derived from experimental SAXS data.

Regarding the deviations between the expected molecular weight and the results obtained form the molecular weight determination by SAXS, raises the question of the accuracy of this method. In recent years Mylonas [114] and Akiyama [138] addressed this question and state an accuracy of this method of about 8 % to 10 %. It is to point out, that all of their SAXS experiments were performed

at synchrotron beamlines under ideal conditions, directly after purification of the protein samples. The acquisition time for each sample is reported to have been in the range of several minutes. In contrast, the scattering experiments reported in this thesis were all performed at a laboratory beamline, generating less flux and thus requiring an extended acquisition time (24 h) and providing results with a higher statistical error. Hence, an obtainable accuracy in the range of 15 % to 20 % can be assumed.

In summary, the reported molecular weight determination by SAXS leads to the assumption that 3H3 actually forms dimers. As a further verification of this result, simulations were done and compared with the experiment.

6.4 | *Calculations based on homologous protein structures*

6.4.1 Calculation of SAXS data using CRYSOL

To verify the conclusion based on the different determinations of the molecular weight and to reveal the structure of 3H3 in solution in an independent way, the theoretical scattering curves of different possible oligomers were calculated by CRYSOL and compared with the experimental data. The calculations were based on crystal structures of homologous proteins listed in the PDB.

Exemplarily the structure of 1X9Q [139] was used for an scFv, for the diabody 1LMK [66] and for the triabody the structure of 1NQB [140] served as the starting point (Fig. 36). Figure 36 shows the experimental data for the protein solution at a concentration of $5\,\mathrm{mg\,ml^{-1}}$ and the results from the simulated scattering curves.

The calculated curve of the diabody structure nicely describes the measured data, whereas both of the other curves show strong deviations. This can also be substantiated when regarding the χ^2-values listed in table 8. Here the results of the comparison of the scattering intensities obtained at three concentrations ($2.5\,\mathrm{mg\,ml^{-1}}$, $5\,\mathrm{mg\,ml^{-1}}$ and $10\,\mathrm{mg\,ml^{-1}}$) with the simulated scattering patterns are given.

It can be clearly seen, that the values obtained with the diabody structure are the smallest ones, corresponding to the best fit. Hence, these model calculations perfectly show that 3H3 forms a diabody in solution.

Single- (s), Dia- (d), Triabody (t)			
	1LMK (d)	1X9Q (s)	1NQB (t)
$10\,\mathrm{mg\,ml^{-1}}$	22.18	272.07	25.75
$5\,\mathrm{mg\,ml^{-1}}$	1.60	85.74	16.32
$2.5\,\mathrm{mg\,ml^{-1}}$	3.28	58.64	9.15

Table 8: χ^2-values obtained with CRYSOL (ATSAS6) when comparing the experimental scattering data of the protein samples with the calculated scattering curves based on the PDB-structure files of a singlebody (1X9Q), a diabody (1LMK) and a triabody (1NQB) structure.

As a next step the experimental scattering patterns were compared with the calculated scattering data of four diabody structures from the PDB (1LMK [66], 1MOE [74], 4Y5X [75], 4Y5Y [75]) in order to identify the structurally most fitting model for 3H3. In figure 37 it can be seen, that all the calculated curves almost fit with the measured data in the given range. Table 9 summarises the results

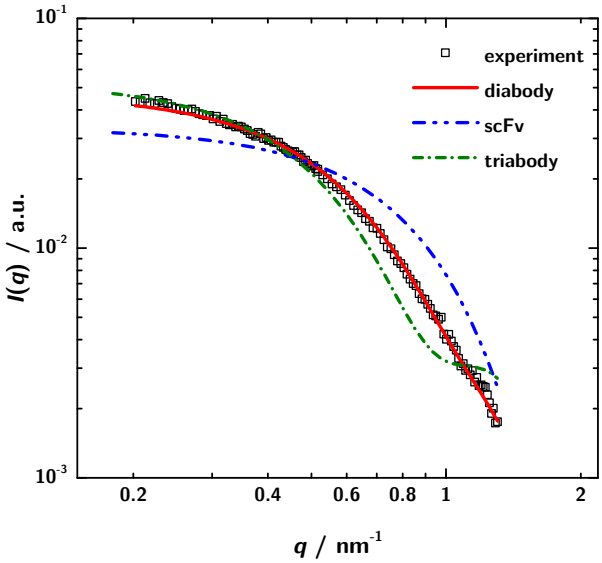

Figure 36: Simulation of the scattering patterns of an scFv (1X9Q), a diabody (1LMK) and a triabody (1NQB) with CRYSOL compared with the data obtained from SAXS measurements (5 mg ml^{-1})

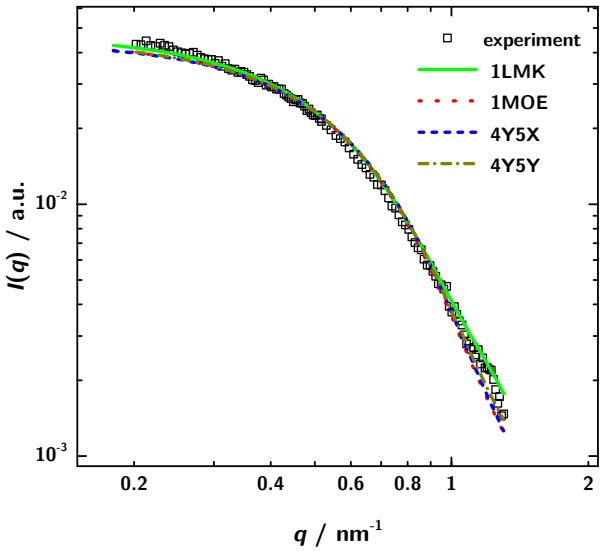

Figure 37: Simulation of the calculated scattering patterns of different diabodies with the data obtained from measurement (5 mg ml^{-1})

of this comparison. The best matching of the experimental data is obtained using the model based on the diabody structure 1LMK with a χ^2-value of 1.60 (referred to the scattering experiment with the concentration of 5 mg ml^{-1}). Thus the structure of 1LMK is suited best as a starting point for a further refinement against the experimental scattering data.

6.4.2 Outlook: Possible structure optimisation using SASREF

SASREF provides the possibility of calculating an optimised structure model by a rigid body refinement approach against the scattering data taking constraints into account [89]. As rigid bodies,

Diabodies				
	1LMK	1MOE	4Y5X	4Y5Y
10 mg ml^{-1}	22.18	36.56	41.06	37.05
5 mg ml^{-1}	1.60	4.05	4.93	4.14
2.5 mg ml^{-1}	3.28	2.34	2.98	2.93

Table 9: χ^2-values obtained with CRYSOL (ATSAS6) when comparing the experimental scattering data of the protein samples with the calculated scattering curves based on four of the available PDB-structure files of diabody structures.

the two heads on both sides of the linker were used (more details given below). The length of the linker between these two regions of the protein was taken to define the constraints for the structure optimisation. With a linker length of five amino acids and an assumed mean length of 3.5 Å per amino acid, the distance of both chain ends was restricted to 17.5 Å. In the rigid body refinements the 1LMK structure was optimised against the experimental scattering data from the 10 mg ml^{-1}, 5 mg ml^{-1}and 2.5 mg ml^{-1} protein solution, respectively. Following this structure optimisation a comparison with the experimental scattering intensities was done. The resulting χ^2-values calculated by CRYSOL are given in table 10 together with those based on the original 1LMK structure. It can be seen that the major improvement (72.45%) of the coincidence with the experimental values is achieved with the 10 mg ml^{-1} sample. Whereas the best fit with a χ^2 value of 0.94 is reached by the refinement with the 5 mg ml^{-1} sample. Here an improvement of 41.25% is obtained. The smallest effect of the refinement against the scattering data, with an improvement of 32.01%, can be observed for the 2.5 mg ml^{-1} sample.

Figure 38 exemplarily presents the result of the structure refinement with SASREF of 1LMK against the SAXS data of the measurement of the 5 mg ml^{-1} solution. Comparing both structures (1LMK and 1LMK-optimised), a shift of the heads away from each other and a rotation along the longitudinal axis of the molecule is observed. As an optical aid, two identical C_α atoms are coloured in each structure (1LMK in red, optimised structure in blue).

