# 1 Introduction

Crystallization of proteins was in past decades mostly used to prepare large and highly pure crystals, to determine the structure of the protein crystals by X-ray measurements [Des11]. With the increase of industrial applications of proteins [Bre08], the desired amount of high purity proteins is increasing, too. Crystallized proteins offer an improved shelf life with a very high degree of preservation of the protein enzyme activity. A further important advantage of protein crystallization is purification at low cost in industrial mass production [Hun03; Jon14]. A large number of studies on the crystallization of proteins with the aim to optimize purity, yield, shape or activity were performed, e. g. by Durbin and Feher [Dur96] or McPherson [McP90]. In research papers on protein crystallization, lysozyme is the most often described protein because it is easily available and it is of fairly low cost. In most cases, however, only the crystallization processes are investigated, the dissolution is neglected even though dissolution of crystals is of most importance for the use of the proteins after storage. Dissolution with respect to the mechanism is known to be the reverse of the crystal growth process (see e. g. Ulrich and Stelzer [Ulr11]).
Müller and Ulrich [Mül12a] described the first time an unusual dissolution behavior for protein crystals which is different from a conventional dissolution behavior, and cannot be explained in the same way as the traditionally known dissolution processes. For a better controlled use of crystallized proteins in future, it is necessary to observe and understand those dissolution mechanisms. One reason for the importance of an understanding of the prediction of the dissolution behavior is bioavailability of pharmaceuticals. Due to the high price most proteins could be examined in only a low quantity. Therefore, new technologies of the examination need to be developed. Furthermore, the term “protein crystal” is not well understood. Usually it is assumed that the crystallization of a protein follows the same rules as a crystallization of conventional crystals. A protein crystal contains, how-

ever, not only protein but also water, buffer and salt (see e. g. Ulrich and Pietzsch [Ulr15]). Also, the molecule size of a protein is much larger than inorganic and even most of the organic compounds. A protein delivers many functional groups distributed over the molecule chain length.
All different components in a crystal, when they are a matter of change, can if big enough change the crystal lattice. Definitely, a change of the amount of a component in a crystal will change its chemistry. This is very much like in solvates or hydrates, but it can also be imagined that the change of the environment (composition of the *liquid phase* around the crystal) of a protein crystal might not change its chemistry but just its lattice, then a classical polymorph would be faced. In both cases (solvate or polymorph) the physical properties (density, solubility, dissolution rate, color, hardness, etc.) can be changed. A phase diagram would be extremely helpful in order to determine the stability ranges (the respective metastable zones). There are, however, not much of such data known.
It is necessary to know whether a crystal pattern of one crystal modification stays the same even if there is e. g. a slight deviation in pH value existing during crystal growth or not. With all of these information's it should be possible to present a new understanding in the crystallization of proteins and on the dissolution mechanism, too, as well as their metastable zone widths.

# 2 State of the art

The number of papers on protein research respectively protein crystallization, is large. In this Chapter only the literature is cited which is necessary to explain and understand the crystallization of the protein lysozyme and the chemistry to understand structural conditions. This Chapter introduces the fundamentals in order to be able to follow the discussion in this work. Most of the information can be found in textbooks (e. g. [Wie02; Jon14; Mul01]).

## 2.1 Proteins

A protein is a biological macromolecule. Those macromolecules are built of amino acids connected by peptide bonds. All proteins are built based on 20 different amino acids (e. g. see Karlson et al. [Kar94]). The resulting protein is a long peptide chain of amino acids with varying number and sequence.
The structure of a protein is not only a long straight line. The protein structure is described by the primary, secondary, ternary and quaternary structure. The primary structure is the sequence of amino acids. The secondary structure describes the local substructure due to hydrogen bonds between CO and NH groups. The ternary structure describes the three-dimensional structure of a protein molecule. In addition to the hydrogen bonds of the secondary structures, some more bonds are responsible for the formation and stabilization of such a structure. Those bonds are hydrophobic interactions, disulfide bonds between cysteine groups by dehydration of HS groups, and some ion bonds due to positive and negative charged side chains. The quaternary structure is formed by the arrangement of more than one protein molecule. Existing complexes of two domains as dimers and of several domains are called multimers. Depending on the properties of the side chains, e. g. the charges of side chains, the domains

are organized in a specific order. Many proteins do not build a quaternary structure, they function as monomers. An overview of the build-up of the different structure elements of a protein can be seen in Figure 2.1 according to Karlson et al. [Kar94].

