We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Rational and Irrational Approaches to Convince a Protein to Crystallize

André Abts, Christian K. W. Schwarz, Britta Tschapek,
Sander H. J. Smits and Lutz Schmitt
*Institute of Biochemistry, Heinrich-Heine University, Düsseldorf,
Germany*

1. Introduction

The importance of structural biology has been highlighted in the past few years not only as part of drug discovery programs in the pharmaceutical industry but also by structural genomics programs. Mutations of human proteins have been long recognized as the source of severe diseases and a structural knowledge of the consequences of a mutation might open up new approaches of drugs and cure. Although the function of a protein can be studied by several biochemical and/or biophysical techniques, a detailed molecular understanding of the protein of interest can only be obtained by combining functional data with the knowledge of the three-dimensional structure. In principle three techniques exist to determine a protein structure, namely X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR) and electron microscopy (EM). According to the protein data bank (pdb; http://www.rcsb.org) that provides a general and open-access platform for structures of biomolecules, X-ray crystallography contributes more than 90% of all structures in the pdb, a clear emphasis of the importance of this technique.

To perform X-ray crystallography it is essential to have large amounts of pure and homogenous protein to perform an even today still "trail and error"-based screening matrix to obtain well diffracting protein crystals. Therefore, successful protein crystallization requires three major and crucial steps, all of them associate with specific problems and challenges that need to be overcome and solved. These steps are (I) protein expression, (II) protein purification and (III) the empirical search for crystallization conditions. As summarized in Figure 1, every single step needs to be optimized along the long and stoney road to obtain protein crystals suitable for structure determination of your "most-beloved" protein via X-ray crystallography. This chapter will focus on these three steps and suggests strategies how to perform and optimize each of these three steps on the road of protein structure determination.

2. Protein expression (I)

To crystallize a protein, the first requirement is the expression of your protein in high amounts and most importantly on a regular basis. This implies that it is possible to obtain a freshly purified protein at least weekly. In general, it is possible to express a protein either homologously or heterologously (see Figure 1 – (I) expression). Especially for large proteins,
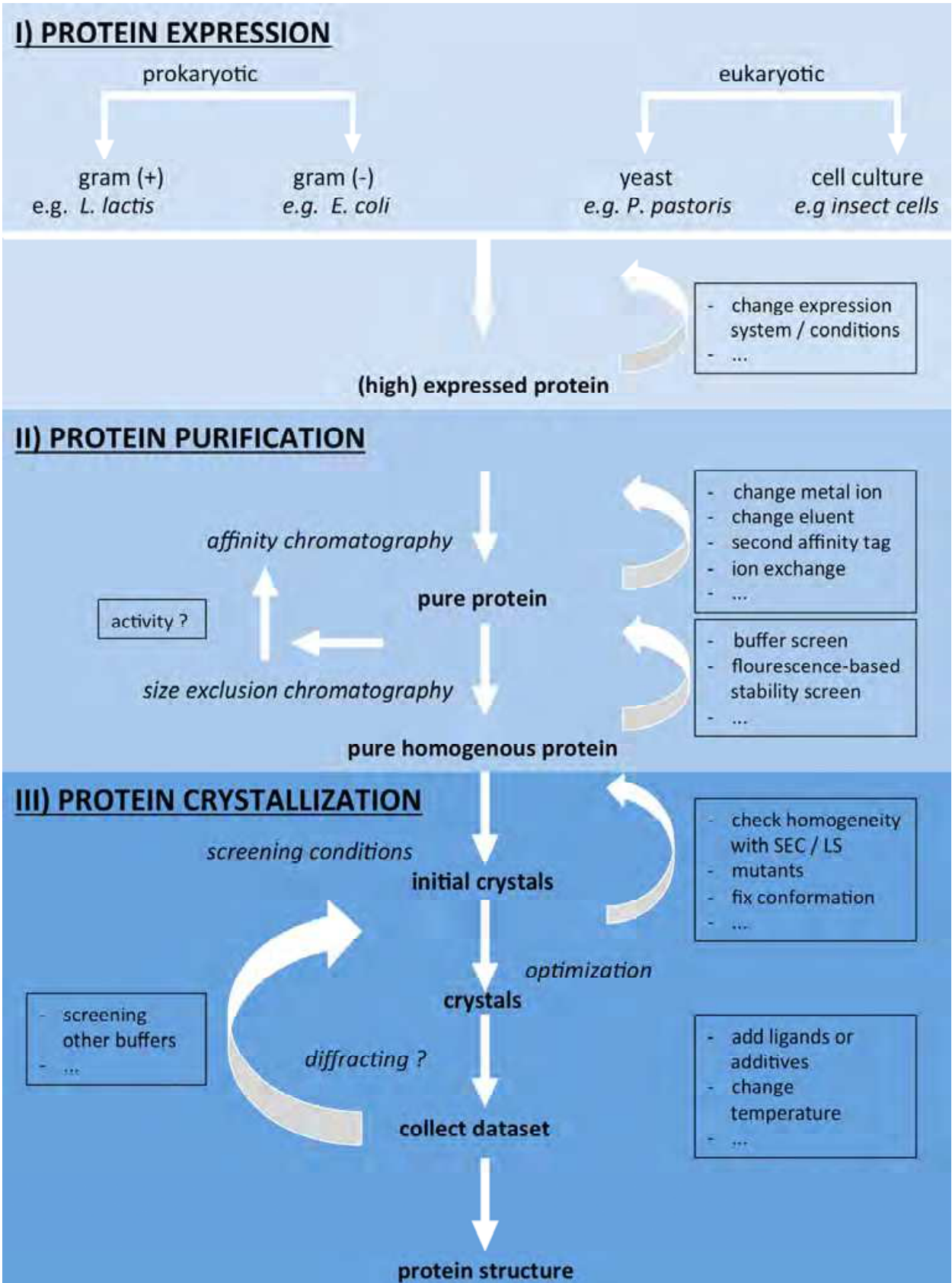
Fig. 1. Schema highlighting the three steps towards a protein crystal. (I) Expression (II) Purification (III) Protein crystallization.

proteins containing a co-factor or a ligand, the natural habitat is likely the best choice to express the protein. However, often the natural host, for example humans, produce only low amounts of protein and suitable overexpression protocolls are not available. To circumvent this problem, several expression strains, cell lines as well as a large number of expression vectors have been developed to allow expression of any protein in a different host (heterologous expression). In general, the used organisms for protein expression can be divided into two different groups: prokaryotic and eukaryotic expression hosts. The natural organism of the protein of interest mainly dictates the choice, which expression system to use. If working with a bacterial protein, it is very likely that also a prokaryotic host is able to express the protein in high amounts. The same holds true for proteins originating from a eukaryotic host, which is likely best overexpressed in a eukaryotic host. These proteins often require posttranslational modification such as glycosylation or disulfide bond formation, which are possible in eukaryotic expression hosts. The most common used heterologous expression host is the gram negative bacteria *Escherichia coli* since it is commercially available and a large number of expression cassettes have been developed. Thus, it is the most widely used expression system with expression rates of several mg/L of culture. The best characterised and understood expression hosts are described in more detail below and the commecially available systems are listed in table 1.

2.1 Expression hosts – Prokaryotic
2.1.1 Gram negative - *E. coli*
As mentioned above, *E. coli* is the most common used expression system (Figure 1 – (I) expression, left side). This is further highlighted by the fact that 80% of all protein structures deposited in the protein data bank were overexpressed in *E. coli* (Sorensen and Mortensen 2005). There are several advantages promoting *E. coli* as expression host: (A) Cultivation of *E. coli* is simple and a doubling time of 30 minutes is rather quick allowing the fast generation of biomass, (B) genetics are well understood and any genetical manipulation is well established, (C) expression levels of up to 60 % of the total protein mass within the cell make the next step, protein purification rather straight forward and finally (D) the cultivation does require only standard equipment normally present in every biochemical laboratory and therefore expression using *E.coli* is relatively cheap. In the last decades, many different plasmid based expression systems have been developed such as the pET vector systems, which contain several different expression plasmids with a choice for the affinity tag on either termini of the protein as well as the possiblity to use a dual cassette when expressing two or more proteins at once. The selection pressure derived from different antibiotics, and the resistance genes encoded on these plasmids further simplify laboratory practice. Only cells harbouring the right plasmids are able to grow and therefore expenditure on sterility is low.

The typical *E. coli* expression system is plasmid-based, which can be transferred to different *E. coli* strains ((Sorensen and Mortensen 2005), Novagen pET vector table). An *E. coli* expression vector consists mainly of five important parts: the replicon, a resistance marker, a promotor and a so-called multiple cloning site (MCS) (Baneyx 1999; Jonasson, Liljeqvist et al. 2002). The replicon is the crucial part of a plasmid to maintain it inside a cell. It is recognized and duplicated by the replication machinery (Baneyx 1999). The selection marker allows the identification of cells carrying a plasmid as it encodes for a resistance, e.g. against antibiotics (see above)(Sorensen and Mortensen 2005). The promotor sequence is the recognition site for the RNA polymerase, however it is inactive under initial cultivation conditions. The addition of an inducer (sometimes also a temperature change) switches the

promotor from 'off' to 'on' whereby the expression is initiated (Jana and Deb 2005). Common inducers are isopropyl-β-D-thiogalactopyranosid (IPTG) in the pET system or arabinose for pBAD vectors (Invitrogen™). The multiple cloning site (MCS) is a short DNA segment combining many (up to 20) restriction sites. This feature simplified the insertion of genes into the plasmid enormously and made cloning procedures very convenient. However, new cloning strategies, which are independent of restriction enzymes and ligases, are emerging and will replace the standard approaches some day (see for example Li and Elledge 2007).