Diabody - Optimisation			
	10 mg ml^{-1}	5 mg ml^{-1}	2.5 mg ml^{-1}
1LMK	22.18	1.60	3.28
▷ optimised	6.11	0.94	2.23

Table 10: χ^2-values obtained with CRYSOL (ATSAS6) when comparing the experimental scattering data of the protein samples with the calculated scattering curves of the original 1LMK structure and 1LMK structure after optimisation with SASREF.

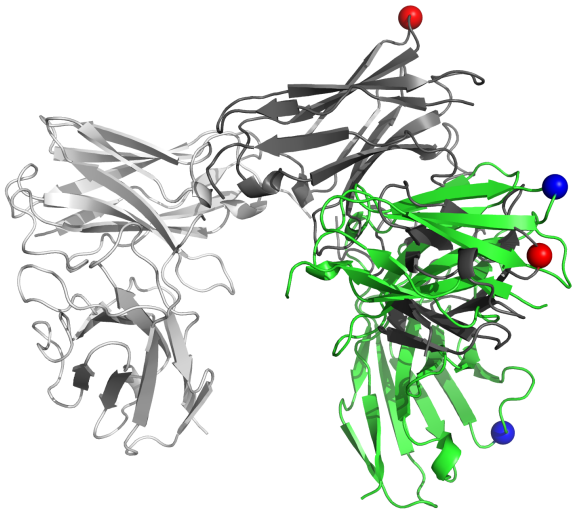

Figure 38: Rigid body refinement results of the 3H3 structure model, 1LMK in light/dark grey, 1LMK after refinement in light grey and green.

The last step was to make a prediction as to the structure of the 3H3 diabody using the 1LMK structure as a template together with the sequence of the 3H3 diabody. This homologous modelling was performed using the online tool 3D-JIGSAW [141, 142]. The resulting raw structure was then divided into two parts at the position of the linker, and a structure optimisation with SASREF was performed. In the non-optimised case the calculated scattering intensities of the predicted 3H3 structure are in poor agreement with the experimental scattering data. Here, even the results of the non-optimised 1LMK comparison show a better accordance (Tab. 11).

Regarding the χ^2-values computed by comparing the optimised structures with the experimental scattering data, the 3H3 structure results, in the case of the $10\,\mathrm{mg\,ml^{-1}}$ and $5\,\mathrm{mg\,ml^{-1}}$ experiment, in an improvement of the χ^2-values compared with the optimised 1LMK structure (Tab. 10, 11).

Diabody, 3H3 and optimised 3H3			
	$10\,\mathrm{mg\,ml^{-1}}$	$5\,\mathrm{mg\,ml^{-1}}$	$2.5\,\mathrm{mg\,ml^{-1}}$
1LMK	22.18	1.60	3.28
3H3	41.62	15.51	27.19
▷ optimised	5.36	0.50	2.51

Table 11: χ^2-values obtained with CRYSOL (ATSAS6) when comparing the experimental scattering data of the protein samples with the calculated scattering curves of the original 1LMK structure, the 1LMK structure with the 3H3 sequence aligned, and after optimisation with SASREF.

With respect to the 2.5 mg ml^{-1} experiment, a slight deterioration can be observed. This could be explained by the increasing statistical error of the experimental scattering data with decreasing protein concentration when using the same measuring time. Due to the limited stability of the protein sample, it was not possible to enhance the measuring time at lower protein concentration. To cope with this challenge, experiments at a synchrotron radiation source are necessary. There, a dramatic decrease in measurement time goes along with an increase in statistics because of the higher flux provided by this type of source. Scattering data with improved statistics and with a possibly extended q-range, furthermore allow the construction of a structure model based on the refinement of all diabody structures measured at several concentrations followed by an averaging of the results. On the other hand beam damage might become an important issue in a synchrotron based experiment. This is not a problem in our inhouse measurements.

6.4.3 Preliminary data treatment steps

The PDB data file of 1LMK consists of the four chains A-C-E-G, whereby the chains A/E and C/G are equivalent. As the first step a data file containing only the chains A and C (corresponding to one dimer) was generated using the SWISSPDB Viewer. In a further step each of these chains was renamed in a manner that in the end each half part of the chain had its own Chain-ID (e.g. A becomes AA and AD, C becomes CC and CB) with WinCOOT (Fig. 39). The first half part contains the amino acids 1 to 123, the second half part the amino acids 124 to 238 .

The rigid body refinement requires at least two sets of sequences which then are moved around each other. The most flexible part of a diabody is the linker region, therefore the structure was divided at the linker into two parts (left part: AA and CC, right part: AD and CB).

For each of these parts the partial scattering amplitudes for all atoms (ALM-file) [83] were calculated using CRYSOL.

The rigid body refinement in SASREF was carried out using the two structure files, the corresponding ALM-files, the experimental scattering data file and the constraints file, which provides the distance restrictions between the rigid bodies.

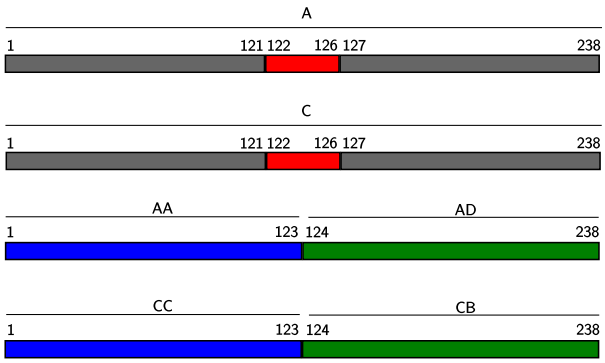

Figure 39: Schematic drawing of the chains (grey, linker in red) and the renamed chains. For the rigid body refinement, files containing the chains AA + CC and AD + CB were generated.

The initial data treatment when using the 3H3 sequence, mainly covered the same steps as used for the 1LMK sequence. After extracting chain A and C from the 1LMK structural data file, the sequence of the 3H3 antibody fragment was aligned to them using SWISS-MODEL [143]. Due to the fact that the linker in 3H3 consists of only four amino acids, the final files contain the amino acids 1 to 132 and 133 to 157 (including the final HIS-tag).

Table 12: Structural parameters derived from experimental SAXS data and modelling.

Parameter	Experiment			Calculation	
	10 mg ml^{-1}	5 mg ml^{-1}	2.5 mg ml^{-1}	1LMK	opt.
$R_{g,Guinier}$ / nm	3.66 ± 0.18	3.29 ± 0.16	3.27 ± 0.16	2.9	2.98
$R_{g,real}$ / nm	3.13 ± 0.16	3.03 ± 0.15	2.89 ± 0.14		
$I_{Guinier}(o)$ / cm^2 mg^{-1}	0.056	0.052	0.059		
$I_{via\ Lys}(o)$ / cm^2 mg^{-1}	0.20	-	-		
M_{abs} / kDa	70.0	66.1	76.2		
$M_{via\ Lys}$ / kDa	45.8	43.3	49.8		
$M_{via\ MoW}$ / kDa	68.6	69.4	–		

6.5 | *Calculation of the hydrodynamic data using HYDROPRO*

In a further evaluation step the program HYDROPRO [94, 97] was used to calculate the hydrodynamic properties of an scFv, a diabody and a triabody. The parameters needed for this calculation such as temperature and solvent density were obtained experimentally. The solvent viscosity, the partial specific volume and the molecular weight were calculated using different online tools as mentioned in the appendix. An overview of the values used in the

calculations is given in table 14. For all parameters not mentioned, the default values recommended in the software were used. To calculate the hydrodynamic radius from the results, the equation 43 was used.

Regarding the values obtained for the shape factor ρ_{sf}, the scFv appears as a perfect sphere, having a value of 0.77 [126, 144], whereas the results obtained for the dia- and triabody structure suggest the formation of coil-like structures. Here it is important to note that these calculations are based on crystal structures, representing only one structural arrangement of these flexible molecules. In general it is assumed that all of these molecules can be treated as spheres when investigating them in solution.

The values obtained for the radius of gyration are significantly different to each other which is very important to enable a proper differentiation when using this method. By comparing the calculated hydrodynamic radii of the three structures - 2.7 nm, 3.6 nm, and 4.3 nm, for scFv, diabody and triabody respectively - with those of the experimental results, an excellent agreement of the experimental value and the calculated value for the diabody within the experimental error can be ascertained. Regarding the translational diffusion coefficient, also a good agreement with the experimental value could be reached in the case of the 1LMK structure.