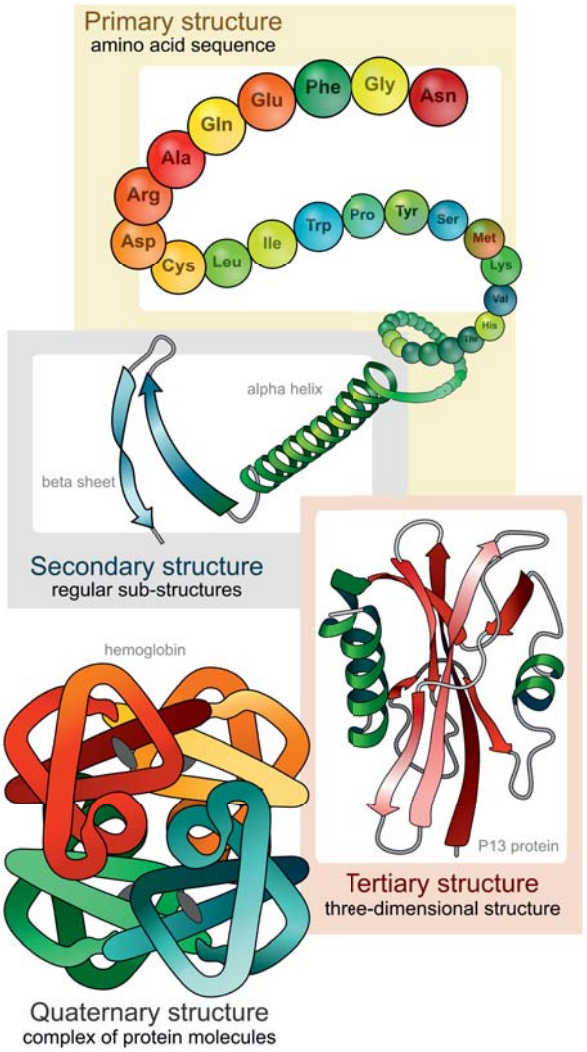
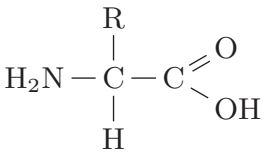



**Figure 2.1:** Overview of the structure of proteins according to Karlson et al. [Kar94]

### 2.1.1 Amino acids

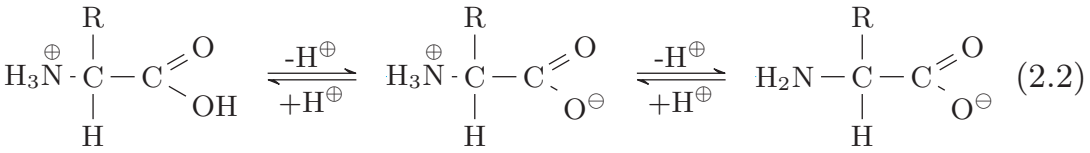
Amino acids consist of a central alpha carbon atom which connects an amino group, a carboxyl group and a mutable side chain to each other. This can be seen in the structural Formula 2.1. The residue side chain

is characteristic for different amino acids. There are non-polar aliphatic and aromatic amino acids, and amino acids with uncharged polar residual groups available. But there are also amino acids existing with positive and negative charged side chains.




R=mutable side chain (2.1)

The charge of this side chains is impressionable by pH value. The equilibrium of each charge is defined by the pKa value. The effect of changed charges by the pH value on the ends of the protein chain is shown in reaction Equation 2.2. Also, the residual is often able to be charged.

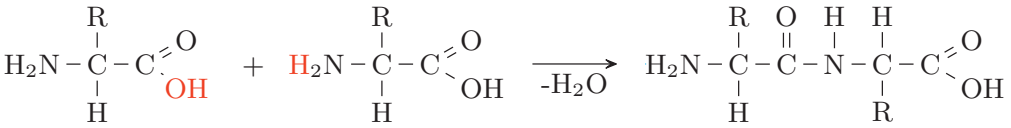



(2.2)

positive charged neutral negative charged

## 2.1.2 Peptide bond

A peptide bond which is used to form the primary structure of a protein is a strong covalent bond. The formation of a peptide bond is shown in the reaction Equation 2.3.




