Structural and biophysical investigations of insulin formulations for subcutaneous injection

Master Thesis

Written by *Anna Freja Hansen*

December 31, 2020

Supervised by

Professor Lise Arleth and Specialist Malin Margareth Zackrisson Oskolkova

University of Copenhagen

UNIVERSITY OF COPENHAGEN

Name of Institute:	The Niels Bohr Institute
Name of Department:	X-ray and Neutron Science
Author(s):	Anna Freja Hansen
Email:	fsj413@alumni.ku.dk
Title and subtitle:	Structural and biophysical investigations of insulin formulations for subcutaneous injection -
Supervisor(s):	Professor Lise Arleth and Specialist Malin Margareth Zackrisson Oskolkova
Handed in:	December 31st 2020
Defended:	January 28th 2021

Name _____

Signature _____

Date _____

Acknowledgements

First and foremost, I would like to acknowledge my supervisors Lise Arleth, Malin Margareth Zackrisson Oskolkova and Carolina Cragnell. Thank you for making this project possible and for all the help and guidance I have received during this project. Thank you for taking the time to discuss theory and results, as well as analysing the SAXS data. Furthermore, I would like to acknowledge all the help I have received from Ryan Oliver. Thank you for all the help with the samples, the SAXS instrument and most of all for believing and aiding in the CG-MALS coming back to life.

Secondly I would like to acknowledge all the help I have received from Saskia Bucciarelli. Thank you for testing and expanding my knowledge as well as proofreading parts of this report. I would also like to acknowledge Claurice Comfort, Ida Malm, Mette Villum Juulsgaard Mathiasen and Lone Hansen for making long days in the lab and office bearable. Thank you for helping me keeping up the good spirit during this project, also in stressful times. A thank you is owed to Agnete Braband Seitzberg for contributing her CG-MALS measurements to this work. Furthermore, I would like to acknowledge the Biophysical Characterization Team in department 205. Thank you for making me feel welcome and for helping me with anything from understanding equipment to showing me how to install software. I also owe a thank you to Hanne David Dondera for supplying chemicals. Furthermore, a thank you to Camilla E. H. Jørgensen is owed, for preparing the insulin product.

Lastly, I would like to acknowledge Novo Nordisk A/S for this opportunity and the possibility to use facilities and equipment.

Abstract

In this project, the role of temperature and preservatives on the protein-protein interactions and the stability of human insulin is investigated. This is done using small angle X-ray scattering and static light scattering. A thorough understanding of the active pharmaceutical ingredient in medicine is crucial for dosage determination. Knowledge of the phase behaviour sheds light on which conditions the medicine can withstand during formulation, storage and usage.

The full insulin drug product Actrapid, is used for investigations. The active pharmaceutical ingredient in Actrapid is human insulin. Part of the experimental work of this study, is storing Actrapid samples, with a variety of concentrations between 1.7 and 155 mg/mL, in temperatures between -10 and 60°C. At 60°C samples of concentrations above 20 mg/mL are found to reach a gel-like formation, which is irreversible. At 5°C samples of concentrations above 100 mg/mL are found to form a reversible birefringent sediment. Addition of salt to the drug product induces precipitation at or below room temperature, which in most cases is reversible with increasing temperature.

Temperature effects on insulin structure, stability as well as interactions are investigated at three different temperatures; 5°C, 25°C and 40°C. From the second virial coefficients, obtained by static light scattering, the repulsive nature of human insulin is found to increase with increasing temperatures, from 25°C to 40°C. Decreasing temperatures from 25°C to 5°C is found to have no affect on the structure the insulin hexamer. However, the decrease in temperature does induce repulsion as well as decrease the average particle distance.

Preservatives have an impact on the phase behaviour of proteins. Furthermore, preservatives affect the stability and conformational state of the insulin hexamer. Small angle X-ray scattering and light scattering experiments of actrapid formulated with and without preservative show changes in protein-protein interactions as well as changes in stability. Removing preservatives results in less repulsive interactions and slight aggregation in the drug product. There is no apparent change to the conformation of the hexamer.

List of Abbreviations and Symbols

α	Constant
α	Electronic polarisabilty
A	Absorbance
A	Hamaker constant
A_2	Second virial coefficient
A_3	Third virial coefficient
A_i	I'th virial coefficient
$A(q)$	Scattering amplitude
a	Radius
API	Active pharmaceutical ingredient
aRI	absolute refractive index
β	Instrument function
β	London dispersion constant
C	Mass per unit volume
c	Concentration
c_i	Partial free i-mer concentration
c_{total}	Total concentration
c_1	Partial free monomer concentration
CG-MALS	Composition-Gradient Multi-Angle Light Scattering
D_{max}	Maximal particle size
d	distance
DLS	Dynamic Light Scattering
dn/dc	Refractive index increment
dn/dw	Refractive index increment with respect to weight concentration
dRI	Differential refractive index
E	Electric field
E	Energy
E	Photon energy
e	Elementary electron charge
ϵ	Dielectric permittivity
ϵ	Extinction coefficient
ϵ_0	Permittivity of free space
F	Force
G	Gibbs free energy

γ_i	Inerfacial free energy
H	Enthalpy
h	Planck constant
Δh	Latent heat
H-bonds	Hydrogen bonds
$I(q)$	Scattering intensity
$I(0)$	Forward scattering intensity
IDF	International Diabetes Foundation
K	Constant
K_D	Dissociation constant
K_i	Equilibrium association constant
K_{LS}	Optical constant
K_{opt}	Optical constant
$\mathbf{k}$	Wavevector
k_B	Boltzmann constant
k_{inst}	Instrumental constant
κ	Debye-Hükel parameter
$1/\kappa$	Debye length
λ	Wavelength
λ_0	Wavelength of light in vacuum
l	Pathlength
l_B	Bjerrum length
LS	Light Scattering
M	Molecular mass
M_w	Molecular weight
MQ	Milli-Q water
MWCF	Molecular weight cut-off
μ	Chemical potential
μ_N	Mean chemical potential
μ_N^0	Mean interaction free energy per molecule
μ_∞^0	Energy per molecule of the bulk
N	Number of species
N_A	Avogadro constant
n_0	Refractive index of solvent
ν	Characteristic frequency
Ω	Spherical orientation

p	Pressure
p_c	Critical pressure
$P(q)$	Form factor
$p(r)$	Pair distance distribution function
$P(\theta)$	Structure factor
Π	Osmotic pressure
ϕ	Volume fraction
ψ	Electrostatic potential
ψ_0	Surface potential
PDB	Protein Data Bank
Q	Charge
Q_t	Total charge
q	Length of the scattering vector
R	Universal gas constant
$R(\theta)$	Excess Rayleigh ratio
r	Radius
r	Particle center of mass
r_{eff}	Effective radius
r_g	Radius of gyration
r_i	Location of i'th scattering element
ρ	Density
$\Delta\rho$	Contrast
$\rho(\mathbf{r})$	Electron density distribution
$\Delta\rho(\mathbf{r})$	Excess scattering length density
rpm	Revolutions per minute
S	Entropy
$S(\mathbf{q})$	Structure factor
$S'(\mathbf{q})$	Apparent structure factor
SAXS	Small Angle X-ray Scattering
SLS	Static Light Scattering
σ	Collision diameter
σ	Surface charge density
T	Temperature
T_c	Critical temperature
θ	Scattering angle
$u(d)$	Potential energy

V	Volume
v	Partial specific volume
v_{eff}	Effective specific volume
$v(r)$	Interaction potential
ν	Specific volume
v_p	Molecular volume
w	Weight concentration
w^{tot}	Total weight concentration
$w(d)$	Binding energy
WAXS	Wide Angle X-ray Scattering
WHO	World Health Organization
X_N	Molecular concentration
$X_{1,c}$	Critical monomer concentration
Z	Interaction constant
z	Valency

Contents

1 Introduction

The World Health Organization (WHO), estimates the number of people with diabetes millitus worldwide, has risen from 108 million in 1980 to 422 million in 2014. WHO also estimates that in 2016, 1.6 million deaths were caused by diabetes [1]. According to the IDF diabetes atlas, by the International Diabetes Federation (IDF), 463 million adults were living with diabetes in 2019. Furthermore 1.1 million children and adults under 20 years of age live with type I diabetes. By 2045 the number of people living with diabetes is estimated to reach 700 million [2]. Thus diabetes millitus poses a serious health problem in the global population.

Diabetes millitus is a chronic disease characterized by a high blood glucose level. Two types of diabetes exists. In type I the high level of blood glucose is due to insulin deficiency, while in type II diabetes the higher levels of blood glucose are due to insulin resistance and damaged regulation of insulin production. 90-95% of diabetics have type II diabetes. Type I diabetes is an autoimmune disorder, where the immune system destroys the β-cells of the pancreas where insulin is produced. In treatment of both types of diabetes millitus the aim is normoglycemia, a normal level of blood glucose, at all times. A normal level of blood glucose lies between 4 and 6 mM [3, 4]. Diabetics may suffer from hyperglycemia or hypoglycemia, both are dangerous conditions. Hyperglycemia is too high blood glucose levels, while hypoglycemia is too low blood glucose levels [5].

Diabetes is treated by subcutaneous injection of insulin drug products, with the aim of mimicking the time-action profile of blood glucose levels in healthy individuals. A basal level of glucose in the blood, is obtained by intermediate and long acting insulin analogues. The increasing blood glucose levels occurring after meals, are mimicked by fast acting insulin analogues [4, 5, 6, 7].

Understanding the self-association of proteins at high concentrations is important for drug product development [8]. Characterising the phase behaviour of the active pharmaceutical ingredients in drug products is of utmost importance. The recommended dosage of drug products is based on the amount of bioavailable ingredients. Unintended dosages may be a risk to patients, as was the case in 1998, where the drug product Ritonavir was found to be polymorphous and was removed from the marked [9, 10].

Insulin is a globular hormone protein, which self-associates into hexamers at concentrations above 2 mM [4, 11, 12]. The insulin hexamer exists in three conformational states, dependant on formulation conditions. Insulin drug products are prepared with preservatives, which induce

the so called R_6 conformational state [13]. This conformation of the hexamer is believed to be the most stable [14]. Insulin has been found to aggregate into fibrils and spherulites at pH around 2 and temperatures from 37 - 100°C [15, 16]. At pH 4 insulin frequently precipitates [4]. Furthermore, insulin fibrils have been found to form a turbid gel, at pH 1.6 with 0.04 M HCl [17].

The aim of this work is to investigate how the stability of human insulin hexamers as well as the inerparticle interactions of human insulin are affected by temperature, salt and preservatives at various concentrations. This is done using the fast acting drug product Actrapid produced by Novo Nordisk A/S. The temperatures of choice are 5°C, as this is the recommended storage temperature of the drug product, 25°C because this is the temperature at which the patient injects the product and lastly 40°C as this temperature may be reached during drug formulation or storage, and this temperature is fairly close to physiological temperatures. The interpaticle interactions of insulin are investigated by small angle x-ray scattering (SAXS), to detect effects on structure or stability of the protein in solution. Furthermore the insulin is investigated by composition-gradient multi-angle light scattering (CG-MALS), to see if temperature and excipients alter the interactions of the particles through the second virial coefficient, A_2. Protein concentration, salt and preservatives as well as temperature among other things alter the phase behaviour of proteins. Therefore, the effect of temperature and salt on the phase behaviour of insulin is also investigated. This is done by heating or cooling samples of various concentrations and detecting visible changes to the sample.

2 Insulin

Insulin is a globular hormone protein, produced as proinsulin in the β-cells of the pancreas. When in the β-cells the insulin monomers form hexamers around two zinc molecules. The zinc(II) hexamers then form granules. The insulin hormone regulates the cellular glucose uptake in humans through the insulin receptor located on the surface of liver cells. The normal range of glucose levels is 4-6 mM. After a meal glucose levels may rise to values of about 10 mM, which is when the pancreas discharges more insulin to once again reach normal glucose levels [4].

Binding of insulin to the receptors of liver cells triggers the process of glucose uptake from the blood. It takes about 10 seconds for the insulin to leave the pancreas and arrive at the liver [14]. The receptor is activated by an insulin monomer [6]. The dissociation rate of the insulin hexamer, is proportional to the rate of action for the insulin [18].

The insulin receptor is a dimeric protein, consisting of heterodimers of two α-chains and two β-chains. The receptor is composed of an extracellular portion where the insulin binds, activating the receptor, a single transmembrane helix and an intracellular tyrosine kinase domain. It has been suggested by models and crystal structures of the receptor and insulin molecules, that two insulin molecules bind to the receptor, at two potential binding sites [19, 20].

Insulin is an anionic crystalline protein. At physiological pH, insulin has a net charge of -4. The isoelectric point of insulin is 5.4 [4]. The diffusion coefficient of human insulin is 14.3 cm^2/s [21].

2.1 Structure of Insulin

The insulin monomer consists of two polypeptide chains named A and B, which are linked by two disulfate bridges. The human insulin monomer has a molecular mass of 5808 daltons, with an A-chain of 21 amino acids and a B-chain of 30 amino acids [4, 22]. The disulfate bridges links amino acids A7 with B7 and A20 with B19. Furthermore the A-chain includes a disulfate loop between amino acids A6 and A11, see Figure 2.1 [12]. The formation of the A-chain as well as the helix of B9 to B19 is structurally stable, while the structural segments of B1 to B8 and B25 to B30 may vary in conformation [11].

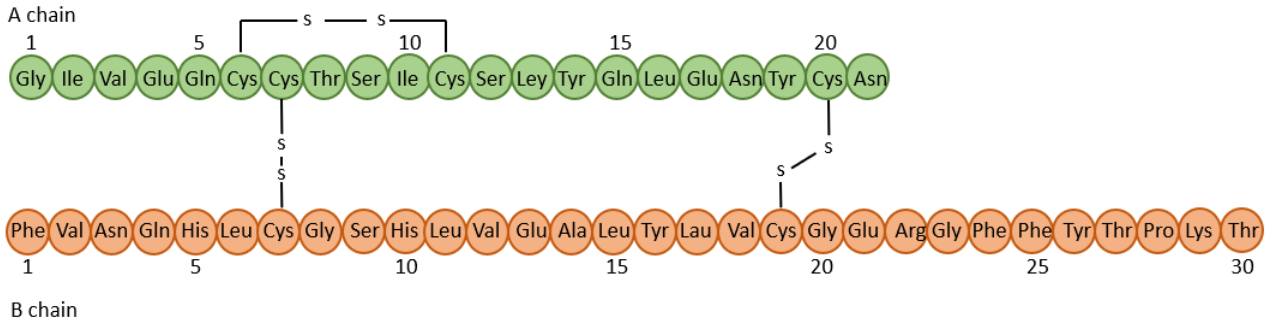

Figure 2.1: Illustration of the primary structure of a human insulin monomer. The green amino acids mark the A chain while the orange amino acids mark the B chain. Included are the three disulfate bridges, between amino acids A7-B7, A20-B19 and A6-A11. The first two bridges link the two chains. The illustration is inspired by Brange et al. [11, 12]

When the two polypeptide chains fold into a three dimensional structure, the A-chain forms two α-helices which are almost antiparallel. The first helix from amino acids A2 to A8 and the second from A13 to A20. The B-chain consists of an α-helix from B9 to B19, then comes a turn followed by a β-strand from amino acids B21 to B30. The tertiary structure leaves the insulin protein with a hydrophobic core. Furthermore the tertiary structure is stabilized

by non-covalent interactions between amino acids of both chains. Both non-polar and polar residues are present on the surface of the insulin monomer [11].

2.2 Self Association of Insulin

The insulin monomer has a hydrophobic core around which there are two extensive non-polar surfaces. As dimers are formed, the antiparallel β-sheet structure buries one of the non-polar surfaces. As hexamers are formed, the other non-polar surface is buried [20]. The two monomers in the dimer are held together by an antiparallel β-sheet structure formed by four hydrogen bonds, between amino acids B24 and B26, and non-polar forces. An insulin dimer in aqueous solutions is stable in the pH range 2 to 8 [11, 17]. The same surfaces of the insulin used to activate the receptor is used for self association [20]. See Figure 2.2 for cartoons of a monomer, dimer and hexamer.

Insulin hexamers are formed by the assembly of three dimers. The hexamer is almost spherical, with a 50 Å diameter and about 35 Å height. At neutral pH and concentrations above or equal to 2 mM the hexamer will form. When zinc ions are present dimers form hexamers around two zinc ions, at concentrations above 0.01 mM. The zinc ions are located in the centre of the hexamer at amino acid B10 [11, 12]. Two to four zinc ions stabilise the insulin hexamer [17]. The hexamer association is complete beyond $\sim$ 0.3 mg/mL [23]. Phenolic ligands bind specifically to each monomer in the hexamer, thus six ligands are present in each hexamer. [11, 12]. The bound phenolic ligands keeps dimers in the hexamer together, thereby stabilizing the hexamer [4, 6, 14]. The interactions of the monomers in the dimer are considered stronger than those of the three dimers in a hexamer [12].

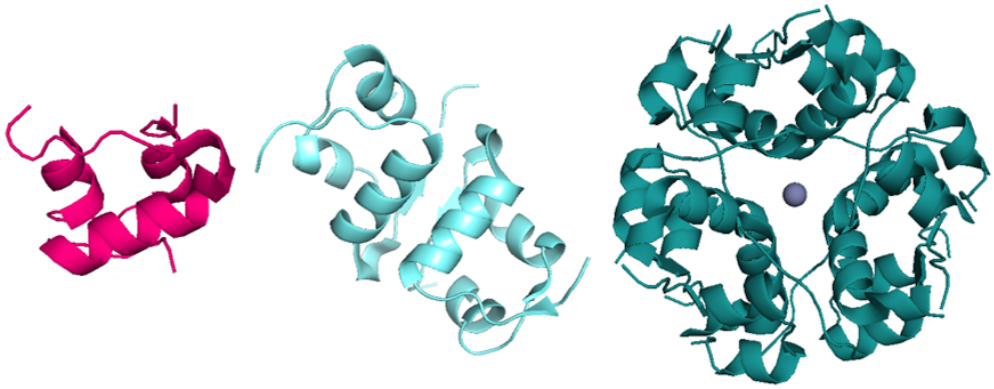

Figure 2.2: Cartoons of an insulin monomer, magenta, an insulin dimer, light blue, and a T_6 insulin hexamer with a zinc ion in the center, blue. Structures from the PDB IDs 3i40, 4ey1 and 1MSO respectively.

2.3 Stability of Insulin

The insulin hexamer may consist of monomers in two different conformations, namely the T or R conformation. [4, 5, 13, 14, 20]. The T conformation describes an insulin monomer with the first eight amino acid residues of the B-chain extended. The R conformation describes a monomer where the first eight B-chain residues have an α-helical structure, which means the B-chain consists of a continuous α-helix from B1 to B19. These conformations depend on chemical additions upon formulation. In the presence of zinc alone, hexamers of only T conformation monomers are formed, this is the above described II zinc hexamer. When chloride is present during formulation, a hexamer with half of the monomers in the R conformation and half in the T conformation is formed. Such a hexamer is referred to as a hexamer with a T_3R_3 conformation. Lastly formualtion with the addition of phenolic ligands leads to a hexamer with all six monomers in the R conformation [13]. The shift in conformations are induced by ligand concentrations above 1 mM [4]. The T conformation hexamer is more prone to dissociation than the R conformation hexamer, since the extended α-helix in the R conformation limits the zinc ions from diffusing out of the centre of the hexamer [6]. The highest order oligomer of the R conformation insulin is a dodecamer [4].

The T_6 hexamer has a cylindrical shape, with a diameter of ~ 4.7 nm and height of ~ 3.4 nm. While the R_6 hexamer has a more spherical shape with one longer edge of ~ 4.8 nm and a shorter edge of ~ 4.0 nm [21]. In general, the R_6 hexmer is more stable than the T_3R_3 hexamer, which is much more stable than the T_6 hexamer [14].

2.4 Insulin Aggregation

Insulin aggregation is presumed to be driven by a hydrophobic force [17]. Hydrophobic surfaces, such as the air and water interface, increases the tendency for insulin to fibrillate [11, 12, 17]. Amyloid fibrils are elongated protein aggregate structures, believed to form via partial protein misfolding and forming stable nuclei, on which monomers attach [24, 25]. When insulin in solution is exposed to heat and mechanical stress, it often results in non-native aggregation and fibrillation, which will precipitate. This is most common for monomeric insulin as compared to the hexamer [6].

Insulin fibrils are formed by inactive aggregates of linear non-native insulin molecules interacting. Formation of fibrils occur in three steps, which are nucleation, growth and precipitation. Nucleation is a number of monomers binding to form an organised structure, which then upon further monomer addition, leads to elongated growth into fibrils. For insulin, the initial step is

assumed to be hydrophobic residues, which are normally buried in the dimer or hexamer, being exposed to the surrounding solvent [17]. Insulin fibrils, are through an electron microscope seen as long fibers with a diameter of 10 - 50 nm, while fibrils in dynamic light scattering appear to have a mean diameter of 200 nm. The insulin fibrils are mostly insoluble, except in alkaline medium with a pH above 11 or strong organic acids [12].

Further self-assembly of amyloid fibrils may lead to spherulite formation, which are larger aggregated structures, in the micron range [25]. Spherulites are spherical aggregates with a core-shell structure, and fibrils with a β-sheet structure extended from the core [15, 24, 25]. Insulin forms spherulites as well as amyloid fibrils upon aggregation, and the two types of aggregates may coexist [24]. At low pH, around pH 2, and high temperatures, 37 - 100°C, insulin forms spherulites [15, 16]. Spherulites are characterized by a Maltese cross structure, visible by the use of crossed polarisers and optical microscopes [25]. Insulin amyloid fibrils are positively birefringent, meaning the slow optical axis is parallel to the fibril axis. Thereby, fibrils are radially arranged in the spherulites [16].

In the work by Krebs et al., [16], 1 mM bovine insulin dissolved in deionized water, was found to form spherulites at pH 2, after storage at 37°C and 65°C with a most common sizes of $\sim 50\mu$m. Furthermore, they found a lack of birefringence in the centre of their sperulites, indicative of a lack of orientation of the polypeptide chains [16]. In later work by Krebs et al., [26], the formation of spherulites of bovine insulin was studied under the same conditions as in [16] and with additional 0.1 M NaCl. The formed sperulites were found to have the same structure as those formed without 0.1 M NaCl, however the sperulites had smaller nonbirefringent cores. Furthermore Krebs et al. found an increase in the birefringence of samples as the amount of spherulites increased [26].

At pH 4, insulin has a tendency to precipitate [4]. Insulin precipitates in amorphous or crystalline from, upon reduction in the solubility. This may be done by altering the pH or by addition of zinc ions or other divalent metal ions. Isoelectric precipitation occurs in the range of pH 4.5 to 6.5 as the isoelectric point of insulin is 5.4 [4, 12].

In the work of Nielsen et al., [17], they studied bovine insulin fibrillation and found their fibrillated insulin solution at concentrations above 5 mg/mL was a firm turbid gel. For concentrations below 5 mg/mL the solution was viscous with visible aggregates. The bovine insulin was dissolved in 0.04 M HCl at pH 1.6 [17].

2.5 Insulin Formulation

For pharmaceutical purposes the insulin drug product is formulated with a number of excipients, from preservatives to isotonic agents. The Actrapid drug products investigated in this work are formulated with an original concentration of 100 units/mL $\sim$ 3.46 mg/mL at pH 7.3 [27].

Preservatives in the form of phenolic ligands are used in the insulin drug product as antimicrobial agents [4, 18]. Phenolic ligands bind hydrophobic pockets in the attachment between the two adjacent insulin dimers, in the area of the disulfate bridge between amino acids A6 and A11. The hydroxyl group of the ligand forms a hydrogen bond to the carbonyl of A6 and the NH group of A11. As there are three of said binding pockets present in the R_3 trimer, the hexamer of a T_3R_3 conformation binds three ligands while an R_6 hexamer binds six ligands [14, 21]. The ligands stabilise the α-helices of amino acids B1-B8. This stabilisation aids in the stability of the insulin hexamer, as it keep the zinc ions in the hexamer centre [18].

Chloride is used in the formulation of insulin as an isotonic agent, meaning chloride aids in keeping the osmotic pressure intact upon injection. A hexamer with four zinc molecules is generated at high chloride concentrations [20]. The chloride monovalent anion binds to the anion sites on the insulin hexamer. There is one site on the T_3R_3 hexamer, and two binding sites on the R_6 hexamer [14].

Zinc is used in the formulation process to obtain hexamer formation. The two binding cites of zinc in the insulin hexamer, are the HIS B10 amino acids [14, 21]. The binding sites are approximately 16 Å apart [14]. Attri et al. [28, 29], investigated the self-association of bovine insulin with and without zinc using concentration-gradient static and dynamic light scattering. The results show that bovine insulin is hexameric at concentrations above 0.3 mg/mL, in the presence of zinc. The hexamers may further self-associate to higher order oligomers [28]. In the absence of zinc in acetic acid at pH 1.95, Attri et al. observed monomers. Dimers were observed in HCl pH 1.85 [29].

2.6 Insulin Analogues

Insulin analogues are insulin molecules which have an altered amino acid sequence compared to native insulin. This gives the analogues different properties aiding in the treatment of diabetes mellitus. In healthy persons the time-action profile of the insulin levels in the blood rises after a meal and slowly decreases, until it reaches a certain basal level. In treatment of diabetic patients this time-action profile is mimicked using both fast, intermediate and long acting insulin analogues. The fast acting analogues mimic the rises of insulin levels caused by

food intake, while the long acting insulin analogues are used to mimic the basal level of insulin [4, 5, 6, 7]. When altering the amino acids in native human insulin, to create insulin analogues, amino acids B27-B30 are advantageous as these amino acids are involved in dimer formation and do not affect insulin receptor binding [7].

2.7 Scattering of Insulin

The scattering profile of insulin differs, depending on the oligomeric state of the protein. In Figure 2.3 are the theoretical scattering curves of insulin in monomeric, dimeric and hexameric state. These curves are generated using the program Crysol [31] and three different PDB structures, 3i40, 4ey1 and 1ev6 for the monomer, dimer and R_6 hexamer respectively.

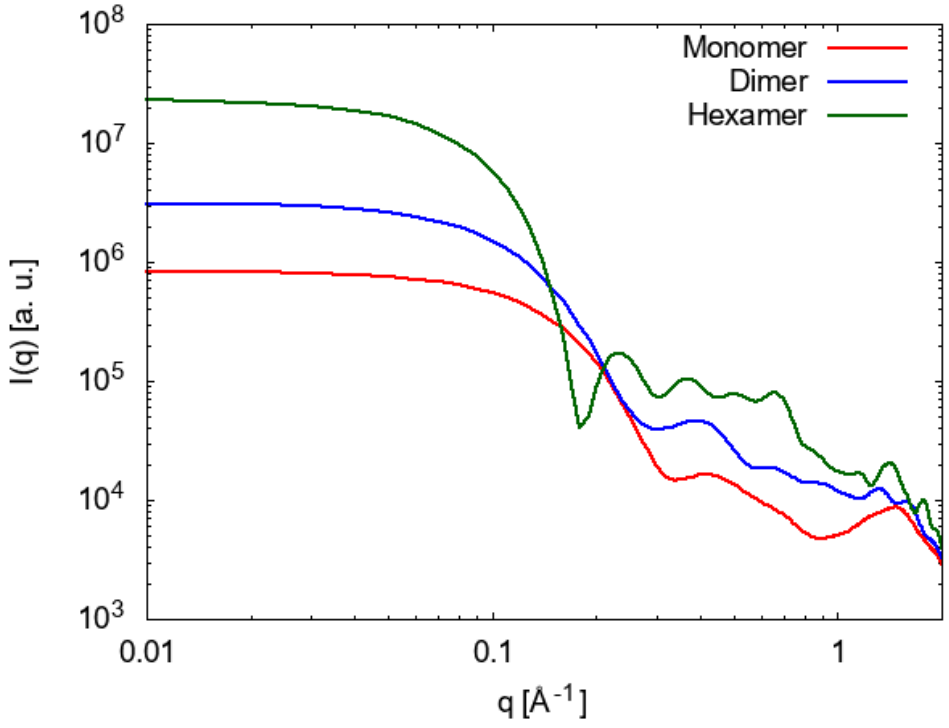

Figure 2.3: Theoretical scattering curves of an insulin monomer, red curve, dimer, blue curve and an R_6 hexamer, green curve on a double logarithmic scale. The curves are generated using Crysol, [31], and PDB structures 3i40, 4ey1 and 1ev6 for the monomer, dimer and hexamer respectively.

Furthermore, the scattering profile of the insulin hexamer shows signs of the hexameric conformational state. The R_6, T_6 and T_3R_3 hexamers all have differing profiles, as may be seen in the curves generated from the PDB structures 1mso, 1trz and 1ev6 in Figure 2.4. As can be seen in the Figure, the first minimum at a q-value of about 0.2 Å, changes with the hexamer conformation. The R_6 hexamer, green curve, has the most pronounced minimum and the T_6

hexamer, red curve, has the least. The theoretical scattering curves will in this work be used to determine the oligomeric species in the experimental scattering curves using OLIGOMER [32].