2.1.2 Gram positive - *L. lactis*

Within the prokaryotic expression system also some gram (+) bacteria are used for protein overexpression (Figure 1 – (I) expression left side). Here, lactic acid bacteria play a privotal role. They are used in food industry and are known since 1873 when Joseph Lister isolated the first strain (Teuber 1995; Mierau and Kleerebezem 2005). For the overexpression of recombinant proteins, there are many lactic acid bacteria around, *lactococci, lactobacilli, streptococci* and *leuconostocs*. (Gasson 1983; van de Guchte, Kok et al. 1992; de Vos and Vaughan 1994). The best characterized and most widely used host is *Lactococcus lactis*, which is famous for its usage within food fermentation and like for *E. coli* the genome, metabolisms and molecular modifications are well known and established (Bolotin, Wincker et al. 2001; Guillot, Gitton et al. 2003). Thus, it has been called the 'bug of the next millennium'(Konings, Kok et al. 2000). All established gram (+) bacteria expression hosts are able to overexpress proteins homologously or heterologously. Since only one, namely the cyctoplasmic membrane is present (Kunji, Slotboom et al. 2003), this host is in comparison to the two-membrane system of gram (-) bacteria, a good choice to express eukaryotic as well as prokaryotic membrane proteins or proteins with membrane anchors (Kunji, Slotboom et al. 2003). The promoter used for expression in *L. lactis* is induced by the external addition of nisin. Nisin is an antimicrobial active peptide, which interacts with lipid II in the cytoplasmic membrane of gram (+) bacteria and causes cell lysis. Interestingly, nisin is produced by *L. lactis* itself. The expression strain is deleted of the nisin producing genes and therefore external nisin can be used as inductor. Nisin binds to NisK, which as part of a two-component system, phosphorylates NisR, which in turn binds to the promotor P_{nisA} thereby allowing synthesis of the protein located downstream on the plasmid. Since nisin is also active against *L. lactis* itself, the concentration range of nisin used in such expression studies is relatively narrow to circumvent killing of the *L. lactis* expression strain. This is clearly a draw back of this expression system since expression can basically only turned on with a certain nisin concentration. Between the inducer concentration and the expressed protein a linear behaviour is observed. Unfortunately the nisin concentration range between the minimal and maximum nisin concentration is very small, nisin concentration higher than 25 ng per liter of cells, cause cell death.

However, *L. lactis* has been proven to be a very efficient expression system. Kuipers et al. created many expression hosts and plasmids to produce any protein of interest by cloning it downstream of the P_{nisA} promotor. With this nisin inducible (NICE) expression system, it is now possible to induce the protein production with minimal concentration (0.1 – 5 ng) of nisin (de Ruyter, Kuipers et al. 1996; Kuipers, de Ruyter et al. 1998). The amount of produced recombinant protein can reach up to 50 % of the total intracellular proteins (Kuipers, Beerthuyzen et al. 1995; de Ruyter, Kuipers et al. 1996). Following a few examples for expressed proteins in *L. lactis* whose structures have been solved: an ECF-type ABC transporter (PDB:3RLB)(Erkens, Berntsson et al. 2011), a peptide binding protein OppA

(PDB:3RYA)(Berntsson, Doeven et al. 2009) and the multidrug binding transcriptional regulator LmrR (PDB:3F8B)(Madoori, Agustiandari et al. 2009).

2.2 Eukaryotic expression hosts

The great benefit of choosing a eukaryotic host for overexpression of a protein of interest are the availability of a posttranslational modification system as well as the frequently enhanced protein folding (Midgett and Madden 2007). Eukaryotic proteins tend to misfold or lack biological activity when expressed in prokaryotic expression systems such as *E. coli* (Cregg, Cereghino et al. 2000; Midgett and Madden 2007). To overexpress these proteins, different yeast strains, insect cells or even mammalian cell lines have been developed as expression hosts (Figure 1 - (I) expression, right side). Eukaryotic expression systems are often more expensive, provide low expression levels and are sometimes hard to handle, when compared to bacterial systems. However, the genetic and cellular contexts are more similar to the original protein-expressing organism (Midgett and Madden 2007). In the following sections, some of the commonly employed eukaryotic expression systems will be described.

2.2.1 Yeast expression systems - *Saccharomyces cerevisiae* and *Pichia pastoris*

The most widely used yeast strains to express protein are *Saccharomyces cerevisiae* and *Pichia pastoris*, which offer the major advantage of a posttranslational modification system for glycosylation, proteolytic processing as well as disulfide bond formation, which for some proteins are essential for the function and/or correct folding (Cregg, Cereghino et al. 2000; Midgett and Madden 2007). The handling of yeast expression systems is similar to prokaryotic systems with respect to the genetic background and cultivation. Similar to the bacterial vector systems, expression in yeast starts with a plasmid-based cloning part which can be performed in *E. coli* (Cregg 2007). Afterwards the expression cassette gets integrated into the genome by simple homologous recombination in the yeast. One major advantage in *P. pastoris* is the insertion of multiple copies of the protein DNA-sequence into genomic DNA, which increases expression yield.

The biggest advantage of yeast as expression system is that well established protocols for fermentation are available. Optimal fermentation of *P. pastoris* can end up with more than 130 gram of cells per liter of culture. Even if expression levels in the cell are not that high the mass of cells easily compensates for this disadvantage (Wegner 1990; Cregg, Cereghino et al. 2000; Hunt 2005; Cregg 2007; Midgett and Madden 2007). Examples of crystal structures from proteins expressed in *P. pastoris* are a human monoamine oxidase B (PDB:3PO7) (Binda, Aldeco et al. 2010) and a protein involved in cell adhesion NCAM2 IG3-4 (PDB:2XY1)(Kulahin, Kristensen et al. 2011).

2.2.2 Insect cells

The expression system in insect cells is beside yeast a well-characterised alternative to express eukaryotic proteins (Midgett and Madden 2007). As insect cells are higher eukaryotic systems their posttranslational modification machinery can carry out more complex alterations than yeast strains. They also have a machinery for the folding of mammalian proteins. The most commonly used vector system for recombinant protein expression in insect cells is baculovirus, which can also be used for gene transfer and expression in mammalian cells (Smith, Summers et al. 1983; D., L.K. et al. 1992; Altmann, Staudacher et al. 1999). A few examples of proteins expressed in insect cells that resulted in

crystal structures are the transferase Ack1 (PDB:3EQP)(Kopecky, Hao et al. 2008), a human hydrolase (PDB: 2PMS)(Senkovich, Cook et al. 2007) and myosin VI (PDB:2BKI)(Menetrey, Bahloul et al. 2005).

2.2.3 Mammalian cell lines

The expression of proteins in mammalian cell lines is the most expensive and complex alternative. Especially for human membrane proteins this expression system has been proven to express the most active protein (Tate, Haase et al. 2003; Lundstrom 2006; Lundstrom, Wagner et al. 2006; Eifler, Duckely et al. 2007). The resulting protein amount, however, obtained from mammalian cell lines is mostly only sufficient for functional studies. Using mammalian cells lines is the most challenging variant of protein overexpression and therefore only choosen if any of the other expression system described failed. Some examples of protein structures expressed in mammalian cell lines are the hydrolase PCSK9 (PDB:2QTW)(Hampton, Knuth et al. 2007) and the acetylcholine receptor AChBP (PDB:2BYQ)(Hansen, Sulzenbacher et al. 2006).

Table 1 sums up advantages and disadvantages of the above mentioned overexpression systems used for protein crystallography.

Expression system	Pros	Cons
Prokaryotic		
Gram negative *E. coli*	- costs - simplicity (genetic/culture) - yield	- inclusion bodies - protein folding - posttranslational modifications - protein secretion
Gram positive *L. lactis*	- costs - protein secretion - one membrane	- posttranslational modifications
Eukaryotic		
Yeast	- high cell densities - costs - simple cultivation - posttranslational modifications	- hyper glycosylation, non-native - lipid composition is different to mammalian cells
Insect cells	- more native lipid environment - good track record of functional proteins	- costs - non-native lipid environment - glycosylation pattern different to mammalian cells - protein amount
Mammalian cells	- native conditions for human membrane proteins to investigate diseases - posttranslational modifications - lipid environment - good track record of functional protein	- costs - difficult to establish - protein amount

Table 1. Overview of expression systems. Summarized are the advantages and disadvantages.

3. Purification

After having expressed your protein of interest, the race for crystals is by no means finished. The next step on the long road to structure determination is to isolate the protein or - phrasing it differently - to remove all other proteins present in the cell (Figure 1 – (II) purification). An elegant method to do so is the genetic attachment of an affinity tag on either site of the protein or in some cases on both sides (Waugh 2005). This affinity tag has the possibility to bind high affine to a immobilized ligand on a matrix, while all other proteins have a much more reduced binding affinity and therefore flow through the matrix (Figure 1 – (II) purification 1st step). This allows a one-step purification, which in almost all cases is relatively harmless for the protein and likely does not interfere with folding and/or overall structure of the protein. There are a lot of affinity tags available as well as matrix materials (Terpe 2003). The well known and most often used affinity tag is the poly-histidine tag (Porath, Carlsson et al. 1975; Gaberc-Porekar and Menart 2001), which can vary in length as well as in position but the overall purification strategy is the same. From all the structures solved nowadays, almost 60 % of the proteins are purified via a histidine tag; mainly due to the great purification efficiency, which can be as large as 90% after a single purification step (Gaberc-Porekar and Menart 2001; Arnau, Lauritzen et al. 2006). Therefore, most commercially available expression systems and methods contain a his-tag encoded on the plasmid. Besides the his-tag, there are other tags avaible and used for protein purification, of which the Strep-, CBP-, GST-, MBP-tag are described below.