Table 13: Shape factors ρ_{sf} of different geometries [126, 144].

Formfactor	ρ_{sf}
sphere	0.774
coil	0.816
rod	1.732

	1X9Q	1LMK	1NQB	experiment
$R_{element}$ / Å	5.9	5.9	5.9	-
Type of calculation	1	1	1	-
T / °C	20	20	20	20
η / mPa s	1.016	1.016	1.016	1.016
M / kDa	29.447	52.086	77.006	-
$\overline{\nu}$ / cm g^{-3}	0.725	0.731	0.730	-
ρ / g cm^{-1}	1.006 77	1.006 77	1.006 77	-
$D^{o,\,theo}$ / 10^{-11} m^2 s^{-1}	7.74	5.94	4.90	D^o: 6.295
$R_{g,\,theo}$ / nm	2.1	2.9	3.5	-
$R_{h,o,\,theo}$ / nm	2.7	3.6	4.3	$R_{h,o}$: 3.4
$R_{g,\,theo}/R_{h,\,theo}$	0.77	0.81	0.81	-

Table 14: Calculated hydrodynamical properties of an scFv (1X9Q), diabody (1LMK), triabody (1NQB) compared with the experimental results, $R_{element}$ consists of 2.9 Å (recommended in HYDROPRO [94, 97]) for one element and 3 Å for the first hydration shell, the calculation is based on the: "atomic-level primary model, shell calculation (type 1)".

A three-dimensional model of the 3H3 diabody was generated based on the structure of 1LMK and the sequence of 3H3 using the program 3D-JIGSAW. This model was than used to calculate the

probable hydrodynamic properties of the 3H3 diabody. The results are given in table 15 and show no significant differences compared to those of 1LMK.

	1LMK	1LMK-3H3	experiment
R_{element} / Å	5.9	5.9	-
Type of calculation	1	1	-
T / °C	20	20	20
η / mPa s	1.016	1.016	1.016
M / kDa	52.086	50.614	-
$\overline{\nu}$ / cm g^{-3}	0.731	0.728	-
ρ / g cm^{-1}	1.006 77	1.006 77	-
$D^{\text{o, theo}}$ / 10^{-11} m^2 s^{-1}	5.94	5.98	D^{o}: 6.295
$R_{\text{g, theo}}$ / nm	2.9	2.9	-
$R_{\text{h,o, theo}}$ / nm	3.6	3.5	$R_{\text{h,o}}$: 3.4
$R_{\text{g, theo}}/R_{\text{h, theo}}$	0.81	0.81	-

Table 15: Comparison of the calculated hydrodynamic properties of a diabody structure (1LMK) and a generated 3H3 structure with the data obtained in an experiment. R_{element} consists of 2.9 Å (recommended in HYDROPRO [94, 97]) for one element and 3 Å for the first hydration shell, the calculation is based on the: "atomic-level primary model, shell calculation (type 1)".

7 Conclusion and Outlook

Scattering techniques are powerful tools which can be used to investigate the structure of proteins in solution. They allow the comparison of the solution structure to structural data from x-ray crystallography and are especially interesting in the context of structure determination of proteins which do not crystallise. For the purposes of the research area of this thesis, these methods were the ideal choice in order to allow the study of the IgG fragment 3H3 and to obtain an approximate structure of it.

3H3 was expected to form diabodies in solution but SEC results suggested that only unimers exist in solution. In order to further closely examine this aspect, this PhD thesis combines SAXS and PCS experiments with model calculations. The results obtained from the application of the combination of these techniques clearly prove the existence of the expected diabodies in mechanically unperturbed solution e.g. the absence of shear stress as present in an SEC run. Both SAXS based and PCS based approaches lead to the same result. Moreover, the additional use of PCS and SLS allow the ruling out of the presence of larger aggregates. Such aggregates would perturb the SAXS experiments and it would also be difficult to analyse them on SAXS.

That the 3H3 protein is a diabody could also be supported by hydrodynamic model calculations based on the methods introduced by J. Garcia de la Torre [94, 97, 133, 134, 145] using high resolution structures of homologous proteins existing as scFv, diabody or triabody. These results, when compared with the experimentally determined hydrodynamic properties from PCS measurements, clearly reaffirm the formation of diabodies.

A step further was ventured. The atomic-level structure of 3H3 was predicted based on the homologous diabody structure and the sequence of 3H3 with the program 3D-JIGSAW [91, 92]. Hence an even more sound basis on which to conduct further calculations

was provided. Model calculations were again carried out using the thus created, very close to reality, structure of the 3H3 protein. These further results clearly reaffirmed the foregoing calculations which had also shown the 3H3 protein to be a diabody.

Moreover, SLS was utilised which also enabled the calculation of the molecular weight of the diabody structure. The value derived from using the Zimm formalism is in a very good agreement with the molecular weight calculated from the amino acid sequence.

Until now it was not possible to crystallise the 3H3 diabody, thus a high resolution structure is not available. Combining the results from the x-ray scattering experiments with the sequence of the 3H3 and a high resolution structure of a homologous diabody, it is possible to predict an approximate crystal structure.

What comes next?

The next step on the road to a more detailed structure determination would be to perform scattering experiments at a synchrotron beamline providing a much higher flux resulting in a much shorter measurement time. Hereby, a better signal-to-noise ratio can be obtained compared to inhouse SAXS systems.

Furthermore, the setup of such beamlines allows the coverage of a wider q-range, whereby especially measurements in the low q-region can be performed. Detailed scattering data in this region allows the analysis of a possible structure factor contribution which might be important when analysing samples with higher protein concentration.

A coupling with a size exclusion chromatography system (SEC-SAXS) would allow the investigation of freshly purified protein. This setup has the potential of preventing the protein from aggregation during storage time.

All in all you can gather very sound information about 3H3 in its native state, which allows a more detailed structure determination using a rigid body approach or even using *ab-initio* shape determination methods.

A completely different method for determining the structure of proteins is the nuclear magnetic resonance (NMR) spectroscopy [146, 147]. In recent years the potential of this method has evolved from the application in structure determination of small molecules

to also include biological macromolecules. NMR spectroscopy also allows the investigation of proteins in solution. The measurement of the spin relaxation rates allows the determination of the rotational correlation time which correlated with the molecular weight of the investigated species [148–152].

Moreover, the use of heteronuclear more-dimensional NMR spectroscopy offers the possibility of a complete structure determination on the atomic level. Applying this method to proteins requires the expression of the desired 3H3 in ^{15}N containing medium to perform a ^{15}N isotopic labelling of the structure. In addition the desired molecular weight in the range of 55 kDa will push this method to its limits.

Another sophisticated, but rarely available, method for the determination of the spin relaxation rate is depolarised dynamic light scattering (DDLS) [153–155]. Using this method, the same prerequisites of sample preparation as for dynamic light scattering experiments are needed.

8 List of Publications

Patchy worm-like micelles: solution structure studied by small-angle neutron scattering

Sabine Rosenfeldt, **Frank Lüdel**, Christoph Schulreich, Thomas Hellweg, Aurel Radulescu, Joachim Schmelz, Holger Schmalz, and Ludger Harnau. *Phys. Chem. Chem. Phys.*, 14, 12750–12756, 2012. DOI: 10.1039/c2cp41231d

Extracts of this thesis are to be published in the following article:

Distinguishing between monomeric scFv and diabody using light and small angle x-ray scattering

Frank Lüdel, Sandra Bufe, Willem M. Bleymüller, Hugo de Jonge, Luisa Iamele, Hartmut H. Niemann, and Thomas Hellweg. *to be submitted.*

9 Bibliography

[1] Chaudhuri, B. N., 2015. Emerging applications of small angle solution scattering in structural biology. *Protein Sci*, **24**, 267.

[2] Hura, G. L., A. L. Menon, M. Hammel, et al., 2009. Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods*, **6**, 606.

[3] Grant, T. D., J. R. Luft, J. R. Wolfley, et al., 2011. Small angle X-ray scattering as a complementary tool for high-throughput structural studies. *Biopolymers*, **95**, 517.