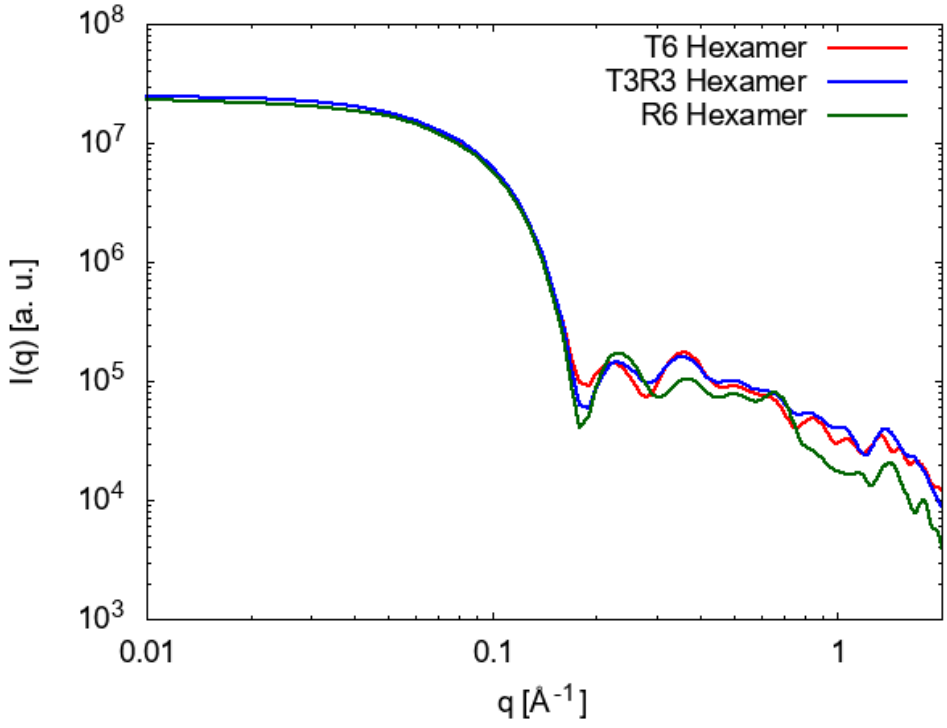

Figure 2.4: Theoretical scattering curves of three different hexamers in the T_6 conformation, red curve, the T_3R_3 conformatin, blue curve, and the R_6 hexamer, green curve on a double logarithmic scale. The curves are generated using Crysol, [31], and PDB structures 1mso, 1trz and 1ev6 for the T_6, T_3R_3 and R_6 hexamer respectively.

The radius of gyration, r_g and the maximal particle size, D_{max} obtained from SAXS in the work of Palmieri et al. [33] shows an r_g of 19 Å and D_{max} of 50 Å. These values are found both for human insulin and the analogue insulin aspart. Furthermore in their work, they found oligomeric fits to their SAXS curves yield best results at human insulin composition of 64% hexamer and 36% dimer. For insulin aspart the composition was found to be 55% hexamers, 25% dimers and 20% monomers. All their work is done using the formulation concentration of 100 units/mL and a dilution to 33 units/mL [33]. In the work of Ahmad et al., [34], they found the radius of gyration of the native insulin hexamer at pH 7.4 to be 20.3 Å. In their work they state, an insulin monomer has r_g of 11 Å [34].

3 Particle Interactions

In particle interactions there are numerous forces. A way of classifying said forces is by their origin. Purely electrostatic forces, arise from Coulomb forces between charges, such as ions. Electrostatic forces also include polarization interactions from dipole moments induced by the electric fields of charges in close proximity to molecules in solution. Other forces are purely entropic, such as osmotic forces and thermal fluctuation forces. Forces such as covalent or chemical bonds are quantum mechanical in origin. These have discrete energy levels [35]

3.1 Covalent Bonding Forces

Covalent bonds are interatomic bonds formed between molecules coming together to form an atom. In these bonds, molecules share electrons, with a valency depending on the specific molecules. Thus one to three electrons may be shared between two molecules. The covalent bonds may be oriented at a certain angle giving the bonds a direction. It is the covalent bonds which dictate how the atoms go together to form the lattice of macromolecules or crystalline solids. The stoichiometry of an atom is the amount of covalent bonds the atom can form with other atoms. This is also termed atomic valency, z. For example hydrogen atoms have valency of 1 [35].

The covalent forces, the forces binding atoms together, are short range forces. The distance at which covalent forces act is of the order 0.1 - 0.2 nm, which is interatomic separation. As the length of a covalent bond increases the strength of the bond decreases [35].

3.2 Coulomb Forces

The Coulomb force is the force arising from an electric field, E at distance d from a charge Q_1 acting on a second charge Q_2 and is given by

$$F(d) = \frac{Q_1 Q_2}{4\pi\epsilon_0\epsilon d^2}$$

where ϵ is the dielectric permittivity of the medium and ϵ_0 the permittivity of free space. The Coulomb force or interaction may take place between ions in a solution. The binding energy, $w(d)$ of ions in aqueous solution is given by

$$w(d) = \frac{z_1 z_2 e^2}{4\pi\epsilon_0\epsilon d}$$

with z_i being the ionic valency and e the elementary electron charge. Coulomb interactions are strong, long range and opposed to the covalent bonds, coulomb interactions do not have specified orientation [35].

Consider a sphere of radius r and total charge $Q = 4\pi e r^3/d^2$ where d is the mean distance between the surface charges. The Coulumb energy, E of binding an ion with charge ze and radius a to the sphere is given by

$$E = \frac{zeQ}{4\pi\epsilon_0\epsilon(r+a)} \tag{3.1}$$

This energy leads to the binding energy, w between ion and sphere, with r $\gg$ a

$$w \approx \frac{zeQ}{4\pi\epsilon_0\epsilon r} \tag{3.2}$$

This result implies that the binding energy is greater for larger spheres, or for smaller distances or higher valency, z [35].

Consider a globular protein of radius r, with all hydrophobic residues shielded by polar residues on the surface, each able to donate one charge. Assume the polar residues all have equal size with radius a and the total charge of the protein is $Q_t = Ne$, the energy of the protein is given by the Boltzmann constant, the temperature, the Bjerrum length, l_B and the radii of the sphere and residues as follows [36]

$$E = k_b T l_B \frac{8r^3}{a^4} \tag{3.3}$$

The Bjerrum length, l_B is the distance, r, between centres of two unit charges as their Coulomb energy is equal to the thermal energy, [35]

$$\frac{e^2}{4\pi\epsilon_0\epsilon a} = k_B T \leftrightarrow l_B = \frac{e^2}{4\pi\epsilon_0\epsilon k_B T} \tag{3.4}$$

Where ion charges are assumed to be monovalent, thus $Q = e$, and ϵ is the relative permittivity of the medium the protein is in [35, 36].

3.3 Hydrogen Bonds

Hydrogen bonds (H-bonds) are formed between hydrogen atoms and electronegative atoms, such as oxygen and chloride. One hydrogen atom may bind more than one atom, making the H-bonds a bridge, or stronger bond, between two electronegative atoms. The hydrogen bond

is mainly electrostatic and weaker than covalent bonds. H-bonds are orientation-dependent. Hydrogen bonds may be found both inter- and intramolecularly, also in non-polar surroundings. Hydrogen bonds play an important role in the folding of proteins, where the bonds link sections inside the protein [35].

3.4 Hydrophobic and Hydrophilic Interactions

Hydrophobic interactions are strong attractions between hydrophobic molecules in water. It is important to note that hydrophobic interactions are attractions, not bonds, between molecules. Some molecules are hydrophilic, meaning the molecule prefers contact with water rather than other molecules. Zwitterions are for example hydrophilic. Both hydrophobic and hydrophilic interactions are determined by the structure of the hydrogen bonds of water around dissolved groups [35].

3.5 Van der Waals Forces

Neutral molecules may have an instantaneous electric dipole which fluctuates. Van der Waals forces are the attractive intermolecular forces acting between two molecules approaching each other, arising from the interactions between the dipole of each molecule. The potential energy of the van der Waals forces between two alike particles, with centres a distance d apart, in vacuum is given by

$$u(d) = -\frac{3\alpha^2 h\nu}{4d^6(4\pi\epsilon_0)^2} \tag{3.5}$$

where α is the electronic polarisability of the particle, h is the Planck constant, ν is the characteristic frequency of the particle. This equation holds true as long as the distance d is shorter than the characteristic wavelength $2\pi/\nu$ [37].

The van der Waals attractive energy, E for two particles with radius r and the surface-to-surface distance d, where $d \ll r$, is

$$E = -\frac{Ar}{12d} \tag{3.6}$$

with A being the effective Hamaker constant given by the Hamaker constant of the particles, A_{11} and the Hamaker constant of the medium, A_{22} as

$$A = \left(A_{11}^{1/2} - A_{22}^{1/2}\right)^2 \tag{3.7}$$

The Hamaker constant of any medium or material is determined by the London dispersion constant, β and the number density of atoms per unit volume, ρ [35, 37]

$$A = \pi^2 \rho^2 \beta^2 \tag{3.8}$$

The force between the two spheres a distance D apart is then, determined by $F = -\mathrm{d}E/\mathrm{D}$

$$F = \frac{-AR}{12D^2} \tag{3.9}$$

The Hamaker constant of water, interacting with water at room temperature in vacuum is 3.7 $\cdot\ 10^{-20}$ J [35].

The van der Waals force acting between two alike particles in a medium is always attractive. If van der Waals forces acted alone the particles would coagulate and precipitate out of the medium. This coagulation and precipitation does not occur when the medium in which particles are suspended, has a high dielectric constant, as this gives rise to repulsive electrostatic forces [35].

3.6 Electrostatic Forces

It is only in some simple systems, such as interactions in vacuum, where the van der Waals forces act alone. In systems which are more complex, long-range electrostatic forces are also to be found, along with the van der Waals forces [35].

Surfaces in a liquid may become charged by three different things. The first being ionization or dissociation of surface groups. The second being adsorption or binding of ions from solution. The last being charges exchanging between two diverse surfaces. The final surface charge, coming from so called co-ions, is balanced by a region of counterions, equal in magnitude however carrying opposite charge as the co-ions [35].

Coulomb interactions occur between charged particle surfaces and electrolyte ions. These interactions lead to higher concentrations of ions with opposite charge as that of the particle. This cloud of electrolyte ions, near the particle surface, and the surface form an electrical double layer. The interaction of colloidal particle depends strongly on the diffusive double layer. The layer is diffusive because of thermal motion of the ions [37].

Consider a system of colloids suspended in water, where the only ions in the water, besides ions from dissociated water (H_3O^+ and OH^-), are those dissociated from the surfaces of the

colloids. In this case, the chemical potential, of any ion in solution is given by

$$\mu = ze\psi + k_B T \log \rho \tag{3.10}$$

where z is the valency, ψ is the electrostatic potential and ρ is the number density of ions at any point, x, between two surfaces. The electric field, E is equal to $-\mathrm{d}\psi/\mathrm{d}x$ Using the equilibrium condition, that chemical potentials must be equal, the Nernst equation is obtained. The Nernst equation is the Boltzmann distribution of counterions at any point x, given by

$$\rho = \rho_0 \mathrm{e}^{-ze\psi/k_B T} \tag{3.11}$$

Where ρ_0 is the density at the midplane between the two surfaces, where for simplicity $x = 0$. The Poisson equation for the total surplus charge density at position x is

$$ze\rho = -\epsilon_0 \epsilon \left(\frac{\mathrm{d}^2 \psi}{\mathrm{d}x^2} \right) \tag{3.12}$$

Combining these two equations leads to the Poisson-Boltzmann equation given by

$$\frac{\mathrm{d}^2 \psi}{\mathrm{d}x^2} = -\left(\frac{ze\rho_0}{\epsilon_0 \epsilon} \right) \mathrm{e}^{-ze\psi/k_B T} \tag{3.13}$$

Solving the Poisson-Boltzmann equation, gives at any point x the potential, ψ, electric field, E and density of the counterions, ρ [35].

Equation 3.13 can be solved using the boundary conditions, that at the midplane, $x = 0$, the electric field must be zero, such that $-(\mathrm{d}^2\psi/\mathrm{d}x^2)_0 = 0$ and overall electroneutrality, such that the counterions are of the same magnitude, but opposite charge as the charges of the surfaces. This last condition suggest that the electric field at the surface E_s, is given by

$$E_s = -\frac{\sigma}{\epsilon_0 \epsilon} \tag{3.14}$$

where σ is the surface charge density in units of $\mathrm{C/m^2}$. It is worth noting that the above expression, for the electric field at the surface, is independent of the distance between the two surfaces [35].

The Poisson-Boltzmann equation, eq. 3.13, can, as mentioned, also be solved for the counterion density, ρ. This is done by differentiating the Nernst equation, eq. 3.11, with respect to

x, resulting in the following expression for ρ at any position x

$$\rho_x = \rho_0 + \frac{\epsilon_0 \epsilon}{2k_B T} \left(\frac{\mathrm{d}\psi}{\mathrm{d}x} \right)_x^2 \tag{3.15}$$

where ρ_0 is the counterion density at the midplane, $x = 0$, and $\left(\frac{\mathrm{d}\psi}{\mathrm{d}x}\right)^2$ is at position x. At the surface, where $x = d/2$, with d being the distance between the two surfaces, the density of counterions ρ_s is expressed as

$$\rho_s = \rho_0 + \frac{\sigma^2}{2\epsilon_0 \epsilon k_B T} \tag{3.16}$$

This expression shows that even for isolated surfaces, where $\rho_0 \to 0$, ρ_s is never less than $\sigma^2/2\epsilon_0 \epsilon k_B T$. Most counterions are present within the first couple of Ångström from the surface. This distance gives rise to the terminology double-layer [35].

Charged macromolecules, such as proteins in electrolyte solutions, will obtain an ionic cloud, resulting in a neutral net charge of the protein. The total charge of the screening cloud is $e\Delta cA/\kappa$. Where Δc is a measure of how the far-field concentration is perturbed by the protein within the cloud, defined as $\Delta c = c_+ - c_-$. A is the surface area of the protein and $1/\kappa$ the width of the ionic cloud which is the Debye length [36, 37]. κ is the Debye-Hückel parameter given by the ion valency and ion charge density in the solution, as follows [35, 37]

$$\kappa = \left(\frac{1}{\epsilon \epsilon_0 k_B T} \sum_{i=1}^{n} z_i^2 e^2 \rho_{\infty i} \right)^{1/2} \tag{3.17}$$

The charge distribution in the electrolyte solution is uniform beyond the Debye length [36]. The Debye length of pure water is 960 nm [35].

Double layer repulsion occurs when charged particles, having ionic clouds in a layer, and the same charge, are the distance d apart, and d is much smaller than twice the double layer thickness [37]. The energy, E, between the double layers of two spherical particles with radii r at a distance d apart, where $d \ll r$, is

$$E = \frac{1}{2} r Z \mathrm{e}^{-\kappa d} \tag{3.18}$$

In the above equation, Z is the interaction constant, given by

$$Z = 64\pi\epsilon_0 \epsilon \left(\frac{k_B T}{ze} \right)^2 \tanh^2 \left(\frac{ze\psi_0}{4k_B T} \right) \tag{3.19}$$

Where ψ_0 is the surface potential, of an isolated surface. Note that Z mainly depends on surface properties, besides the valency. The remaining elements of the energy are the radii, separation distance and the Debye-Hükel parameter, κ which is dependent on properties of the solution. The force of the double layer interaction is determined by $F = -dE/dD$, giving

$$F = \frac{\kappa}{2} r Z \mathrm{e}^{-\kappa d} \tag{3.20}$$

3.7 DLVO Theory

Derjaguin-Landau-Verwey-Overbeek (DLVO) theory, explains the stability of colloidal systems of charged particles. DLVO theory states the stability is determined by a balance between the van der Waals interactions and electrostatic interactions between the particles [37]. Thereby DLVO theory link van der Waals forces with the electrostatic forces of the double layer. The van der Waals interaction potential is nearly unchanged by changes in pH or electrolyte concentration, as opposed to the electrostatic interaction potential. Because of the power law in the attraction of the van der Waals force, it must always be greater than the repulsion of the double layer [35].

Consider the interaction between two identical spherical particles with radius r in an electrolyte medium. The distance between the particles is d. Both particles have a double layer. The potential energy between the double layers of the particles is given by equation 3.18, and the van der Waals attraction energy by equation 3.6. The total energy of interaction is the sum of these two energies,

$$E_{tot} = \frac{1}{2} r Z \mathrm{e}^{-\kappa r} - \frac{Ar}{12r} \tag{3.21}$$

This expression for the total energy of interaction will have a primary and secondary minimum as well as a maximum, under certain conditions [37].

4 Factors Influencing Interactions

4.1 pH

pH is for aqueous solutions defined as $\mathrm{pH} = -\log_{10}[\mathrm{H}^+]$. Where $[\mathrm{H}^+]$ is the concentration of hydrogen ions in solution. Amino acid side chains of proteins in aqueous solution either donate or receive hydrogen ions [36]. Therefore, the pH value determines the charge of the side chains, as well as the net charge of the protein in solution. This effects electrostatic interactions, which

may affect protein stability. The first being by nonspecific repulsions from charged parts of a highly charged protein, which may occur at pH values well away from the pI [38]. The pI value is a point in pH named the isoelectric point. At this point the total surface charge of a protein is zero. Thus there are equal amounts of positive and negative surface charges. There may still be local regions with high negative or positive surface charges [35]. Altering the pH value may lead to an increase in charged groups on a protein, which gives rise to repulsion within the protein. This repulsion results in destabilization of the protein conformation arising from a greater charge density on the folded protein, than on the unfolded protein. Thus the unfolded state has lower electrostatic free energy and is therefore favoured [38].

4.2 Buffer Solutions

Buffer solutions are solutions with a certain pH value that is not altered by small additions of strong acids or bases [39]. The Henderson-Hasselbalch equation can be used to determine the pH value of a buffer solution. To derive it, consider the chemical reaction between a macromolecule, for example a protein, P, and hydrogen ions in solution

$$\mathrm{HP} \leftrightharpoons \mathrm{H^+} + \mathrm{P^-} \tag{4.1}$$

The dissociation constant, K_d for this reaction is given

$$K_\mathrm{d} = \frac{[\mathrm{H^+}][\mathrm{P^-}]}{[\mathrm{HP}]} \tag{4.2}$$

The dissociation constant is used in the definition of the pK value, $\mathrm{p}K = -\log_{10} K_\mathrm{d}$. p$K$ is a measure of the tendency of a molecule to dissociate. The pK value is part of the Henderson-Hasselbalch equation which is given by

$$\mathrm{pH} = \mathrm{p}K + \log_{10} \frac{[\mathrm{P^-}]}{[\mathrm{PH}]} \tag{4.3}$$

Where [$\mathrm{P^-}$] and [HP] are the concentrations of protein and protein-hydrogen complex in the buffer solution respectively [36].

4.3 Osmotic Pressure

Osmotic pressure arises from a mechanical force [36]. Osmosis is the flow of solvent molecules through a semipermeable membrane, separating pure solvent and solution. The flow of solvent molecules arises from the chemical potentials, as molecules tend to flow toward lower potentials.

Applying pressure to the solution side of the membrane prevents osmosis from occurring. The amount of pressure required is dictated by the chemical potential of the solvent on each side of the membrane, which must be equal [40]. The osmotic pressure is the pressure difference between the pressure of the solution, p_2, and the pressure of the pure solvent, p_1. The osmotic pressure of a dilute solution is given by the van't Hoff formula

$$\Pi = p_2 - p_1 = \frac{N_s}{V} k_B T \tag{4.4}$$

Where N_s is the number of solutes and V is the solution volume [36, 40].

5 Protein Aggregation

In a system of molecules which form aggregates in solution, equilibrium thermodynamics dictates equality between the chemical potential of each aggregated state. The chemical potential, μ is given by,

$$\mu = \mu_N = \mu_N^0 \frac{k_B T}{N} \log\left(\frac{X_N}{N}\right) = \text{constant} \tag{5.1}$$

In the above equation μ_N is the mean chemical potential, which is constant, μ_N^0, is the mean interaction free energy per molecule in aggregated state N, X_N is the molecular concentration. All variables mentioned are of a molecule in aggregated state N, where N is a positive integer. $N = 1$ is a monomer in solution. k_B is Boltzmanns constant and T is the temperature [35].

Aggregates form if the mean interaction free energy per molecule, μ_N^0 decreases with number of molecules in the aggregate N. For a spherical aggregate, with radius r, N is proportional to the volume of the sphere. The amount of unbound surface molecules is proportional to the area of the sphere, which can be expressed as $N^{2/3}$. μ_N^0 is then,

$$\mu_N^0 = \mu_\infty^0 + \frac{\alpha k_B T}{N^{1/3}} \tag{5.2}$$

where μ_∞^0 is the energy per molecule of the bulk and α is a constant that depends on the strength of the intermolecular interactions in the sphere,

$$\alpha = \frac{4\pi r^2 \gamma_i}{k_B T} \tag{5.3}$$

Where r is the radius of the molecule and γ_i the interfacial free energy per unit area. The critical concentration at which further addition of monomers leads to the formation of more

aggregates is given by

$$X_{1,c} \approx \exp\left(-\frac{\mu_1^0 - \mu_N^0}{k_B T}\right) \tag{5.4}$$

where $X_{1,c}$ is the critical monomer concentration. Using the expression of μ_N^0 in equation 5.2, the critical concentration simplifies to

$$X_{1,c} \approx \mathrm{e}^{-\alpha} \tag{5.5}$$

Above this critical concentration, the monomer concentration of the solution is practically unchanged, regardless of the additional amount of monomers [35].

Proteins may aggregate into a form named amyloid fibrils. The formation of fibrils is initiated by a nucleus [41]. Amyloid fibrils are large needle-like assemblies of polymers consisting of multiple copies of one single protein with a predominantly cross-β structure [26, 41]. Fibrils are usually longer than 1 μm with diameters between 10 - 12 nm [26]. In order to form amyloid fibrils, proteins must be partially unfolded. However fibrillation is not the only fate of a partially unfolded protein. Refolding or aggregation may also occur [26]. At low pH values, globular proteins tend to be partially unfolded [42].

The formation of spherulites is a common form of self-assembly, which is found both for some proteins and during crystallization of polymers [16]. Spherulites are characterized by a Maltese cross which is caused by radially oriented amyloid fibrils in the spherulite [25, 26]. The cross is visible when spherulites are studied under a polarized light microscope [26]. In the cross-β structure, the core is organized and consists of β-sheets with strands perpendicular to the fibril axis [42].

6 Phase Behaviour

In the pharmaceutical industry it is highly relevant to understand and control the state of products. An injectable drub product, such as insulin, may be a transparent liquid under optimal conditions. In the case of condition changes, giving rise to crystal formation or aggregation, the drug product will be compromised. This could, worst case scenario, be a risk to patients' health and safety. An example of a risk, is the drug Ritonavir used to treat HIV. In 1998 the drug was temporarily removed from the marked, due to polymorphism in the crystal phase, giving rise to multiple crystal states with different solubilities. Ritonavir was thought to have one type of crystal, which the drug was based on. In 1998 another more stable and thereby less

soluble, crystal of Ritonavir was discovered. The second crystal form, altered the dissolution of the drug product, which lead to the intended dosage being compromised [9, 10].

The physical state of a substance depends on temperature, pressure, the substance itself, pH, salt and other ambient conditions. All substance are subject to state changes, even molecules, ranging from small hydrocarbons, such as ethane [43], to colloidal particles, which may be referred to as superatoms [44, 45], and complex biomolecules, such as globular proteins [46]. Changes of state are governed by thermodynamics, where the lowest free energy dictates the state of the substance at the given set of conditions [40]. A pure substance may experience equilibrium between fluid phases, if the particle interaction in the substance has an element of attraction [47].

Phase transitions are divided in order, dependent on the derivative of the Gibbs free energy of the specific transition. If the first derivative of the Gibbs free energy is discontinuous the transition is referred to as a first order phase transition. If the first derivative is continuous, and higher order derivatives are discontinuous the transition is a higher order phase transition [48, 49].

6.1 Phase Diagrams

A way of representing the phase behaviour of a substance is through a phase diagram, which is often displayed as a plot of temperature versus pressure, showing the different states of a given substance [40]. An example of a simple phase diagram of a one component system is displayed in Figure 6.1. In a phase diagram one finds regions representing which of the different states is the most stable, given the specific conditions represented on the axes of the diagram. The more complex the substance, the more complexity will be found in the phase behaviour. This can be seen in the solid phase, where for complex molecules one can find more than one kind of solid, due to the fact that there are multiple ways molecules can stack to form different crystal structures [39, 40].

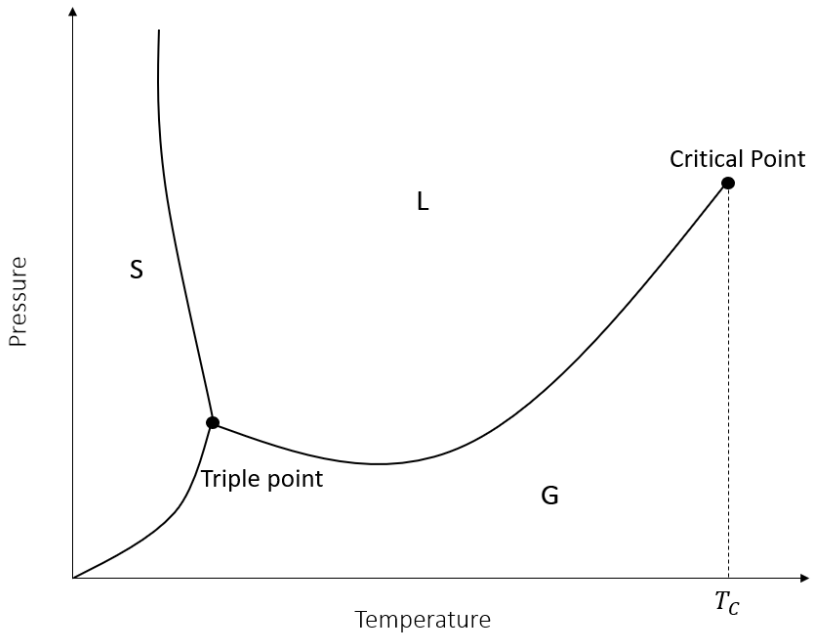

Figure 6.1: Schematic phase diagram of a one component system displayed with temperature versus pressure. Shown are the three states; solid (S), liquid (L), and gas (G), along with the triple point and the critical point. Marked on the temperature axis is the critical temperature, T_c. See the main text for details of these points. Schematic inspired by Kondepudi et al., [48], and Reichl, [49].

Phase boundary lines, also known as coexistence curves, divide the phase diagram into regions. These shed light on how two phases of the substance behaves given the conditions of the axes in the phase diagram. On the boundaries, the substance will be in equilibrium between the two states. The equilibrium condition is equality of the chemical potentials of the different states. For example in a Temperature-pressure (T-p) digram, where temperature and pressure are the axes conditions, the boundary between the vapour and liquid or solid states, will show how temperature affects the vapour pressure. Likewise the phase boundary between the solid and liquid states will show how the melting temperature is affected by pressure. Applying pressure to a substance will, in most cases, raise the melting temperature, but not always. If the solid state is the least dense phase, then the melting temperature will decrease [39, 40, 48, 49]. When describing elements of phase diagrams in the following, the T-p diagram is used as an example, therefore the elements in a phase diagram are described as dependent on pressure and temperature. Of course, these elements may depend on other conditions and not explicitly temperature and pressure.

The slope of any phase boundary is related to the entropies and volumes of the two separated phases by the Clausius-Clapeyron relation. Moving along the coexistence curve the Gibbs free

energy of the two states must change equally. Using the Gibbs-Duhem equation and the fact that the chemical potentials of state one and two are equal along the curve, the Clausius-Clapeyron relation is obtained

$$-S_1 dT + V_1 dP = -S_2 dT + V_2 dP \Leftrightarrow \tag{6.1}$$

$$\frac{dp}{dT} = \frac{S_1 - S_2}{V_1 - V_2} = \frac{\Delta S}{\Delta V} = \frac{\Delta h}{T \Delta V} \tag{6.2}$$

Where the left hand side is the slope of a coexistence curve and $\Delta h = T\Delta S$ is latent heat required to convert the substance from one phase to the other [40, 49].

In a phase diagram one might find a triple point, for an example see Figure 6.1. The triple point is where three boundaries intersect giving rise to an equilibrium between three different states. This equilibrium will only occur at the specific pressure and temperature given by the location of the triple point in the phase diagram, and dictates equality between the chemical potential of all three states [39, 40, 48, 49].

When raising the temperature of a system in equilibrium between vapor or gas and liquid, the pressure also needs to increase to ensure the liquid does not vaporize. When increasing the pressure, the vapour will become more dense, which means there will be less of a difference between the vapour and liquid. This ends in a point where there is no intermittent change from liquid to vapour, this is the critical point, see Figure 6.1 [40, 48]. The critical point defines the critical pressure, p_c and critical temperature T_c of the given substance. At the critical values, distinction between the two phases ends [48]. Beyond this temperature and pressure the substance will be a supercritical fluid. A supercritical fluid is a gas with a density equal to that of a liquid [39].

6.1.1 Simple Hard Spheres

The simplest possible system for modelling particle interactions is the hard sphere system. In the hard sphere system the interaction potential is infinitely repulsive at contact between two spheres and zero otherwise, see Figure 6.2 [45, 50, 51].