3.1 Choice of the right tag
3.1.1 Polyhistidine-tag (his-tag)
As mentioned above the polyhistidine-tag is the most common affinity tag and the required affinity resins and chemicals are relatively inexpensive. The purification step is a so-called immobilized metal ion affinity chromatography (IMAC) (Porath, Carlsson et al. 1975). Here, a matrix is able to bind bivalent metal ions. For example nitrilotriacetic acid (NTA), which is a chelator and binds metal ions like Ni^{2+}, Zn^{2+}, Co^{2+} or Cu^{2+} (Hochuli, Dobeli et al. 1987). These metal ions have a high affinity to the imidazole group of the amino acid histidine. A stretch of histindines in a row with for example an *E. coli* protein is very unusual. Thus, the genetical introduction of several, in most cases 6- 10 histidines in a row selects for specific binding of this protein. As eluant very elegantly imidazole can be used, which competes with the histidine tag and elutes the protein of interest. When used in low concentrations, imidazole can also be used to remove unspecifically bound proteins, which bind with low affinity to the matrix (Hefti, Van Vugt-Van der Toorn et al. 2001). Normally, a protein with a 6-10 histidine tag should be bound to the matrix relative strongly and 100-250 mM imidazole in the buffer is required to elute the protein from the resin. In contrast, proteins with a low affinity to the matrix can already be eluted with 10-50 mM imidazole (the "impurities" of *E. coli*). Therefore, a linear imidazol gradient, for example, separates the protein of interest and impurities (Hochuli, Dobeli et al. 1987; Gaberc-Porekar and Menart 2001). Although the polyhistidine-tag is the most common and mostly an efficient variant, there are a few applications where the his-tag can cause problems. Metalloproteins can interact either directly with the his-tag or with the ions immobilized on the matrix. In comparison to some other affinity-tags, the specificity of the his-tag is not that high and in some cases this results in the co-purification of other proteins (Waugh 2005).

3.1.2 Strep-tag

In comparison to the his-tag, which binds to immobilized metal ions, the strep-tag II constists of a small octapeptide (WSHPQFEK), which binds to the protein streptavidin (Schmidt, Koepke et al. 1996). The commercial available matrix is a streptavidin variant and is called Strep-Tactin. This variant is able to bind the Strep-tag II octapeptide under mild buffer conditions and can be gently eluted with biotin derivates such as desthiobiotin (Schmidt, Koepke et al. 1996; Voss and Skerra 1997). Especially for metal-ion containing enzymes it is a promising alternative to the his-tag (Groß, Pisa et al. 2002). However, as chemicals are more expensive and the matrix has a lower binding capacity, compared to NTA resins, it is often not the first option choosen. Moreover, it cannot be used under denaturating conditions since Strep-Tactin denatures and will not bind the tag anymore (Terpe 2003; Waugh 2005). Examples of proteins crystallized after a Strep-tag purification are OpuBC (PDB:3R6U)(Pittelkow, Tschapek et al. 2011) and *Af*ProX (PDB:3MAM)(Tschapek, Pittelkow et al. 2011) as well as the sodium dependent glycine betain transporter BetP from *Corynebacterium glutamicum* (PDB:2WIT)(Ressl, Terwisscha van Scheltinga et al. 2009).

3.1.3 CBP-tag

Another peptide tag, is the calmodulin binding peptide, first described in 1992 (Stofko-Hahn, Carr et al. 1992). This peptide is prolonged compared to the Strep-tag II, consisting of 26 amino acids and binds with nanomolar affinity to calmodulin in the presence of Ca^{2+} (Blumenthal, Takio et al. 1985). It is derived from the C-terminus of the skeletal-muscle myosin light-chain kinase, which makes the system an excellent choice for proteins expressed using a prokaryotic expression system, since in prokaryotic systems nearly no protein interacts with calmodulin. This allows extensive washing to remove impurities and elution with EGTA, which complexes specifically Ca^{2+}, and a protein recovery around 90 % can be achieved (Terpe 2003). A drawback of this tag however is that the CBP tag can only be fused to the C-terminus of the protein since it has been shown that CBP on the N-terminus negatively influences the translation and thereby the expression rate (Zheng, Simcox et al. 1997).

3.1.4 GST-tag

With respect to the length of the tags, the his-tag contains only a few amino acids, the Strep-tag II and the CBP-tag already contain 8 – 26 amino acids, but it is possible to fuse whole proteins with 26 – 40 kDa to a recombinant protein. Here, the high affinity binding of the protein to their substrate is used to purify the protein of interest (Smith and Johnson 1988). In the case of the glutathione S-transferase (GST, 26 kDa) the protein specifically binds to immobilized glutathione. To elute the fusion protein from the resin, non-denaturating buffer conditions employing reduced glutathione are used (Terpe 2003). The tag can help to protect the recombinant protein from degradation by cellular proteases. It is recommended to cleave off the GST-tag after purification with a specific protease like thrombin or TEV (Tobacco Etch Virus) protease (Terpe 2003).

3.1.5 MBP-tag

Another affnitiy tag, which can be fused to the protein of interest, is the maltose binding protein (MBP) from *E. coli*. This protein has a molecular weight of 40 kDa and has the ability to bind to a cross-linked amylose matrix. The binding affinity is in the micro molar range

and the tag can be used in a pH range from 7.0 – 8.0, however, denaturating buffer conditions are not possible (di Guan, Li et al. 1988). The elution of the recombinant protein is recommended with 10 mM maltose. A great opportunity of the MBP-tag is the increasing solubility effect of the recombinant protein in prokaryotic expression systems and even more pronounced in eukaryotic systems (Sachdev and Chirgwin 1999). Like the CBP-tag, a fusion at the N-terminal side might influence translation and expression rates (Sachdev and Chirgwin 1999).

3.1.6 Tag position and double tags

As described above, the position of the tag either at the N- or C-terminus has a considerable influence on translation and expression rate as well as on the biological function (Arnau, Lauritzen et al. 2006). If information regarding activity of the protein is already available especially about the location of interaction sites, this should be included in the protein design, meaning tag position etc. In general, the tag should be placed at the position of the protein, which is less important for interactions and/or expression. To minimize the influence of the tag on folding and/or activity in some cases it helps to create a linker region of a few amino acids between the tag and the protein (Gingras, Aebersold et al. 2005). A very efficient and sophisticated solution is, the addition of amino acids between tag and protein of interest, which functions not only as an accessibility increasing factor, but, also encodes for a recognition site for proteases like thrombin or TEV. Due to this arrangement the tag – protein interaction is minimized and the tag can be cleaved off if necessary (Arnau, Lauritzen et al. 2006). In some special cases a combination of two affinity tags results in enhanced solubility and more efficient purification. To enhance the purity of a protein, often a construct of two different short affinity tags like his-tag and Strep-tag or CBP-tag can be engineered (Rubio, Shen et al. 2005). Also a combination of two his-tag or two strep-tag kept apart by a linker region enhances the binding affinity extremely. This allows more stringent washing steps prior to elution of the protein (Fischer, Leech et al. 2011).

3.2 Size exclusion chromatography and ion exchange chromatography

Despite the usage of affinity tags a second purification step is sometimes required (Figure 1 – (II) purification). Which kind of purification procedure is required depends on the nature of impurities. If these impurities differ in molar mass compared to the protein of interest, a method based on size separation can be applied. Size exclusion chromatography (SEC) also separates different oligomeric species of the protein from each other, which otherwise would strongly inhibit crystallization and also allows analysis of stability and monodispersity of the protein (Regnier 1983a; Regnier 1983b).

However, in many cases, SEC is not sufficient to remove all impurities. Then separation by overall charge of the protein might be an option. Depending on the isoelectric point of the protein either anion or cation exchange chromatography can be performed. The protein binds to a matrix under very low ionic strength and is eluted afterwards either by increasing the ionic strength or by pH variation. Similar results can be achieved by hydrophobic interaction chromatography. Here, proteins with different surface properties show differences in their binding strength and binding of the protein is done inversely as during ion exchange chromatography. High ionic strength favors protein binding to a hydrophobic matrix and elution takes place when reducing the ionic strength. Although there are many other possibilities to increase the purity of a protein, the above mentioned techniques are without any doubt the most widely used and general applicable methods.

3.3 How to get a homogenous protein solution

In some cases isolated proteins are stable and homogenous at high concentrations after the purification and can be directly used for crystallization experiments. Often, however, the protein does not behave ideal and precipitates at high concentrations or forms aggregates or inhomogenous, oligomeric species; all of them prohibit crystal growth. SEC is a very elegant method to visualize the stability and oligomeric state of a protein. If the stability or the homogeneity of a protein sample is critical, you need to adapt your purification protocol and search for an optimized procedure. Different approaches are summarized below, for example a buffer screen to enhance protein solubility, multi-angle light scattering experiments to determine the absolute mass and the oligomeric state of the protein sample or fluorescence-based experiments to investigate the stability of the protein of interest.

3.3.1 Purified proteins – An *in vitro* system

After a protein is expressed in a soluble form, the subsequent purification procedure changes the environment of the protein dramatically. The cytoplasm of the cells, where the overexpression takes place, is packed with macromolecules. In *E. coli*, for example, the concentrations of proteins, RNAs and DNAs are about 320 mg/mL, 120 mg/mL, and 18 mg/mL, respectively (Cayley, Lewis et al. 1991; Zimmerman and Trach 1991; Elowitz, Surette et al. 1999) resulting in an overall concentration of macromolecules of above 450 mg/mL. During cell lysis and the first purification step, likely an IMAC (see above), the protein is separated from almost all other cell components. This rigorous procedure is accompanied with a severe change of the environment into an *in vitro* system. As a result proteins often tend to aggregate, precipitate or form inhomogeneous oligomeric states that prevent the formation of crystals in further experiments. Therefore one of the biggest challenges in structural studies is the preparation of protein solutions with high concentrations (as a rule of thumb 10-20 mg/mL) in a homogenous state. To fulfill these requirements, the *in vitro* system needs to be optimized with respect to different parameters as highlighted in *Figure 1 – (II) purification*. If a sufficient protein sample cannot be obtained, different strategies are available to increase the important characteristics of the protein: purity and homogeneity. As mentioned above, the usage of different metal ions during IMAC, ion exchange, a second affinity chromatography etc. can be sufficient to enhance purity. This might also lead to an increased stability. However, if the stability and/or homogeneity of a protein is still a problem, screening for a new buffer composition is essential to succeed during crystallization trials.