[4] Sander, B., G. Tria, A. V. Shkumatov, et al., 2013. Structural characterization of gephyrin by AFM and SAXS reveals a mixture of compact and extended states. *Acta crystallographica section d-biological crystallography*, **69**, 2050.

[5] Perry, J. J. P. and J. A. Tainer, 2013. Developing advanced X-ray scattering methods combined with crystallography and computation. *Methods*, **59**, 363.

[6] Koch, C., G. Tria, A. J. Fielding, et al., 2013. A structural model of PpoA derived from SAXS-analysis-Implications for substrate conversion. *Biochimica et biophysica acta-molecular and cell biology of lipids*, **1831**, 1449.

[7] Liu, X., M. Hammel, Y. He, et al., 2013. Structural insights into the interaction of IL-33 with its receptors. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, 14918.

[8] Shtykova, E. V., L. A. Baratova, N. V. Fedorova, et al., 2013. Structural Analysis of Influenza A Virus Matrix Protein M1 and Its Self-Assemblies at Low pH. *PLoS One*, **8**.

[9] Otrelo-Cardoso, A. R., M. A. da Silva Correia, V. Schwuchow, et al., 2014. Structural Data on the Periplasmic Aldehyde

Oxidoreductase PaoABC from Escherichia coli: SAXS and Preliminary X-ray Crystallography Analysis. *Int. J. Mol. Sci.*, **15**, 2223.

[10] Skerlova, J., V. Kral, M. Kachala, et al., 2015. Molecular mechanism for the action of the anti-CD44 monoclonal antibody MEM-85. *J Struct Biol*, **191**, 214.

[11] Kachala, M., E. Valentini, and D. I. Svergun, 2015. Application of SAXS for the Structural Characterization of IDPs. In Felli, IC and Pierattelli, R, ed., *INTRINSICALLY DISORDERED PROTEINS STUDIED BY NMR SPECTROSCOPY*, vol. 870 of *Advances in Experimental Medicine and Biology*, 261–289.

[12] Sauter, A., F. Zhang, N. K. Szekely, et al., 2016. Structural Evolution of Metastable Protein Aggregates in the Presence of Trivalent Salt Studied by (V)SANS and SAXS. *J. Phys. Chem. B*, **120**, 5564.

[13] Frka-Petesic, B., D. Zanchi, N. Martin, et al., 2016. Aggregation of Antibody Drug Conjugates at Room Temperature: SAXS and Light Scattering Evidence for Colloidal Instability of a Specific Subpopulation. *Langmuir*, **32**, 4848.

[14] Dadinova, L. A., E. V. Shtykova, P. V. Konarev, et al., 2016. X-Ray Solution Scattering Study of Four Escherichia coli Enzymes Involved in Stationary-Phase Metabolism. *PLoS One*, **11**.

[15] Rubio-Marrero, E. N., G. Vincelli, C. M. Jeffries, et al., 2016. Structural Characterization of the Extracellular Domain of CASPR2 and Insights into Its Association with the Novel Ligand Contactin1. *J. Biol. Chem.*, **291**, 5788.

[16] Drebes, J., M. Kuenz, B. Windshuegel, et al., 2016. Structure of ThiM from Vitamin B1 biosynthetic pathway of Staphylococcus aureus - Insights into a novel pro-drug approach addressing MRSA infections. *Sci. Rep.*, **6**.

[17] Greving, I., C. Dicko, A. Terry, et al., 2010. Small angle neutron scattering of native and reconstituted silk fibroin. *Soft Matter*, **6**, 4389.

[18] Clarke, M. J., J. B. Artero, M. Moulin, et al., 2010. Investigation of gamma E-crystallin target protein binding to bovine lens alpha-crystallin by small-angle neutron scattering. *Biochimica et biophysica acta-general subjects*, **1800**, 392.

[19] Abramo, M. C., C. Caccamo, M. Calvo, et al., 2011. Molecular dynamics and small-angle neutron scattering of lysozyme aqueous solutions. *Philos. Mag.*, **91**, 2066.

[20] Spinozzi, F., P. Mariani, I. Micetic, et al., 2012. Quaternary Structure Heterogeneity of Oligomeric Proteins: A SAXS and SANS Study of the Dissociation Products of Octopus vulgaris Hemocyanin. *PLoS One*, **7**.

[21] Hennig, J., I. Wang, M. Sonntag, et al., 2013. Combining NMR and small angle X-ray and neutron scattering in the structural analysis of a ternary protein-RNA complex. *J. Biomol. NMR*, **56**, 17.

[22] Göbl, C., T. Madl, B. Simon, et al., 2014. NMR approaches for structural analysis of multidomain proteins and complexes in solution. *Prog. Nucl. Magn. Reson. Spectrosc.*, **80**, 26.

[23] Shtykova, E. V., L. A. Baratova, N. V. Fedorova, et al., 2013. Structural Analysis of Influenza A Virus Matrix Protein M1 and Its Self-Assemblies at Low pH. *PLoS ONE*, **8**, e82431.

[24] Dadinova, L. A., E. V. Rodina, N. N. Vorobyeva, et al., 2016. Structural investigations of E. Coli dihydrolipoamide dehydrogenase in solution: Small-angle X-ray scattering and molecular docking. *Crystallography Reports*, **61**, 414.

[25] Fischer, S., C. Hartl, K. Frank, et al., 2016. Shape and Interhelical Spacing of DNA Origami Nanostructures Studied by Small-Angle X-ray Scattering. *Nano Letters*, **16**, 4282.

[26] Ce Shi, Y.-q. L., 2015. Progress on the Application of Small-angle X-ray Scattering in the Study of Protein and Protein Complexes. *Acta polymerica sinica*, 871.

[27] Svergun, D. I., M. H. J. Koch, P. A. Timmins, et al., 2013. *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules*. Oxford University Press (OUP).

[28] Tsukamoto, S., T. Yamashita, Y. Yamada, et al., 2009. Non-native α-helix formation is not necessary for folding of lipocalin: Comparison of burst-phase folding between tear lipocalin and β-lactoglobulin. *Proteins: Structure, Function, and Bioinformatics*, **76**, 226.

[29] Matsumura, Y., M. Shinjo, A. Mahajan, et al., 2010. α-Helical burst on the folding pathway of FHA domains from Rad53 and Ki67. *Biochimie*, **92**, 1031.

[30] Ortore, M. G., F. Spinozzi, S. Vilasi, et al., 2011. Time-resolved small-angle x-ray scattering study of the early stage of amyloid formation of an apomyoglobin mutant. *Physical Review E*, **84**.

[31] Spinozzi, F., E. Maccioni, C. V. Teixeira, et al., 2003. Synchrotron SAXS Studies on the Structural Stability of Carcinus aestuarii Hemocyanin in Solution. *Biophys. J.*, **85**, 2661.

[32] Saio, T., X. Guan, P. Rossi, et al., 2014. Structural Basis for Protein Antiaggregation Activity of the Trigger Factor Chaperone. *Science*, **344**, 1250494.

[33] Simon, A. C., J. C. Zhou, R. L. Perera, et al., 2014. A Ctf4 trimer couples the CMG helicase to DNA polymerase α in the eukaryotic replisome. *Nature*, **510**, 293.

[34] Peisley, A., B. Wu, H. Xu, et al., 2014. Structural basis for ubiquitin-mediated antiviral signal activation by RIG-I. *Nature*, **509**, 110.

[35] Hanlon, A. D., M. I. Larkin, and R. M. Reddick, 2010. Free-Solution, Label-Free Protein-Protein Interactions Characterized by Dynamic Light Scattering. *Biophysical Journal*, **98**, 297.

[36] Lewis, E., W. Qi, L. Kidder, et al., 2014. Combined Dynamic Light Scattering and Raman Spectroscopy Approach for Characterizing the Aggregation of Therapeutic Proteins. *Molecules*, **19**, 20888.

[37] Wu, D. and A. P. Minton, 2013. Quantitative Characterization of the Compensating Effects of Trimethylamine- N -oxide and Guanidine Hydrochloride on the Dissociation of Human

Cyanmethmoglobin. *The Journal of Physical Chemistry B*, **117**, 9395.

[38] Murphy, K. P., C. A. Janeway, P. Travers, et al., 2011. *Janeway's immunobiology*. Garland Science, London, 8. ed.