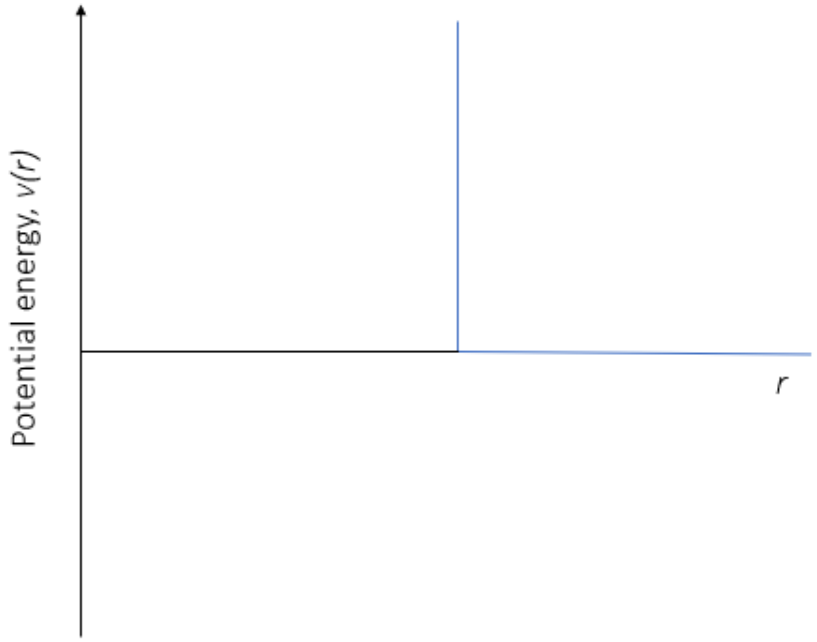

Figure 6.2: Schematic of the interaction potential, $v(r)$, for the simple hard sphere system. The interaction energy of two spheres, is infinite as the two spheres touch and zero otherwise. r is the distance between the spheres. Schematic inspired by Poon et al., [51]

In the hard sphere system the only variable is the volume fraction, ϕ_{HS}, which is given by the number of hard spheres N, with radius r, in the volume, V as follows [52]

$$\phi_{HS} = \frac{4\pi r^3 N}{3V} \tag{6.3}$$

The simple, or pure, hard sphere system only has two distinct phases, see Figure 6.3. Entropy dictates at which volume fraction the hard sphere system will begin to form crystals. This volume fraction is termed the freezing volume fraction, $\phi_f = 0.494$. Below ϕ_f, all spheres are fluid. Above the melting volume fraction, $\phi_m = 0.545$, all spheres are in crystal form. Thus between ϕ_f and ϕ_m crystal and fluid phases coexist [45, 53]. When the volume fraction increases the hard sphere particles will be progressively caged by neighbours. A critical volume fraction is reached, when the caging stops long-range particle motion. At this volume fraction, $\phi_g \approx 0.58$, the hard sphere system is considered a glass or glassy. A glass is a solid state with no long-range spatial order. This phase may not crystallize, thus there is an arrest in the kinetics of the system [45, 52, 53, 54]. Two types of glass exist, making the glass state reappear for higher concentrations, as found by Pham et al., [52]. One glass is dominated by attraction and the other by repulsion [47, 52, 55].

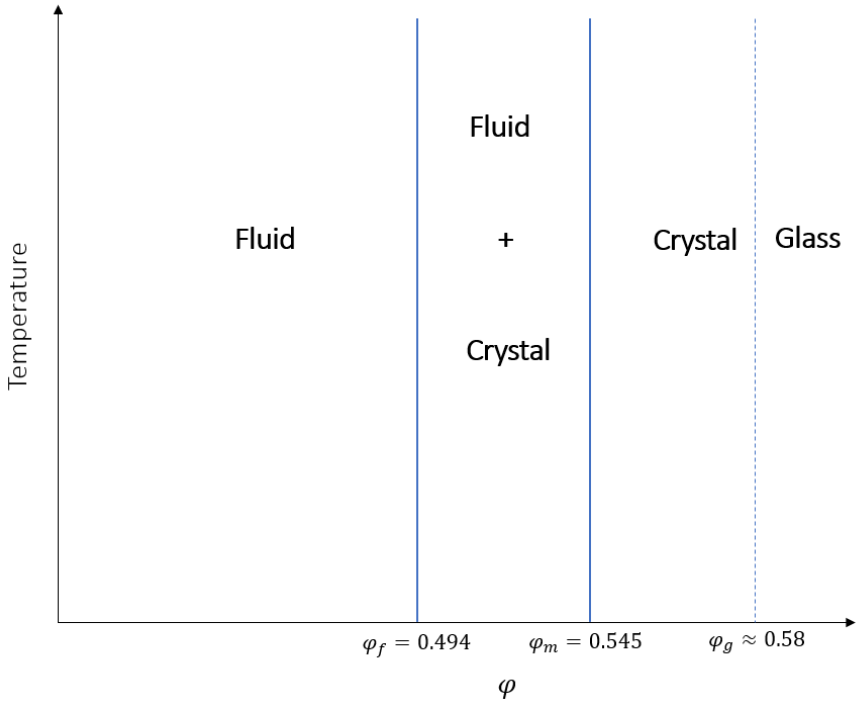

Figure 6.3: Schematic phase diagram of hard spheres in suspension where the two phases are fluid (F) and crystal (C). Indicated is the freezing, melting and glass volume fractions, which are described in more detail in the main text. The schematic is inspired by Anderson et al., [53], and Poon et al., [51].

When adding attractions to the hard sphere system a three-phase equilibrium will occur, which is similar to atomic systems. The simplest way to model attractions is the square well potential [47]. Therefore when modelling an atomic system, hard spheres with long-range attractions are used. This gives rise to gas, liquid and crystal phases, as seen in Figure 6.4. Modelling the hard sphere system with short-range attractions leads to similar phase diagrams as those of some globular proteins, see Figure 6.6. This gives rise to equilibrium between crystal and gas, as well as a metastable liquid-liquid transition, which will be described in more detail in the following sections [53].

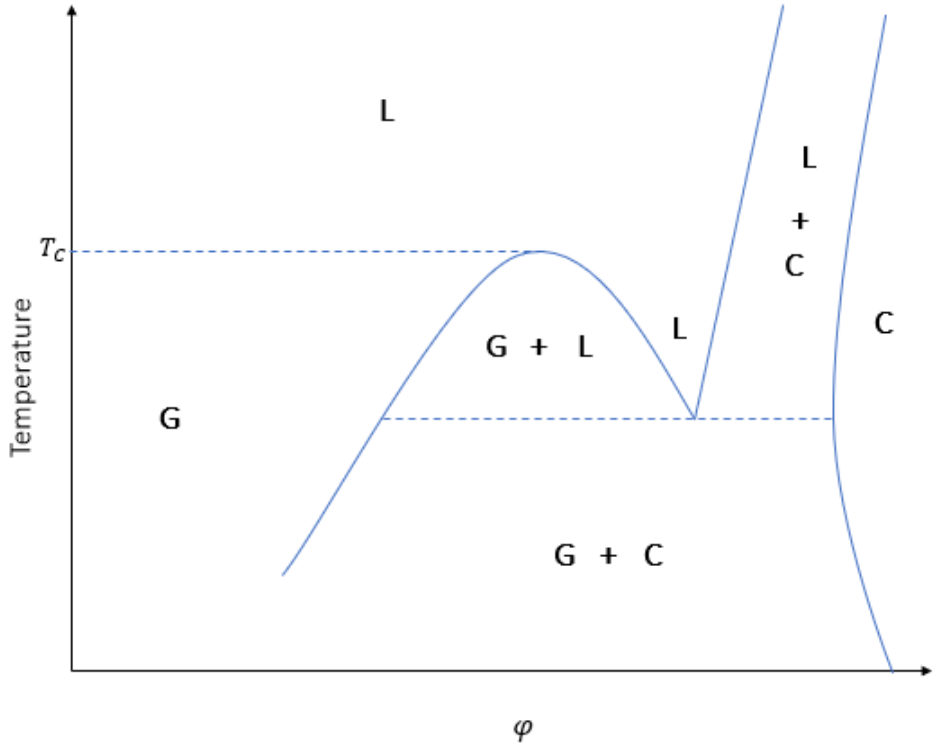

Figure 6.4: Schematic phase diagram of long-range attractive hard spheres in suspension where gas (G), liquid (L) and crystal (C) phases exist. This phase behaviour is also seen in atomic systems. The schematic is inspired by Anderson et al., [53].

6.1.2 Colloidal Systems

A good approximation to hard sphere systems, which cannot be found in the physical world, are spherical colloidal particles in suspension. The phase behaviour of such particles display similar states as those found in simple liquids and atomic systems, from gas to glass. Simple liquids are usually defined by having an interaction potential energy as the Lennard-Jones potential energy, which is given by

$$v(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{6.4}$$

In equation 6.4, σ is the collision diameter, defined as the particle separation when the potential is zero. r is the distance between particles and ϵ is the depth of the potential, at the minimum $v(r)$ value [50].

Colloidal systems are one type of small particles in a continuous matrix of another particle type. An example of a colloidal system is milk, where liquid fat is suspended in a water-based liquid solution. Colloidal systems are worth studying for a number of reasons. For one,

interactions in the system can be manipulated. Furthermore, the accessibility of glass in an atomic system takes rapid cooling, while it is reached by other means in colloidal systems [51, 54]. A phase diagram of spherical colloidal particles, such as those studied by Pusey et al., [54], can be theoretically obtained by an effective hard-sphere model [54].

Increasing the colloid concentration, leads to coexistence of liquid and crystal states, similar to the coexistence of a liquid and solid phase of atoms. Upon further increase in concentration, pure crystallized samples will appear. A colloidal glass state will appear when the colloid concentration is further increased, and the crystal phase is not obtained. The colloid glass is a viscous fluid, with amorphously arranged particles [54].

The addition of polymers to a colloidal system leads to attraction between the colloidal particles. The polymers, which cannot penetrate the colloids, are excluded from the area between two colloids in close proximity, giving rise to a depletion force, pulling the colloids together. In other words, an osmotic pressure is exerted on the colloid by the polymers, as they are excluded from the volume, which lead to colloid attraction [52, 53, 56]. The attraction leads to phase behaviour with similar features as that of molecules. For example, coexistence of gas and crystal, and three phase coexistence, also known as a triple point. Furthermore gel-like states can occur in polymer-colloid systems. When the polymers are small, the phases observed in the colloid system are fluid and crystal, as seen in Figure 6.3. When the polymers are larger, a colloid liquid, critical and triple points are reached [45, 51, 53].

Colloidal and polymer systems can undergo solid arrests in two forms. One is caging, where the colloidal particles are trapped, due to a cage formed by neighbouring particles. As such, one particle is both trapped and trapping neighbouring particles. This is the colloidal glass, which occurs around the volume fraction of 0.58, as previously mentioned for the hard sphere system [45, 56]. The other solid arrest is gel formation. Here short-range attractions give rise to aggregation into larger clusters, where the colloidal particles form bonds, which are reversible. The bonds between the larger groups of colloids make up the gel. Gelation has a sharp boundary in the colloidal phase diagram. Gelation may occur at low volume fractions as well, and it is sample preparation independent [56].

6.1.3 Globular Proteins in Solution

Proteins in solution also undergo phase transitions which depends on temperature, protein concentration along with specifics of the solution, which may be the pH, ionic strength, the buffer identity and additives. Studying phase diagrams of proteins is important in a variety of pro-

cesses including drug delivery, protein crystallization and industrial separations. A usual phase diagram for a globular protein in aqueous solution, is a two dimensional graph with the protein concentration or volume fraction as a function of one other parameter, while the remaining are held fixed [57]. Colloids and globular proteins share overall similar features in their phase diagrams. Proteins in solution may also experience gelation [53]. However, it is important to note that proteins are not perfect spheres nor isotropic, thus proteins are orientation dependent giving rise to aelotopic interactions, which leads to complexity as compared to colloids [50, 53]. For example a liquid-liquid phase separation is found for proteins, but not for colloidal systems [53].

The volume fraction of a protein solution can be determined by

$$\phi = \frac{c v_p N_A}{M} \tag{6.5}$$

Where c is the concentration of protein in solution, v_p is the molecular volume in solution, N_A is the Avogadro constant and M is the molecular mass of the protein [58].

The solubility curve is part of protein phase diagrams. Solubility is the concentration of protein in solution, at equilibrium, where a crystal gains and looses protein molecules with the same rate [57]. Thereby the solubility line marks the equilibrium between proteins in solution and the crystal phase [59]. Solubility depends on solution conditions. Below the solubility curve is the undersaturated region, where protein crystals will dissolve. The supersaturated region lies above the solubility line, here protein crystals will grow. Crystal growth is initiated with a nucleus, which is a cluster of proteins. Cluster formation occurs when an energy barrier is overcome. Thereby crystal growth does not experimentally appear just above the solubility curve, but at even higher protein concentrations. Thus the supersaturated region is divided into three zones, see Figure 6.5. Figure 6.5 is a schematic phase diagram of a protein in solution, displaying the solubility curve (blue curve) and the zones of the supersaturated region. The phase diagram shows how the solubility depends on precipitant concentration of the solution which the protein is dissolved in.

The zone right above the solubility curve is termed metastable zone. Here nucleus formation takes quite some time, such that protein crystal formation will not occur within reasonable timescales, as seen by experimentalists. In the labile zone, spontaneous nucleation formation occurs, as the supersaturation is large enough. When supersaturation becomes too large, formation of disordered structures occurs. These disordered structures may be aggregates or precipitates. At this supersaturation one finds the precipitation zone. The boundaries dividing

the supersaturated region into the three zones are not well defined, as these depend on kinetic phenomena, as compared to the solubility curve which is equilibrium dependent [57].

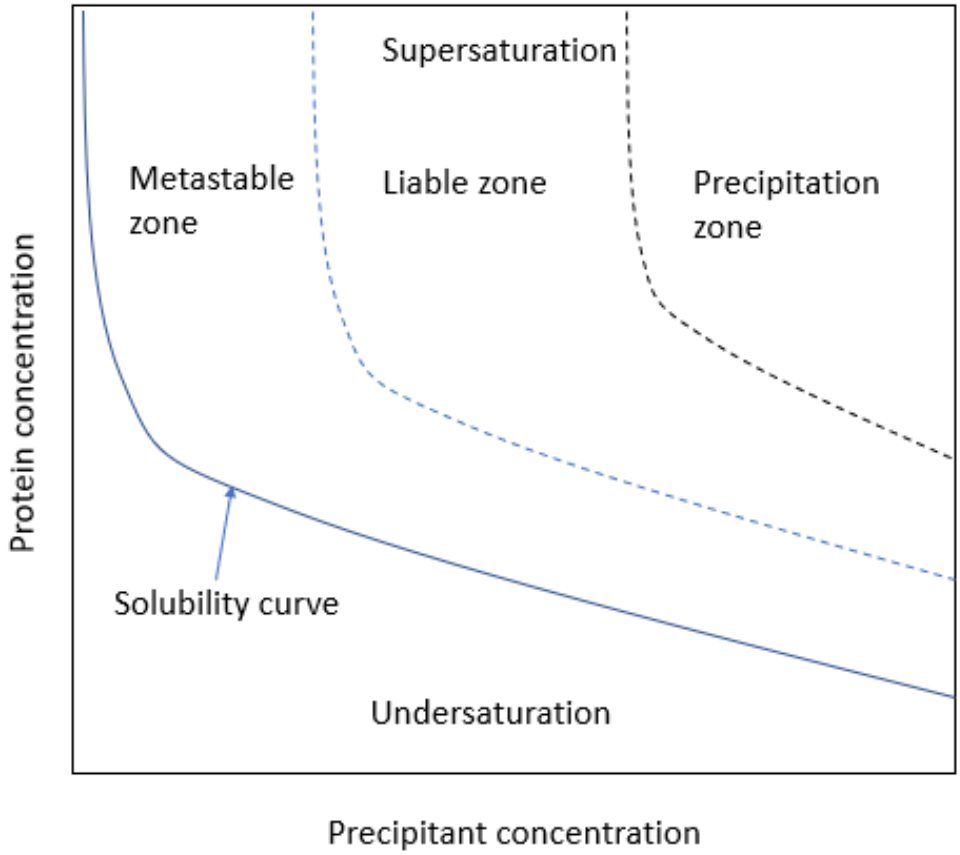

Figure 6.5: Schematic phase diagram of a protein in solution, displaying the solubility curve (blue line) along with the zones in the supersaturated region. Here the solubility curve is displayed as a function of the precipitant concentration in the protein solution. Schematic inspired by Asherie, [57].

Precipitants in protein solutions aid in the formation of crystals. However, sometimes upon addition of a precipitant, droplets may form. These droplets may become visible by altering solution conditions such as pH or temperature. Drops have a higher protein concentration than the remaining solution, and may separate from the solution due to gravity. The separation will eventually lead to the formation of two different liquid phases in the sample, one with a higher protein concentration than the other, thus none of the two liquids have the concentration of the initial protein solution. This is known as liquid-liquid phase separation and it insulates effective attractive interactions between the proteins in solution. In Figure 6.6 is a schematic curve of a phase diagram for proteins, including the liquid-liquid coexistence curve, displaying how the protein concentration varies with temperature. Above the coexistence curve, which is also known as the binodal, the protein solution is homogeneous, here no droplets will form

[60]. On the binodal in Figure 6.6, are two open circles which represent two different protein concentrations which are in equilibrium. Above these is a black circle which marks the critical point. At this point the two liquid phases are identical, and it marks the end point of phase separation in the liquid, given the protein concentration and temperature. The liquid-liquid phase separation is a metastable state. In the metastable state there is a possibility of crystal growth in the liquid with the higher protein concentration [57].

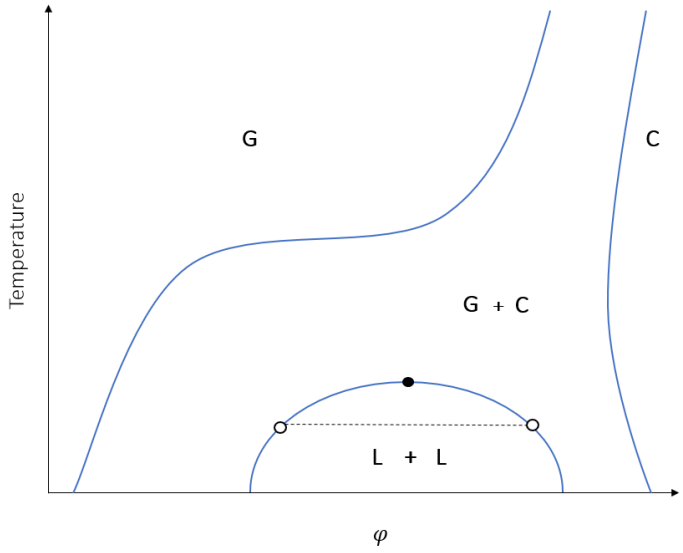

Figure 6.6: Schematic phase diagram for protein systems or short-range hard sphere systems. Here gas (G), liquid (L) and crystal (C) phases occur, as well as the metastable liquid-liquid coexistence. On the liquid-liquid coexistence curve, the two open circles indicate two concentrations of protein in equilibrium. The black circle indicates the critical point, which is where the two liquids in equilibrium will be identical and above this point, the solution is homogeneous. Schematic inspired by Anderson et al., [53] and Asherie [57].

The curves and boundaries of phase diagrams lead to more information than when crystals form, when the liquid is inhomogeneous or where the different states are most stable. From the solubility curve and the binodal numerical values of the entropy and enthalpy of the protein in solid and liquid form, can be determined [61]. For a dilute solution a fit using the van't Hoff equation may be performed on the solubility curve, in order to determine entropy, ΔS, and enthalpy, ΔH, of the crystal phase formation

$$\ln \phi = \frac{\Delta G}{RT} = \frac{\Delta H}{RT} - \frac{\Delta S}{R} \tag{6.6}$$

In equation 6.6, ϕ is the volume fraction of protein in solution, which can be expressed by the concentration c with units of mg/mL, and the specific volume ν in the units mL/mg, $\phi = c\nu$. ΔG is the Gibbs free energy of crystal formation, R the universal gas constant and T the

temperature [59].

Positions of the phase boundaries can be changed by additives, which provides information about the protein-additive interactions. Additives can be ions from salts, excipients etc. [61]. As the positions of the solubility and coexistence curves are related, one can estimate how additives affect phases based on movement of one curve or the other [57]. The addition of glycerol and salt to lysozyme has been found, by Gögelein et al. [61], to move the coexistence curve and the solubility curve to lower temperatures, as well as increasing the gap between the two curves [61].

Aggregation is a process which also occurs, however it happens deep in the phase space. Aggregates form clusters within the sample in different manners. One is named diffusion-limited cluster aggregation. This type of aggregation has a monodisperse distribution of cluster size and is governed by diffusion. Another is named reaction-limited cluster aggregation, which is polydisperse in cluster size. Here particles have to overcome an energy barrier before bonding. Aggregation may result in gelation with one large cluster spanning the entire sample [53].

7 Scattering

The two kinds of scattering experiments performed in this work are X-ray scattering and static light scattering. The scattering experiments yield structural information, such as shape of the insulin particles in solution, for example the protein tertiary structure, as well as information about the particle interactions. Light scattering also sheds light on details of the particles in the sample such as, molar mass, binding affinity and stoichiometry of interactions, for example the second virial coefficient, which may also be obtained from X-ray scattering [23]. In this section, the theory behind the scattering techniques is explained, along with elements of the different instruments used for experiments.

The atoms within an object illuminated by a monochromatic plane wave will become sources of spherical waves. Some of the radiation of the incoming wave passes through unaffected, while some of the radiation will be scattered. Scattering occurs from fluctuations in the sample medium. Light scattering is associated with variations in refractive index, while X-rays are scattered by electrons in the sample [62].

The incoming monochromatic wave can be described by the wavevector $k_i = |\mathbf{k_i}| = 2\pi/\lambda$. When scattering is purely elastic energy is conserved, i.e in the Rayleigh regime, the modulus of the outgoing spherical wave, k_s, is equal to k_i. This is Thomson scattering. Between the

wavevectors, one finds the scattering angle, 2θ, see the schematic in Figure 7.1. The scattering vector, $\mathbf{q} = \mathbf{k_s} - \mathbf{k_i}$, has modulus

$$q = \frac{4\pi \sin(\theta)}{\lambda} \tag{7.1}$$

In equation 7.1, θ is half of the scattering angle, 2θ, and the wavelength, λ, is of the incident beam [63]. The scattering vector, $\mathbf{q}$, describes changes in direction of the scattered beam [64]. The scattering vector is in units of inverse length [63].

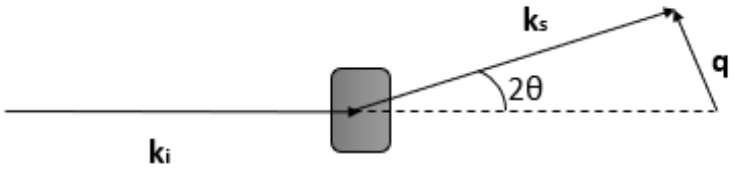

Figure 7.1: Schematic of a scattering experiment, with the incoming and outgoing waves, the scattering vector $\mathbf{q}$ and the scattering angle 2θ. Schematic inspired by Svergun et al. [63].

In a scattering experiment, a detector is placed at the scattering angle, 2θ, where it detects the intensity of the scattered waves. In general, scattering intensity fluctuates in time, therefore the detected intensity is a function of both scattering angle and time, $I(\theta, t)$. The region of the sample which is illuminated by radiation and seen by the detector is referred to as the scattering volume, V [62]. Proteins in solution are oriented randomly, which means the detected scattering from a sample is an average over all possible orientations of the proteins. The scattering intensity, therefore, is not dependent on the direction of the scattering vector $\mathbf{q}$ but the magnitude [64]. The interference pattern created by all the scatterers of the scattering volume, is the form factor, $P(\mathbf{q})$ [65]. Both the incoming and scattered waves of scattering experiments are typically shaped by optics, slits etc. [62].

7.1 Small Angle X-ray Scattering

Small angle X-ray scattering, SAXS, is a scattering technique which uses the interaction of electromagnetic waves with the electron clouds of atoms, to determine structural information of matter. When interaction occurs, wavelets are generated by the electrons, giving rise to coherent or elastic scattering, which can interfere either constructively or destructively. Small angle X-ray scattering is the constructive interfering wavelets from the electrons in a single molecule either in solution or in crystals [66]. In this study SAXS is used to determine the structure of insulin, a macromolecule, in solution.

X-rays are electromagnetic waves, usually set to travel in the z-direction perpendicular to the electric and magnetic fields. The wavelength of X-rays are in the Ångström region, which is 10^{-10} m [67]. The wavelength of the photons is given in terms of the photon energy E, as $\lambda = 1.256/E$. With the energy in keV and the wavelength in nm. When using X-rays for structural studies, the X-ray energy is usually about 10 keV, giving a wavelength between 0.10 - 0.15 nm [63]. The angles of small angle scattering are typically 0.1 to 10 degrees [65].

A volume element of the sample, $d\mathbf{r}$, at a distance $\mathbf{r}$, contributes to the scattered waves with a spherical phase factor and the electron density distribution $\rho(\mathbf{r})$. The form factor of this electron is given by $-r_0 P(\mathbf{q}) = -r_0 \int \rho(\mathbf{r}) e^{i\mathbf{q}\cdot\mathbf{r}}$. Where $r_0 = 2.82 \cdot 10^{-5}$ Å, is the Thomson scattering length, also called the classical radius of an electron, which is a fundamental length scale in SAXS. As can be seen, the form factor is the Fourier transform of the distribution of electrons. The form factor of a sample of several atoms is then given by [67]

$$P(\mathbf{q}) = \int \rho(\mathbf{r}) e^{i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} \tag{7.2}$$

The amplitude of the scattered waves is given by [63]

$$A(\mathbf{q}) = \int_V \Delta\rho(\mathbf{r}) e^{i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} \tag{7.3}$$

Where $\Delta\rho(\mathbf{r})$ is the excess scattering length density, which is the difference in scattering length between the sample and buffer solution, as the buffer solution also scatters. A constant scattering length density of the buffer is assumed. Since an ensemble of molecules with random orientation is illuminated during a SAXS measurement, one must determine the orientational average. The phase factor thus becomes [63]

$$\langle e^{i\mathbf{q}\cdot\mathbf{r}} \rangle_\Omega = \frac{\sin(qr)}{qr} \tag{7.4}$$

Where $\langle ... \rangle_\Omega$ denotes orientational average. The scattering intensity is then given by

$$I(q) = \langle A(\mathbf{q}) A^*(\mathbf{q}) \rangle_\Omega = 4\pi \int_0^{D_{max}} r^2 \gamma(r) \frac{\sin(qr)}{qr} dr \tag{7.5}$$

where $\gamma(r)$ is the spherically averaged autocorrelation function of the excess scattering density, given by

$$\gamma(r) = \left\langle \int \Delta\rho(\mathbf{u}) \Delta\rho(\mathbf{u} - \mathbf{r}) d\mathbf{u} \right\rangle_\Omega \tag{7.6}$$

When the particle size becomes larger than the maximal value, D_{max}, $\gamma(r)$ is zero. The inverse Fourier transform of $\gamma(r)$, yields the pair distance distribution function, $p(r)$ [67]

$$p(r) = \frac{r^2}{2\pi r^2} \int_0^\infty q^2 I(q) \frac{\sin(qr)}{qr} dr \tag{7.7}$$

The shape of the $p(r)$ function indicates the structure of the sample molecules [65, 66].

For a dilute monodisperse suspension of N identical particles, in the volume V, the scattering intensity is proportional to the form factor, which is characteristic of the shape of the particles [62, 65]

$$I(\mathbf{q})_{dil} = N_{dil} P(\mathbf{q}) \tag{7.8}$$

When the particle concentration of the sample is increased, the interparticle distances and intra-particle distances become similar in magnitude [65]. Therefore the scattering intensity is expressed with an extra term, as follows

$$I(\mathbf{q})_{conc} = N_{conc} P(\mathbf{q}) S(\mathbf{q}) \tag{7.9}$$

Where $S(\mathbf{q})$ is the structure factor, given by

$$S(\mathbf{q}) = 1 + \frac{1}{N} \left\langle \sum_{i=1}^{N} \sum_{j=1}^{N} e^{i\mathbf{q}\cdot(\mathbf{r}_j - \mathbf{r}_i)} \right\rangle \tag{7.10}$$

The average in the structure factor is an average of orientations and configurations [62, 68]. The particle-particle interactions are visible in the structure factor, at low q-values [65]. Experimentally the structure factor is obtained from the ratio of the intensity of a dilute sample, where $S(\mathbf{q}) = 1$, as well as a concentrated sample [62]

$$S(\mathbf{q}) = \frac{I(\mathbf{q})_{conc}}{I(\mathbf{q})_{dil}} \frac{N_{dil}}{N_{conc}} \tag{7.11}$$

The structure factor may be corrected by assuming particle sizes and positions are uncorrelated. This is known as decoupling approximation, which relates the true structure factor, $S(\mathbf{q})$, to the apparent structure factor $S'(\mathbf{q})$. The apparent structure factor is expressed as

$$S'(\mathbf{q}) = 1 = \beta(\mathbf{q})[S(\mathbf{q}) - 1] \tag{7.12}$$

where $\beta(\mathbf{q}) = |\langle F(\mathbf{q}) \rangle|^2 / \langle |F(\mathbf{q})|^2 \rangle$ and $\langle |F(\mathbf{q})|^2 \rangle = P(\mathbf{q})$. Decoupling approximation is valid

for spherical systems with low polydispersity [69, 70].

During a SAXS experiment, the intensity of the scattered waves is measured as a function of the scattering vector, **q**. The detected scattering data is two dimensional. This is then integrated around the centre of the beam, yielding a one dimensional curve, the scattering profile, which is the scattering intensity, $I(q)$, versus q. [64].