3.3.2 Buffer composition

Many examples illustrate the importance of an adequate buffer composition for protein stability, homogeneity, conformation, and activity (Urh, York et al. 1995; Holm and Hansen 2001; Jancarik, Pufan et al. 2004; Collins, Stevens et al. 2005). Some buffers are very frequently used and recommended by manufactures (see for example Qiagen, Roche, New England BioLabs, Fermentas, etc.). All of them contain a buffer reagent that keeps the pH constant in a well-defined range. Well-known examples are phosphate, tris (hydroxymethyl) aminomethane (Tris), or HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) that buffer at the physiological relevant pH range of 6- 9 (Durst and Staples 1972; Chagnon and Corbeil 1973; Tornquist, Paallysaho et al. 1995). In recent years, the development of other buffer systems has been quite successful (Taha 2005) (for a list of buffers and corresponding

pH ranges, see for example: http://delloyd.50megs.com/moreinfo/buffers2.html). Next to the well-defined pH, the stability and homogeneity of proteins depend on many other parameters, for example ionic strength, the presence of ligands and/or co-factors, divalent ions, glycerol, etc. The appropriate buffer composition cannot be predicted so far and needs to be identified by trial-and-error approaches.

3.4 Protein purification – How to overcome problems

In this part we would like to present some pitfalls that might occur during protein purification and provide some ‚rationales’ to overcome these problems. As usual, the crucial step of solving a problem is its identification. Here, we are trying to sensitize the reader to indications, which might point towards problems related to instability and/or inhomogeneity of the protein sample. Moreover, such problems cannot always be recognized without the adequate technique(s). Therefore, we are introducing techniques that are capable to visualize the state of proteins.

3.4.1 Visible protein precipitations during IMAC

A very obvious stability problem is the formation of precipitations in the elution fractions of a chromatography step (see Figure 2). In this example, the his-tagged protein was eluted with a linear imidazole gradient from 10 to 500 mM imidazole and eluted at about 250 mM imidazole. Protein precipitation occured immediately after elution (Figure 2A and B) and continued (Figure 2C) resulting in a low amount of soluble protein. This aggregation can be reduced by dilution with a IMAC buffer (typically lacking imidazole) immediatly after the elution. Thereby, dilution hinders the concentration-dependent aggregation. In many cases, this rational is not sufficient to prevent precipitation. After applying, for example, a *buffer screen* (see Figure 1 – (II) purification) the new defined buffer is used for the chromatography or the eluting protein is diluted into the new buffer (see Figure 2D).

Other elution strategies of his-tagged proteins from an IMAC column are available. As described before, competing the poly-histidine from the IMAC column by imidazole is the most common elution strategy, however, for some proteins other strategies are superior, for example, replacing imidazole by histidine. Imidazole is only a mimic for histidine. If one uses histidine instead of imidazole aggregation can be avoided as concentration of the eluent can be reduced by a factor of ten. An example for a protein sensitive to imidazole concentration is shown in Figure 3B. Here a comparative SEC chromatogram is shown. After elution from the IMAC column with imidazole only a very small amount of the protein elutes at the volume corresponding to the size of a monomer or the dimer, respectively (Figure 3B, continuous line). Most of the protein passes the column very fast and elutes at the void volume indicating large radii meaning aggregated protein. Yields of dimeric (at about 150 mL) and monomeric (at about 180 mL) proteins are strongly increased after an elution with histidine (dashed line) compared to an elution with imidazole (continous line) and only the monomeric species could be crystallized (data not shown). The choice of the eluent in IMAC might therefore be an important step in a purification protocol. Another elution strategy of his-tagged proteins is a pH change from 8 to 4. In an acidic environment, histidines become positively charged and are therefore released from the column matrix. This strategy results in a sharp elution from the matrix and the protein is eluted highly concentrated. Although this strategy is recommended by the manufacturers (see GE Healthcare, Qiagen, etc.) the desired protein needs to retain activity at acidic pHs. The

bivalent metal ions (Ni^{2+}, Co^{2+}, Zn^{2+},…, see above) that complex the his-tag can be removed from the matrix by chelating reagents as ethylenediaminetetraacetic acid (EDTA) as another elution strategy (Muller, Arndt et al. 1998)
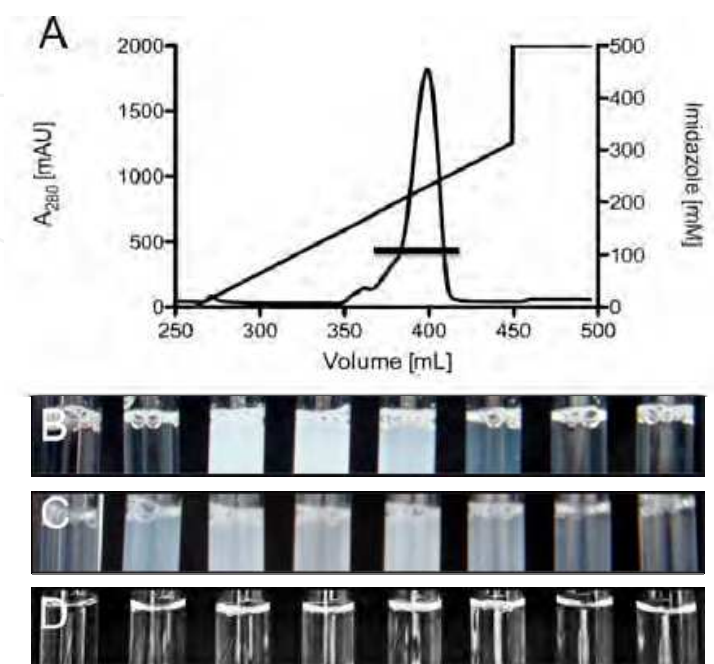

Fig. 2. Elution fractions of an IMAC. The protein was eluted via a linear imidazole gradient from 10 to 500 mM and the absorption at 280 nm was recorded. The elution fractions were collected and photographed. A: IMAC chromatogram of the his-tagged protein. Elution fractions containing the desired protein (indicated by a bar) are collected and shown in B – D. B and C: Elution fractions of the protein in 50 mM Tris-HCl, 150 mM NaCl, pH 8.0 immediately and 10 min after the elution, respectively. D: The elution fractions were immediately mixed in a 1 to 1 ratio with a buffer that enhances the protein stability (50 mM citrate, 50 mM LiCl, pH 6.00) evaluated during a solubility screen.

3.4.2 Invisible aggregations
Sometimes aggregation of proteins in solution can not be detected directly by eye. This inhomogeneity of protein samples can be visualized SEC, a method that separates proteins by their hydrodynamical radius (see above). Protein aggregates are eluting at the void volume, since they are clumbed together resulting in a big hydrodynamical radius (see Figure 3A and B). If invisible aggregation is detected the buffer composition needs to be adjusted. In one case we applied this technique to visualize the state of a protein after an IMAC, and the resulting elution profile is shown in Figure 3A (continous line). Comparable to the imidazole-induced precipitation described above, the protein aggregated and elutes within the void volume of the column (about 40 mL). Moreover, several other protein species elute from 55 to 80 mL indicating a highly inhomogeneous protein sample. The running buffer of the SEC was 50 mM Tris-HCl, pH 8.0 and 150 mM NaCl. Remarkably, a simple change to a new buffer (20 mM HEPES, 150 mM NaCl, pH 7.0) Resulted in a stable and homogenous protein sample (Figure 3A, dotted line), which was suitable for

crystallization trials. Next to the rigorous change in the homogeneity of the protein, the biological activity of the protein could only be determined in the new buffer system. The influence of the buffer composition for the protein activity is a well-known phenomenon (Urh, York et al. 1995; Holm and Hansen 2001; Zaitseva, Jenewein et al. 2005) and in many cases the activity goes hand in hand with an optimal buffer for the purification. Mentionable, the new buffer was not found by trial-and-error approaches. We searched for literature dealing with homologous proteins, especially for established purification protocols. This literature search revealed the new buffer, illustrating that not every step towards a protein structure determination must be a trial-and-error process.
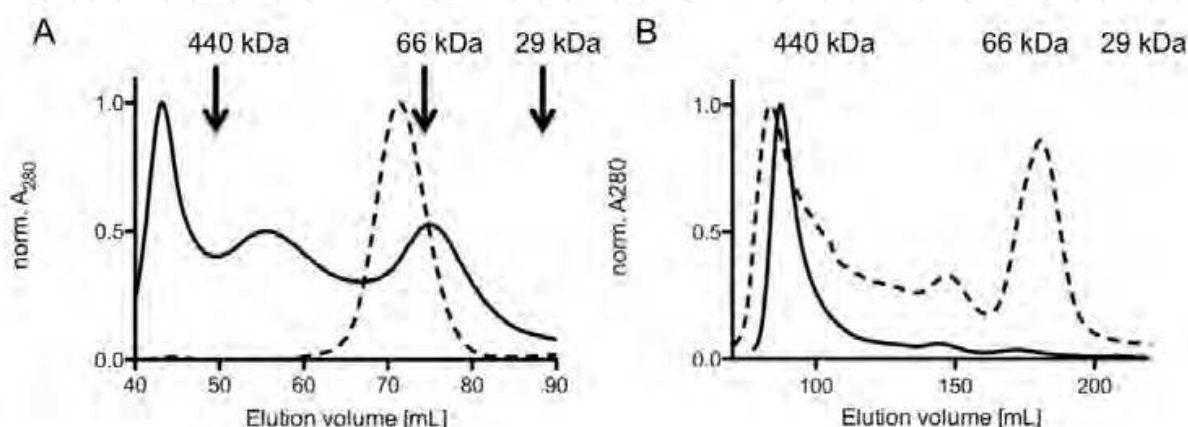
Fig. 3. Size exclusion chromatograms (UV 280nm) of proteins in different buffers. A: The homogeneity of a protein was analyzed in two different buffers; continuous line: 50 mM Tris-HCl, 150 mM NaCl, pH 8.00; dotted line: 20 mM Hepes, 150 mM NaCl, pH 7.00. B: The protein was eluted of the IMAC column either with imidazole (continous line) or with histidine (dotted line), concentrated and applied to the SEC.