[39] Lehninger, A. L., D. L. Nelson, and M. M. Cox, 2008. *Lehninger principles of biochemistry*. W.H. Freeman, New York, 5. ed.

[40] Silverton, E. W., M. A. Navia, and D. R. Davies, 1977. Three-dimensional structure of an intact human immunoglobulin. *Proc. Natl. Acad. Sci. U. S. A.*, **74**, 5140.

[41] Plückthun, A. and P. Pack, 1997. New protein engineering approaches to multivalent and bispecific antibody fragments. *Immunotechnology*, **3**, 83.

[42] Hudson, P. J. and A. A. Kortt, 1999. High avidity scFv multimers; diabodies and triabodies. *J. Immunol. Methods*, **231**, 177.

[43] Edelman, G. M., 1973. Antibody Structure and Molecular Immunology. *Science*, **180**, 830.

[44] Porter, R. R., 1973. Structural Studies of Immunoglobulins. *Science*, **180**, 713.

[45] Davidson, G., E. Daniels, H. Nunan, et al., 1989. Passive immunisation of children with bovine colostrum containing antibodies to human rotavirus. *The Lancet*, **334**, 709.

[46] Schaffer, A. C. and J. C. Lee, 2008. Vaccination and passive immunisation against Staphylococcus aureus. *Int. J. Antimicrob. Agents*, **32**, S71.

[47] Weiner, H. L. and D. A. Hafler, 1988. Immunotherapy of multiple sclerosis. *Ann. Neurol.*, **23**, 211.

[48] Meuth, S., S. Bittner, and H. Wiendl, 2009. Immunotherapy of multiple sclerosis. *Acta neuropsychiatrica*, **21**, 27.

[49] Molina, M. A., J. Codony-Servat, J. Albanell, et al., 2001. Trastuzumab (herceptin), a humanized anti-Her2 receptor monoclonal antibody, inhibits basal and activated Her2 ectodomain cleavage in breast cancer cells. *Cancer Res.*, **61**, 4744.

[50] Goldenberg, M. M., 1999. Trastuzumab, a recombinant DNA-derived humanized monoclonal antibody, a novel agent for the treatment of metastatic breast cancer. *Clin. Ther.*, **21**, 309.

[51] Hortobagyi, G. N., 2001. Overview of treatment results with trastuzumab (Herceptin) in metastatic breast cancer. In *Seminars in oncology*, vol. 28, 43–47. Elsevier.

[52] Baum, R., M. Lorenz, R. Senekowitsch, et al., 1987. Klinische Ergebnisse der Immunszintigraphie und Radioimmuntherapie. *Nuklearmedizin*, **26**, 68.

[53] Hudson, P. J., 1998. Recombinant antibody fragments. *Curr. Opin. Biotechnol.*, **9**, 395.

[54] Hudson, P. J., 1999. Recombinant antibody constructs in cancer therapy. *Curr. Opin. Immunol.*, **11**, 548.

[55] Le Gall, F., S. M. Kipriyanov, G. Moldenhauer, et al., 1999. Di-, tri- and tetrameric single chain Fv antibody fragments against human CD19: effect of valency on cell binding. *FEBS Lett.*, **453**, 164.

[56] Della Cristina, P., M. Castagna, A. Lombardi, et al., 2015. Systematic comparison of single-chain Fv antibody-fusion toxin constructs containing Pseudomonas Exotoxin A or saporin produced in different microbial expression systems. *Microb. Cell Fact.*, **14**, 19.

[57] Dall'Acqua, W. and P. Carter, 1998. Antibody engineering. *Curr. Opin. Struct. Biol.*, **8**, 443.

[58] Plückthun, A. and P. Pack, 1997. New protein engineering approaches to multivalent and bispecific antibody fragments. *Immunotechnology*, **3**, 83.

[59] Nelson, A. L., 2010. Antibody fragments: hope and hype. *mAbs*, **2**, 77.

[60] Herrington-Symes, A. P., M. Farys, H. Khalili, et al., 2013. Antibody fragments: Prolonging circulation half-life special issue-antibody research. *Advances in Bioscience and Biotechnology*, **04**, 689.

[61] Hudson, P. J. and C. Souriau, 2003. Engineered antibodies. *Nat. Med.*, **9**, 129.

[62] Bird, R. E., K. D. Hardman, J. W. Jacobson, et al., 1988. Single-chain antigen-binding proteins. *Science*, **242**, 423.

[63] Huston, J. S., D. Levinson, M. Mudgett-Hunter, et al., 1988. Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.*, **85**, 5879.

[64] Rodrigo, G., M. Gruvegård, and J. V. Alstine, 2015. Antibody Fragments and Their Purification by Protein L Affinity Chromatography. *Antibodies*, **4**, 259.

[65] Turner, D. J., M. A. Ritter, and A. J. George, 1997. Importance of the linker in expression of single-chain Fv antibody fragments: optimisation of peptide sequence using phage display technology. *J. Immunol. Methods*, **205**, 43.

[66] Perisic, O., P. A. Webb, P. Holliger, et al., 1994. Crystal structure of a diabody, a bivalent antibody fragment. *Structure*, **2**, 1217.

[67] Atwell, J. L., L. A. Pearce, M. Lah, et al., 1996. Design and expression of a stable bispecific scFv dimer with affinity for both glycophorin and N9 neuraminidase. *Mol. Immunol.*, **33**, 1301.

[68] Holliger, P., T. Prospero, and G. Winter, 1993. " Diabodies": small bivalent and bispecific antibody fragments. *Proc. Natl. Acad. Sci. U. S. A.*, **90**, 6444.

[69] Arndt, K. M., K. M. Müller, and A. Plückthun, 1998. Factors influencing the dimer to monomer transition of an antibody single-chain Fv fragment. *Biochemistry (Mosc.)*, **37**, 12918.

[70] Kortt, A. A., M. Lah, G. W. Oddie, et al., 1997. Single-chain Fv fragments of anti-neuraminidase antibody NC10 containing five-and ten-residue linkers form dimers and with zero-residue linker a trimer. *Protein Eng.*, **10**, 423.

[71] Atwell, J. L., K. A. Breheney, L. J. Lawrence, et al., 1999. scFv multimers of the anti-neuraminidase antibody NC10: length of the linker between VH and VL domains dictates precisely the transition between diabodies and triabodies. *Protein Eng.*, **12**, 597.

[72] Berman, H. M., 2000. The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235.

[73] Root, A., W. Cao, B. Li, et al., 2016. Development of PF-06671008, a Highly Potent Anti-P-cadherin/Anti-CD3 Bispecific DART Molecule with Extended Half-Life for the Treatment of Cancer. *Antibodies*, **5**, 6.

[74] Carmichael, J. A., B. E. Power, T. P. Garrett, et al., 2003. The crystal structure of an anti-CEA scFv diabody assembled from T84. 66 scFvs in V L-to-V H orientation: implications for diabody flexibility. *J. Mol. Biol.*, **326**, 341.

[75] Moraga, I., G. Wernig, S. Wilmes, et al., 2015. Tuning Cytokine Receptor Signaling by Re-orienting Dimer Geometry with Surrogate Ligands. *Cell*, **160**, 1196.

[76] Weiskopf, K., N. S. Jahchan, P. J. Schnorr, et al., 2016. CD47-blocking immunotherapies stimulate macrophage-mediated destruction of small-cell lung cancer. *J. Clin. Invest.*, **126**, 2610.

[77] Banatao, D. R., D. Cascio, C. S. Crowley, et al., 2006. An approach to crystallizing proteins by synthetic symmetrization. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 16230.

[78] Birchmeier, C., W. Birchmeier, E. Gherardi, et al., 2003. Met, metastasis, motility and more. *Nat. Rev. Mol. Cell Biol.*, **4**, 915.

[79] Bukowska, M. A. and M. G. Grütter, 2013. New concepts and aids to facilitate crystallization. *Curr. Opin. Struct. Biol.*, **23**, 409.

[80] Niemann, H. H., 2013. Structural basis of MET receptor dimerization by the bacterial invasion protein InlB and the HGF/SF splice variant NK1. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, **1834**, 2195.

[81] Bufe, S., 2014. *Charakterisierung und Kristallisation eines agonistischen anti-MET Diabody*. Master's thesis, Bielefeld University, Bielefeld.