The scattering intensity, $I(q)$, is an even function, which may be expanded as a series of even powers of q. This was first done by Guinier in 1939, therefore named the Guinier approximation, in the range $r_g^2 q^2 \ll 1$

$$I(q) = I(0)\left[1 - \frac{1}{3}r_g^2 q^2 + O(q^4)\right] \simeq I(0)\exp\left(-\frac{1}{3}r_g^2 q^2\right) \tag{7.13}$$

The Guinier approximation is only valid for very small angles, where for globular proteins $qr_g < 1.3$ [63, 66, 71]. For an ideal monodisperse sample, the Guinier plot, a plot of $\ln(I(q))$ versus q^2, should be linear in the Guinier regime [33, 63]. If the Guinier plot is not linear, it indicates interparticle attraction or repulsion in the sample, which occurs from Coulombic repulsion. The intercept of the Guinier plot, yields forward scattering, $I(0)$, while the slope is related to the radius of gyration, r_g. r_g is characteristic of the particle size [63]. The forward scattering, can be expressed using the number of scattering particles per unit volume, N, the contrast $\Delta\rho$ and the sample volume V

$$I(0) = N\Delta\rho^2 V^2 = \frac{C\Delta\rho^2 v^2 M_w}{N_A} \tag{7.14}$$

In the last expression of the forward scattering above, C is the mass per unit volume, v is the partial specific volume, M_w the molecular weight and N_A is the Avogadro constant. Using this relation, the molecular weight of the scattering material can be determined [66].

The forward scattering, as well as the scattering intensity can be used to determine the second virial coefficient. This is done through the Zimm approximation,

$$\frac{I(0)}{I(q)} \cong \frac{1}{P(q)} + 2A_2 Mc + 3A_3 Mc^2 + ... \tag{7.15}$$

Where M_w is the molecular weight, c is sample concentration, A_i is the i'th virial coefficient. Virial coefficients are thermodynamic properties which describe interactions in the sample. [72, 73]. A negative value of A_2 indicates net attraction while a positive value indicates net repulsive interactions [74, 75].

Interparticle interactions are concentration dependent. Therefore, it is good practice to measure multiple concentrations of the same sample in order to detect if trends of the interactions are in fact trends and not protein aggregation or noise in the data. Protein aggregation may be detected in the I(q) vs q curve, the pair distance distribution function or the forward scattering. In the I(q) vs q curve, on a double logarithmic scale, an upturn of the curve at low-q values can be a sign of aggregation. The pair distance distribution function, $p(r)$, should approach zero when, r approaches D_{max}. If that is not the case, it indicates aggregation of the protein in the sample. If D_{max} is difficult to determine, it is indicative of interparticle interactions [66]. The forward scattering of attractive systems shows a maximum in values of I(0) with respect to sample concentration [77]. An increase in I(0)/c with increasing sample concentration, indicates aggregation or oligomerization in the sample, while a decrease indicates interparticle interactions [66].

Scattering data may also be represented by a dimensionless Kratky plot, a plot with $(\mathrm{qr}_g^2)\mathrm{I(q)/I(0)}$ vs qr_g. The shape of the Kratky plot indicates if sample domains are folded and if flexibility is possible. A bell-shaped curve, with a maxima around $\mathrm{qr}_g = 1.75$, with a value of 1.1 corresponds to proteins with folded domains. With increasing particle unfolding or elongation, the maxima shifts to higher qr_g values. If the tail of the curve, at higher qr_g, has an increasing upward slope, it also indicates particle unfolding [78].

Since everything scatters, the intensity of a SAXS measurement is a signal of everything in the path of the X-ray beam. This includes the capillary, solution and macromolecules. Therefore, to only analyse the scattering of the macromolecules in solution, background is subtracted. Thus it is very important to measure the buffer solution, and that the buffer solution is a perfect match for the sample. Furthermore, as SAXS measurements may be performed on in-house sources as well as at large scale facilities it is convenient to be able to compare data. This is possible with absolute scaling of the data. One way of absolute scaling requires a measurement of the empty sample capillary as well as a measurement of water [62, 66, 76].

7.1.1 Experimental Setup

The SAXS and WAXS experiments presented in this work were conducted using a BioSAXS-2000$^{\mathrm{nano}}$, see Figure 7.2. X-rays are generated by an FR-X, which is an Ultrahigh-Intensity Microfocus Rotating Anode X-ray Generator. Both parts are from Rigaku. The rotating anode in the FR-X is copper, producing X-rays with a wavelength of 1.54 Å [79, 80]. Flux of the FR-X is 8.4 $\cdot 10^9$ photons/s. The collimation in the BioSAXS is 2D Kraty. The beamstop is a PIN

diode detector. The SAXS settings of the BioSAXS has a q-range of 0.006 $Å^{-1}$ to 0.065 $Å^{-1}$. The WAXS settings have a q-range of 0.03 $Å^{-1}$ to 3.6 $Å^{-1}$. The BioSAXS detector is a HPC detector with a pixel size of 100 μm [81]. The BioSAXS has a two-dimensional Kratky camera which, for isotropic scattering, gives point collimation without intensity loss. The beam is focused by multilayer optics onto the detector [80]. For instrument calibration when measuring small angles, silver behenate is used. Silver behenate is suitable for observing Bragg peaks in the SAXS q-range [82]. For wide angle scattering calibration, silica particles are used.

Figure 7.2: Photo of the BioSAXS-2000 used for all SAXS and WAXS experiments of this work.

7.2 Light Scattering

Light scattering is a classical technique used for determination of molar mass and radius of gyration of biological and synthetic macromolecules in solution. By analysing composition dependence of scattering, information on the interactions of macromolecules in solution may be obtained [83]. Static light scattering, uses a larger wavelength than SAXS, and therefore goes to smaller q-values, between about 0.025 to 0.35 $Å^{-1}$ [75].

7.2.1 Static Light Scattering

During a static light scattering experiment, the average intensity of light, $I(c, \theta)$, as a function of concentration c and scattering angle θ, is detected. This intensity is light scattered by the sample in the scattering volume. $I(c, \theta)$ is related to the excess Rayleigh ratio, R, by the intensity of light scattered by pure solvent, $I(0, \theta)$, as follows

$$R(c, \theta) = k_{inst}[I(c, \theta) - I(0, \theta)] \tag{7.16}$$

Here, k_{inst} is an instrumental constant which can be determined through calibration. In the ideal limit, particles in a dilute solution behave as point particles. For particles in this limit the excess Rayleigh ratio is angular dependent and $R(c,\theta)$ is given by [83, 84, 85]

$$R(c,\theta) = K_{opt}\left(\frac{dn}{dc}\right)^2 McP(\theta)\left[1 - 2A_2P(\theta)c + O(c^2)\right] \tag{7.17}$$

Where dn/dc is the solute refractive index increment, M is solute molecular mass, A_2 is the second virial coefficient. K_{opt} is an optical constant determined by the wavelength of the incoming light in vacuum λ_0, Avogadros number N_A, and the refractive index of the solvent, n_0 as follows [74]

$$K_{opt} = \frac{4\pi^2 n_0^2}{\lambda_0^4 N_A} \tag{7.18}$$

$P(\theta)$ in equation 7.17 is a structure factor given by

$$P(\theta) = 1 - \frac{16\pi^2 n_0^2}{3\lambda_0^2}\langle r_g^2\rangle \sin^2\left(\frac{\theta}{2}\right) + O\left(\sin^4\left(\frac{\theta}{2}\right)\right) \tag{7.19}$$

r_g is the radius of gyration, $\langle r_g^2\rangle$ is the mean square radius of the solute. $\langle r_g^2\rangle$ is defined as

$$\langle r_g^2\rangle \equiv \sum_{i=1}^{n} \frac{(r_i - r)^2}{n} \tag{7.20}$$

With r_i being the location of scattering element i and r the solute particle center of mass. In the case where the scattering particles dimensions is much smaller than λ_0, $P(\theta) \approx 1$. In that case, the scattering molecules are considered isotropic scatteres [74, 84]. When K_{opt} and dn/dc are known, the molecular mass and second virial coefficient may be determined by extrapolating data, of various concentrations and scattering angles, to infinite dilution and a scattering angle of zero [8, 74].

Light scattering of specific particle interactions, such as site-specific association, is modelled as a composition-dependent polydisperse solution, to include both complexes and monomers of protein in solution. For reversible association, in the ideal limit, where particles of species i, in solution, do not interact the following expression, another version of equation 7.17, is used for modelling

$$\frac{R(\theta)}{K} = \sum_i \frac{R_i(\theta)}{K} = \sum_i M_i c_i P(\theta)\left(\frac{dn}{dc_i}\right)^2 \tag{7.21}$$

Where $R(\theta)$ is the excess Rayleigh ratio, as a function of scattering angle, θ. R_i is the excess Rayleigh ratio from each species, i. K is a constant, which depends on the system. M_i is the molar mass of species i, c_i is the concentration of species i and dn/dc_i is the refractive increment of species i [23]. n_0 of water with 150 mM NaCl, at 25°C is 1.3340 [86]. For proteins with a molecular weight less than a few hundred kDa, $r_g \ll \lambda_0/n_0$ and $P(\theta) \approx 1$, thus we again have isotropic scatterers [23, 74].

Non-specific particle interactions are modelled as thermodynamic non-ideality. Such interactions are not mutually exclusive and may occur simultaneously. These are usually weak interactions, compared to the specific interactions. For proteins in concentrations below ~ 1 g/L, the non-specific interactions are negligible [23, 84].

Static light scattering data is often represented in a two dimensional Zimm plot, with $\sin^2(\theta/2)$ along the first axis and $K_{LS}c/R(\theta)$ along the second axis. K_{LS} is the optical constant, K_{opt}, multiplied by $(dn/dc)^2$. This plot allows the molecular mass, M, radius of gyration, r_g, as well as the second virial coefficient, A_2, to be determined. An example of a Zimm plot is shown in Figure 7.3. In a Zimm plot, data from varying scattering angles and concentrations is presented, giving a pattern of vertical and horizontal lines. Vertical lines represent scattering angles while horizontal lines represent concentrations. The y-intercept is the reciprocal of the molecular mass of the scattering molecules, M. The slope of the vertical line representing zero scattering angle, is proportional to A_2. The slope of the line representing zero molecular concentration is proportional to the radius of gyration [75].

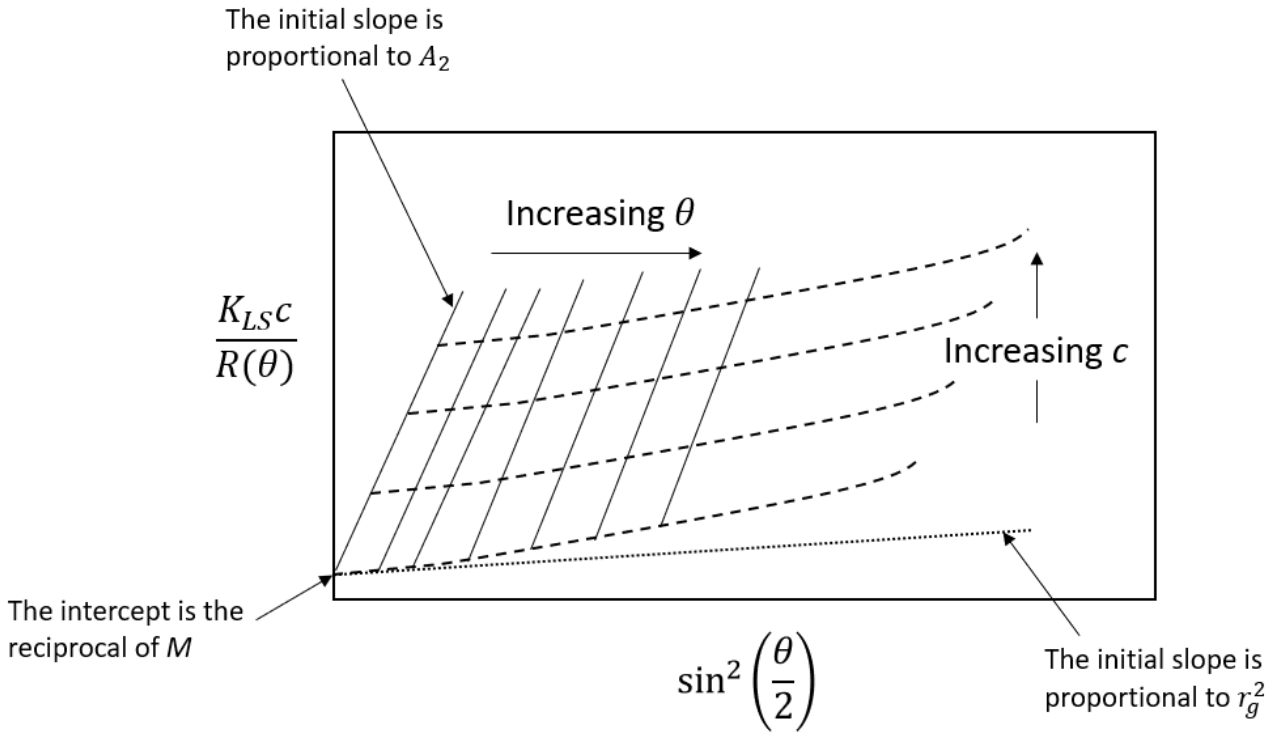

Figure 7.3: A schematic representation of a Zimm plot. The vertical lines represent scattering angles, with values increasing to the right, while the horizontal dashed lines represent concentrations, increasing upwards. The dotted line represents the limit of zero concentration. The slope of this line is proportional to the radius of gyration r_g The y-axis intercept value is the reciprocal value of the molecular mass, M. The slope of the first vertical line, representing zero scattering angle, is proportional to the second virial coefficient, A_2. Schematic inspired by Glatter, [75].

7.2.2 CG-MALS

Composition-Gradient Multi-Angle Light Scattering, CG-MALS, is an equilibrium technique, using differing compositions to analyse stoichiometry and affinity. This technique is used for light scattering measurements in this work. The CG-MALS technique involves automatic sample mixing, with sample pumped through two detectors, a light scattering detector and a concentration detector [8]. As the sample in the CG-MALS setup is freshly mixed, it allows for detection of protein association, dissociation or aggregation, assuming the process equilibrates within a detectable time frame. Optimal CG-MALS measurements performed on 10 kDa molecules, have an upper limit to the dissociation constant, K_D, of 10 μM [23].

The CG-MALS technique can be used for analysis of protein-protein interactions as well as protein self-association. A way of quantifying these associations in the ideal limit, that is for sample concentrations below 1 mg/mL, is through three equations. The first equation is

conservation of mass, which for self-association is as follows

$$c_{total} = \sum_i ic_i \tag{7.22}$$

Where c_{total} is the total protein concentration, this includes both free and bound monomers, i is the degree of self-association for each oligomeric state and c is molar concentrations of protein. The second equation is an equation of mass action

$$c_i = K_i c_1^i \tag{7.23}$$

Here c_i is the partial concentration of free i-mer while c_1 is the partial concentration of free monomers and K_i is the equilibrium association constant for the association of monomer-i-mer. Lastly is the equation of ideal light scattering,

$$R(0) = K_{opt}\left(\frac{dn}{dw}\right)^2 \sum_i (iM)^2 c_i \tag{7.24}$$

In the above equation, $R(0)$ is the excess Rayleigh ratio extrapolated to zero angle, the constant K_{opt} is defined in equation 7.18, dn/dw is the protein refractive index increment with respect to weight concentration, w, and M is the molar mass of the monomer [8, 83].

For sample concentrations above 1 mg/mL, one has to take thermodynamic non-ideality into account, as mentioned above. This non-ideality comes from non-specific particle interactions such as hard-core molecular repulsion, van der Waals interactions, dipoles, electrostatic interactions and so on. To account for non-ideality when modelling, virial expansion is taken into account. For a non-associating single species the virial expansion of $R(0)$ is given by [8, 84]

$$\frac{R(0)}{K_{opt}} = \left(\frac{dn}{dw}\right)^2 \frac{Mw}{1 + 2A_2Mw + 3A_3Mw^2 + ...} \tag{7.25}$$

The non-specific interactions of proteins, may be modelled as effective hard spheres, with effective specific volume v_{eff}, given by

$$\frac{R(0)}{K_{opt}} = \left(\frac{dn}{dw}\right)^2 \left(\frac{Mw}{1 + 8v_{eff}w + 30(v_{eff}w)^2 + ...}\right) \tag{7.26}$$

where the effective specific volume is determined from the effective radius, r_{eff} and mass, as

follows

$$v_{eff} = \frac{4\pi N_A}{3M} r_{eff}^2 \tag{7.27}$$

From comparing equations 7.25 and 7.26, an expression for the second virial coefficient of hard spheres in solution is obtained

$$A_2 = \frac{16\pi N_A r^3}{3M^2} = \frac{4v_{eff}}{M} \tag{7.28}$$

The effective hard sphere model, is not suited for modelling unscreened long-range charge-charge interactions. However it does suit many non-association proteins as well as the non-specific interacting proteins. The hard-sphere model can be simplified by assuming monomers and oligomers have the same effective specific volume, v_{eff}, using the total weight concentration w^{tot}

$$\frac{R(0)}{K_{opt}} = \left(\frac{dn}{dw}\right)^2 \left(\frac{\sum_i iMw_i}{1 + 8v_{eff}w^{tot} + 30(v_{eff}w^{tot})^2 + ...}\right) \tag{7.29}$$

Equation 7.29 describes light scattering from non-ideal self-associating proteins [8, 84].

The protein refractive index increment, dn/dc, is a value used to determine protein concentration. It describes the change in refractive index between the pure buffer and sample solutions, which arises from the protein concentration, c [84]. It is generally accepted that dn/dc for proteins is 0.185 mL/g. However Zhao et al., [86], argues that the protein refractive index increment of various human proteins have a distribution resembling a Gaussian, with a mean of 0.1888 mL/g and a standard deviation of 0.0025 mL/g for proteins larger than 100 kDa. For proteins smaller than 10 kDa, such as human insulin, the distribution of refractive index increments has a larger spread. The mean is 0.1902 mL/g and the standard deviation is 0.0053 mL/g [86]. Analysis results of the light scattering measurements of this work, analysed with 0.185 mL/g and 0.1902 mL/g are in very good agreement and almost equal. Therefore the generally accepted value of 0.185 mL/g is used for data analysis, for better comparison with other work.

7.2.3 Experimental Setup

The setup used to measure the static light scattering experiments presented in this work is a CG-MALS setup from Wyatt Technology, see Figure 7.4. The setup consists of a Calypso II pump system, which is connected to a DAWN HELEOS II 8, followed by an Optilab T-rEX

Differential Refractometer. The DAWN HELEOS II 8, a multi angle light scattering detector, consists of eight detectors, all located on the same side of the laser beam. The static signal is detected in seven different angles while the dynamic signal is detected at 108° [87]. The detector at 108° is a quasi elastic light scattering, QELS, detector. In the DAWN HELEOS II is a flow cell of fused silica. The laser wavelength is 661 nm. The Optilab T-rEX is a refractometer with extended range. The T-rEX contains both a sample and reference cell, which is used to measure the differential refractive index, dRI, between the flowing liquid and a reference. Furthermore the T-rEX measures the absolute refractive index, aRI [88]. The Calypso II pump system has three pumps, which are controlled by the Calypso software. In the pump system, inline filters are used to improve signal-to-noise by eliminating larger aggregates and dust particles [23].

Figure 7.4: Photo of the CG-MALS setup, used in this work. To the left is the Calypso II pump system, leading sample to the DAWN HELEOS II 8, detector (top of the two detectors on the right). Sample flows from the DAWN to the T-rEX Differential Refractometer, and out to the waste bottle in the middle. To the right of the detectors is the computer controlling the pumps and collecting data (not shown in the photo).

8 Experimental Design

In this work, the phase behaviour of insulin, induced by temperature and salt is investigated. Furthermore investigations of temperature effects as well as the effect of the product preservative is investigated by SAXS and CG-MALS measurements. For the temperature effects, the chosen temperatures are 5°C, 25°C and 40°C. The reasoning for this choice is to investigate effects at the temperature of storage, product handling as well as temperatures close to physiological

temperatures. In the following sections the sample preparation, data analysis and results are presented.

8.1 Sample Preparations

The laboratory assistant Camilla E. H. Jørgensen prepared 4 liters of full product Actrapid, with human insulin as the active pharmaceutical ingredient (API). This batch was prepared on March 10th 2020. On May 27th, the first preparations with the batch was started. With the aid of Camilla E. H. Jørgensen, two batches of 1 liter were prepared on July 14th 2020. One batch was full product Actrapid, the other was Actrapid without meta-cresol. All batches have a concentration of 100 units/mL, which is equivalent to 3.46 mg/mL [27]. Preparations with these batches started on August 28th.

All Milli-Q water, from here on MQ, used in all laboratory work is filtered and purified by a Millipore water purification system.

8.1.1 Concentrating

Actrapid samples are concentrated using Amicon Ultra Centrifugal filter units from Millipore, with a molecular weight cut off of 3 kDa. The filters contain remains of glycerine. Therefore, prior to use, filters are washed with MQ and centrifuged for 10 minutes with 10 mL MQ. Thereafter, filters are filled with 10 mL sample. The first $\sim$5 mL filtrate, is discarded due to dilution from excess MQ in the filter. All centrifugations are run at 4000 rpm, 25°C. Both filtrate and sample is collected from the centrifugal filter units. The filtrate from the centrifugations is used as buffer.

The Actrapid sample is centrifuged multiple times to reach the desired sample concentration. After a few centrifugations of the entire sample, agglomerates form. To remove the agglomerates and avoid further formation, the sample is manually filtered through 0.22 μm Millex GV filter units from Millipore, using 3 mL syringes. After the second or third centrifugation of the entire sample, the sample in the filter is homogenized when collecting it, in order to make sure all protein is collected. This is done using plastic pasteur pipettes, creating movement in the fluid until it is visibly homogeneous. A reversible turbid phase is reached, after the third or fourth centrifugation of the entire sample. The turbid phase appears as a white gel-like state in the bottom of the filter. The sample is homogenized using pasteur pipettes.

8.1.2 Dilution Series

When performing a dilution series, all pipetted volumes are weighed on a XSE105 DualRange Metler Toledo balance. The balance has been calibrated in 2018, showing an uncertainty of 8 μg. Weighing all pipetted volumes is done in order to determine the correct concentrations, once the stock solution concentrations have been measured with a NanoDrop. Before then, concentrations and volumes are calculated using the equation

$$c_1 v_1 = c_2 v_2 \tag{8.1}$$

which relates the initial volume, v_1, and concentration, c_1, with the final volume, v_2, and concentration, c_2, of a sample.

Dilutions are prepared from two different stock solutions, concentrated separately. One from a stock believed to be ∼100 mg/mL the other from a stock assumed to be ∼200 mg/mL. The table below shows all prepared sample concentrations

1.8 mg/mL	3.6 mg/mL	30 mg/mL	50 mg/mL	80 mg/mL	100 mg/mL
150 mg/mL	160 mg/mL	170 mg/mL	180 mg/mL	190 mg/mL	200 mg/mL

From the second stock, a dilution of 2 mg/mL is also prepared. The concentration of the 2 mg/mL sample is used to determine the stock solution concentration on the NanoDrop.

8.1.3 NanoDrop Measurements

The concentrations of the stock solutions are determined using a NanoDrop One C from Thermo Scientific. The NanoDrop One C uses a path length of 10 mm [89]. For all measurements a sample volume of 2 μL is used. One dilute sample of each stock is measured. Samples are well mixed before measuring, using a vortex. Each sample is measured in triplicates. If the absorbances are far apart in value, more measurements are performed.

8.1.4 Dialysis

Dialysis is performed to add salt to the insulin samples. Initially the salt concentration is 100 mM. 7 mL of the 100 mg/mL Actrapid sample is dialysed using three Slide-A-Lyzer Dialysis Cassettes from Thermo Scientific. The cassettes have a molecular weight cut off of 2.000, and a capacity of 0.5 - 3 mL. All three cassettes are placed in the same beaker with a 1 L volume, on a magnetic stirrer, running at 200 rpm. However, the sample became opaque. Therefore the salt concentration is diluted to 75 mM. This did not change the transition back, thus another

dilution is performed. At 50 mM salt, the beaker is temperature monitored using an IKA RCT basic stirrer and a contact thermometer. The temperature is set to 30°C overnight, without change. Therefore, the temperature is raised to 35°C. Within three days, the samples became homogeneous again, and were collected from the cassettes.

Since the dialysis of 100 mM NaCl ended with precipitation, a small experiment, is conducted, to test how much NaCl the Actrapid can handle. 300 μL of 100 mM NaCl solution is mixed with 300 μL of 200 mg/mL and 50 mg/mL sample, giving two 50 mM NaCl samples with concentrations of 25 mg/mL and 100 mg/mL. These were kept at room temperature.

8.1.5 SAXS Sample Preparation

For all SAXS measurements using the automated sample changer, 100 μL of sample is measured. Samples are loaded in 0.2 mL PCR tubes, in strips of eight, from Bio-Rad. When using the static sample holder, 40 μL of sample is manually loaded into a 1 mm capillary cartridge. Before manually loading sample in a capillary, it is washed with fresh MQ water, 2% Helmanex and more MQ water. Then the capillary is dried using compressed air TL-676 from Kema, and inspected for dust or other small particles before sample is loaded.

For SAXS experiments conducted with temperatures different from room temperature, sample temperatures are controlled using two Julabo F25-MA Heating Circulators. One circulator is connected to the sample storage of the automatic sample changer. The other is connected to the capillary sample holder in the chamber.

Prior to SAXS experiments for the study of temperature effects, samples are dialysed against buffer, to avoid buffer mismatch. For the 25°C samples, 0.5 mL concentrated Actrapid is dialysed, while for the 5°C samples 2 mL of concentrated Actrapid is dialysed. Both samples are dialysed against 500 mL buffer for 24 hours at room temperature, using Slide-A-Lyzer Dialysis Cassettes from Thermo Scientific. During dialysis the buffer is gently stirred.

8.1.6 CG-MALS Sample Preparation

Prior to CG-MALS experiments, all sample is filtered through 0.22 μm filter units from Millex GV using luer-lock syringes. Sample concentrations are 3.46 mg/mL for all measurements. For the two experiments with Actrapid at 25°C all buffer is also filtered using Stericup 250 mL Millipore Express filters of 0.10 μm. Such filters require vacuum which is provided by a vacuum pump from Buch Holm. On the Calypso II, in-house filters from Merck Millipore are inserted in filter housings of the pumps, which will be used in experiments. The filters are 0.1 μm PC

membrane filters, with a diameter of 13 mm.

For CG-MALS experiments conducted with a sample temperature of 40°C, the temperature is set on both detectors. Furthermore samples are placed in an eppendorf ThermoMixer C, from which the Calypso pumps load samples into the detectors.

8.1.7 Temperature Studies

For the temperature studies, 1 mL of each concentration of both series are stored in 5°C and 37°C for a week. The concentration series diluted from the 100 mg/mL sample are stored for two weeks in 30°C, before being moved to 37°C. After being stored at 37°C for a weeks time, the samples were removed.

A 60°C experiment is conducted with all samples using an eppendorf ThermoMixer C, with a 50 mL falcon tube holder, with room for four tubes. The 50 mL falcon tubes are filled with approximately 6 mL of water, and the 1 mL sample vials are placed in the water. The 1.8 - 160 mg/mL samples were tested at 50°C and 55°C before heating at 60°C.

A -10°C study is conducted using a Julabo F25-MA Heating Circulator, and glycol in the form of Antifreeze from Q8. The Julabo circulator has an uncertainty of 0.02°C [90]. A 1:1 mix of MQ and glycol is filled in the Julabo, which is then cooling the water mixture. The samples are all in small glass vials. These are held below the water level. The 1.8 and 150 mg/mL sample were held at -10°C for five minutes, while checked for changes every minute. The 200 mg/mL and 25 mg/mL 50 mM NaCl samples were held at -10°C for 15 minutes, and checked every 5 minutes.

8.2 Data Treatment

8.2.1 NanoDrop Measurements

The absorbance at 280 nm measured on the NanoDrop One C, is used to determine concentrations of the dilute samples. To do so, the method Protein A205 is used with scopes and a baseline correction of 340 nm. Prior to measuring sample, the NanoDrop is blanked with buffer. The absorbance values are normalized to a 1 cm pathlength equivalent [89]. As mentioned samples are measured in triplicates. Therefore, absorbances are averaged. Hereafter, the absorbance is used to determine the concentration through the Beer-Lambert law given by,

$$A = \epsilon l c \tag{8.2}$$

Where A is the absorbance, ϵ the extinction coefficient, l the pathlength of the beam and c is the concentration of the sample [91]. The extinction coefficient of insulin at 280 nm is 5734 $cm^{-1} \cdot M^{-1}$ [92]. The NanoDrop one uses a 1 cm path length [89]. Uncertainties on the dilute concentrations are determined using error propagation. The uncertainty of a NanoDrop measurement is 0.1 mg/mL for proteins in solutions of 0.1 - 10 mg/mL [89].

The concentration of the dilute sample is multiplied by the dilution factor, to give the concentration of the stock solutions. The dilution factor is the ratio between the volume of the dilute sample and the volume of stock added to the dilute sample. Upon sample preparation volumes are weighed. From the weights, concentrations of the remaining samples are determined using equation 8.1. The uncertainty of the weight is 0.00008 g.

8.2.2 CG-MALS Data

Data measured by the CG-MALS system, may be collected both using the CALYPSO and ASTRA software. Both softwares are developed by Wyatt Technology. In the ASTRA software, data from the QUELS, or dynamic light scattering detector may be analysed. In the CG-MALS set-up used, detector 6 is the QUELS detector. Both softwares allows analysis on data from all other detectors as well as the refractive index from the T-rEX instrument. In this work, data is only collected through CALYPSO version 2.2.4, therefore there is no QUELS data preseneted.