Another example for the influence of the buffer composition was published bei Mavaro et al. (Mavaro, Abts et al. 2011). Instead of the buffer agent, the ionic strength of the buffer was the crucial determinant. Purification of the protein in low-salt buffer resulted in an inhomogenous protein sample containing a mixture of aggregates, dimers and monomers without biological activity. However, a simple change to high-salt buffer allowed the purification of homogenous dimeric protein, that was able to bind its substrate.

3.4.3 Overcoming protein instability

In the previous sections different strategies were mentioned to enhance the stability and the homogeneity of purified proteins and in all cases the buffer composition was the solution. Still, the essential question how to determine the optimal buffer to make a protein feel happy *in solution* is not answered? Some rationales and experiences are listed above: different elution strategies for IMAC purifications, the usage of frequently used buffer agents and a literature research for established purification protocols of related proteins. However, in many cases these approaches do not solve the problems occuring during the purification. But, is there a general methodology to overcome the problems? Unfortunately, the answer is as frustrating as challenging - there is not a general panacea around for the right buffer composition of a protein. If a new buffer needs to be found, trial-and-error

approaches have to be applied. A lot of different parameters are influencing the state of a protein, i. e. the buffer agent, the salt concentration, presence of metal ions with different valences, the hydrophobicity, and even the temperature of the buffer. The analysis of the protein in different buffers can be done by SEC and/or light scattering experiments. However, screening of all the different variables is very labor- and cost-intensive, and time-consuming, moreover only combinations of two or more additives might be sufficient to enhance the solubility and homogeneity of the protein. Therefore high-throughput methods are needed that handle a lot of different conditions simultaneously using as few as possible protein sample.

3.4.4 Buffer screen – Enhancing the solubility

Many publications are available suggesting methods for a solubility screening to allow the crystallization of initially inhomogeneous, aggregating protein samples (Jancarik, Pufan et al. 2004; Zaitseva, Holland et al. 2004; Collins, Stevens et al. 2005; Sala and de Marco 2010; Schwarz, Tschapek et al. 2011). In all of these methods aggregating protein samples are mixed with commercially available crystallization screens incubated for a period of time, and analyzed for precipitation visually using a light microscope. Screening conditions resulting in no precipitations are analyzed upon their composition, and protein samples are further examined with respect to their solubility and homogeneity under these conditions by SEC or light scattering experiments. This technique allows high-throughput screening in a 96-well format, where an automated pipetting system mixes only 50-200 nL of protein solution with 50-200 nL of buffer solutions to minimize the needed protein sample and increase the screening efficiency. Several buffer screens are commercially available that cover many different buffer agents, salt concentrations and other buffer parameters (i. e. from Hampton Research, Molecular Dimensions, Sigma, Jena Bioscience, Qiagen). After a solubility screening was applied, we were able to stabilize a previously unstable protein at concentrations above 3 mg/mL (see above "Protein precipitations during IMAC" and Figure 2D) at concentrations of up to 100 mg/mL for weeks (Schwarz, Tschapek et al. 2011). Typically, the new buffer (50 mM citrate, 50 mM $LiCl_2$, pH 6,00) should be used during the entire purification procedure starting with cell lysis. In the described case, the new buffer contains citrate, which is incompatible with an IMAC purification. Therefore the protein was immediately mixed with the new buffer after the elution of the IMAC column.

3.4.5 Size-exclusion chromatography versus light scattering experiments

Size-exclusion chromatography (SEC) and light scattering experiments (LS) are very helpful tools to analyze the homogeneity (Collins, Stevens et al. 2005) and the molecular mass of proteins; however both of them have advantages and disadvantages compared to each other. In SEC experiments proteins are separated based on their hydrodynamic radius by partitioning between a mobile phase and a stationary liquid within the pores of a matrix. All SEC columns are characterized by the volumes V_0, the liquid volume in the interstitial space between particles, V_i, the volume contained in the matrix pores and V_T, the total diffusion volume ($V_0 + V_i$) (Regnier 1983a; Regnier 1983b). In dependency of the hydrodynamic radius molecules are eluting at specific retention volumes in between V_0 and V_T with big molecules eluting first. After a calibration of a SEC column with proteins of known

molecular weight (i. e. Sigma-Aldrich, "Kit for Molecular Weights") the molecular mass of the protein of interest can be roughly estimated; the elution volume is correlated to the $\log_{10}$ of the molecular weight (therefore, the hydrodynamic radius is considered to be proportional to the molecular weight). However, many extraneous mechanisms such as adsorptive, hydrophobic and ionic effects are further limiting the correlation between the retention volume and the molecular mass giving sometimes rise to wrong estimations.

Light scattering (LS) experiments can be applied to overcome these disadvantages and investigate the exact molecular weight of the protein sample. The rayleigh scattering of particles of monochromatic light depends directly on the molar mass of the particle. If you know the exact number of particles you can calculate the average molar mass of these particles. This technique is very powerful when used online after separation of the protein depending on their hydrodynamic radius, meaning SEC. This technique is always superior to normal SEC but requires special equipment and especially more time. However, if the protein fold is not really globular or other effects occur (see above: ionic, hydrophobic, etc.) assumption on size and oligomeric state based on SEC is not possible at all. For protein crystallization information about monodispersity, which can be provided by such an experiment, is an additional benefit.

3.4.6 Analysis of the homogeneity – High-throughput methods

Despite the development of various sophisticated methods, a bottleneck of homogeneity screening is high-throughput analysis. As mentioned above, proteins need to be analyzed by SEC and/or LS experiments after visual read-out of the protein-buffer droplets. Therefore, fluorescence-based solubility screens were developed that allow the high-throughput analyzes of many samples in a 96-well format (Ericsson, Hallberg et al. 2006; Alexandrov, Mileni et al. 2008; Kean, Cleverley et al. 2008). All these assays use fluorophores as reporters of the protein state. A suitable fluorophore is, for example, Sypro Orange, which exhibits different fluorescence properties as a function of its environment. This dye is almost dark in hydrophilic environment, however, after binding to hydrophobic molecules, it emits light at 570 nm. In inhomogenous and unfolded protein samples hydrophobic amino acids are exposed to the surface of proteins (Murphy, Privalov et al. 1990). An increase in the fluorescence signal of Sypro Orange correlates therefore with unfolding events of proteins. The homogeneity screening can be performed in basically two ways: temperature- or time-dependent. For the first setup the protein sample is heated gradually in distinct steps (i. e. 1 °C) and the emission is monitored at 570 nm. Hereby, a "melting" temperature is determined, which is characterized by 50% fluorescence of the maximal fluorescence at the highest temperature; the higher the melting temperature, the higher the stability of the protein (Ericsson, Hallberg et al. 2006). Secondly, the protein sample is incubated at a specific temperature (i. e. 40°C) and the fluorescence is measured for a period of time. The "half-life" time, at which 50 % fluorescence of the maximum fluorescence in one sample is detected, can be compared to all buffer conditions. In Figure 4 an example of the time-dependent approach is shown. Here, the protein is incubated in different buffers with various salt concentrations. The emission of Sypro Orange is recorded each minute at 570 nm. An analysis of all time-dependent fluorescence plots indicates that the protein is most stable in buffers containing 125 mM NaCl but unfolds fast in 1 M ammonium sulfate. These assays result in qualitative indications about a favourable environment of proteins that enhance the stability. Ericsson et al. proved the concept of this method by applying it to

different proteins (Ericsson, Hallberg et al. 2006). The stability optimization yielded a twofold increase in initial crystallization leads. Moreover these assays enable the search for putative ligands of the protein. Upon binding of a substrate in the binding pocket or an inhibitor, the stability of the protein increases, which can be detected experimentally.
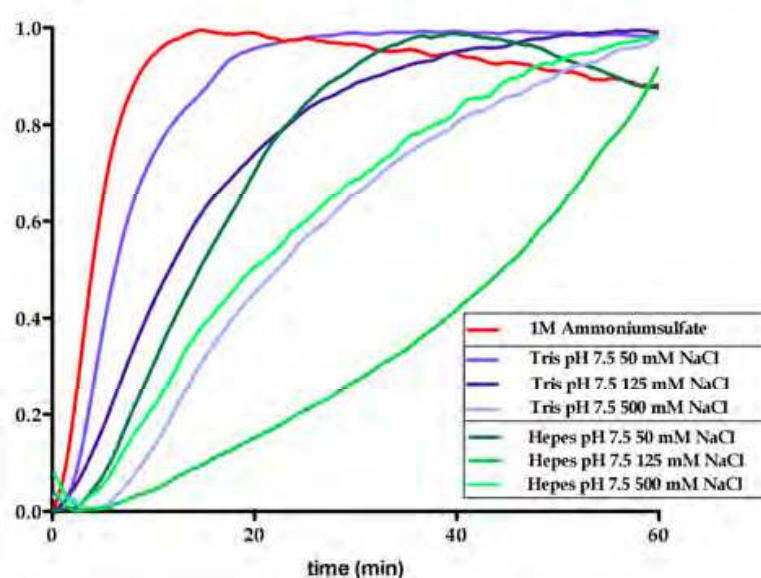

Fig. 4. Time-dependent stability optimization screen using Sypro Orange as reporter. The protein is diluted 1:50 into each test buffer containing Sypro Orange, excited with 490 nm and the fluorescence at 570 nm is measured for 60 minutes automatically with a PLATE READER (Fluorostar, BMG Labtech). Normalized fluorescence is plotted against the time.