[82] Büttner, C. R., I. Sorg, G. R. Cornelis, et al., 2008. Structure of the Yersinia enterocolitica Type III Secretion Translocator Chaperone SycD. *J. Mol. Biol.*, **375**, 997.

[83] Svergun, D., C. Barberato, and M. Koch, 1995. CRYSOL-a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.*, **28**, 768.

[84] Stuhrmann, H. B., 1970. Interpretation of small-angle scattering functions of dilute solutions and gases. A representation of the structures related to a one-particle scattering function. *Acta Cryst Sect A*, **26**, 297.

[85] Stuhrmann, H. B., 1970. Ein neues Verfahren zur Bestimmung der Oberflächenform und der inneren Struktur von gelösten globulären Proteinen aus Röntgenkleinwinkelmessungen. *Z. Phys. Chem.*, **72**, 177.

[86] Lattman, E. E., 1989. Rapid calculation of the solution scattering profile from a macromolecule of known structure. *Proteins: Struct., Funct., Genet.*, **5**, 149.

[87] Lottspeich, F., ed., 2012. *Bioanalytik*. Springer Spektrum, Berlin, 3rd ed.

[88] Petoukhov, M. V., D. Franke, A. V. Shkumatov, et al., 2012. New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.*, **45**, 342.

[89] Petoukhov, M. V. and D. I. Svergun, 2005. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.*, **89**, 1237.

[90] Svergun, D., 1997. Restoring three-dimensional structure of biopolymers from solution scattering. *J. Appl. Crystallogr.*, **30**, 792.

[91] Bates, P. A., L. A. Kelley, R. M. MacCallum, et al., 2001. Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins: Struct., Funct., Bioinf.*, **45**, 39.

[92] Contreras-Moreira, B. and P. A. Bates, 2002. Domain fishing: a first step in protein comparative modelling. *Bioinformatics*, **18**, 1141.

[93] Schrödinger, LLC, 2015. The PyMOL Molecular Graphics System, Version 1.8.

[94] Ortega, A., D. Amorós, and J. García de la Torre, 2011. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic-and residue-level models. *Biophys. J.*, **101**, 892.

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

[96] Bloomfield, V. A. and D. P. Filson, 1968. Shell model calculations of translational and rotational frictional coefficients. *J Polym Sci Pol Sym*, **25**, 73.

[97] García de la Torre, J., M. L. Huertas, and B. Carrasco, 2000. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.*, **78**, 719.

[98] García de la Torre, J., G. del Rio Echenique, and A. Ortega, 2007. Improved Calculation of Rotational Diffusion and Intrinsic Viscosity of Bead Models for Macromolecules and Nanoparticles. *J. Phys. Chem. B*, **111**, 955.

[99] Hoffmann, M., C. S. Wagner, L. Harnau, et al., 2009. 3D Brownian Diffusion of Submicron-Sized Particle Clusters. *ACS Nano*, **3**, 3326.

[100] Hellweg, T., W. Eimer, E. Krahn, et al., 1997. Hydrodynamic properties of nitrogenase — the MoFe protein from Azotobacter vinelandii studied by dynamic light scattering and hydrodynamic modelling. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, **1337**, 311.

[101] Lindner, P. and T. Zemb, eds., 2002. *Neutrons, x-rays and light. scattering methods applied to soft condensed matter*. North-Holland delta series. Elsevier, Amsterdam, 1st. ed.

[102] Brumberger, H., ed., 2013. *Modern Aspects of Small-Angle Scattering*. Springer Netherlands.

[103] Perret, R. and W. Ruland, 1972. Glassy carbon as standard for the normalization of small-angle scattering intensities. *J Appl Cryst*, **5**, 116.

[104] Fan, L., M. Degen, S. Bendle, et al., 2010. The Absolute Calibration of a Small-Angle Scattering Instrument with a Laboratory X-ray Source. *J. Phys.: Conf. Ser.*, **247**, 012005.

[105] Guinier, A., 1939. La diffraction des rayons X aux très petits angles: Application a l'étude de phénomènes ultramicroscopiques. *Annales de Physique (Paris)*, **12**, 161–237. Out of print.

[106] Guinier, A. and G. Fournet, 1955. *Small angle scattering of X-rays*. J. Wiley & Sons, New York.

[107] Glatter, O., 1977. A new method for the evaluation of small-angle scattering data. *J Appl Cryst*, **10**, 415.

[108] Konarev, P. V. and D. I. Svergun, 2015. A posteriori determination of the useful data range for small-angle scattering experiments on dilute monodisperse systems. *IUCrJ*, **2**, 352.

[109] Shannon, C. E. and W. Weaver, 1949. *The mathematical theory of communication*. University of Illinois press.

[110] Marks, R., 1991. *Introduction to Shannon sampling and interpolation theory*. Springer Science & Business Media.

[111] Moore, P. B., 1980. Small-angle scattering. Information content and error analysis. *J. Appl. Crystallogr.*, **13**, 168.

[112] Jacques, D. A., J. M. Guss, D. I. Svergun, et al., 2012. Publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution. *Acta Cryst D*, **68**, 620.

[113] Orthaber, D., A. Bergmann, and O. Glatter, 2000. SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. *J. Appl. Crystallogr.*, **33**, 218.

[114] Mylonas, E. and D. I. Svergun, 2007. Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *J. Appl. Crystallogr.*, **40**, 245.

[115] Petoukhov, M. V., D. I. Svergun, P. V. Konarev, et al., 2003. Quaternary Structure of Azospirillum brasilense NADPH-dependent Glutamate Synthase in Solution as Revealed by Synchrotron Radiation X-ray Scattering. *J. Biol. Chem.*, **278**, 29933.

[116] Fischer, H., I. Polikarpov, and A. F. Craievich, 2004. Average protein density is a molecular-weight-dependent function. *Protein Sci*, **13**, 2825.

[117] Fischer, H., M. de Oliveira Neto, H. B. Napolitano, et al., 2009. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J Appl Cryst*, **43**, 101.

[118] Svergun, D., 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.*, **25**, 495.

[119] Berne, B. J. and R. Pecora, 1976. *Dynamic light scattering: with applications to chemistry, biology, and physics*. Courier Corporation.

[120] Chu, B., 1974. *Laser Light Scattering*. Quantum Electronics. Academic Press, New York.

[121] Schärtl, W., 2007. *Light Scattering from Polymer Solutions and Nanoparticle Dispersions*. Springer-Verlag GmbH, Berlin, Heidelberg, New York.

[122] Provencher, S. W., 1982. CONTIN: a general purpose constrained regularization program for inverting noisy linear algebraic and integral equations. *Comput. Phys. Commun.*, **27**, 229.

[123] Einstein, A., 1905. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann. Phys.*, **322**, 549.

[124] Yadav, S., T. M. Scherer, S. J. Shire, et al., 2011. Use of dynamic light scattering to determine second virial coefficient in a semidilute concentration regime. *Anal. Biochem.*, **411**, 292.

[125] Zeiser, M., 2013. *Designing microgels: where particle-architecture and thermo-response meet*. Ph.D. thesis, Bielefeld University, Bielefeld.

[126] Rajagopalan, R. and P. C. Hiemenz, 1997. *Principles of colloid and surface chemistry*. Marcel Dekker, New York, 3rd. ed.

[127] Zimm, B. H., 1948. The Scattering of Light and the Radial Distribution Function of High Polymer Solutions. *J. Chem. Phys*, **16**, 1093.

[128] Skoog, D. A., F. J. Holler, and S. R. Crouch, 2013. *Instrumentelle Analytik: Grundlagen-Geräte-Anwendungen*. Springer Spektrum, Berlin.

[129] Weber, K., J. Pringle, and M. Osborn, 1972. Measurement of molecular weights by electrophoresis on SDS-acrylamide gel. In *Methods in Enzymology*, vol. 26, 3–27. Elsevier BV.

[130] Roland Winter, C. C., Frank Noll, 2011. *Methoden der Biophysikalischen Chemie*. Vieweg+Teubner Verlag, Wiesbaden.

[131] Konarev, P., M. Petoukhov, and D. Svergun, 2001. MASSHA-a graphics system for rigid-body modelling of macromolecular complexes against solution scattering data. *J. Appl. Crystallogr.*, **34**, 527.