In the CALYPSO software, DAWN parameters as well as calibration parameters of the instruments are used in the analysis process. The solvent is set to water, with refractive index 1.33. The Zimm Formalism analysis type is used, with order in concentration and angle set to one. For the setup used, the concentration channel is the refractive index data. The dn/dc value used for all samples is 0.185 mL/g. The baseline and normalization of data is manually selected. For the normalisation, standard R_g is set to three, and one plateau of the data is used. All plateaus of the data are used for A_2 and molecular weight calculations as well as the generated Zimm plot. The plateau width is set to 10. The software also outputs a plot of R/K vs concentration and a plot of apparent molecular weight vs concentration. If one plateau has a larger peak in the middle it is not used for analysis. Values of A_2 and the molecular weight, are used in this work, the generated plots are checked but not used.

During a CG-MALS experiment, the calypso pumps are controlled to produce a stepwise gradient of solute concentration, from the stock solution concentration diluted to the buffer concentration. This produces a result as displayed in Figure 8.1. In the Figure the signal is calibrated using the above described method for background and normalisation selections.

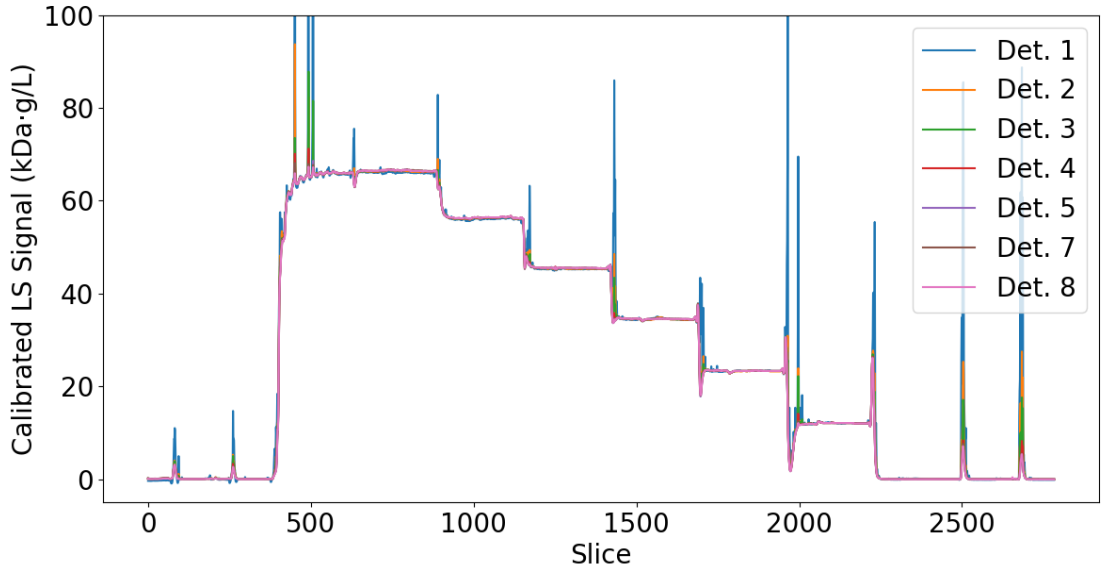

Figure 8.1: Example of calibrated light scattering signals from the seven static light scattering detectors of the DAWN HELEOS II 8 instrument. The data has been calibrated using the CALYPSO software from Wyatt Technology. Such signals are used for data analysis, as described in the main text.

8.2.3 SAXS Data

When measuring SAXS data, the image display is refreshed every five minutes, giving a total of six images, or curves, for each 30 minute measurement. Each curve is inspected to make sure they overlap, giving a correct average. When that is not the case, the intensity on the detector throughout the measurement is checked. If there is a bubble in the sample, it becomes visible as a peak in the intensity, which affects one or more images. If the sample is drifting out of focus, it is visible as a large upturn in the intensity on the detector. Once the average is approved, the data is background subtracted, using the SAXSLab software, version 4.0.2b4, by Rigaku. Before background subtraction all curves are normalised with the transmission factor, also in the SAXSLab software.

Data is absolute scaled using measurements of the empty capillary and MQ. The absolute scaling is performed using DATABSOLUTE from the ATSAS software package version 2.8.4. Concentration normalisation is performed by running DATOP, also from ATSAS. The structure factor of data is obtained by dividing scattering intensities with the intensity of a merged scattering profile. A merge of the low-q range of the lowest sample concentration with the high-q region of a sample concentration above 5 mg/mL. The division is performed using DATOP. Concentration normalized data is used for further analysis. The scattering intensity

at q = 0.0117 $Å^{-1}$, is used as the forward scattering.

SAXS data is fitted against PDB structures using OLIGOMER, [32], from ATSAS. For dilute samples the fitting starts at q = 0.02 $Å^{-1}$, while for samples around or above 3.46 mg/mL, fitting starts at q = 0.05 $Å^{-1}$. Theoretical scattering curves of the different hexamers, and the monomer and dimer is obtained from CRYSOL, [31]. Fitting hard sphere models to structure factors and scattering curves is performed in SasView version 5.0.3. For the polydispersity a Gaussian distribution is used. The Guinier and Kratky analysis is performed using BioXTAS RAW version 2.0.3 [93]. Pair distance distribution function are obtained using PRIMUS from ATSAS [32]. When determining pair distance distribution functions, the value of D_{max} is varied to see how the p(r) curve behaves with D_{max}, in order to try and limit over- or underestimations of D_{max}.

The normalized scattering profiles of Actrapid measured at 5°C and 25°C are compared with scattering profiles of Actrapid with known concentrations. This is done to scale curves with concentrations in order to eliminate concentration errors from the uncertain NanoDrop measurements due to the presence of m-cresol.

9 Results

All scattering curves are absolute scaled and normalized by sample concentration, unless otherwise stated. Data from CG-MALS measurements are only collected using the CALYPSO software. Therefore, there is no dynamic light scattering data to present.

9.1 Phase Transitions

Sample concentrations are determined from one dilute sample of each stock, the dilution factor and volumes of stock used to dilute. In the table below, concentrations of all samples are listed. Errors are determined from error propagation, based on uncertainties of the balance and the NanoDrop.

Sample	Concentration (mg/mL)
1.8	1.70 ± 0.17
3.6	3.22 ± 0.33
30	26.79 ± 2.74
50	48.93 ± 4.49
80	70.02 ± 7.16
100	87.45 ± 8.94
150	117.06 ± 11.59
160	124.51 ± 12.33
170	131.74 ± 13.05
180	138.98 ± 13.76
190	146.75 ± 14.53
200	154.63 ± 15.31

Table 9.1: Concentrations of the twelve Actrapid samples, determined from NanoDrop measurements. The errors are based on error propagation, with uncertainties from the balance used to weigh pipetted volumes and the NanoDrop.

After concentrating 1 litre of Actrapid, the sample was stored in the fridge. It should be noted, that the sample had not been filtered through 0.22 μm filters, during concentrating. On the third day of storage, the sample is observed to be a little turbid, with spherical formations, as can be seen in Figure 9.1. After being at room temperature and gently stirred for about ten minutes, the sample is completely clear and homogeneous again. Therefore the 155 mg/mL Actrapid sample is not stored in the fridge, but at room temperature on the lab bench.

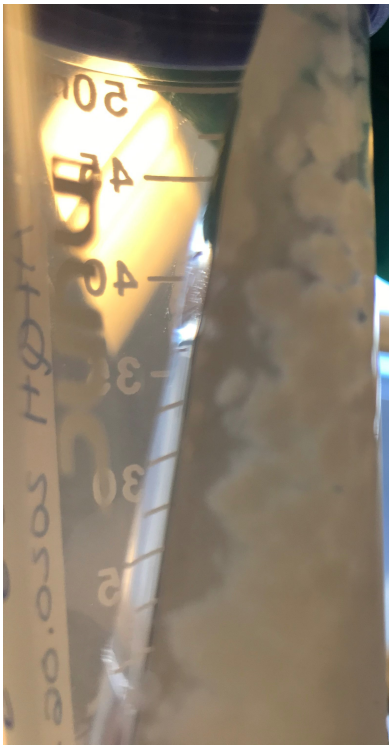

Figure 9.1: Photograph of the 155 mg/mL Actrapid sample after sample storage in 5°C for three days. As can be seen there appears to be spherical formation in the more turbid part of the sample.

9.1.1 Temperature Induced Transitions

One mL of samples with concentrations from 1.7 to 88 mg/mL were stored at 30°C for two weeks, without visible changes to the insulin solutions. Thereafter, the samples were moved to 37°C, where they were stored for a week. One mL of samples with concentrations from 117 to 155 mg/mL were also placed at 37°C. There were no visible changes to any of the samples. Thus, the 60°C experiment was started. Changes started to be observed in the higher concentration samples after about 48 hours, where samples appeared translucent. After an additional 24 hours samples turned turbid. Some samples then transitioned into gel, after about a week at 60°C. All samples except the 1.7 and 3.2 mg/mL samples started the transition to a gel, not all reached the gel stage before the experiments ended.

The turbid sample of 139 mg/mL, stored at 60°C has been measured in the SAXS. This was one of the samples that did not reach complete gelation at 60°C. In Figure 9.2 the scattering curve can be seen. From the curve, it is apparent by the upturn in the low-q region, that the sample has aggregated. Looking at the higher-q region, the sample still appears to have some structure, resembling that of a hexamer, as seen in the theoretical scattering curves in Figure 2.4. To check the sample for an amyloid peak at q = 1.3 $Å^{-1}$ a WAXS measurement was

performed. There is no amyloid peak in the scattering profile, which is shown in Appendix A.

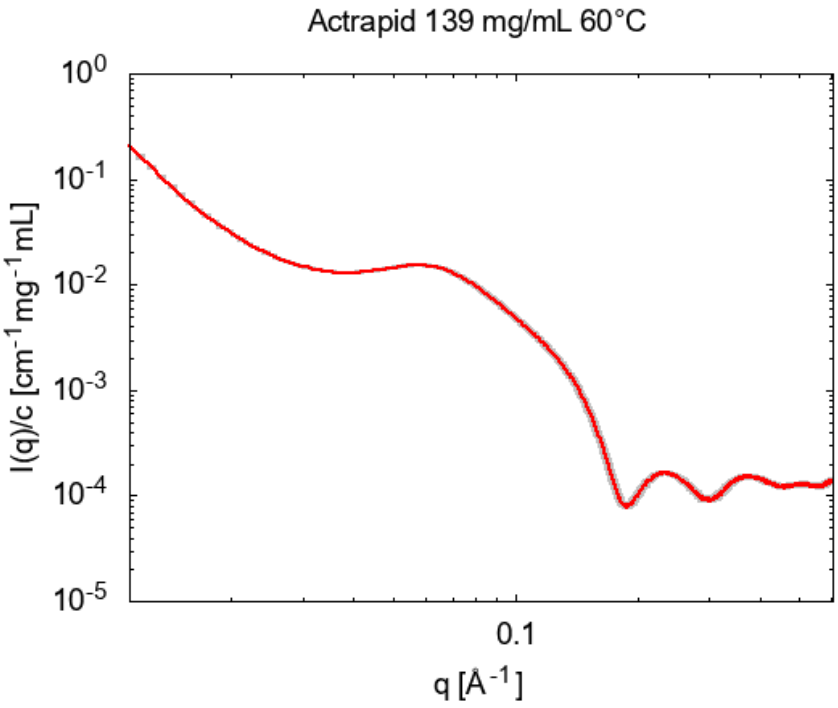

Figure 9.2: Scattering pattern of the turbid 139 mg/mL Actrapid sample stored at 60°C. Data is absolute scaled and concentration normalized.

One mL of sample of each concentration listed in table, 9.1 were stored at 5°C. Within two days the 155 mg/mL sample formed the same spherical structures as seen during sample preparation, see Figure 9.1. Within a weeks time the same formation was visible in the 147 mg/mL sample. After two weeks time, the 139 mg/mL sample had also started to form the same spherical structures, as the two other samples. Upon moving the samples from 5°C to room temperature, the samples became homogeneous within 5 to 10 minutes. There were observable changes in samples of 117, 125, 132 and 139 mg/mL after months time stored at 5°C. Samples were investigated in cross-polarized light, where the spherical formations appear birefringent.

The temperature study at -10°C was performed on samples with concentrations of 1.7, 117, 132 and 155 mg/mL. There were no visible changes in the samples.

All the above mentioned observations are plotted in a phase diagram, shown in Figure 9.3. In the Figure are two photographs. One of the transition found at 5°C, taken when the sample is in the cross-polarized light, where the sample appears birefringent, and the birefringent parts

of the sample appear somewhat spherical in formation. The other photograph is of the gel like transition found at 60°C. In the Figure the birefringent sediment is represented by blue triangles, the gelation is represented by green rhombi, while all observed temperatures and concentrations at which samples are homogenous are represented by dark blue dashes.

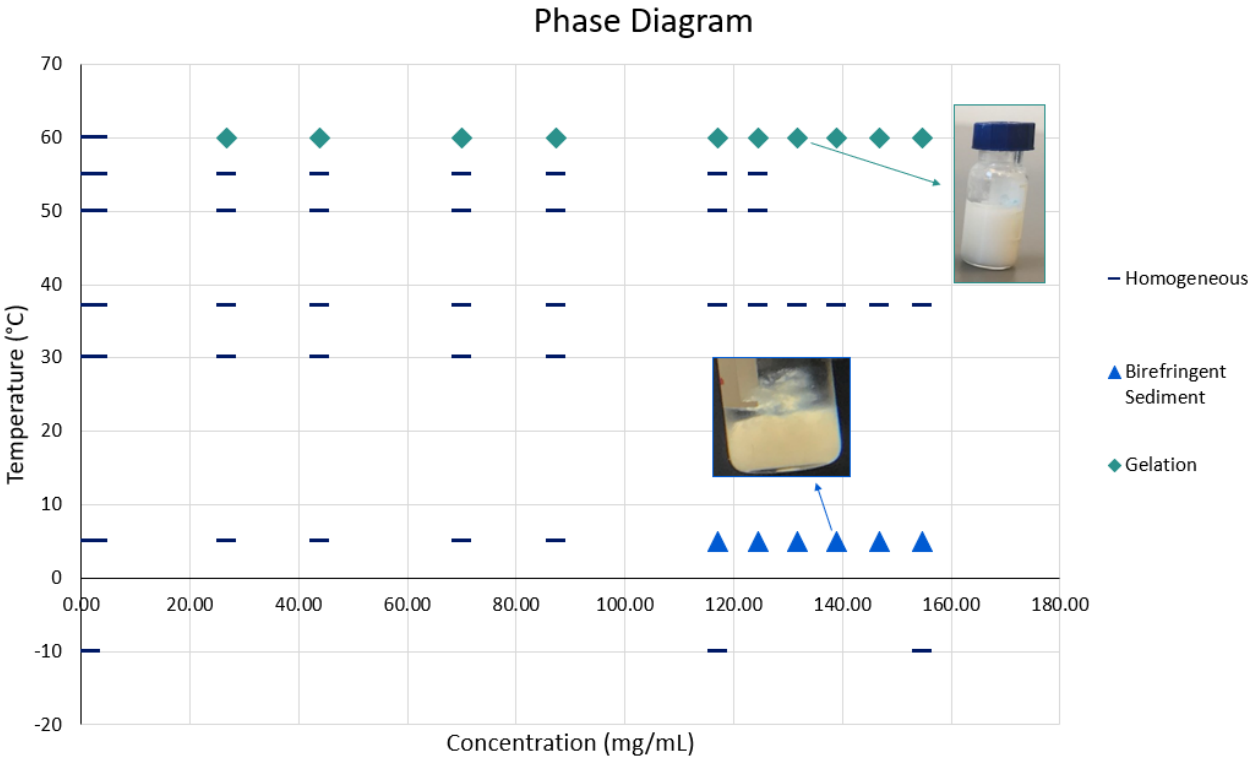

Figure 9.3: Phase diagram of the transitions observed in the insulin samples, including photographs of the birefringent sediment observed at 5°C and the gelation observed at 60°C. For the specifics of the transitions see the main text.

9.1.2 Salt Induced Transitions

During dialysis of the 88 mg/mL sample against buffer with 100 mM NaCl, the sample precipitated at room temperature. First the sample, divided in three dialysis discs, became translucent, then turbid and then precipitation was visible at the bottom of the dialysis cassettes. In order to reverse this, the salt concentration was decreased to 50 mM. As this change had no effect on the sample, the temperature was increased to 35°C. Within three days, the sample was homogeneous once again. The sample volume had increased from 7 to 9 mL. Therefore the final sample concentration is unknown. Storing the sample at 5°C had no transition effects.

A small salt study was conducted, mixing 44 and 155 mg/mL samples with 100 mM NaCl resulting in two samples of 22 and 77 mg/mL respectively, with 50 mM NaCl. Within the first 5 - 10 minutes after mixing, the sample of 77 mg/mL became turbid. After about an hour the

sample was more translucent, and about six hours after mixing, precipitation had occurred. There was no change in the 22 mg/mL sample. Both samples were kept at room temperature for the entirety of the experiment. The 22 mg/mL sample with the 50 mM NaCl, was also tested for phase transitions in -10 °C. After 15 minutes, the sample became slightly translucent. Afterwards, the sample was stored at 5°C. Within 15 hours the sample had precipitated, as found with the 77 mg/mL sample at 25°C. As the 88 mg/mL sample in the dialysis discs became homogeneous at 35°C, both precipitated samples of 22 and 77 mg/mL were placed at 40°C. After about 40 minutes, the 22 mg/mL sample was homogeneous. The 77 mg/mL sample did not turn fully homogeneous after three days.

The above mentioned observed transitions are plotted in a phase diagram which can be seen in Figure 9.4. Included in the diagram are photographs of the turbid 88 mg/mL sample in a dialysis disc as well as three photographs of the transitions of the 77 mg/mL sample, from turning turbid to precipitation. In the diagram, the concentrations and temperatures where samples turned homogeneous are also marked, that is the blue dashes, which are samples with 50 mM NaCl. The green rhombi represent the precipitated samples with 50 mM NaCl and the red circles represent precipitated samples with 100 mM NaCl.

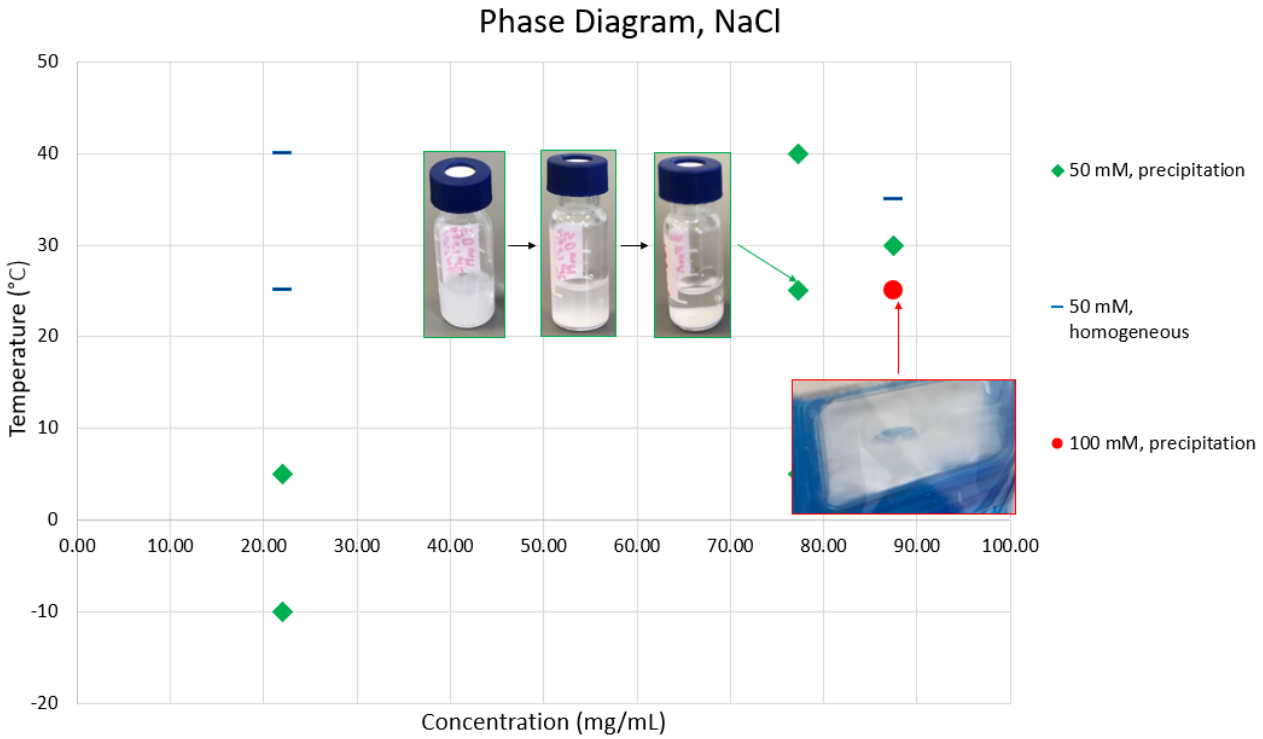

Figure 9.4: Phase diagram of the transitions observed in the insulin samples with additional NaCl of 50 and 100 mM. For some transitions, heating resulted in reversibility of the precipitation observed, as described in the main text. The photographs of the small glass vials, show the three steps of the precipitation of the 77 mg/mL sample at room temperature, starting with the sample turning turbid after 5 - 10 minutes after mixing, and ending in precipitation. The photo of the dialysis disc, shows the sample before precipitation where it has turned turbid, as the 77 mg/mL sample.

9.2 Temperature Effects

Four CG-MALS experiments at 40°C have been conducted. Furthermore two measurements of Actrapid from an older batch have been performed at room temperature. In Figure 9.5 are the second virial coefficients of each experiment. As can be seen from the Figure, all virial coefficients are positive in value, thereby net interactions in the Actrapid are repulsive. The second virial coefficient is larger in magnitude for all samples measured at 40°C, as compared to the Actrapid at 25°C indicating the temperature increase induces repulsion in the Actrapid sample. All A_2 values and errors are obtained from data analysis in the CALYPSO software.

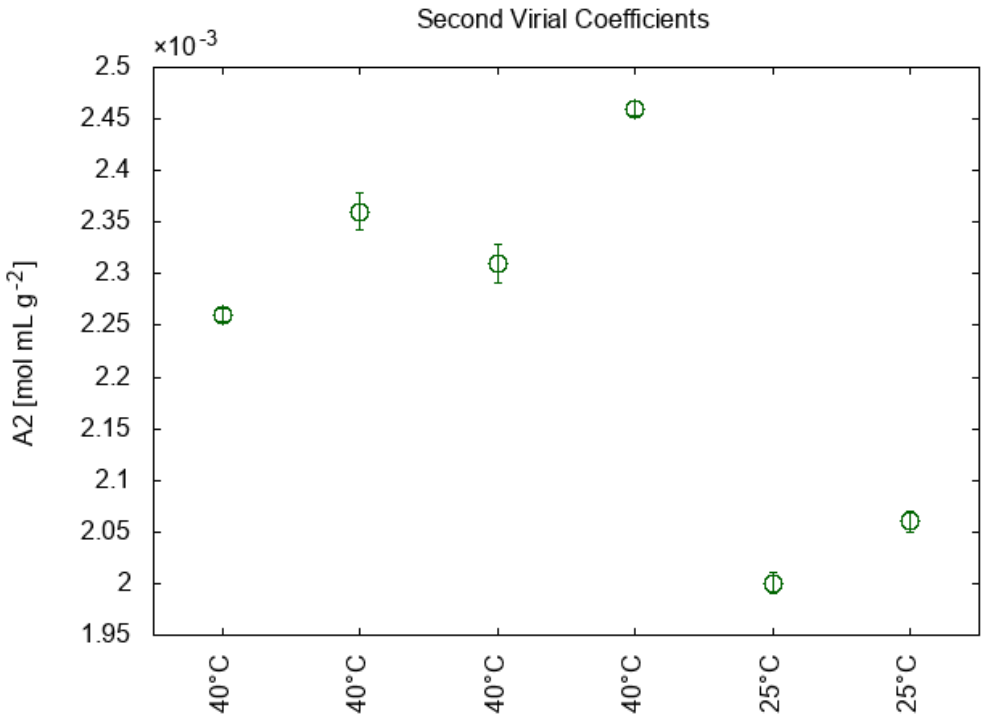

Figure 9.5: Graph of the second virial coefficients of Actrapid samples measured at 25°C and 40°C obtained from CG-MALS measurements. Values and uncertainties obtained from data analysis using the CALYPSO software.

The molecular weights from the CG-MALS measurements are listed in table 9.2. The theoretical molecular weight of an insulin hexamer is 34.85 kDa. Thus the obtained molecular weights in table 9.2 all indicate the oligomeric state of the insulin is hexameric. All weights and uncertainties are obtained from data analysis in the CALYPSO software.

Sample	Molecular Weight (kDa)
40°C	34.63 ± 0.04
40°C	36.18 ± 0.12
40°C	35.81 ± 0.12
40°C	35.98 ± 0.04
25°C	32.54 ± 0.06
25°C	33.08 ± 0.06

Table 9.2: Molecular weights of each Actrapid sample, determined from CG-MALS measurements at 25 and 40°C.

SAXS experiments of 25°C and 5°C have been performed on samples dialysed against buffer, to avoid buffer mismatch. Sample concentrations of these two experiments are listed in table 9.3.

The 3 mg/mL sample measured at 25°C has not been dialysed against buffer, as this sample has been measured at an earlier stage than the remaining 25°C measurements. Concentrations are initially measured on the NanoDrop. Thereafter the analysed scattering curves are compared to Actrapid scattering curves of samples with known concentrations and scaled accordingly.

25 °C	5 °C	
2	1	67
3	3	73
9	7	82
56	22	97
66	46	114
111	62	133
133		

Table 9.3: List of sample concentrations all in mg/mL, for samples measured in the SAXS at temperatures of 25°C and 5°C.

In Figures 9.6 and 9.7 the scattering curves of the 25°C and the 5°C SAXS measurements are shown. From the curves, it appears in all samples of concentrations above 3 mg/mL, that the characteristic shape of a hexamer as seen in the theoretical scattering curves of Figure 2.4 is visible. Furthermore, in the mid-q region, a peak starts to form and grows with increasing concentration. This peak, as well as the behaviour of the low-q region is more visible in Figures 9.8 and 9.9 showing a zoom of the lower q-regions of the scattering. In the 25°C experiment, Figure 9.8, the peak appears in the samples of 66, 111 and 133 mg/mL, and does not appear to change in shape or maximum with increasing sample concentration. For the 5°C samples, the peak maximum shifts towards higher q values with increasing concentration. Furthermore, the peak intensity is at a maximum for the 82 mg/mL sample, and decreases slightly at higher concentrations.

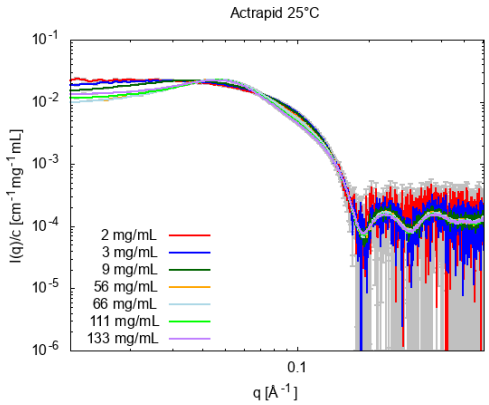

Figure 9.6: Scattering curves of the 25°C Actrapid scattering experiment.

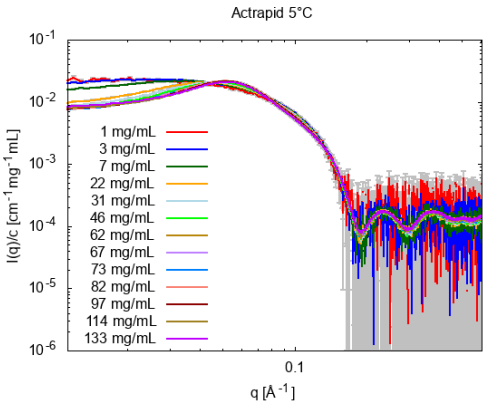

Figure 9.7: Scattering curves of the 5°C SAXS experiment.

From the zoom of the scattering curves in Figures 9.8 and 9.9, the behaviour of the curves in the low-q region is better visible. In the 25°C experiment, Figure 9.8, the curves of concentrations above 2 mg/mL all indicate repulsion in the sample, as is in agreement with the CG-MALS A_2 results. This is indicated by the decrease in the low-q region of the curves. The decrease in low-q increases with concentration until, 111 mg/mL where, the decrease takes an upturn. A feature which is not as expressed in the 5°C sample in Figure 9.9. In the 5°C data, all curves above 3 mg/mL show repulsion in the samples. There is a slight upturn in the 133 mg/mL sample, compared to the rest of the curves.