4. Protein crystallization: Introduction

Protein crystals suitable for X-ray diffraction experiments and usable for subsequent structure determination are normally relatively large with a size of at least 10 to 100 μm. In contrast to crystals of mineral compounds, protein crystals are rather soft and sensitive to mechanical stress and temperature fluctuations. These properties are due to weak interactions between single proteins within the crystal, their high flexibility as well as the size of the macromolecules. The periodic network of building blocks is held together by dipole-dipole interactions, hydrogen bonds, salt bridges, van der Waals contacts or hydrophobic interactions. All of them have binding energies in the low kcal/mol range.

Especially the limited number of crystal contacts and their directionality are the largest difference to the high interactions generally observed in salt crystals. An example of the interactions within a protein crystal is shown in Figure 5. This picture highlights the main pitfalls in protein crystallization. A protein is a highly irregular shaped and flexible macromolecule which allows weak and stinted interactions at very specific locations of its surface. All vacuity is filled with buffer, in general not contributing to any kind of interactions between the protein molecules. Figure 5A shows a protein of around 30 kDa, which crystallizes in a rather small unit cell (shown in black). Only one protein monomer is located in the asymmetric unit of the unit cell, the other shown monomers represent symmetry related proteins. Figure 5B highlights the three-dimensional packing of protein molecules within a crystal.
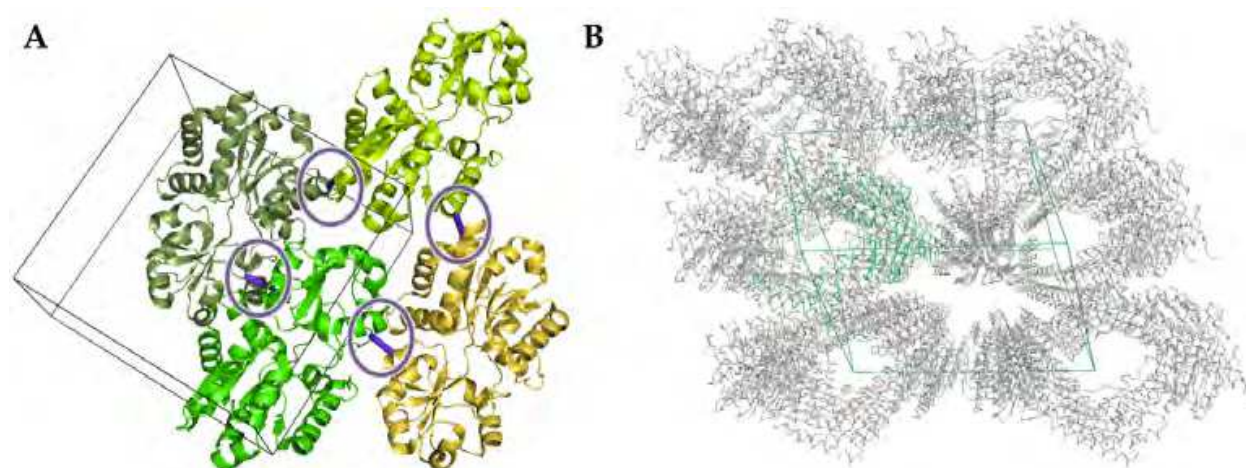
Fig. 5. Example of the packing within a crystal. A: The unit cell is shown in black, crystal contacts are highlighted with purple circles and lines. B: Three-dimensional crystal packing of a different protein. The unit cell as well as one protein monomer are depicted in green.

The flexibility as well as the other mentioned characteristics of proteins are responsible for the problems occuring during crystallization trials and despite extensive efforts not every protein is suitable for crystallization. If one cannot generate crystals one has to move back several steps and change the properties of the protein, e. g. surface properties by mutation of single amino acids, truncation of the protein or sometimes only changing buffer compositions that result in a more suitable protein for crystallization (see Figure 1 and also below). There are several prediction servers available that help choosing the 'right' protein and modification (Linding, Jensen et al. 2003; Goldschmidt, Cooper et al. 2007). However, protein crystallization still remains an empirical approach, sometimes called voodoo, while crystallography is science.

4.1 Phase diagram

The conditions or protocols for obtaining good crystals are still poorly understood and despite all progress and efforts protein crystallization is a trail-and-error approach. However, a step towards a better understanding of crystal growth can be achieved by analyzing the phase diagram of a protein-water mixture. The phase diagram is a simple illustration to help understanding how protein crystals are formed. Mostly, it is shown as a function of two ambient conditions that can be manipulated, i. e. the temperature and the concentration. Three-dimensional diagrams (two dependent parameters) have also been reported (Sauter, Lorber et al. 1999) and even a few more complex ones have been determined as well (Ewing, Forsythe et al. 1994). Figure 6 shows a schematic phase diagram for a protein solution as a function of protein concentration and precipitant concentration. The phase diagram is broken down into four distinct zones (Rosenbaum and Zukoski 1996; Haas and Drenth 1999; Asherie 2004):

1. Undersaturated zone: Under these condition the protein will stay in solution as neither the concentration of the protein nor of the precipitant is high enough to reach supersaturation.
2. Precipitation zone: Is the protein concentration or the precipitant concentration too high, the protein precipitates out of solution; this kind of solid material is not useful for crystallographic studies.

3. Labil zone: This is the most important configuration of the two parameters, as nucleation and initial crystal growth take place under these conditions.
4. Metastable zone: After initial crystals are formed and start growing in the labil zone, protein concentration decreases in the drop and the metastable zone will be reached. Here the crystal can grow further to its final maximum size.
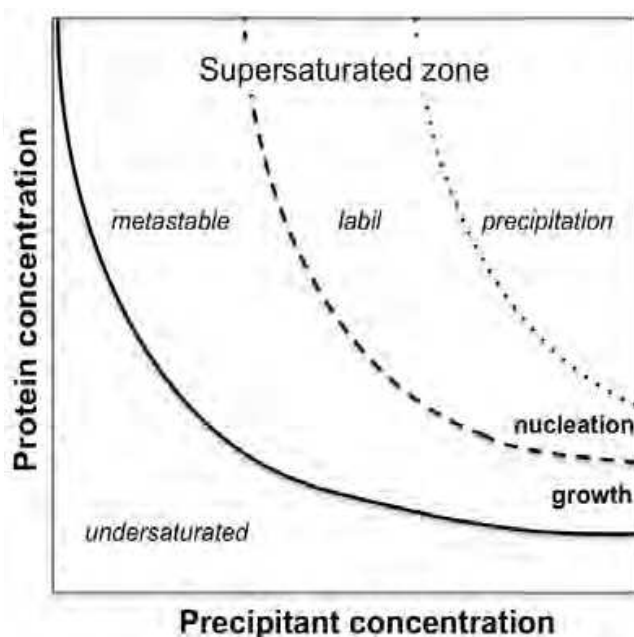

Fig. 6. A basic solubility phase diagram for a given temperature (adapted from (Rupp 2007).

The curve separating the undersaturated zone from the supersaturated one is called *solubility curve*. If conditions are chosen below the solubility curve, the protein will stay in solution and never crystallize. This means when a protein crystal is placed in a solvent, which is free of protein, it will start to dissolve. If the volume of the droplet is small enough it will not dissolve completely: it will stop dissolving when the concentration of the protein in the droplet reaches a certain level. At this concentration the crystal loses protein molecules at the same rate at which protein associate to the crystal – the system is at equilibrium. Determination of the solubility of the protein of interest might be a helpful information at the beginning of crystallization experiments. This can be done in a two-dimensional screen varying for example ammonium sulfate concentrations as well as the protein concentration.

4.2 Crystallization techniques

Crystallization is a phase transition phenomenon. Protein crystals grow from a supersaturated aqueous protein solution. Varying the concentration of precipitant, protein and additives, pH, temperature and other parameters induce the supersaturation. However, as mentioned before, prediction of this kind of phase diagrams is *a priori* impossible.

Protein crystallization can be divided into two main steps:
1. Generating initial crystals: 'Searching the needle in a haystack'
2. Empirical optimization of these crystallization condition

The first step is mostly based on experiences from other crystallization trials with different proteins. Nowadays several supplier offer crystallization screens that contain solutions for

initial experiments that were used successfully in the past for crystallization trials (Jancarik and Kim 1991), so-called "sparse matrix screens". There are also some trials around to use more systematic approaches (Brzozowski and Walton 2000) to get more information about solubility prior and simultaneous to crystallization (incomplete factorials, solubility assays). Both kinds of screens can be applied to different crystallization techniques.
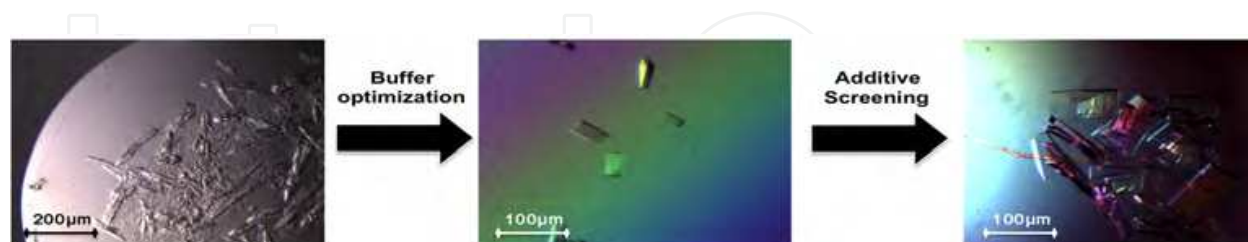

Fig. 7. Crystal optimization. First steps in crystal optimization are shown. Initial protein crystals look weak and fragile, after screening around this initial buffer composition crystal evaluation by eye results in less fragile, homogeneous looking crystals. However, diffraction quality was poor. Therefore an additive screening was performed that resulted in a different crystal form. These crystals finally were able to diffract X-rays to a reasonable resolution.