(2.3)

## 2.1.3 Surface charge

Based on the above stated it is possible to assume that if an amino acid is able to have a point which can be electrostatically charged, the built

protein molecules have several charges distributed over the whole molecule chains due to the formation of polypeptide bonds between the amino acids. In science this is called a polyelectrolyte [Pod06].
The different charges with different polarity, and a formation of disulfide and hydrogen bonds, form the secondary and ternary structure of the protein molecules. Some of the charges of different polarity neutralize each other already, other charges are located at the exterior of the protein molecule. Those have to be neutralized by the solvent molecules in the *liquid phase*, or so called "counter ions" in the *solid state*.

### 2.1.4 Isoelectric point

The pI (isoelectric point) describes the pH value at which the positive and negative charges will balance and the net charge will be zero. Usually in literature the protein solubility is minimal at the protein's isoelectric point [Ric92]. If there is a charge at the protein surface, the protein prefers to interact with water, rather than with other protein molecules. The charge makes the protein more soluble.

### 2.1.5 Active site

In chemistry, an active site is called the point at which the catalyzed reaction of a catalyst occurs. In the field of enzymes, it describes the area of the enzyme which is responsible for the catalyzed reaction. Catalytic residues of the site interact with the substrate to lower the activation energy of a reaction and therefore to make it proceed faster. There are two models of binding the substrate to the enzyme, the "lock and key hypothesis" and the "induced fit" [Sul08]. Due to this property, a protein molecule is able to increase the reaction rate of a specific reaction. This is called the "activity of a protein" which can be measured by methods depending on the chosen protein. If the structure of a protein is influenced by the pH, the activity of a protein is affected, too. Here are still some open questions. If there are different modifications existing which are crystallized at different pH regions, is also the enzymatic activity influenced by the modification? Will the modification have an influence on the shelf life of the activity? Furthermore, there are no reasonable information about

the shelf life of crystallized proteins by means of activity preservation with changing humidity and temperature conditions in literature given yet.

### 2.1.6 UV/Vis activity

UV/Vis activity describes a phenomenon of a substrate, which will be illuminated by light with a wavelength in region of UV up to visible light, is able to decrease the light emission at a small wavelength region. This is due to movement of electrons from a lower energy level to a higher level. The wavelength can be calculated by use of the plank constant and the energy difference between both levels. Different energy levels exist for organic molecules which are unsaturated. Depending on the concentration of those substrate the absorbance is changed. With use of several substrate concentration a calibration curve can be plotted, to subsequently measure the concentration of any substrate with a UV/Vis spectrophotometer. In 1969, Aune and Tanford [Aun69] described this phenomenon for the first time for the protein lysozyme. In 1989, Gill and von Hippel [Gil89] presented a method to calculate the extinction coefficient of any protein by the known amino acid sequence. This is possible due to the fact that many of the known amino acids are able to be UV active depending on the residual chain. A protein is a combination of those amino acids (see above) and by formation of the peptide bond between amino acids and therefore also UV active.

## 2.2 Buffersystems in protein crystallization

### 2.2.1 pH buffer

Per definition a buffer is a mixture of substances with a pH which changes very little when a small amount of strong acid or base is added to it, and thus it is used to prevent changes in the pH of a solution. A buffer solution (also called pH buffer) is an aqueous solution consisting of a mixture of weak acid and its conjugated base, or vice versa.
With use of the Handerson-Hasselbalch equation, the context between the pH value and the point of the equilibrium of the acid-base-reaction of an acid and its corresponding base is described. With this equation, it is

possible to calculate the pH value out of the concentration of buffer components. More information on the concept of acid and base can be found in textbooks (e. g. see Holleman and Wiberg [Hol07] or Riedel [Rie04]).

### 2.2.2 Precipitants

With precipitants, the solubility of proteins can be affected in different ways depending on the used ions. Hofmeister ranked the efficiency of a variety of electrolyte types to salt-in and salt-out of proteins in 1888 [Hof88]. Due to this effect, a new solvent system is created and thus has a different solubility on the solved protein and other substances.

#### Salting-in

The salting-in can be explained by the Debye-Huckel theory. The electrolyte concentration affects the value $\kappa$, the inverse Debye length. An increase of the electrolyte concentration causes an increase in the ionic strength, which increases $\kappa$ and thus, protein solubility. This theory is working with low ionic strength but fails for moderate and high electrolyte concentrations. Also, it fails to predict the commonly observed salting-out behavior of proteins [Wie02].