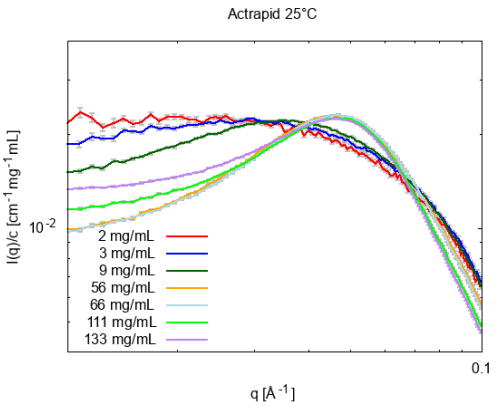

Figure 9.8: A zoom of the lower q-region of the scattering curves measured at 25°C

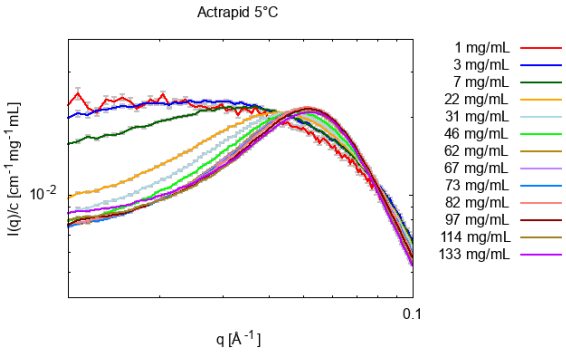

Figure 9.9: A zoom of the lower q-region of the scattering curves from the 5°C SAXS measurement.

In Figure 9.10 the forward scattering of the 5°C, green points, and 25°C, orange points, is plotted against sample concentration. From this Figure, the behaviour of the 25°C curve is further visible. The forward scattering has a minimum around 60 mg/mL, between the 56 and

66 mg/mL samples. The forward scattering of the 5°C curves, show a slight increase in I(0) for samples of concentrations above 100 mg/mL. The increase is not as visible in the data, as compared to the 25°C data.

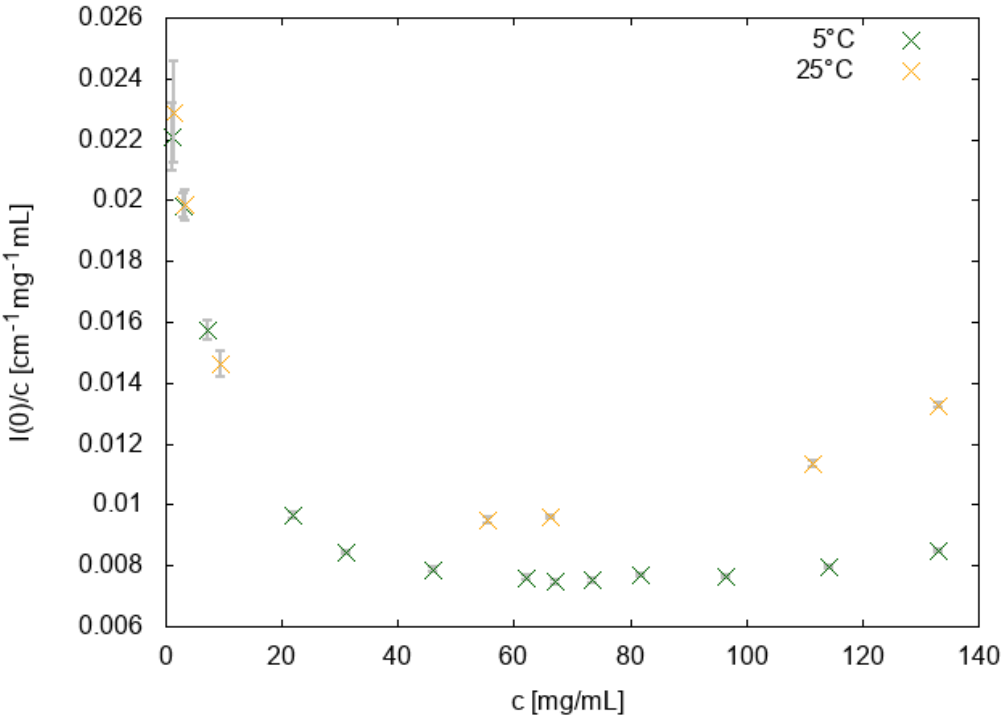

Figure 9.10: Plot of the I(0) values for the 5°C, green points, and 25°C, orange points measurements. I(0) values and uncertainties from scattering intensities at q = 0.0117 Å^{-1}.

The above mentioned peak at mid-q in the scattering curves, is also visible in the structure factors of the scattering curves. The structure factors are shown in Figures 9.11 and 9.12. In the structure factors of the 25°C data, Figure 9.11, the peak hardly changes in shape with increasing concentration. The minimum of the structure factors after the peak, decreases with concentration. The two samples with concentrations above 100 mg/mL, have the lowest minima. In the 5°C structure factors in Figure 9.12, the peak grows with concentration. In this representation of the data, it is not clearly visible if the peak maximum is shifting with concentration. The minimum after the peak, decreases with concentration. The minimum is not as pronounced in the 5°C data as compared to the 25. data.

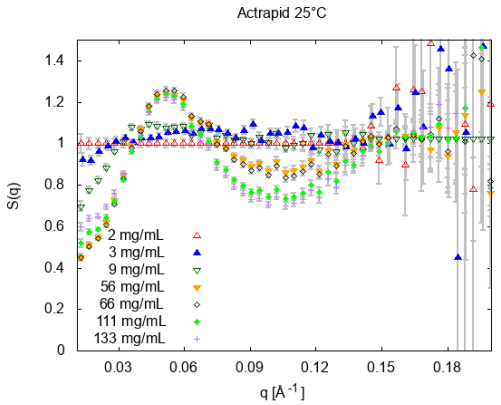

Figure 9.11: Structure factors of the 25°C SAXS data.

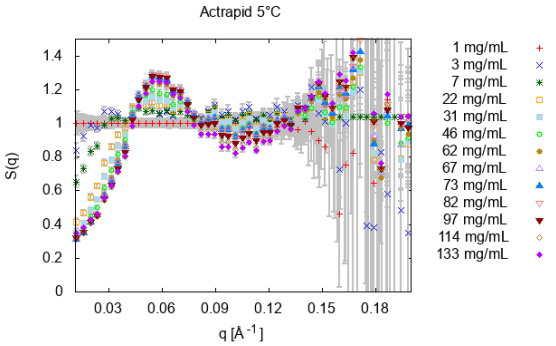

Figure 9.12: Structure factors of the 5°C SAXS data.

In Figures 9.13 and 9.14, the Guinier region of the scattering data is shown. From the Guinier linear fits, the radius of gyration and forward scattering is obtained. These values are listed in Appendix B. Samples with concentrations below 10 mg/mL, have a linear Guinier region, indicating monodisperse solutions. Samples above 10 mg/mL have a smaller linear Guinier region due to interparticle repulsion.

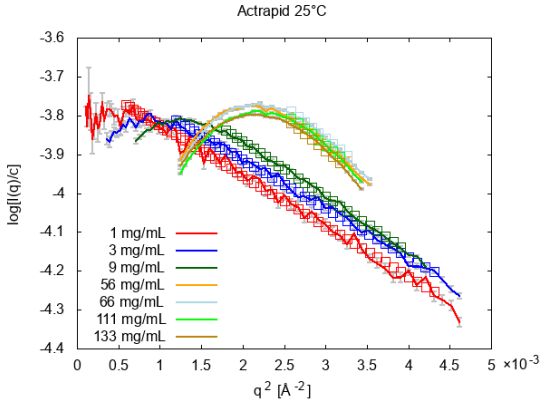

Figure 9.13: Plot of the Guinier region of the 25°C scattering data, along with Guinier fits (open squares).

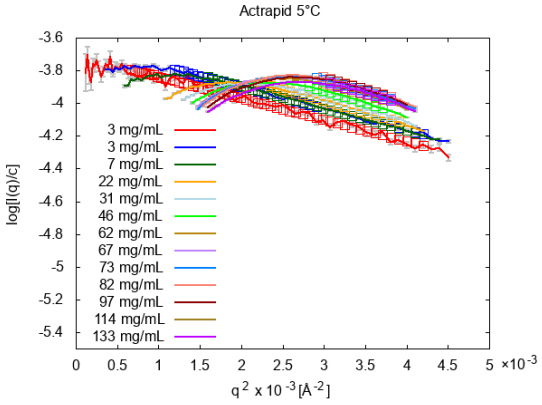

Figure 9.14: Plot of the Guinier region of the 5°C scattering data, along with Guinier fits (open squares).

Normalized pair distance distribution functions of the 2 (red curve), 3 (blue curve), and 9 mg/mL (green curve), 25°C scattering data are shown in Figure 9.15. The shape of all three curves indicate the insulin has a spherical structure. As the sample concentration increases from 3 to 9 mg/mL so does the maximum size in the sample, D_{max}. The p(r) functions of the 5°C data of 1, 3 and 7 mg/mL are shown in Figure 9.16. These curves also indicate a spherical structure of the insulin. D_{max} increases slightly with increasing sample concentration.

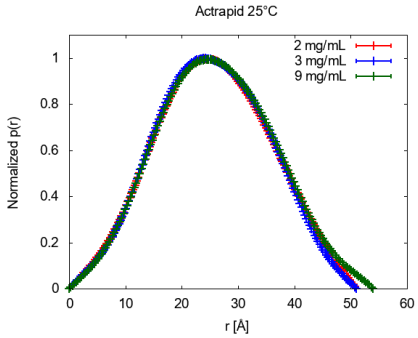

Figure 9.15: Normalized pair distance distribution functions of the 2 mg/mL (red curve), 3 mg/mL (blue curve) and 9 mg/mL (green curve) samples from the 25°C scattering experiment. Functions obtained using GNOM.

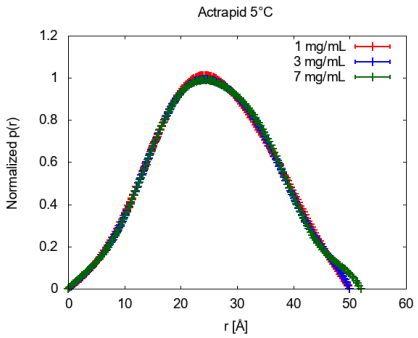

Figure 9.16: Normalized pair distance distribution functions of the 1 mg/mL (red curve), 3 mg/mL (blue curve), 7 mg/mL (green curve), 22 mg/mL (yellow curve) and 31 mg/mL (light blue curve) samples from the 25°C scattering experiment. Functions obtained using GNOM.

Kratky plots of the 25°C and 5°C are shown in Figures 9.17 and 9.18. Errors are not shown, for clarity of the curves. The bell shape of the curves indicate that the samples are mainly folded structures. The tail at higher qR_g values indicates there is flexibility in the samples.

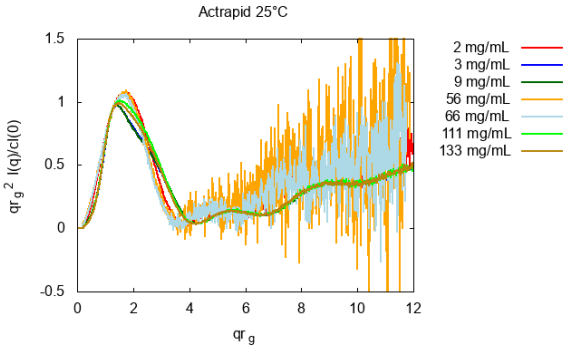

Figure 9.17: Kratky plot of the 25°C scattering data, currently obtained from PRIMUS.

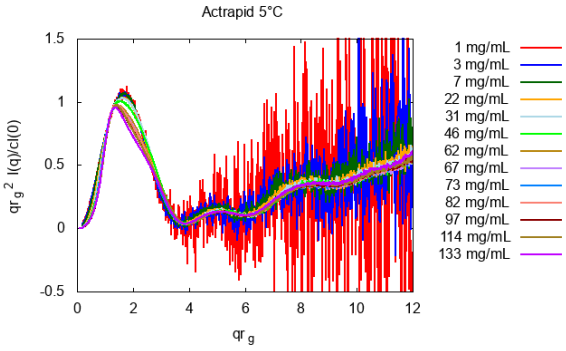

Figure 9.18: Kratky plo of the 5°C scattering data, currently obtained from PRIMUS.

Hard sphere structure factor fits to curves of the 9 and 133 mg/mL 25°C samples, are shown in Figures 9.19 and 9.20. The green curves are the fits, and the blue curves the structure factors. As can be seen in both Figures, the scattering data is not well fit against the structure factor of a hard sphere. The 9 mg/mL fit is shown in Figure 9.19. The fit has a volume fraction, ϕ, of 0.047 and a radius of 67. The volume fraction corresponds to a concentration of 67 mg/mL. The χ^2 of the fit is 9.66. The hard sphere fit to the structure factor of the 133 mg/mL data in

Figure 9.20, has a χ^2 of 140, a radius of 56 and a volume fraction of 0.087, corresponding to a concentration of 124 mg/mL.

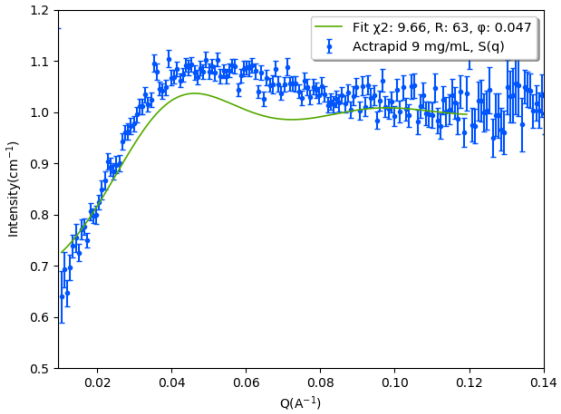

Figure 9.19: Hard sphere structure factor fit (green curve) to the 9 mg/mL 25°C data.

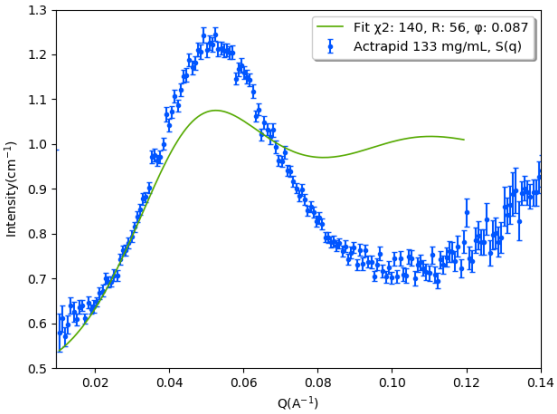

Figure 9.20: Hard sphere structure factor fit (green curve) to the 133 mg/mL 25°C data.

The full scattering curves, of the 2 and 9 mg/mL 25°C data have been fit against a hard sphere structure factor with polydispersity. The fits are shown in Figures 9.21 and 9.22. The 2 mg/mL fit is displayed in Figure 9.21, where it can be seen that the hard sphere model fits quite well to most of the curve, until the higher-q region, where the scattering intensity lacks distinct structure as compared the fit (orange curve). The polydisperse fit has a χ^2 of 1.36, a radius of 23.97, a polydispersity of 0.083 and a volume fraction of 0.0016 corresponding to a concentration of 2.29 mg/mL. The fit against the 9 mg/mL data is shown in Figure 9.22, where it is apparent the polydisperse hard sphere is not an improvement as compared to the hard sphere against the structure factor, in 9.19. The hard sphere fit (orange curve) does not grasp the low-q part of the curve as well as the higher-q curvature. The χ^2 of the fit is 270. The fit has a volume fraction of 0.022 corresponding to a concentration of 31.43 mg/mL, a radius of 24.45 and a polydispersity of 0.03.

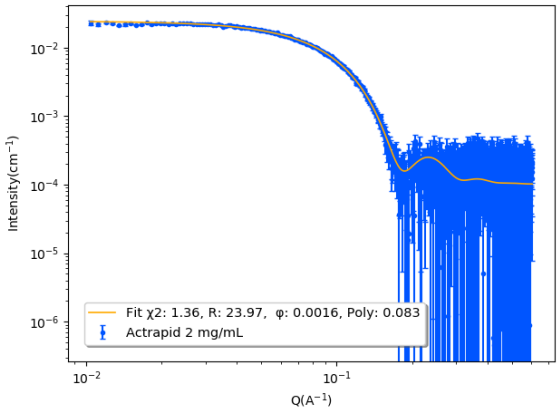

Figure 9.21: Hard sphere polydisperse fit (orange curve) to the 2 mg/mL 25°C I(q) curve.

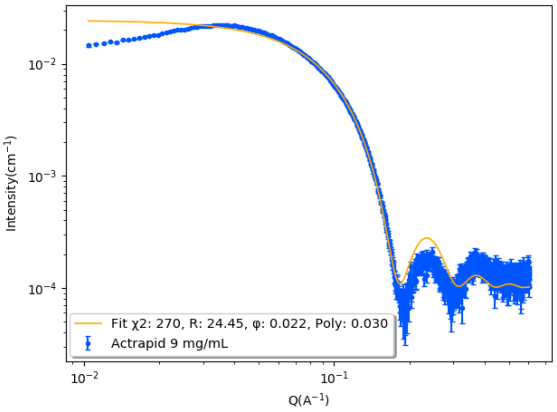

Figure 9.22: Hard sphere polydisperse fit (orange curve) to the 9 mg/mL 25°C I(q) curve.

The scattering curves of the 3 mg/mL sample of both 5 and 25°C have been fit against a monomer, dimer and hexamers of the T_3R_3 and R_6 conformations. In Figures 9.23 and 9.24 are the fits of the 25°C and 5°C against a T_3R_3 hexamer, respectively. The 25°C fit has a χ^2 of 1.25, and does not grasp the first minimum of the higher-q range very well. From the fit the sample appears to consist of 91% hexamer and 9% dimer. The 5°C fit has a χ^2 of 1.17, and fits the higher-q region of the curve better than the 25°C fit. The fit results yield a sample composition of 93% hexamer and 7% dimer.

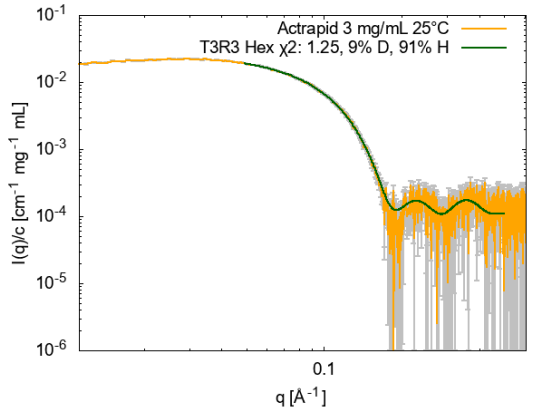

Figure 9.23: 3 mg/mL 25°C SAXS curve, fitted with a monomer, dimer and T_3R_3 hexamer using oligomer. The blue curve is the fit.

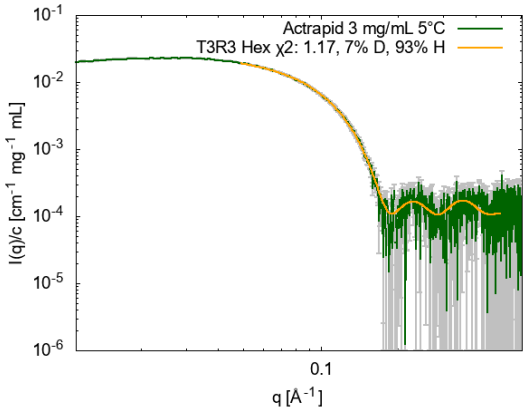

Figure 9.24: 3 mg/mL 5°C SAXS curve, fitted with a monomer, dimer and T_3R_3 hexamer using oligomer. The blue curve is the fit.

The fits to the 3 mg/mL scattering curves against a monomer, dimer and an R_6 hexamer are shown in Figures 9.25 and 9.26. For the 25°C 3 mg/mL scattering curve, in Figure 9.25, the fit has a χ^2 of 1.47. This fit does also not grasp the first minimum at the high-q region

very well. However it does fit the second minimum better than the fit of the T_3R_3 hexamer in Figure 9.23. According to the fit the sample consists of 89% hexamer and 11% dimer. The R_6 fit to the 5°C data in Figure 9.26 has a χ^2 of 1.32. The fit does not grasp the first minimum at higher-q very well, however does fit better to the second minimum at high-q as compared to the T_3R_3 fit in Figure 9.24. According to the fit, the sample consists of 91% hexamers and 9% dimers.

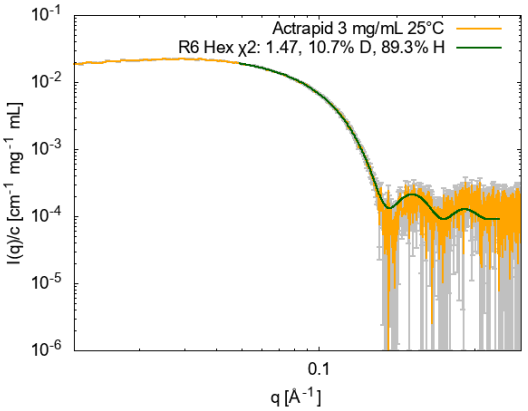

Figure 9.25: 3 mg/mL 25°C SAXS curve, fitted with a monomer, dimer and R_6 hexamer using oligomer. The blue curve is the fit.

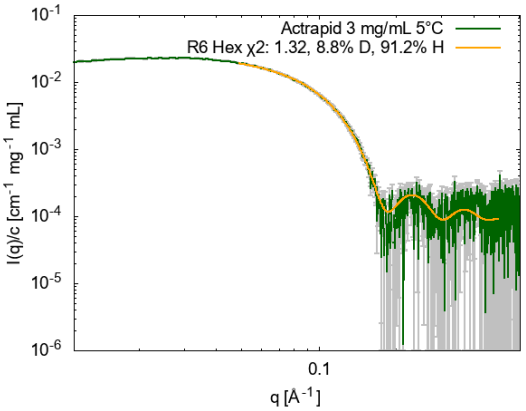

Figure 9.26: 3 mg/mL 5°C SAXS curve, fitted with a monomer, dimer and R_6 hexamer using oligomer. The blue curve is the fit.

9.3 Preservative Effects

Two CG-MALS measurements of the Actrapid samples with and without m-cresol have been conducted. The resulting second virial coefficients are shown in Figure 9.27. From the Figure, data suggests that the insulin in the Actrapid sample with m-cresol is more repulsive compared to the sample without m-cresol, indicating that m-cresol increases repulsion in Actrapid.

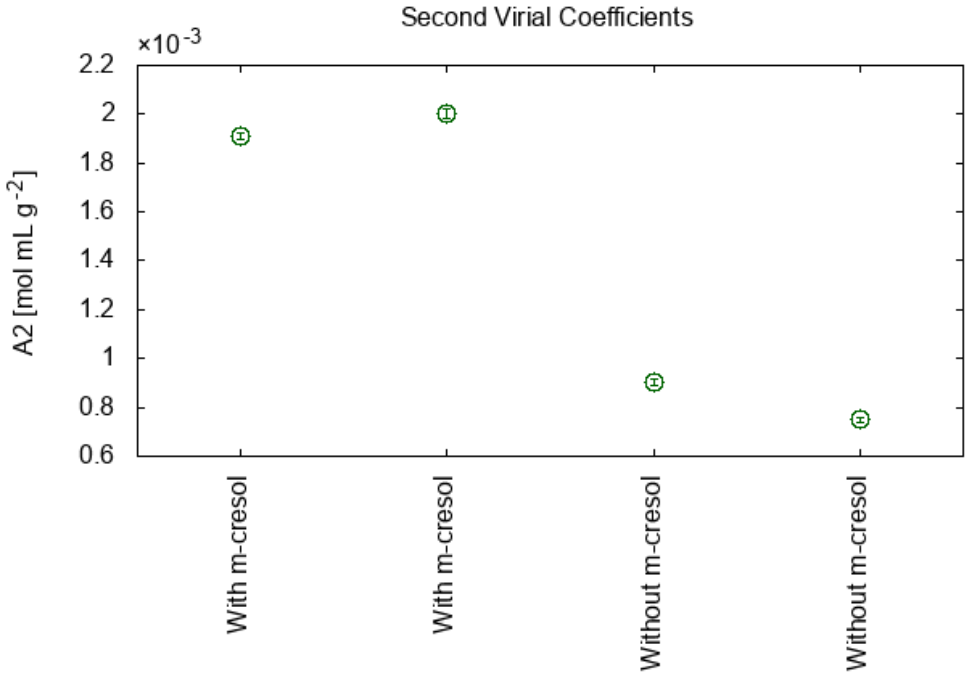

Figure 9.27: Graph of the second virial coefficients for Actrapid samples with and without m-cresol, obtained from CG-MALS measurements. Values and uncertainties obtained from data analysis using the CALYPSO software.

The molecular weights from the CG-MALS measurements are listed in table 9.4. These weights indicate that the insulin is hexameric. Values of the second virial coefficient as well as the molecular weights and uncertainties are obtained from the CALYPSO software.

Sample	Molecular Weight (kDa)
Without m-cresol	35.90 ± 0.08
Without m-cresol	35.52 ± 0.06
With m-cresol	35.23 ± 0.08
With m-cresol	35.49 ± 0.06

Table 9.4: Molecular weights of each Actrapid sample with and without m-cresol, determined from CG-MALS measurements.

SAXS experiments of the samples with and without m-cresol, have been performed on two dilutions of 1.73 mg/mL and 0.865 mg/mL and the product concentration of 3.46 mg/mL, at 25°C. In Figures 9.28, 9.29 and 9.30 the scattering curves are shown. Actrapid with m-cresol are the red curves while Actrapid without m-cresol are blue curves. The scattering curves of the dilutions do not appear to differ with m-cresol, while the samples at 3.46 mg/mL in Figure

9.30 do appear slightly different. In the low-q region, the Actrapid sample with m-cresol (red curve) shows signs of repulsion, while the sample without m-cresol does not. Considering the second virial coefficients in Figure 9.27, the repulsion is stronger in the sample with m-cersol, thus the scattering curves are in agreement.

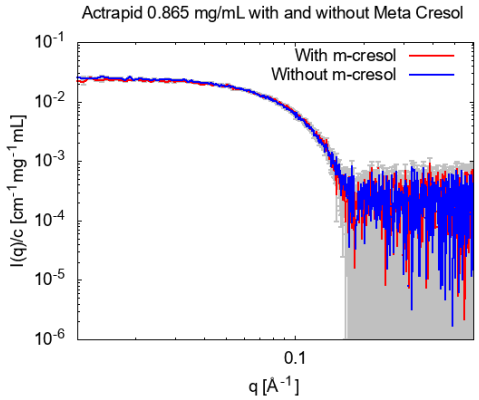

Figure 9.28: Scattering curves of 0.865 mg/mL Actrapid samples with (red curve) and without (blue curve) m-cresol.

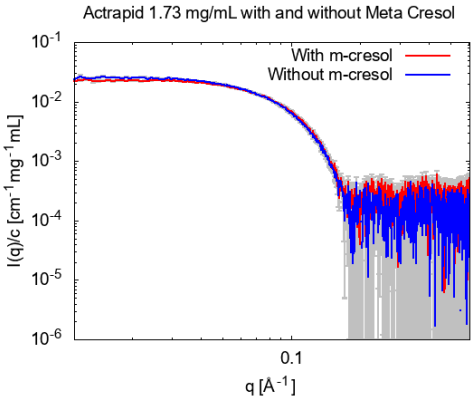

Figure 9.29: Scattering curves of 1.73 mg/mL Actrapid with (red curve) and without (blue curve) m-cresol.

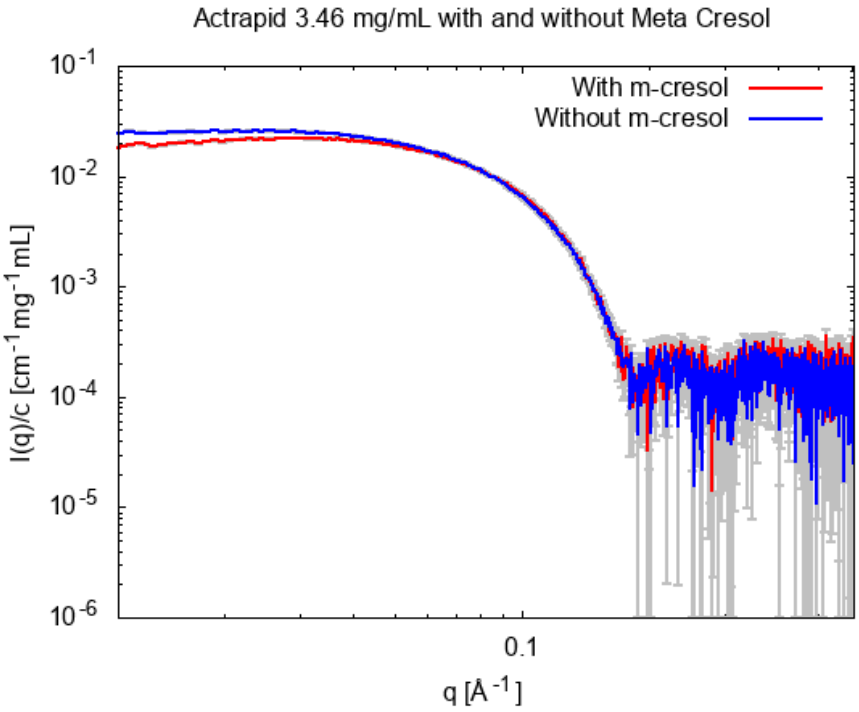

Figure 9.30: Scattering curves of 3.46 mg/mL Actrapid with (red curve) and without (blue curve) m-cresol. Both curves are absolute scale and normalised by the concentration.

From the scattering curves, the intensity at q = 0.0117 $Å^{-1}$ is used as a value of the forward

scattering. The forward scattering of all six scattering curves shown above, are displayed as a function of concentration in Figure 9.31. The forward scattering of the sample without m-cresol (blue) does not change with sample concentration. The repulsion seen in the scattering curve of the 3.46 mg/mL Actrapid sample with m-cresol is visible in the forward scattering (red), as a lower value compared to the two dilute samples. The forward scattering of the sample without m-cresol is a little higher in intensity compared to the sample with m-cresol.