[132] Konarev, P. V., M. V. Petoukhov, and D. I. Svergun, 2001. *MASSHA* – a graphics system for rigid-body modelling of macromolecular complexes against solution scattering data. *J. Appl. Crystallogr.*, **34**, 527.

[133] García de la Torre, J. and V. A. Bloomfield, 1977. Hydrodynamics of macromolecular complexes. III. Bacterial viruses. *Biopolymers*, **16**, 1779.

[134] García de la Torre, J. and V. A. Bloomfield, 1981. Hydrodynamic properties of complex, rigid, biological macromolecules: theory and applications. *Q. Rev. Biophys.*, **14**, 81.

[135] Czuryło, E. A., T. Hellweg, W. Eimer, et al., 1997. The size and shape of caldesmon and its fragments in solution studied by dynamic light scattering and hydrodynamic model calculations. *Biophys. J.*, **72**, 835.

[136] Glatter, O. and O. Kratky, eds., 1982. *Small Angle X-ray Scattering*. Academic Press, London.

[137] Kratky, O. and G. Porod, 1949. Röntgenuntersuchung gelöster Fadenmoleküle. *Recl. Trav. Chim. Pays-Bas*, **68**, 1106.

[138] Akiyama, S., 2010. Quality control of protein standards for molecular mass determinations by small-angle X-ray scattering. *J Appl Cryst*, **43**, 237.

[139] Midelfort, K., H. Hernandez, S. Lippow, et al., 2004. Substantial Energetic Improvement with Minimal Structural Perturbation in a High Affinity Mutant Antibody. *J. Mol. Biol.*, **343**, 685.

[140] Pei, X. Y., P. Holliger, A. G. Murzin, et al., 1997. The 2.0-Å resolution crystal structure of a trimeric antibody fragment with noncognate VH–VL domain pairs shows a rearrangement of VH CDR3. *Proc. Natl. Acad. Sci. U. S. A.*, **94**, 9637.

[141] Contreras-Moreira, B., P. W. Fitzjohn, M. Offman, et al., 2003. Novel use of a genetic algorithm for protein structure prediction: Searching template and sequence alignment space. *Proteins: Struct., Funct., Genet.*, **53**, 424.

[142] Soding, J., 2004. Protein homology detection by HMM-HMM comparison. *Bioinformatics*, **21**, 951.

[143] Biasini, M., S. Bienert, A. Waterhouse, et al., 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.*, **42**, W252.

[144] Burchard, W. and W. Richtering, 1989. Dynamic light scattering from polymer solutions. In *Progress in Colloid and Polymer Science*, 151–163. Springer Science and Business Media.

[145] Carrasco, B. and J. García de la Torre, 1999. Hydrodynamic properties of rigid particles: comparison of different modeling and computational procedures. *Biophys. J.*, **76**, 3044.

[146] Wider, G., 1998. Technical aspects of NMR Spectroscopy with biological macromolecules and studies of hydration in solution. *Prog. Nucl. Magn. Reson. Spectrosc.*, **32**, 193.

[147] Pervushin, K., R. Riek, G. Wider, et al., 1997. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. U. S. A.*, **94**, 12366.

[148] Daragan, V. A. and K. H. Mayo, 1997. Motional model analyses of protein and peptide dynamics using 13C and 15N NMR relaxation. *Prog. Nucl. Magn. Reson. Spectrosc.*, **31**, 63.

[149] Cavanagh, J., W. J. Fairbrother, A. G. Palmer, et al., 2007. *Protein NMR Spectroscopy*. Elsevier LTD, Oxford, 2nd ed.

[150] Lee, D., C. Hilty, G. Wider, et al., 2006. Effective rotational correlation times of proteins from NMR relaxation interference. *J. Magn. Reson.*, **178**, 72.

[151] Brüschweiler, R., 2003. New approaches to the dynamic interpretation and prediction of NMR relaxation data from proteins. *Curr. Opin. Struct. Biol.*, **13**, 175.

[152] de la Torre, J. G., M. Huertas, and B. Carrasco, 2000. HYDRONMR: Prediction of NMR Relaxation of Globular Proteins from Atomic-Level Structures and Hydrodynamic Calculations. *J. Magn. Reson.*, **147**, 138.

[153] Droegemeier, J. and W. Eimer, 1994. Polarized and depolarized dynamic light scattering study on F-actin in solution: comparison with model calculations. *Macromolecules*, **27**, 96.

[154] Patkowski, A., W. Steffen, H. Nilgens, et al., 1997. Depolarized dynamic light scattering from three low molecular weight glass forming liquids: A test of the scattering mechanism. *J. Chem. Phys*, **106**, 8401.

[155] Bolten, M., M. Niermann, and W. Eimer, 1999. Structural Analysis of G-DNA in Solution: A Combination of Polarized and Depolarized Dynamic Light Scattering with Hydrodynamic Model Calculations. *Biochemistry (Mosc.)*, **38**, 12416.

[156] Konarev, P. V., V. V. Volkov, A. V. Sokolova, et al., 2003. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.*, **36**, 1277.

[157] Whitten, A., S. Cai, and J. Trewhella, 2008. MULCh: modules for the analysis of small-angle neutron contrast variation data from biomolecular assemblies. *J. Appl. Crystallogr.*, **41**, 222.

A Appendix

A.1 | Used software

ATSAS package [88] with:	
PRIMUSqt	Basic 1D SAXS data treatment [156]
CRYSOL	Calculation of scattering curves from PDB structure files and fitting to experimental data [83]
GNOM	Calculation of the particle distance distribution function [118]
SASREF	Structure modeling using known subunits against solution scattering data [89]
SHANUM	Determining the useful experimental data range of scattering data sets [108]

Online-Tools:	
SWISS MODEL	Protein structure homology modelling [143] - http://swissmodel.expasy.org/
Protparam	Calculation of different parameters using a known sequence - http://web.expasy.org/protparam/
MULCh	Calculation of the scattering contrast and the partial specific volume based on the sequence [157] - http://smb-research.smb.usyd.edu.au/NCVWeb/
3D-JIGSAW 2.0	Structure prediction based on homologous structures and a sequence [141, 142] - http://bmm.cancerresearchuk.org/~3djigsaw/; https://bmm.crick.ac.uk/ populus/
SAXS MoW	Molecular weight determination using only one scattering curve [117] - http://www.if.sc.usp.br/~saxs

ESRF SAXS - Programms	Initial SAXS data treatment - http://www.esrf.eu/Instrumentation/software/data-analysis/OurSoftware/SAXS

CONTIN	Initial PCS data treatment [122]
HYDROPRO	Calculation of hydrodynamic properties of proteins using the PDB file [97] - http://leonardo.inf.um.es/macromol/programs/hydropro/hydropro.htm
COOT	PDB file viewer and modification tool
PyMOL	PDB file viewer
SwissPDB	PDB file viewer and modification tool

A.2 | Chemical crosslinking

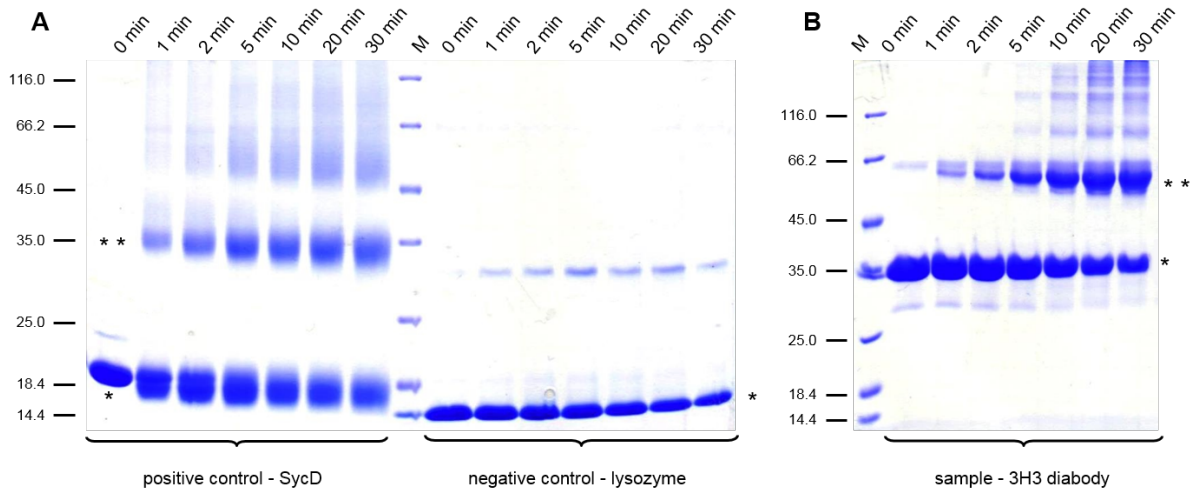

Figure 40: Analysis of chemical crosslinking with glutaraldehyde [81].