#### Salting-out

A popular explanation for this effect relies on the relative hydration of the protein versus bulk electrolyte. The electrolyte is assumed to bind bulk water, as water of hydration near the ions surface. The bulk electrolyte and the protein molecules compete for bulk water to hydrate their respective surfaces. The dehydrated protein molecules fill such exposed dehydrated surface with other protein molecules. Thus, the solubility of the protein is reduced [Wie02].

### 2.2.3 Anti-solvent

An anti-solvent is the addition of a small polar organic solvent, as an example methanol or ethanol. It works in a similar manner as electrolytes do and competes with the protein for water molecules. Furthermore, the solvent tends to lower the dielectric constant of the water and enhances

electrostatic interactions between protein molecules, which reduces their solubility. The lower dielectric constant tends to solubilize the hydrophobic core of the protein molecule. This results in unfolding and denaturation of the protein. (see e. g. Wiencek [Wie02])

### 2.2.4 Polymers

Polymers are also often used to lower the solubility. Especially polyethylene glycols (PEGs) of different molecular weight are used. The molecular weight must be optimized for each protein. At the moment it is not clear how it is working, but it tends to be a similar way as an anti-solvent. It will hydrate and change the dielectric constant of the solvent also. (see e. g. Wiencek [Wie02])

## 2.3 Crystallization

### 2.3.1 Metastable zone

The metastable zone covers the region between two lines in the plot of concentration versus temperature. The first line in this plot is the solubility line. The solubility line is a thermodynamic value, it covers the maximum amount of a substance which can be dissolved in the used solvent. The second line is the nucleation line. This is a kinetic value. This means the position of the nucleation line is influenced by the method which is used to determine this value, and it is depending on process parameters.
The first major investigation on solubility of proteins was presented by Howard et al. [How88]. It covers the temperature dependence over a range of pH values between 4 and 7.5, and also the amount of salt was investigated. The data show that the solubility of lysozyme is strongly dependent upon the sodium chloride concentration. This behavior is consistent with the inverse Hofmeister series behavior (e. g. see Riès-Kautt and Ducruix [Riè91] and Zhang and Cremer [Zha09]) as expected for a basic protein [Jon14].
A more comprehensive study of lysozyme chloride solubility was carried out by Pusey's group [Pus88; Cac91a; Cac91b; For96; For99a] using the column method presented in 1988 [Pus88]. The results show clearly that

protein solubility behaves in a similar manner when compared to the majority of small molecule organic compounds, in that the solubility increases with temperature [Jon14]. The behavior of the solubility as function of the precipitant concentration decreases with increasing salt concentration. For the acetate buffer concentration in the range of 0.01 up to 0.5 M the solubility appears to have a minimum for an acetate concentration of 0.1 M. Already Cacioppo and Pusey [Cac91b] in 1991 showed in their 3D plots the influence by the pH value at a higher solubility at certain pH values. The influence by the pH value is not straight-forward as oscillations were observed within the small observed pH intervals. The overall data show, however, the expected trend for a solubility decrease as the pH value approach the pI (see Chapter 2.1.4). In the paper of Forsythe et al. [For99a], the following sentence is written: *"The effect of pH is not clearly seen and varies unpredictably with pH"*. There is a lack of information on the influence of the pH value on the solubility of protein crystals, since there is no mechanism known to explain the results based on a scientific background. An overview of the effects of the pH value in the field of crystallization (except protein crystals) can be found in the work of Mohameed [Moh96b]

### 2.3.2 Kinetics

The growth rates of protein crystals are typically measured via a microscope with controlled image capturing systems. The crystals grow in a temperature controlled cell. A capturing system takes images and from the length difference of a growing crystal over a period of time. Therefrom, the growth rate is calculated [Dur86]. Due to the limited amount of mother liquor in such a cell and the small space the microscope system is able to capture, only a small portion of crystals can be measured at the same time. But there is GRD (growth rate dispersion) existing, which means each crystal is different from each other, some crystals are growing fast, others are growing slow. An automated computer supported growth rate measurement system is presented by Pusey [Pus93]. This system is able to measure 40 crystals at the same time within one cycle. If 40 crystals can be placed and measured at the same time in such a system, still it is not clear if these are fast or slow growers? To determine the average rates, much more crystals must observed. An overview about this phenomenon