A lead/hit in that initial step might not be a 'real' crystal rather than a crystalline precipitate or just phase separation. In the next step, fine-tuning the buffer composition further optimizes this hit. Varying pH, salt concentration, type and concentration of precipitant and protein concentration are expected to yield larger and hopefully also better-diffracting crystals. In this step, the chemicals used are much more defined and therefore it is a more systematic than empirical screening (see Figure 7).

4.2.1 Vapor diffusion
The most popular and simplest technique to obtain protein crystals is the vapor diffusion method either in the sitting or hanging drop variant (see Figure 8). For both a defined volume (mostly < 1µl) of protein solution is mixed with an equivalent volume of screening solution and then equilibrated against the original precipitant/screening concentration. During this equilibration, the vapor pressure of the solution rises as the protein crystallizes (protein in solution lowers water activity) while the water evaporates to maintain equilibrium, which causes the precipitant concentration to rise. Therefore, if the crystal growth is sensitive to the precipitant concentration, vapor diffusion can rapidly force the mixture to unstable conditions where growth and nucleation are too rapid. This is the main disadvantage of vapor diffusion: Growing large crystals might be problematic!

4.2.2 Micro batch method
In this set-up the protein solution is mixed with screening solution at concentrations required for supersaturation right at the beginning of the experiment. Typical drop sizes of micro batch experiments ranges from 1-2 µl. The drop is then covered with oil, which acts as an inert sealing to protect the drops during incubation from evaporation (see Figure 8).
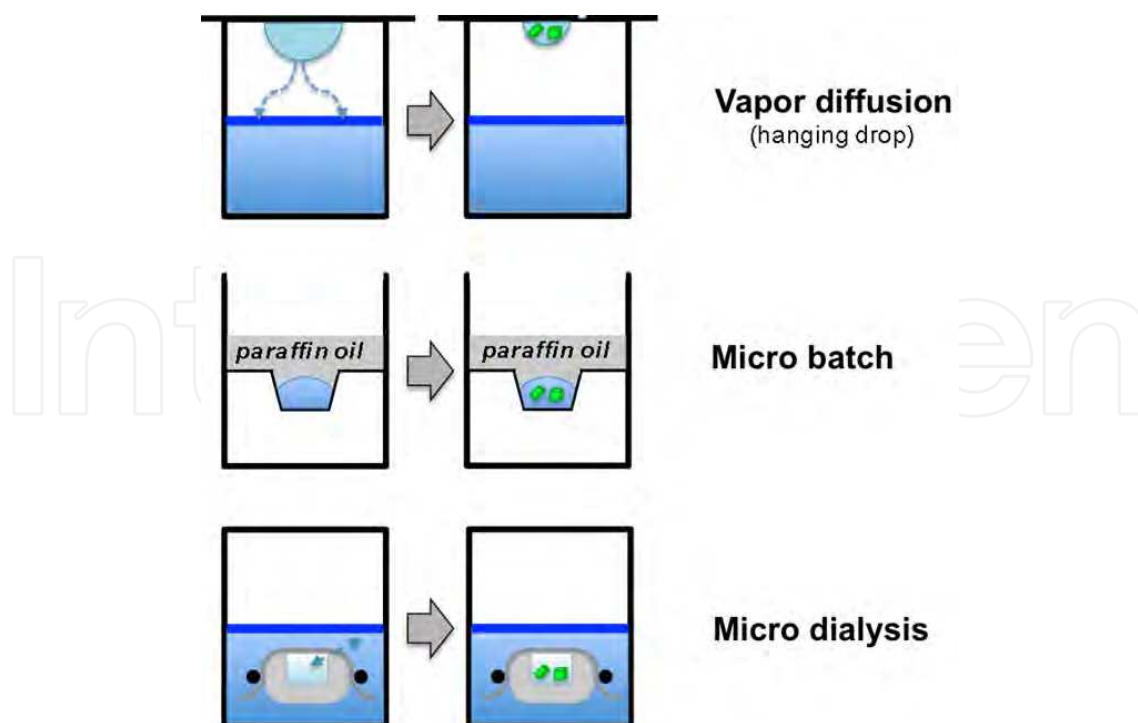
Fig. 8. Protein crystallization techniques. Schematic representation of a) vapor diffusion, b) micro batch and c) micro dialysis crystallization techniques widely used for crystal growth (adapted from (Drenth 2006)).

4.2.3 Micro dialysis

Dialysis is another way to change the buffer composition and increase its concentration in the crystallization experiment gradually (see Figure 8). Micro-dialysis buttons are exposed to different screening buffers. This method requires rather high amounts of protein but might yield large crystals.

After obtaining initial crystal hits in a commercial screen the tough part of crystal optimization starts. By varying pH, salt concentration, temperature, precipitant concentration or protein concentration these initial crystals should be reproduced and become larger, more regular shaped or are simply growing faster. A further improvement of crystal quality might be achieved by the addition of small amounts of so called 'additives'. At this point basically each chemical compound might be sufficient to improve the crystal quality. Luckily, there are some preferable working additives, which have been proven to produce better crystal in more than one case. Especially compounds that are known to reduce undirected interactions in proteins like organic solvents, i. e. DMSO or phenol, or detergents and reducing agents are very often used at this stage and helpful to force more homogeneous well diffracting crystals.

4.3 Crystal nucleation

There are two fundamental steps during protein crystallization: Nucleation and crystal growth. If one cannot obtain single crystals of adequate quality for analysis, this is generally a consequence of problems associated with the growth phase (see above). But failure to obtain any crystals at all or failure to obtain single, supportable nuclei reflects difficulties in the nucleation step. Therefore control of nucleation is a powerful tool to optimize protein crystals

or sometimes it is the only way to get crystals at all. Nucleation can take place either homogeneous meaning in the bulk of the solution, when the supersaturation is high enough for the free-energy barrier to nucleus formation to be overcome or heterogeneous mostly by solid material in the crystallization solution. This can also occur even when the supersaturation is not achieved. Therefore in order to control nucleation one has to work with highly clean solutions to avoid nucleation by the second mentioned possibility. The nucleation zone can be bypassed by insertion of crystals, crystal seeds or other nucleants to the protein/precipitant mixture. Addition of crystals or tiny fragments of crystals is called seeding. This method is then subdivided into macro- and micro-seeding dependent on the size of the nucleant added. In macro-seeding experiments one single, already well-formed but small crystal is placed into a new crystallization solution at lower saturation. Microseeding in contrast requires small fragments of a crystal or almost invisible microcrystalline precipitate. These 'seeds' are then transferred into a fresh crystallization solution either by a seeding wand which is dipped into the microseed mixture to pick up seeds and then touched across the surface of the new drop or by a animal whisker or hair that is stroked over the surface of the parent crystal to trap the nuclei and then is drawn through the new drop. As this method also enhances the speed of crystal growth it can be used with sensitive substrate that undergo decomposition over time. Oswald et al. proved this in 2008 by solving the structure of ChoX from *Sinorhizobium meliloti* in complex with a highly hydrolyzing substrate, acetylcholine (Oswald, Smits et al. 2008). In classical vapor diffusion experiments crystals appear after four weeks but data showed only little electron density in the ligand-binding site and turned out to result from a choline bound instead of acetylcholine. Hydroxylation was favored due to the relatively long time for crystal growth but also because of an acetic pH in the crystallization set-up. To circumvent these problems accelerated crystal growth was required. In this case micro-seeding results in crystals suitable for data collection in less than 24 hours.

Recent years more effort in nucleation control yielded in fancy materials that can be used as nuclei for crystals. These methods use the second way of nuclei formation, as a solid material is introduce into the crystallization solution as an 'universal' nucleant (Chayen, Saridakis et al. 2006). There have been several substances that have been tried more or less successful. Some have been useful for individual proteins, but mostly they were not applicable in general (McPherson and Shlichta 1988; Chayen, Radcliffe et al. 1993; Blow, Chayen et al. 1994). In 2001, Chayen et al. proposed the idea of using porous silicon whose pore size is comparable with the size of a protein molecule. In theory such pores may confine and concentrate the protein molecules at the surface of the silicone and thereby encourage them to form crystal nuclei (Chayen, Saridakis et al. 2001). These nucleants have made it to commercial availability (www.moleculardimensions.com) and have proven to be suitable for different kinds of proteins and even membrane proteins that have not been possible to crystallize before formed nice crystals in the presence of these nucleants.

4.4 Cryoprotection

Exposure of a protein crystal at room temperature results in dramatic radiation damage due to radicals formed by the ionizing X-ray photons. To reduce that harmful disintegration of the protein crystal the crystal is cooled to ≈100K with the help of liquid nitrogen (Low, Chen et al. 1966; Hope 1988; Rodgers 1994; Garman 1999). However, it is common for the cooling process to disrupt the crystal order and decrease diffraction quality. Thus, the crystal must be cooled fast so that the water in the solvent channels is in the vitreous rather than in the

crystalline state at the end of this procedure. As for pure water this cooling has to take place very quick (10^{-5}s, (Johari, Hallbrucker et al. 1987), some water molecules can be replaced by a cryoprotective solution prior to cooling (Juers and Matthews 2004). This exceeds the time window for freezing up to 1-2s (Garman and Owen 2006) however, finding a good 'cryoprotectant' for a special protein crystal again involves substantial screening. Once flash frozen in liquid nitrogen, the crystal must be kept below the glass transition temperature of the cryobuffer at or below 155K at all times (Weik, Kryger et al. 2001).

4.5 What can you do when all efforts did not succeed in crystals?
4.5.1 Buffer composition – Again!

The choice of the right buffer used for crystallization experiments is very crucial. As shown above, every protein needs its own buffer composition to feel kind of happy in this aqueous artificial environment. Especially as high protein concentrations (>10mg/ml) are required for crystallization, one might has to test several buffer compositions again (see also Figure 1). As a rule of thumb you should obtain around 50% of clear drops immediately after mixing protein and buffer solution. If you detect drastically more precipitation in your drops you should think first about lower protein concentration but of course secondly about changing your buffer system again.