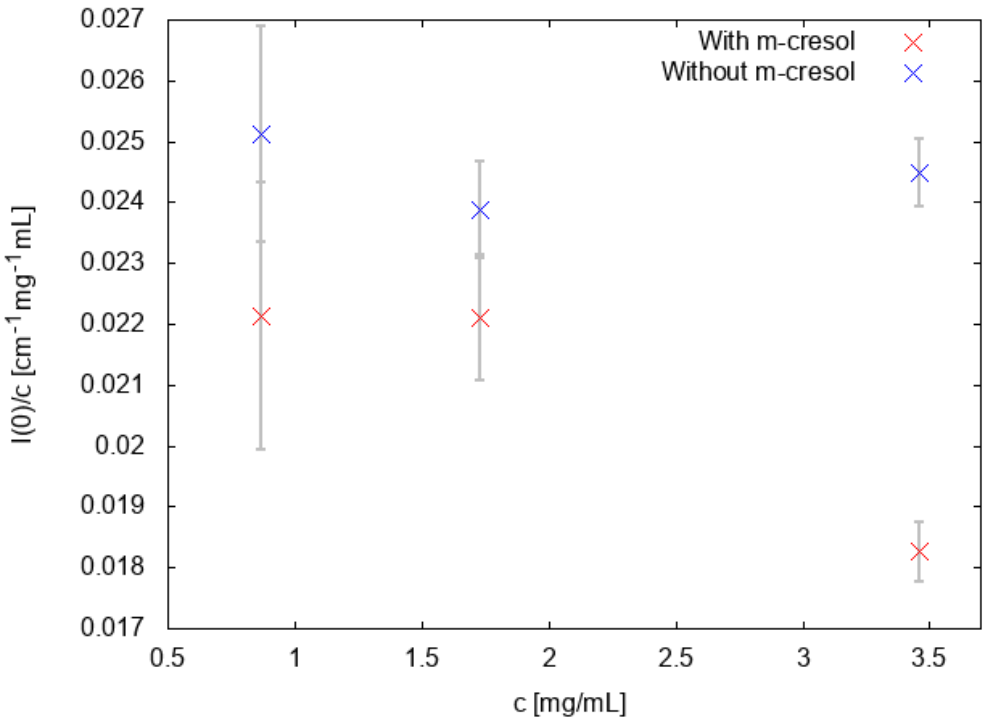

Figure 9.31: Forward scattering of the samples with (red) and without m-cresol (blue). I(0) values and uncertainties from scattering intensities at q = 0.0117 Å^{-1}.

In Figure 9.32 the Guinier regions of curves with m-cresol are shown. Guinier regions of Actrapid without m-cresol are shown in Figure 9.33. Forward scattering and radius of gyration of the Guinier analysis is listed in appendix B. The linearity of the Guinier region is indicative of monodisperse samples.

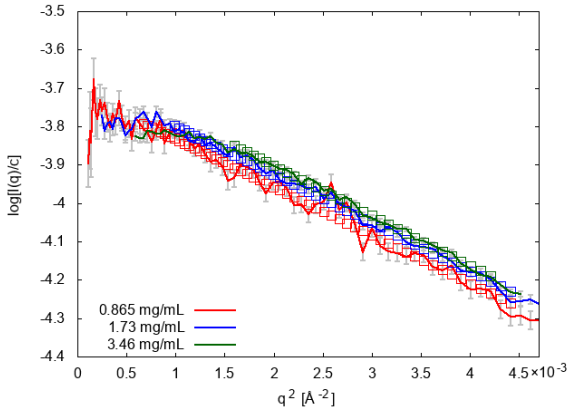

Figure 9.32: Plot of the Guinier region of the scattering curves with m-cresol, with Guinier fits (open squares)

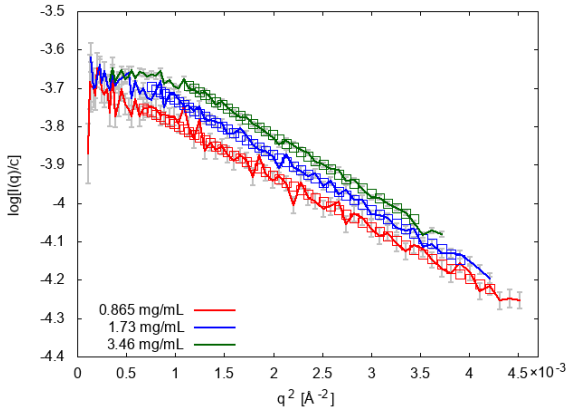

Figure 9.33: Plot of the Guinier region of the scattering curves without m-cresol, with Guinier fits (open squares)

Normalized pair distance distribution functions of the samples with and without m-cresol are shown in Figures 9.34 and 9.35 respectively. Actrapid with m-cresol has a D_{max} of 50 Å, while Actrapid without m-cresol has a D_{max} between 58 and 70. The p(r) functions of data without m-cresol have a shoulder at higher r values, indicative of aggregation.

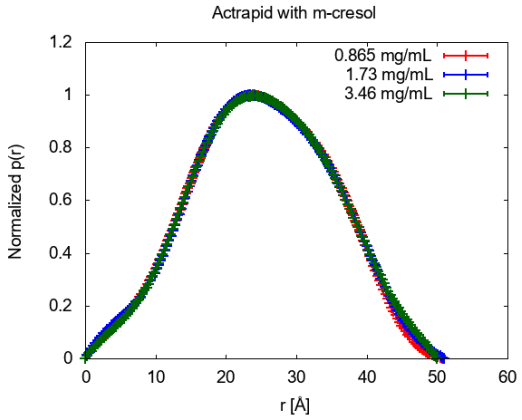

Figure 9.34: Normalized p(r) function of the Actrapid with m-cresol

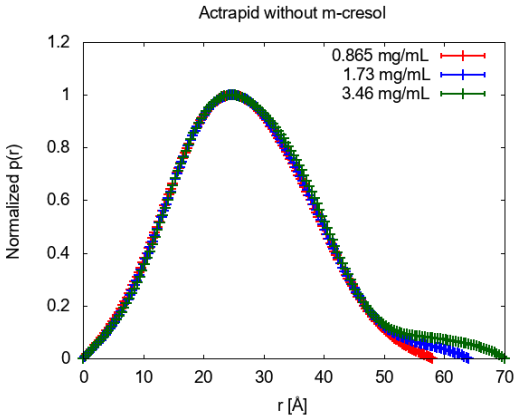

Figure 9.35: Normalized p(r) function of the Actrapid without m-cresol

The scattering curves of the 3.46 mg/mL sample with and without m-cresol have been fit against a monomer, dimer and hexamer of different conformation in OLIGOMER. In Figures 9.36 and 9.37, the scattering curves are fit with a T_3R_3 hexamer. The fit to the sample with m-cresol, Figure 9.36 has a χ^2 value of 1.64. The fit indicates the sample consists of 88% hexamers, 9% dimers and 3% monomers. As can be seen from the fit (blue curve) it does not fully reach the first minimum in the higher-q region of the scattering curve. While the second minimum is fit better. The fit to the sample without m-cresol in Figure 9.37, has a χ^2 of 3, and

the sample consists of 95% hexamers and 5% monomers. The fit (red curve) does not match the scattering curve at mid-q range very well, however does match the first and second minimum better, compared to the fit to data with m-cresol in Figure 9.37.

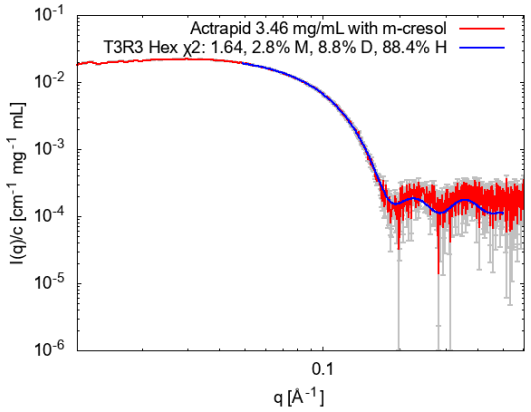

Figure 9.36: Fit with a T_3R_3 hexamer, to the scattering curve of 3.46 mg/mL Actrapid with m-cresol

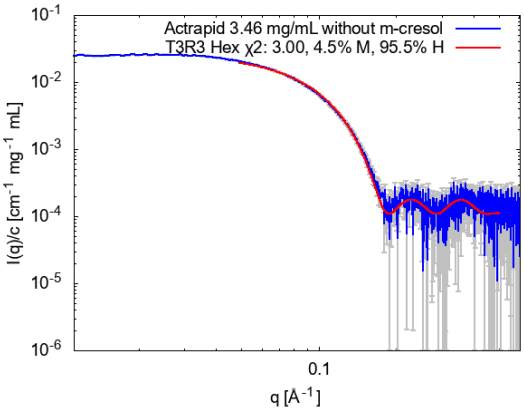

Figure 9.37: Fit with a T_3R_3 hexamer, to the scattering curve of 3.46 mg/mL Actrapid without m-cresol

In Figures 9.38 and 9.39, the OLIGOMER fits to the 3.46 mg/mL scattering curves using an R_6 hexamer are shown. The fit to the data with m-cresol in Figure 9.38, has a χ^2 value of 2.18, and indicates the sample consists of 86% hexamers and 14% dimers. The fit (blue curve) matches the first minimum at high-q better than it does the second minimum, and the local maximum after the second minimum, is also off in the fit. Compared to the T_3R_3 fit in Figure 9.36, this R_6 fit is not matching the higher q-regions well. Looking at the fit to the Actrapid without m-cresol in Figure 9.39, the fit (red curve) has a χ^2 value of 1.78, and indicates the sample consists of 94% hexamers and 6% monomers. The fit matches the first and second minimum in the higher-q region, however does not match the local maximum very well compared to the T_3R_3 fit in Figure 9.37. Both fits to the data without m-cresol do not fully match the mid-q range of the scattering profile, just before the first minimum at high-q as well as the fits to the data with m-cresol.

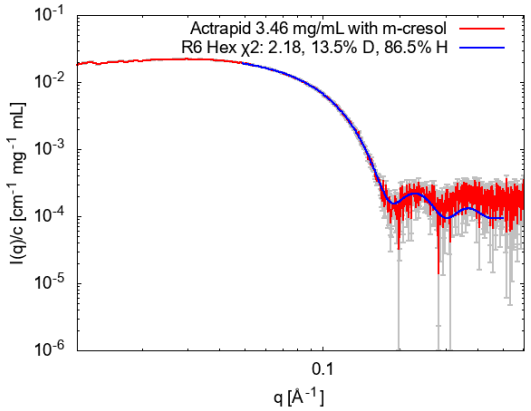

Figure 9.38: OLIGOMER fit with an R_6 Hexamer, to the scattering curve of 3.46 mg/mL Actrapid with m-cresol

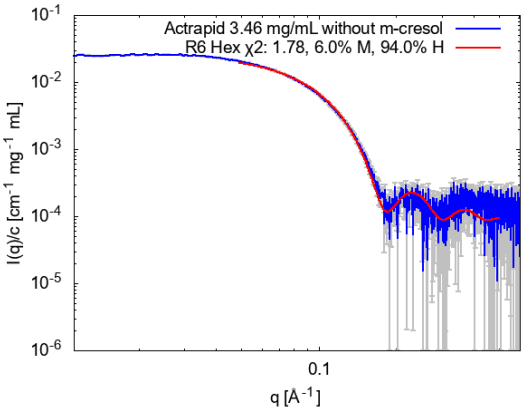

Figure 9.39: OLIGOMER fit with an R_6 Hexamer, to the scattering curve of 3.46 mg/mL Actrapid without m-cresol

10 Discussion

For the interpretation of the CG-MALS results, it should be noted that all experiments are conducted with the same flow-rates. Thus the effect of mixing timescales on the stoichiometry is unknown [83]. In this study, the main output of the CG-MALS experiments are the second virial coefficients, which are not dependent on the timescale of mixing.

10.1 Phase Transitions

Increasing the insulin concentration and altering temperature or salt conditions is found to affect the state and stability of insulin in solution. However, the binodal, under the studied conditions, has not been observed in the phase behaviour of Actrapid. A reason for this, may be the preservatives in the insulin drug product. As found by Gögelein et al. [61], the presence of glycerol and other additives shifts the binodal of lysozyme, to lower temperatures and increases the temperature gap between the binodal and fluid-solid transition. As the Actrapid drug product contains glycerol, it could very well affect the temperature at which the binodal occurs. Furthermore, the samples were prepared in centrifugal filter units, which may contain excess glycerol, despite washing the filters prior to use. Thus samples may have contained trace amounts of excess glycerol from the filters, in addition to the formulation amount of glycerol.

Another reason for the lack of a binodal may be sample concentrations. Unfortunately increasing sample concentrations above 160 mg/mL resulted in precipitation at room temperature, making it very difficult to prepare sample, also because storage in 5°C results in the formation of birefringent sediment. Furthermore, the turbid phase occurring in the centrifugal

filter units became increasingly more difficult to homogenize with increasing protein concentration and decreasing sample volume.

During experiments at temperatures above room temperature, some of the liquid in the samples could have evaporated resulting in higher protein concentrations than expected. However, comparing the normalized scattering curve of the 139 mg/mL gel sample, in Figure 9.2 with the scattering curves of the 25°C experiment in Figure 9.6, there is only a slight difference, within uncertainties of the determined sample concentration, indicating no evaporation took place. A Figure of the scattering profiles in the same plot is shown in appendix C.

10.1.1 Temperature Induced Transitions

The Actrapid samples are found to form a turbid gel at 60°C. This transition is not observed to be reversible. Scattering measurements of the 139 mg/mL sample in the gel-like transition shows aggregation in the sample, as well as features at higher-q indicative of hexamers still present in the sample. Thereby indicating the insulin is not completely denatured at 60°C. From the WAXS measurement, results shown in appendix A, it is apparent that the gel sample does not have an amyloid peak. Thereby, data indicates, the aggregated gel-like state does not show signs of fibrillation. Gelation is found in colloid-polymer systems at low volume fractions, where short-range attractions leads to aggregation into larger space-spanning clusters. The bonds between clusters is reversible. However, Actrapid is repulsive and as found in the second virial coefficients, becomes net more repulsive with increasing temperature. Thus the gel-like transition found in Actrapid is not similar to that of colloid-polymer systems.

At 5°C Actrapid samples of concentrations above 100 mg/mL are found to form birefringent sediment. As seen in the photograph of the phase diagram in Figure 9.3, parts of the sediment appears spherical in shape. The fact that the sediment is birefringent indicates structural ordering in the sample. However, due to the fast reversibility it is difficult to perform structural measurements to investigate sample ordering. Due to the sediment appearing quite fast and dissolving even faster, it is most likely not crystalline. The samples have not been investigated under a microscope to reveal the absence of presence of the characteristic Maltese cross of spherulites. Furthermore, it has not been shown if the sediment affects insulin aggregation.

10.1.2 Salt Induced Transitions

The addition of 50 mM NaCl to Actrapid with various concentrations is found to result in precipitation at temperature and concentration ranges, where the Actrapid without additional salt is found to be homogeneous. Thus the addition of salt increases particle interactions.

Precipitation induced by salt is found for salt concentrations of 50 and 100 mM at various temperatures. Precipitation is reversible for samples with 50 mM NaCl, in the range of the investigated temperatures. It has not been investigated if the reversible precipitation had an impact on the structure or stability of insulin in solution. However, as the precipitate is reversible, it indicates the insulin is not fibrillated.

10.2 Temperature Effects

The second virial coefficients obtained from CG-MALS experiments performed at 25°C and 40°C indicate an increase in the net repulsion of the insulin with increasing temperature. The scattering curves of Actrapid measured at 5°C and 25°C also show signs of repulsive particle interactions in the low-q region for concentrations above 2 mg/mL. With more repulsion, also visible as a larger decrease in the forward scattering of the 5°C experiments. In the structure factors of both experiments, increasing the sample concentration results in a decrease in scattering at low-q. This behaviour is as expected for repulsive interactions. The forward scattering shows a minimum, as is expected for repulsive systems. The downturn in the Guinier region of samples above 10 mg/mL, are due to particle interactions, which may occur from Coulombic repulsion. These indications and the net repulsion of the second virial coefficients are in agreement. The system is repulsive regardless of temperature.

The main differences in the scattering curves of Actrapid measured at 25°C and 5°C is the larger decrease in forward scattering for the 5°C measurement as well as differences in the peak position of the peak appearing in samples above 10 mg/mL around q = 0.05 $Å^{-1}$, best visible in Figures 9.8 and 9.9. In the 5°C scattering data, the peak shifts to higher q-values with increasing concentration, while for the 25°C data, the peak position is unchanged with concentration. The change in peak position indicates that the The larger decrease in the forward scattering of the 5°C Actrapid samples, indicate insulin is in a repulsive minimum around 25°C and repulsion increases with both an increase and a decrease in temperature.

Another difference between the 25°C and 5°C scattering data, is visible in the structure factors. In this representation, data suggest that the minimum around q = 0.1 $Å^{-1}$ changes greater with concentration in the 25°C measurement, than in the 5°C. In the structure factors, the peak at q = 0.05 $Å^{-1}$, is different between the two experiments. For the 5°C experiment, the peak continuously grows with increasing concentration, while for the 25°C data, there is only a slight difference in the peak between different samples. As the structure factor peak is related to nearest-neighbour distances, data indicates the average particle distances decrease with increasing concentration in the 5°C. This suggests growing oligomeric species in the sample.

The molecular weights from the CG-MALS experiments indicates the largest oligomeric species in all samples is hexamers. The bell shape of the normalized pair distance distribution functions in Figures 9.15 and 9.16 indicates the insulin is spherical, for both 25°C and 5°C measurements. Actrapid samples with concentrations below 3 mg/mL for both the 25 and 5 °C experiments, indicate monodispersity in the linearity of the Guinier regions shown in Figures 9.13 and 9.14. At concentrations above 3 mg/mL the D_{max} of the p(r) functions are difficult to determine, which is indicative of interparticle interactions. This is also why higher concentrations are not represented by p(r) functions. The D_{max} values of Actrapid measured at 25°C and 5°C is just above 50 Å, which is in close agreement with the findings of Palmieri et al., [33]. Thus the decrease in temperature does not impact the hexamer shape or the maximal particle size, D_{max}.

The bell shapes of the Kratky curves in Figures 9.17 and 9.18 indicate folded domains in the insulin. With increasing sample concentrations the maximum of the curve shifts to lower qr_g and $qr_g^2 I(q)/cI(0)$ values. This is an effect that could be due to the interparticle interactions in the samples resulting in folded domains, which may be more compact. The tail at qr_g above 7, indicates flexibility in the insulin and does not appear to change in the 5°C scattering data. For the Actrapid measured at 25°C it seems the tail at qr_g above 7, decreases with concentrations above 100 mg/mL, thus indicating concentrations above 100 mg/mL reduces the flexibility of insulin. Besides this, there are no significant differences between the Kratky plots of the 5°C and 25°C scattering experiments.

The structure factors of the 9 mg/mL and 133 mg/mL 25°C scattering data, have been fit against hard sphere structure factors shown in Figures 9.19 and 9.20. The hard sphere fits, underestimates the nearest-neighbour peak of the structure factors in both cases. Furthermore, the minima in the 133 mg/mL data is overestimated by the fit (green curve). From these fits, it is apparent the interactions in the Actrapid at 25°C is not modelled well by hard spheres. As a result of these structure factor fits, the full scattering profiles of 2 mg/mL and 9 mg/mL 25°C, have been fit against polydisperse hard sphere models, as seen in Figure 9.21 and 9.22. The 2 mg/mL fit, matches the scattering profile well, until the higher q-range, where the scattering profile lacks specific structure, meaning any shape of a fit in this region of the curve is off. The polydisperse hard sphere fit to the 9 mg/mL data, only matches the mid-q range. The repulsive decrease at low-q is not incorporated in the fit (orange curve). The high-q region of the scattering profile is mismatched. The first minima at about $q = 0.2$ Å^{-1} is underestimated, while the local maxima after the minima, is overestimated. The scattering profiles of Actrapid measured at 25°C and 5°C, do not differ significantly in shape, between the 1 and 2 mg/mL,

the 7 and 9 mg/mL and the 133 mg/mL. Therefore, scattering profiles measured at 5°C, have not been fit against hard sphere structure factors, as the outcome would yield the same results. The interactions of insulin in the Actrapid drug product cannot be modelled by hard spheres, polydisperse or not.

Neither the T_3R_3 or the R_6 OLIGOMER fits to the 3 mg/mL scattering curves of the 25 and 5°C experiments are very good matches to the higher q-region of the curves. Therefore it is hard to determine which conformation the hexamer is in. As the Actrapid samples contain both m-cresol and zinc, the hexamer should have an R_6 conformation, however the scattering curves are not fit well to the theoretical scattering of the R_6 hexamer. All fits, indicate samples mainly contain hexamers, in equilibrium with dimers.

In the SAXS experimental setup, the sample holder in the chamber of the SAXS instrument is cooled by a Julabo water bath. Outside the chamber, samples are cooled prior to experiments by another water bath from Julabo. However, the sample holder in the chamber is not cooled properly. The sample holder and connecting tubes do not feel cold to the touch after the water in the bath reaches 5°C. The tubes do not show signs of clogging. Thus experiments conducted at 5°C are most likely not measured at 5°C, but at room temperature. Samples were cooled to 5°C prior to measuring, however slowly heat in the chamber during measurements. This slow heating in the chamber, could impact the observed differences in the scattering of Actrapid at 5°C as compared to the scattering at 25°C.

10.3 Preservative Effects

The CG-MALS and SAXS experiments performed on Actrapid with and without meta cresol, indicate that the presence of m-cresol impacts insulin interactions. From the second virial coefficients shown in Figure 9.27, it is apparent that the interactions in Actrapid without m-cresol are half as repulsive, as the interactions in Actrapid with m-cresol. Looking at the scattering curves, of the 3.46 mg/mL Actrapid, both with and without the m-cresol in Figure 9.30, the difference is primarily visible at low-q, where repulsion is visible in the sample with m-cresol. This result is in agreement with the second virial coefficients. As the repulsion in Actrapid with m-cresol is greater than that of Actrapid without, it explains why repulsion is most clearly visible in the sample with m-cresol at this concentration.

The theoretical scattering curves of the hexamer in the three different conformations, T_6, T_3R_3 and R_6, differ significantly in the first minimum around $q = 0.2$ Å^{-1}. The most pronounced minimum is found in the R_6 hexamer. From the shapes of the scattering profiles of 3.46 mg/mL

Actrapid with and without m-cresol, it appears the first minima, at higher-q values, is not different in the two samples. This indicates the hexamer conformation is not changed in the absence of m-cresol.

The normalized pair distance distribution functions indicate that the absence of meta cresol does not affect the spherical shape of the insulin molecules. The molecular weights obtained from the CG-MALS measurements all indicate hexamers, in all four measurements. As the R_6 hexamer has a more spherical shape, compared to the cylindrical T_6 hexamer, and the p(r) functions indicate spherical shapes rather than cylindrical, data indicates that the insulin hexamers are not in the T_6 conformation. However, looking at the OLIGOMER fits to the 3.46 mg/mL scattering curves, data suggests, that for both the T_3R_3 and R_6 hexamers, the theoretical scattering curves are not good fits to data. For the Actrapid without m-cresol, both fits have a steeper decline in the curvature at mid-q just before the first minima, compared to the scattering curve. The two fits do not clearly indicate which conformational state the hexamers in the Actrapid without m-cresol are in. However the T_3R_3 and R_6 fits do indicate that insulin in absence of m-cresol, is mainly hexameric with about 5% monomer present.

Insulin in the presence of phenolic ligands is expected to be in the R_6 hexameric conformation. However the results of the OLIGOMER fits to Actrapid with m-cresol, do not exclusively indicate the hexamer is R_6. The R_6 fit to the 3.46 mg/mL scattering profile shown in Figure 9.38 does not match the characteristic shape of the profile at high-q. The fit using the T_3R_3 hexamer, shown in Figure 9.36 mimics the curvature of the scattering profile at high-q better, however not perfect. The first minimum, is overestimated in the fit. Both fits indicate the Actrapid mainly consists of hexamers with about 10% dimers. This result is in agreement with the molecular weights obtained from the CG-MALS measurements, being equivalent to the molecular weight of a hexamer. This is in agreement with the fact that insulin exists mainly in hexameric form when concentrations are above 2 mM.

The OLIGOMER fits, indicate that the Actrapid with m-cresol contains dimers besides the hexamers, while the second oligomeric species in the Actrapid without m-cresol is monomers. Furthermore, the fits indicate a larger percentage of hexamers in the sample without m-cresol. Despite irregularities in the fits, data support dimer formation induced by m-cresol.

It should be noted, that OLIGOMER fits work best for form factors, and there appears to be structure factor effects in the data, appearing as the repulsion in Actrapid with m-cresol. This is why, the fitting starts at q = 0.05 $Å^{-1}$. Structure factor effects, may explain why OLIGOMER has a hard time matching the scattering profiles of Actrapid without m-cresol

just before the first minimum at mid-q.

The presence of preservatives such as m-cresol in insulin drug products, aid in the stability of the hexamer. This may explain the indication of aggregation in the normalized pair distance distribution functions of Actrapid without m-cresol shown in Figure 9.35.

11 Conclusion

In this work, the phase transitions of human insulin, prepared as the drug product Actrapid, have been investigated by altering salt concentrations and temperatures of samples with various concentrations. Furthermore, temperature effects have been investigated using small angle X-ray scattering and static light scattering. Lastly the influence of preservatives in the Actrapid drug product has been investigated, also by small angle X-ray scattering and static light scattering. Understanding the phase behaviour of the active pharmaceutical ingredient aids in knowledge of which conditions the medicine can withstand during formulation, storage and usage.

Investigations of the phase behaviour of insulin resulted in two transitions. A turbid gel at temperatures of 60°C, and a birefringent sediment found at 5°C. Additional salt in the insulin drug product resulted in precipitation which, in some cases, is reversible at elevated temperatures. It has not been investigated if the reversible precipitation affects the insulin structure or stability.

Increasing temperatures from 25°C to 40°C increases net repulsion in the Actrapid, as seen in the values of the second virial coefficients. Decreasing temperatures from 25°C to 5°C does not have an impact on the structure of te insulin hexamer. However the temperature decrease does reduce protein distances and increases the repulsion with increasing sample concentration.

The absence of m-cresol in the insulin drug product, is found to affect the repulsive interactions as well as the stability of the hexamer. The absence of m-cresol does not affect hexamer formation or conformation. From the second virial coefficients obtained by CG-MALS, it is found that m-cresol induces overall repulsion. Scattering profiles of Actrapid with and without m-cresol are primarily similar apart from repulsion in the sample with m-cresol at formulation concentrations of 3.46 mg/mL. Aggregation is indicated in the pair distance distribution functions of Actrapid without m-cresol. Aggregation is also found in SAXS measurements enduring longer than 30 minutes.

To further understand the phase behaviour of insulin, searching for the binodal is highly relevant. This could be done using Actrapid formulated without glycerol, to investigate if the presence of glycerol, affects the position of the binodal in the phase diagram, as well as the effect glycerol may have on the stability of the insulin. Furthermore investigations of the effect of zinc is also relevant with regards to the phase behaviour as well as stability. Lastly experimenting with the addition of salt to the Actrapid, in samples without m-cresol, glycerol or zinc as well as the full drug product would further broaden the understanding of the self-association and phase behaviour of human insulin. It would also be highly relevant to investigate excipients of the drug product with the aim to understand effects on the active pharmaceutical ingredient.

References

[1] WHO, *Diabetes Fact Sheet. http://www.who.int/mediacentre/factsheets/fs312/en/*, 2020, accessed Dec 28, 2020.

[2] IDF, *IDF Daibetes Atlas. http://diabetesatlas.org/data/en/*, 9th ed., 2019, accessed Dec 28, 2020.

[3] T. M. Chapman, S. Noble, and K. L. Goa, "Insulin aspart. a revie of its use in the management of type 1 and 2 diabetes mellitus," *Drugs*, vol. 62, no. 13, pp. 1945–1981, 2002.

[4] T. Hoeg-Jensen, *Chapter 7 in Peptide and Protein Design for Biopharmaceutical Applications.* John Wiley & Sons, Ltd, 2009, edited by K. J. Jensen, pages 249-286.

[5] L. S. Dreyer, J. Nygaard, L. Malik, T. Hoeg-Jensen, R. Høiberg-Nielsen, and L. Arleth, "Structural insight into the self-assembly of a pharmaceutically optimized insulin analogue obtained by small-angle x-ray scattering," *Molecular Pharmaceutics*, vol. 17, pp. 2809 – 2820, 2020.

[6] J. Brange and A. Vølund, "Insulin analogs with improved pharmacokinetic profiles," *Advanced Drug Delivery Reviews*, vol. 35, pp. 307–335, 1999.

[7] S. Gammeltoft, B. F. Hansen, L. Dideriksen, A. Lindholm, L. Schäffer, T. Trüb, A. Dayan, and P. Kurtzhals, "Insulin aspart: a novel rapid-acting human insulin analogue," *Expert Opinion on Investigational Drugs*, vol. 8, no. 9, pp. 1431–1442, 1999.

[8] D. Some, J. Pollastrini, and S. Cao, "Characterizing reversible protein association at moderately high concentration via composition-gradient static light scattering," *Journal of Pharmaceutical Sciences*, vol. 105, no. 8, pp. 2310 – 2318, 2016.