100 µl of protein solution (1 mg ml^{-1}) was mixed with 5 µl 2.5 % glutaraldehyde. Samples were taken at the time stated in the figure, the reaction was stopped by adding 1 M tris pH 8 and the sample was analysed by SDS-PAGE (12 % gels). 10 µg of protein were loaded per lane and the gels were stained with Coomassie Brilliant Blue solution. Part A above shows the positive control SycD21-163 (previously shown to form dimers, [82]), the negative control lysozyme in 1 x PBS pH 7.4. The positive control shows a band in the region of the dimer (**) that increased with time while that of the monomer is decreasing in intensity. The band in the region of the dimer that can be seen in the negative control can be neglected due to its comparably low and constant intensity. Part B above shows the 3H3 diabody purified by SEC in 20 mM MES with 20 mM NaCl pH 6.5. Due to the crosslinking a band at the size of the dimer can be seen comparable to the positive control [81].

A.3 | Sequence and generated sequence parts of 1LMK

```
>File: 1LMK-A1B2
  1: VQLQQSGTEL MKPGRSLKIS CKTTGYIFSN YWIEWVKQRP GHGLEWIGKI
 51: LPGGGSNTYN DKFKGKATFT ADTSSNIAYM QLSSLTSEDS AVYYCARGED
101: YYAYWYVLDY WGQGTTVTVS SGGGGSDIEL TQSPLSLPVS LGDQASISCR
151: SSQSLVHSNG NTSLHWYLKK PGQSPKLLIY KVSTRFSGVP DRFSGSGSGT
201: DFTLKISRVE AEDLGVYFCS QSTHVPFTFG SGTKLELK

>File: 1LMK-C1-D2
  1: VQLQQSGTEL MKPGRSLKIS CKTTGYIFSN YWIEWVKQRP GHGLEWIGKI
 51: LPGGGSNTYN DKFKGKATFT ADTSSNIAYM QLSSLTSEDS AVYYCARGED
101: YYAYWYVLDY WGQGTTVTVS SGGGGSDIEL TQSPLSLPVS LGDQASISCR
151: SSQSLVHSNG NTSLHWYLKK PGQSPKLLIY KVSTRFSGVP DRFSGSGSGT
201: DFTLKISRVE AEDLGVYFCS QSTHVPFTFG SGTKLELK

>File: dia-1lmk.bearbeitet
  1: VQLQQSGTEL MKPGRSLKIS CKTTGYIFSN YWIEWVKQRP GHGLEWIGKI
 51: LPGGGSNTYN DKFKGKATFT ADTSSNIAYM QLSSLTSEDS AVYYCARGED
101: YYAYWYVLDY WGQGTTVTVS SGGGGSDIEL TQSPLSLPVS LGDQASISCR
151: SSQSLVHSNG NTSLHWYLKK PGQSPKLLIY KVSTRFSGVP DRFSGSGSGT
201: DFTLKISRVE AEDLGVYFCS QSTHVPFTFG SGTKLELKVQ LQQSGTELMK
251: PGRSLKISCK TTGYIFSNYW IEWVKQRPGH GLEWIGKILP GGGSNTYNDK
301: FKGKATFTAD TSSNIAYMQL SSLTSEDSAV YYCARGEDYY AYWYVLDYWG
351: QGTTVTVSSGGGGSDIELTQ SPLSLPVSLG DQASISCRSS QSLVHSNGNT
401: SLHWYLKKPG QSPKLLIYKV STRFSGVPDR FSGSGSGTDF TLKISRVEAE
451: DLGVYFCSQS THVPFTFGSG TKLELK
```

A.4 | Calculation of the hydroynamic properties of the protein after structure optimisation

Protein structure	1LMK	1LMK	experiment
optimised against dataset	-	5 mg ml^{-1}	-
R_{element} / Å	5.9	5.9	-
Type of calculation	1	1	-
T / °C	20	20	20
η / mPa s	1.016	1.016	1.016
M / kDa	53.911	53.911	-
$\overline{\nu}$ / cm g^{-3}	0.731	0.731	-
ρ / g cm^{-1}	1.00677	1.00677	-
$D^{o,\,\text{theo}}$ / 10^{-11} m^2 s^{-1}	5.94	5.83	D^o: 6.295
$R_{\text{g, theo}}$ / nm	2.9	3.0	-
$R_{\text{h,o, theo}}$ / nm	3.6	3.6	$R_{\text{h,o}}$: 3.4
$R_{\text{g, theo}}/R_{\text{h, theo}}$	0.82	0.82	-

Table 16: Comparison of the calculated hydrodynamic properties of a diabody structure (1LMK) and one possible structure optimised diabody structure (1LMK, 5 mg ml^{-1}) with the data obtained in an experiment. R_{element} consists of 2.9 Å (recommended in HYDROPRO [94, 97]) for one element and 3 Å for the first hydration shell, the calculation is based on the: "atomic-level primary model, shell calculation (type 1)".

Table 17: Comparison of the calculated hydrodynamic properties of a diabody structure (1LMK) and three possible structure optimised diabody structures. The optimisation referred to the experimentally obtained datasets. R_{element} consists of 2.9 Å (recommended in HYDROPRO [94, 97]) for one element and 3 Å for the first hydration shell, the calculation is based on the: "atomic-level primary model, shell calculation (type 1)".

Protein structure	1LMK			
optimised against dataset	10 mg ml^{-1}	5 mg ml^{-1}	2.5 mg ml^{-1}	-
R_{element} / Å	5.9	5.9	5.9	5.9
Type of calculation	1	1	1	1
T / °C	20	20	20	20
η / mPa s	1.016	1.016	1.016	1.016
M / kDa	53.911	53.911	53.911	53.911
$\overline{\nu}$ / cm g^{-3}	0.731	0.731	0.731	0.731
ρ / g cm^{-1}	1.00677	1.00677	1.00677	1.00677
$D^{o,\,\text{theo}}$ / 10^{-11} m^2 s^{-1}	5.75	5.83	5.82	5.94
$R_{\text{g, theo}}$ / nm	3.1	3.0	3.0	2.9
$R_{\text{h,o, theo}}$ / nm	3.7	3.6	3.6	3.6
$R_{\text{g, theo}}/R_{\text{h, theo}}$	0.84	0.82	0.82	0.82

Table 18: Comparison of the calculated hydrodynamic properties of a diabody structure (1LMK) and three possible structure optimised diabody structures. The optimisation referred to the experimentally obtained datasets. $R_{element}$ consists of 2.9 Å (recommended in HYDROPRO [94, 97]) for one element and 3 Å for the first hydration shell, the calculation is based on the: "atomic-level primary model, shell calculation (type 1)".

Protein structure	3H3			1LMK
optimised against dataset	2.5 mg ml^{-1}	5 mg ml^{-1}	10 mg ml^{-1}	5 mg ml^{-1}
$R_{element}$ / Å	5.9	5.9	5.9	5.9
Type of calculation	1	1	1	1
T / °C	20	20	20	20
η / mPa s	1.016	1.016	1.016	1.016
M / kDa	53.911	53.911	53.911	52.086
$\overline{v}$ / cm g^{-3}	0.727	0.727	0.727	0.731
ρ / g cm^{-1}	1.00677	1.00677	1.00677	1.00677
$D^{o,\,theo}$ / 10^{-11} m^2 s^{-1}	5.88	5.83	5.77	5.83
$R_{g,\,theo}$ / nm	2.9	3.0	3.1	3.0
$R_{h,o,\,theo}$ / nm	3.6	3.6	3.7	3.6
$R_{g,\,theo}/R_{h,\,theo}$	0.81	0.82	0.83	0.82