[9] S. L. Morissette, S. Soukasene, D. Levinson, M. J. Cima, and Örn Almarsson, "Elucidation of crystal form diversity of the hiv protease inhibitor ritonavir by high-throughput crystallization," *PNAS*, vol. 100, no. 5, pp. 2180–2184, 2003.

[10] J. Bauer, S. Spanton, R. Henry, J. Quick, W. Dziki, W. Porter, and J. Morris, "Ritonavir: An extraordinary example of conformational polymorphism," *Pharmaceutical Research*, vol. 18, no. 6, pp. 859 – 866, 2001.

[11] J. Brange, *Stability of Insulin.* Kluwer Academic Publishers, 1994.

[12] J. Brange and L. Langkjær, *Chapter 11, Insulin Structure and Stability from Stability and Characterization of Protein and Peptide Drugs: Case Histories.* Plenum Press, 1993, edited by Y. John Wang and Rodney Pearlman.

[13] G. D. Smith and G. G. Dodson, "Structure of a rhombohedral r_6 insulin/phenol complex," *PROTEINS: Structure Function, and Genetics*, vol. 14, no. 3, pp. 401–408, 1992.

[14] M. F. Dunn, "Zinc-ligand interactions modulate assembly and stability of the insulin hexamer - a review," *BioMetals*, vol. 18, no. 4, pp. 295–303, 2005.

[15] N. Yagi, N. Ohta, T. Iida, and K. Inoue, "A microbeam x-ray diffraction study of insulin spherulites," *Journal of Molecular Biology*, vol. 326, no. 2, pp. 327 – 333, 2006.

[16] M. R. H. Krebs, C. E. MacPhee, A. F. Miller, I. E. dunlop, C. M. Dobson, and A. M. Donald, "The formation of spherulites by amyloid fibrils of bovine insulin," *PNAS*, vol. 101, no. 40, pp. 14 420 – 14 424, 2004.

[17] L. Nielsen, R. Khurana, A. Coats, S. Frokjaer, J. Brange, S. Vyas, V. N. Uversky, and A. L. Fink, "Effect of environmental factors on the kinetics of insulin fibril formation: Elucidation of the molecular mechanism," *Biochemistry*, vol. 40, pp. 6036 – 6046, 2001.

[18] J. L. Whittingham, D. J. Edwards, A. A. Antson, J. M. Clarkson, and G. G. Dodson, "Interactions of phenol and *m*-cresol in the insulin hexamer, and their effect on the association properties of b28 pro → asp insulin analogues," *Biochemistry*, vol. 37, pp. 11 516 – 11 523, 1998.

[19] G. Scapin, V. P. Dandey, Z. Zhang, W. Prosise, A. Hruza, T. Kelly, T. Mayhood, C. Strickland, C. S. Potter, and B. Carragher, "Structure of the insulin receptor-insulin complex by single-particle cryo-em analysis," *Nature*, vol. 556, pp. 122 – 125, 2018.

[20] P. D. Meyts, "Insulin and its receptor: structure, function and evolution," *Bio Essays*, vol. 26, no. 12, pp. 1351–1362, 2004.

[21] D. B. Steensgaard, G. Schluckebier, H. M. Strauss, M. Norrman, J. K. Thomsen, A. V. Fridericksen, S. Havelund, and I. Jonassen, "Ligand-controlled assembly of hexamers, dihexamers, and linear multihexamer structures by the engineered acylated insulin degludec," *Biochemistry*, vol. 52, pp. 295 – 309, 2012.

[22] M. Zieliński, A. Romanik-Chruścielewska, D. Mikiewicz, N. Łukasiewicz, I. Sokołowska, J. Antosik, A. Sobolewska-Ruta, A. Bierczyńska-Krzysik, P. Zaleski, and A. Płucienniczak,

"Expression and purification of recombinant human insulin from e. coli 20 strain," *Protein Expression and Purification*, vol. 157, pp. 63–69, 2019.

[23] D. Some, "Light-scattering-based analysis of biomolecular interactions," *Biophysical Reviews*, vol. 5, no. 2, pp. 147–158, 2013.

[24] G. D. Luca, D. F. Galparsoro, G. Sancataldo, M. Leone, V. Foderà, and V. Vetri, "Probing ensemble polymorphism and single aggregate structural heterogeneity in insulin amyloid self-assembly," *Journal of Colloid and Interface Science*, vol. 574, pp. 229 – 240, 2020.

[25] D. Cannon, S. J. Eichhorn, and A. M. Donald, "Structure of spherulites in insulin, β-lactoglobulin, and amyloid β," *ACS Omega*, vol. 1, pp. 915 – 922, 2016.

[26] M. R. H. Krebs, E. H. C. Bromley, S. S. Rogers, and A. M. Donald, "The mechanism of amyloid spherulite formation by bovine insulin," *Biophysical Journal*, vol. 88, no. 3, pp. 2013 – 2021, 2005.

[27] N. N. A/S, "Product specifications," *DFP Quality Support*, no. 185190.

[28] A. K. Attri, C. Fernández, and A. P. Minton, "Self-association of zn-insulin at neutral ph: Investigation by concentration gradient-static and dynamic light scattering," *Biophysical Chemistry*, vol. 148, pp. 23 – 27, 2010.

[29] ——, "ph-dependent self-association of zinc-free insulin characterized by concentration-gradient static light scattering," *Biophysical Chemistry*, vol. 148, pp. 28 – 33, 2010.

[30] M. H. Jensen, P.-O. Wahlund, K. N. Toft, J. K. Jacobsen, D. B. Steensgaard, M. van de Weert, S. Havelund, and B. Vestergaard, "Small angle x-ray scattering-based elucidation of the selv-association mechanism of human insulin analogue lys^{B29}($n^{\epsilon}\omega$-caboxyheptadecanoyl) des(b30)," *Biochemistry*, vol. 52, pp. 282 – 294, 2012.

[31] D. I. Svergun, C. Barberato, and M. H. J. Koch, "Crysol – a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates," *Journal of Applied Crystallography*, vol. 28, pp. 768 – 773, 1995.

[32] P. V. Konarev, V. V. Volkov, A. V. Sokolova, M. H. J. Koch, and D. I. Svergun, "Primus: a windows pc-based system for small-angle scattering data analysis," *Journal of Applied Crystallography*, vol. 36, pp. 1277 – 1282, 2003.

[33] L. C. Palmieri, M. P. Fávero-Retto, D. Lourenço, and L. M. T. R. Lima, "A t_3r_3 hexamer of the human insulin variant b28asp," *Biophysical Chemistry*, vol. 173-174, pp. 1–7, 2013.

[34] A. Ahmad, I. S. Millet, S. Doniach, V. N. Uversky, and A. L. Fink, "Partially folded intermediates in insulin fibrillation," *Biochemistry*, vol. 42, no. 39, pp. 11 404 – 11 416, 2003.

[35] J. N. Israelachvili, *Intermolecular and Surface Forces*, 3rd ed. Academic Press, 2011.

[36] R. Philips, J. Kondev, J. Theriot, and H. G. Garcia, *Physical Biology of the Cell*, 2nd ed. New York: Garland Science, Taylor & Francis Group, LLC, 2013.

[37] H. Ohshima and K. Makino, *Colloid and Interface Science in Pharmaceutical Research and Development*. Oxford: Elsevier B. V., 2014.

[38] E. Y. Chi, S. Krishnan, T. W. Randolph, and J. F. Carpenter, "Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation," *Pharmaceutical Research*, vol. 20, no. 9, pp. 1325 – 1336, 2003.

[39] P. Atkins, L. Jones, and L. Laverman, *Chemical Principles*, 6th ed. New York: W H Freeman and Company, 2013.

[40] D. V. Schroeder, *An Introduction to Thermal Physics*, first edition, pearson new international ed. Pearson, 2014, chapter 5 Free Energy and Chemical Thermodynamics.

[41] E. Chatani, R. Inoue, H. Imamura, M. Sugiyama, M. Kato, M. Yamamoto, K. Nishida, and T. Kanaya, "Early aggregation preceding the nucleation of insulin amyloid fibrils as monitored by small angle x-ray scattering," *Scientific Reports*, vol. 5, no. 15485, 2015.

[42] C. M. Dobson, "Protein folding and misfolding," *Nature*, vol. 426, no. 18, pp. 884 –890, 2003.

[43] J. E. Proctor, M. Bailey, I. Morrison, M. A. Hakeem, and I. F. Crowe, "Observation of liquid-liquid phase transitions in ethane at 300 k," *The Journal of Physical Chemistry B*, vol. 122, pp. 101 742 – 10 178, 2018.

[44] E. A. Doud, A. Voevodin, T. J. Hochuli, A. M. Cmapsaur, C. Nuckolls, and X. Roy, "Superatoms in materials science," *Nature Reviews Materials*, vol. 5, pp. 371 – 387, May 2020.

[45] W. K. Kegel and A. von Blaaderen, "Direct observation of dynamical heterogeneities in colloidal hard-sphere suspensions," *SCIENCE*, vol. 287, pp. 290–293, January 2000.

[46] S. Bucciarelli, L. Casal-Dujat, C. D. Michele, F. Sciortino, J. Dhont, J. Bergenholtz, B. Farago, P. Schurtenberger, and A. Stradner, "Unusual dynamics of concentration fluctuations in solutions of weakly attractive globular proteins," *The Journal of Physical Chemistry Letters*, vol. 5, pp. 4470 – 4474, 2015.

[47] M. A. Miller and D. Frenkel, "Phase diagram of the adhesive hard sphere fluid," *The Journal of Chemical Physics*, vol. 121, no. 1, pp. 535 – 545, 2004.

[48] D. Kondepudi and I. Prigogine, *Modern Thermodynamics: From Heat Engines to Dissipative Structures*, 2nd ed. John Wiley & Sons, Ltd, 2015, chapter 7: Thermodynamics of Phase Change, pages 187-205.

[49] L. E. Reichl, *A Modern Course in Statistical Physics*, 4th ed. WILEY-VCH Verlag GmbH & Co. KGaA, 2016, chapter 4: The Thermodynamics of Phase Transitions, pages 87 - 134.

[50] I. R. McDonald and J.-P. Hansen, *Theory of Simple Liquids*, 4th ed. Elsevier Ltd, 2013, chapter 1: Introduction, pages 1-11.

[51] W. Poon, P. Pusey, and H. Lekkerkerker, "Colloids in suspense," *Physics World*, vol. 9, no. 4, pp. 27–32, April 1996.

[52] K. N. Pham, A. M. Puertas, J. Bergenholtz, S. U. Egelhaaf, A. Moussaïd, P. N. Pusey, A. B. Schofield, M. E. Cates, M. Fuchs, and W. C. K. Poon, "Multiple glassy states in a simple model system," *SCIENCE*, vol. 296, pp. 104–106, 2002.

[53] V. J. Anderson and H. N. W. Lekkerkerker, "Insights into phase transition kinetics from colloid science," *Nature*, vol. 416, pp. 811–815, April 2002.

[54] P. N. Pusey and W. van Megen, "Phase behaviour of concentrated suspensions of nearly hard colloidal spheres," *Nature*, vol. 320, pp. 340–342, 1986.

[55] G. Foffi, C. de Michele, F. Sciortino, and P. Tartaglia, "Scaling of dynamics with the range of interaction in short-range attractive colloids," *Physical Review Letters*, vol. 94, p. 078301, 2005.

[56] J. Bergenholtz, W. C. K. Poon, and M. Fuchs, "Gelation in model colloid-polymer mixtures," *Langemuir*, vol. 19, no. 10, pp. 4493 – 4503, 2003.

[57] N. Asherie, "Protein crystallization and phase diagrams," *Methods (San Diego, Calif.)*, vol. 34, no. 3, pp. 266–272, 2004.

[58] D. N. Petsev, X. Wu, O. Galkin, and P. G. Vekilov, "Thermodynamic functions of concentrated protein solutions from phase equilibria," *The Journal of Physical Chemistry B*, vol. 107, no. 16, pp. 3921–3926, 2003.

[59] A. Pande, J. Pande, N. Asherie, A. Lomakin, O. Ogun, J. King, and G. B. Benedek, "Crystal cataracts: Human genetic cataract caused by protein crystallization," *Proceedings of the National Academy of Sciences*, vol. 98, no. 11, pp. 6116–6120, 2001.

[60] A. Stradner and P. Schurtenberger, "Potential and limits of a colloid approach to protein solutions," *Soft Matter*, vol. 16, pp. 307–323, 2020.

[61] C. Gögelein, D. Wagner, F. Cardinaux, G. Nägele, and S. U. Egelhaaf, "Effect of glycerol and dmso on the phase behavior of lysozyme - theory and experiments," *The Journal of Chemical Physics*, vol. 136, 2012.

[62] P. Lindner and T. Zemb, *Neutrons, X-rays and Light. Scattering Methods Applied to Soft Condensed Matter*, 1st ed., ser. North-Holland delta series. Elsevier Science Ltd., 2002.

[63] D. I. Svergun and M. H. J. Koch, "Small-angle scattering studies of biological macromolecules in solution," *Reports on Progress in Physics*, vol. 66, pp. 1735 – 1782, 2003.

[64] S. Skou, R. E. Gillilan, and N. Ando, "Synchrotron-based small-anlge x-ray scattering (saxs) of proteins in solution," *Nature Protocols*, vol. 9, no. 7, pp. 1727 – 1739, 2014.

[65] H. Schnablegger and Y. Singh, *The SAXS Guide, Getting acquainted with the principles*, 3rd ed. Anton Paar GmbH, 2013.

[66] D. A. Jacques and J. Trewhella, "Small-angle scattering for strucutral biology - expanding the frontier while avoiding pitfalls," *Protein Science*, vol. 19, pp. 642–657, 2010.

[67] J. Als-Nielsen and D. McMorrow, *Elements of Modern X-ray Physics, Second Edition*. John Wiley & Sons, Ltd, 2011.

[68] D. G. Greene, D. V. Ferraro, A. M. Lenhoff, and N. J. Wagner, "A critical examination of the decoupling approximation for small-angle scattering from hard ellipsoids of revolution," *Journal of Applied Crystallography*, vol. 49, pp. 1734 – 1739, 10 2016.

[69] ——, "A critical examination of the decoupling approximation for small-angle scattering from hard ellipsoids of revolution," *Journal of Applied Crystallography*, vol. 49, pp. 1734 – 1739, 2016.

[70] J. S. Pedersen, "Analysis of small-angle scattering data from colloids and polymer solutions: modeling and least-square fitting," *Advances in Colloid and Interface Science*, vol. 70, pp. 171 – 210, 1997.

[71] M. Koch, P. Vachette, and D. Svergun, "Small-angle scattering: A view on the properties, structures and structural changes of biological macromolecules in solution," *Quarterly Reviews of Biophysics*, vol. 36, no. 2, pp. 147–227, 2003.

[72] R. Høiberg-Nielsen, P. Westh, L. K. Skov, and L. Arleth, "The effect of glycosylation on interparticle interactions and dimensions of native and denatured phytase," *Biophysical Journal*, vol. 96, pp. 153 – 161, January 2009.

[73] V. Receveur, D. Durand, M. Desmadril, and P. Calmettes, "Repulsive interparticle interactions in a denatured protein solution revealed by small angle neutron scattering," *FEBS Letters*, vol. 426, no. 1, pp. 57 – 61, 1998.

[74] A. P. Minton, "Recent applications of light scattering measurement in the biological and biopharmaceutical sciences," *Analytical Biochemistry*, vol. 501, pp. 4 – 22, 2016.

[75] O. Glatter, *Scattering Methods and their Application in Colloid and Interface Science*. Elsevier Inc, 2018.

[76] D. Orthaber, A. Bergmann, and O. Glatter, "Saxs experiments on absolute scale with kratky systems using water as a secondary standard," *Journal of Applied Crystallography*, vol. 33, pp. 218 – 225, 2000.

[77] W. Li, B. A. Persson, M. Lund, J. Bergenholtz, and M. Z. Oskolkova, "Concentration-induced association in a protein system caused by a highly directional patch attraction," *Journal of Physical Chemistry B*, vol. 120, pp. 8953 – 8959, 2016.

[78] J. Trewhella, A. P. Duff, D. Durand, F. Gabel, J. M. Guss, W. A. Hendrickson, G. L. Hura, D. A. Jacques, N. M. Kirkby, A. H. Kwan, J. Pérez, L. Pollack, T. M. Ryan, A. Sali, D. Schneidman-Duhovny, T. Scwede, D. I. Svergun, M. Sugiyama, J. A. Tainer, P. Vachette, J. Westbrook, and A. E. Whitten, "2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: an update," *Acta Crystallographica section D Structural Biology*, vol. D73, pp. 710 – 728, 2017.

[79] M. H. Mendenhall, A. Henins, L. T. Hudson, C. I. Szabo, D. Windover, and J. P. Cline, "High-precision measurement of the x-ray cu kα spectrum," *Journal of Physics B: Atomic, Molecular and Optical Physics*, vol. 50, no. 11, p. 115004, 2017.

[80] D. I. Svergun, M. H. J. Koch, P. A. Timmins, and R. P. May, *Small Angle X-ray and Neutron Scattering from Solutions of Biological Macromolecules*, 1st ed. Oxford University Press, 2013.

[81] *Macromolecular SAXS*. Rigaku Oxford Diffraction, June 2018, no. 1.

[82] M. Nyam-Osor, D. V. Soloviov, Y. S. Kovalev, A. Zhigunov, A. V. Rogachev, O. I. Ivankov, R. V. Erhan, and A. I. Kuklin, "Silver behenate and silver stearate powders for calibrationi of sas instruments," *Journal of Physics: Conference Series*, vol. 351, p. 012024, 2012.

[83] A. K. Attri and A. P. Minton, "New methods for measuring macromolecular interactions in solution via static light scattering: basic methodology and application to nonassociating and self-associating proteins," *Analytical Biochemistry*, vol. 337, pp. 103 – 110, 2005.

[84] D. Some and S. Kenrick, *Protein Interactions*. Rijeka: IntechOpen, 2012, chapter 20: 'Characterization of Protein-Protein Interactions via Static and Dynamic Light Scattering'.

[85] A. K. Attri and A. P. Minton, "Composition gradient static light scattering: A new technique for rapid detection and quantitative characterization of reversible macromolecular hetero-associations in solution," *Analytical Biochemistry*, vol. 346, pp. 132 – 138, 2005.

[86] H. Zhao, P. H. Brown, and P. Schuck, "On the distribution of protein refractive index increments," *Biophysical Journal*, vol. 100, no. 9, pp. 2309 – 2317, 2011.

[87] *DAWN HELEOS II User's Guide*. Wyatt Technology Corporation, 2015, m3200 Rev B.

[88] *Optilab (U)T-rEX User's Guide*. Wyatt Technology Corporation, 2014, m1510 Rev C.

[89] *NanoDrop One User Guide*. Thermo Fisher Scientific, July 2016, revision B.

[90] *Operating Manual , Refrigerated/Heating Circulators*. Julabo, September 2019, 1.953.2809-V7.

[91] D. F. Swinehart, "The beer-lambert law," *Journal of chemical education*, vol. 39, no. 7, pp. 333 – 335, 1962.

[92] S. C. Gill and P. H. von Hippel, "Calculation of protein extinction coefficients from amino acid sequence data," *Analytical Biochemistry*, vol. 182, pp. 319 – 326, 1989.

[93] J. B. H. R. E. Gillilan and S. Skou, "Bioxtas raw: improvements to a free open-source program for small-angle x-ray scattering data reduction and analysis," *Journal of Applied Crystallography*, vol. 50, pp. 1545 – 1553, 2017.

A 139 mg/mL 60°C WAXS

The scattering profile of the 139 mg/mL Actrapid sample stored at 60°C from WAXS is shown in Figure A.1. The scattering profile does not show an amyloid peak at q = 1.3 $Å^{-1}$, which is clear from the zoom in Figure A.2. The scattering curve is normalized by concentration, but not absolute scaled.

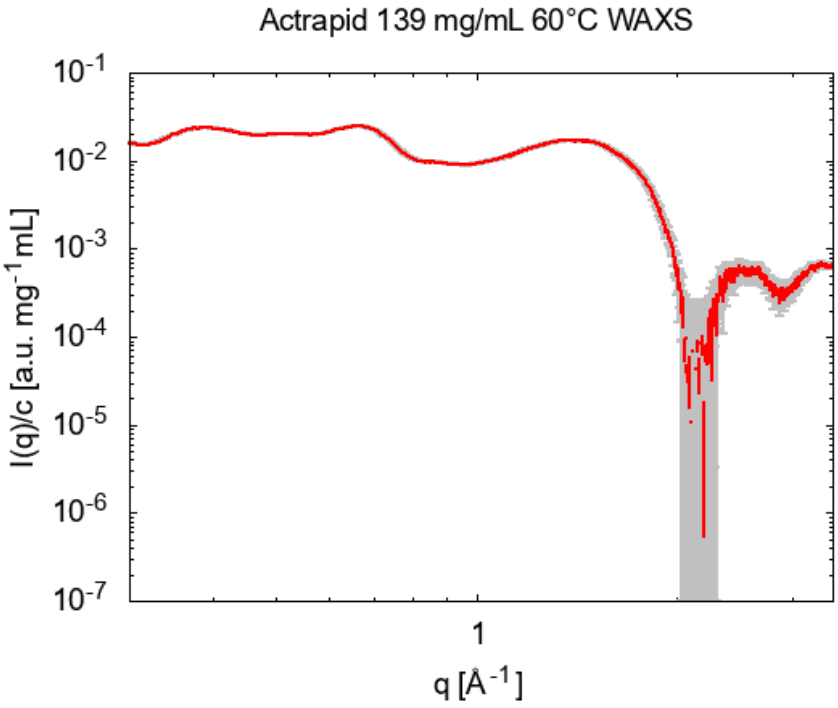

Figure A.1: WAXS scattering profile of the 139 mg/mL Actrapid sample stored at 60°C.

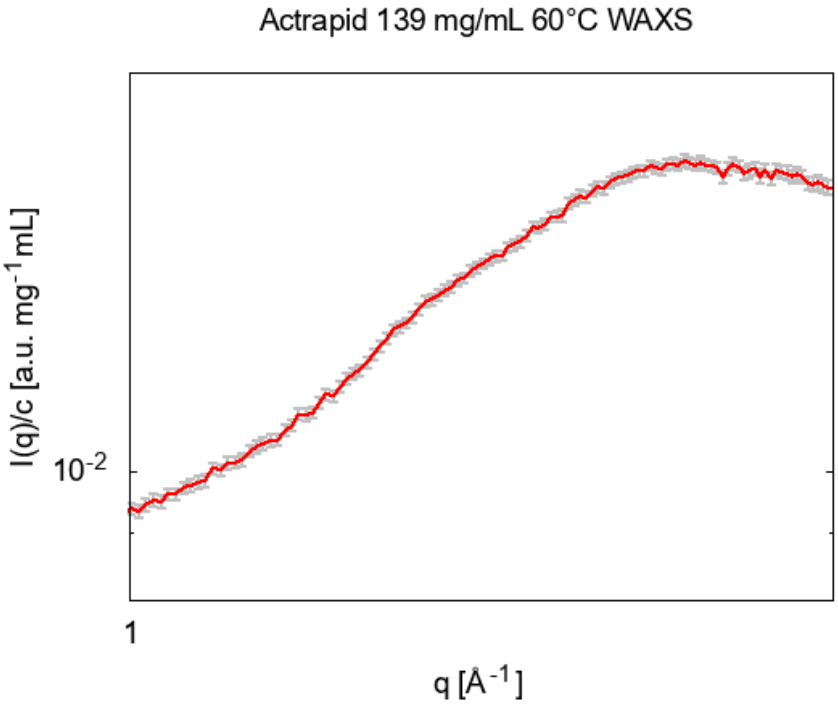

Figure A.2: Zoom of the scattering curve, to get a better view of the region around q = 1.3 $Å^{-1}$.

B Guinier Analysis Results

In the tables below, the forward scattering and radius of gyration of all Guinier analysis are shown, as well as the coefficient of determination values, R^2, of each linear fit. First the 25°C scattering data, second the 5°C data and lastly data of Actrapid with and without m-cresol.

Sample Concentration (mg/mL)	r_g (Å)	I(0) ($cm^{-1}mg^{-1}mL$)	R^2
2	19.70 ± 0.13	0.0248 ± 0.0001	0.99
3	19.39 ± 0.10	0.0257 ± 0.0000	0.99
9	20.63 ± 0.10	0.0278 ± 0.0000	1.0
56	22.68 ± 0.25	0.0348 ± 0.0004	0.99
66	22.49 ± 0.24	0.0350 ± 0.0004	0.99
111	22.69 ± 0.24	0.0346 ± 0.0004	0.99
133	22.89 ± 0.24	0.0345 ± 0.0004	0.99

Table B.1: Forward scattering and radius of gyration from the Guinier analysis of the 25°C scattering data

Sample Concentration (mg/mL)	r_g (Å)	I(0) ($cm^{-1}mg^{-1}mL$)	R^2
1	19.85 ± 0.18	0.0246 ± 0.0001	0.98
3	19.88 ± 0.11	0.0263 ± 0.0001	1.0
7	20.12 ± 0.09	0.0266 ± 0.0000	1.0
22	20.48 ± 0.10	0.0283 ± 0.0001	0.99
31	20.95 ± 0.16	0.0296 ± 0.0002	0.99
46	21.15 ± 0.17	0.0310 ± 0.0002	0.99
62	20.80 ± 0.21	0.0320 ± 0.0003	0.99
67	20.72 ± 0.25	0.0324 ± 0.0004	0.99
73	20.62 ± 0.35	0.0324 ± 0.0005	0.98
82	20.96 ± 0.35	0.0334 ± 0.0006	0.98
97	20.83 ± 0.25	0.0330 ± 0.0004	0.99
114	21.00 ± 0.23	0.0325 ± 0.0004	0.99
133	20.66 ± 0.20	0.0320 ± 0.0003	0.99

Table B.2: Forward scattering and radius of gyration from the Guinier analysis of the 5°C scattering data

	Sample Concentration (mg/mL)	r_g (Å)	I(0) ($cm^{-1}mg^{-1}mL$)	R^2
with m-cresol	0.865	19.39 ± 0.20	0.0245 ± 0.0002	0.97
	1.73	19.43 ± 0.14	0.0276 ± 0.0001	0.99
	3.46	19.76 ± 0.12	0.0262 ± 0.0001	0.99
without m-cresol	0.865	20.02 ± 0.20	0.0257 ± 0.0002	0.98
	1.73	20.68 ± 0.14	0.0276 ± 0.0001	0.99
	3.46	21.29 ± 0.13	0.0296 ± 0.0001	1.0

Table B.3: Forward scattering and radius of gyration from the Guinier analysis of the scattering data with and without m-cresol

C Comparison of Scattering Profiles

In Figure C.1, are the absolute scaled and concentration normalized scattering curves of all samples measured at 25°C along with the scattering curve of the gelled 139 mg/mL Actrapid sample (salmon) shown, to visualize the normalization of the 139 mg/mL curve is in agreement with the 25°C data, and prove that no evaporation occurred in the sample during storage at 60°C.

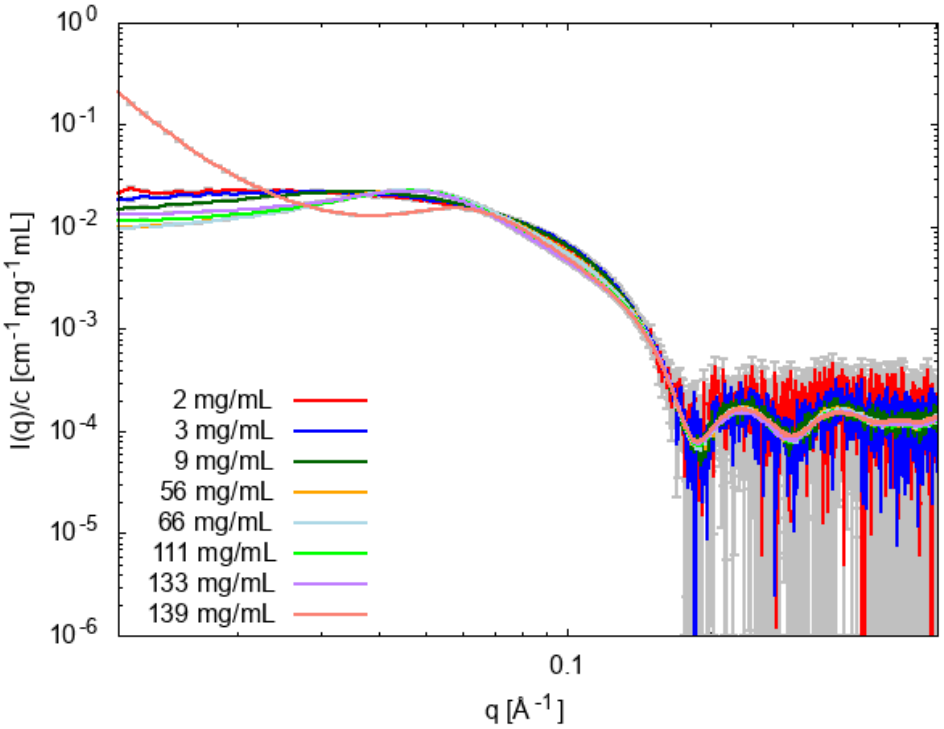

Figure C.1: Scattering profiles of all samples of the temperature study measured at 25°C including the scattering curve of the 139 mg/mL Actrapid sample stored at 60°C.