4.5.2 How to obtain a rigid protein suitable for crystallization?

To overcome the problem of flexibility of some regions in the protein addition of ligands is often a very powerful tool to fix the protein in a single conformation that is more favorable for crystallization. A good example for this strategy is the crystallization of so-called substrate binding proteins (for a recent review see (Berntsson, Smits et al. 2010)). These proteins catch their substrate in the periplasm of bacteria or on the outer membrane of archaea and then deliver it to their cognate transport system located in the membrane. The mechanism of substrate binding is quite well understood. These binding proteins all consist of two domains, which rotate towards each other during the binding event. In solution without substrate they are quite flexible and NMR-studies proved a equilibrium between open and closed conformation (Tang, Schwieters et al. 2007). Analysis of all available structures for this class of proteins showed that more than 95% were crystallized with a ligand bound (Berntsson, Smits et al. 2010). Thus, a stabilization of the two domains seems to simplify crystal contact formation dramatically. Although people always want to obtain a functional conformation of the protein in their crystal structure, it is sometimes helpful to think about how to stop the protein from doing its job. A non-functional protein is in general less flexible and fixed in one conformation. One example for successful implementation of this strategy is the crystal structure of NhaA from *Escherichia coli* solved in 2005 (Hunte, Screpanti et al. 2005). Here, Hunte et al., downregulated the protein activity by working at an acidic pH of 4. Although the protein shows almost no activity at this pH the structure reveals the basis for mechanism of Na+/H+ exchange and also its regulation by pH could be understood.

4.5.3 Rational protein design for crystallization: Surface engineering

The first example of rational protein design that yielded a good diffracting protein crystal is given by Lawson et al. in 1991 (Lawson, Artymiuk et al. 1991). They compared amino acids

involved in crystal contact formation of the rat ferritin protein L. (which is highly homologous to human ferritin H, the target protein) with the amino acids present at that position in human ferritin H. A replacement of Lys86, found in the human sequence, with Glu, which occurs in rat, recreated a Ca^{2+} binding bridge that mediates crystal contacts in the rat ortholog. As this method was successful for several other proteins (McElroy, Sissom et al. 1992; Braig, Otwinowski et al. 1994; Horwich 2000), a general protocol was required. The concept Derewenda et al. proposed in 2004 is based on the general equation for the free energy that drives protein crystallization:

$$\Delta G = \Delta H - T(\Delta S_{protein} + \Delta S_{solvent})$$

As the enthalpy values of intermolecular interactions in a crystal lattice are rather small (see above), crystallization is very sensitive to entropy changes of both protein and solvent. The formation of ordered protein aggregates carries a negative entropy term. This can only be overcome by positive entropy from the release of water bound to the protein. However, large hydrophilic residues (e.g. lysines, arginies, glutamates, glutamines) exposed on the protein surface need to be ordered. Since they are rather flexible this can cause problems. This can be overcome by mutating large amino acids into smaller ones, for example alanines. Among these large amino acids lysines and glutamates play a particular role, as they are always (with only very few exeptions) located on the protein surface (Baud and Karlin 1999). Both lysines and glutamates are typically disfavored at interfaces in protein protein complexes (Lo Conte, Chothia et al. 1999), therefore it is rather straight forward to assume that lysine and glutamate to alanine mutants are good targets for protein crystallization if wildtype protein hardly forms crystals. However this also means that you have to go several steps backwards on road to a protein structure determination (see Figure 1).

4.5.4 Affinity tag removal: Philosophic question???
Another variant in protein crystallization nowadays is the affinity tag used for purification of the desired protein. The decision about position and choice of the affinity tag are mostly made at the beginning of the long way to a crystal structure (see Figure 1). However, it becomes crucial again at the crystallization step. In general most people like to remove the tag before crystallization to prove a physiological conformation. But, there are examples where the tag played a pivotal role in crystallization (Smits, Mueller et al. 2008a). The crystal structure of the octopine dehydrogenase from *Pecten maximus* is shown in Figure 9 (Smits, Mueller et al. 2008b), with the interactions sides/crystal contacts highlighted in green.
In Figure 9A contacts look quite similar to that presented in Figure 5. However when having a closer look on the his-tag, you recognize that it is located in a cavity formed by another monomer of that protein. In that cavity it can perform several hydrogen bonds with amino acids from the other monomer resulting in a very strong interaction which yields good quality crystals.

4.5.5 Crystallization using antibody fragments
A number of ways to stabilize proteins for crystallography have been developed, for example genetic engineering, co-crystallization with natural ligands and reducing surface

entropy (see above). Recently, crystallization mediated by antibody fragments has moved into the focus of crystallographers especially to obtain crystals of membrane proteins (Ostermeier, Iwata et al. 1995; Hunte and Michel 2002). Membrane protein crystallization is even tougher compared to soluble proteins, because of the amphipathic surface of the molecules. As they are located in the lipid bilayer most of their surface is hydrophobic and must be covered to keep them in solution. This is maintained by detergents. The detergent micelles cover the hydrophobic surface and therefore this area is no longer available to form crystal contacts. Crystal contacts can only be formed by the polar surfaces of these proteins. As many membrane proteins contain only relatively small hydrophilic domains, a strategy to increase the probability of getting well-ordered crystals is required. Antibody fragments can play this role. They can be designed for binding at specific regions in the protein and then function as additional polar domain in the membrane protein complex (for example see (Ostermeier, Iwata et al. 1995; Huber, Steiner et al. 2007).
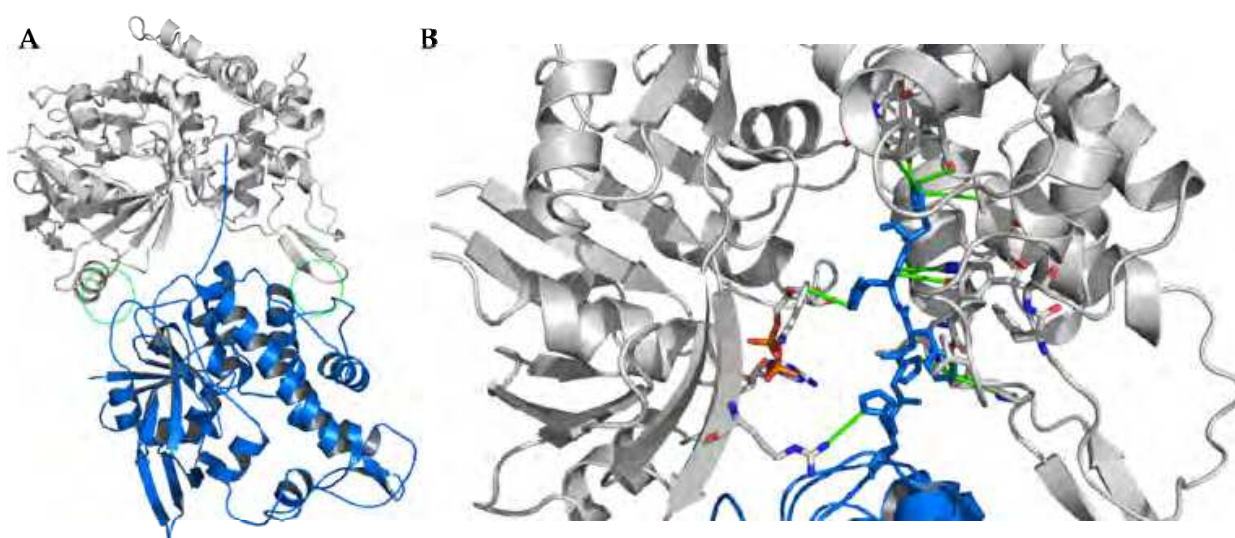
Fig. 9. Crystal contacts in OcDH protein. A: Overall view on two monomers. Surface Crystal contacts are highlighted in the green circles. B: Zoom in on the His-tag of one monomer. The his-tag of one monomer in the crystal structure is located near the binding site in a deep cavity formed by the other monomer. Therefore it is able to form several hydrogen bonds (highlighted in green) with side and main chains of the other protein but also with the ligand bound in this binding site (orange).

5. Conclusion

For what reason do we effort so much work on good quality crystals?

Single good quality crystals constitute an essential prerequisite for structural investigations of biological macromolecules using X-ray diffraction. The harder one works on crystal quality the easier the determination of a reasonable atomic model of the molecule of interest becomes. The vast majority of problems encountered in crystal structure determination can typically be traced back to data-quality issues caused by crystal imperfections. Consequently, although primary focus of structural biology is on the macromolecule that makes up a crystal, there is also considerable interest in the physical properties, nucleation

and growth of the crystals themselves. Statistics of various Structural Genomics Centers proved that protein crystallization is still despite all the progress in the technology of crystallization robotics is still a rather tough field in biological science. Success rate ranges for small prokaryotic proteins from 10-30% and decreases dramatically to a few percent for human proteins. The struggle obtaining crystals for protein structure determination is justified. After all efforst looking at electron density and subsequent the protein structure is still one of the most intriguing as well auspicious parts in structural biology

6. References

Alexandrov, A. I., M. Mileni, et al. (2008). "Microscale fluorescent thermal stability assay for membrane proteins." Structure 16(3): 351-359.

Altmann, F., E. Staudacher, et al. (1999). "Insect cells as hosts for the expression of recombinant glycoproteins." Glycoconj J 16(2): 109-123.

Arnau, J., C. Lauritzen, et al. (2006). "Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins." Protein expression and purification 48(1): 1-13.

Asherie, N. (2004). "Protein crystallization and phase diagrams." Methods 34(3): 266-272.

Baneyx, F. (1999). "Recombinant protein expression in *Escherichia coli*." Current opinion in biotechnology 10(5): 411-421.

Baud, F. and S. Karlin (1999). "Measures of residue density in protein structures." Proc Natl Acad Sci U S A 96(22): 12494-12499.

Berntsson, R. P., M. K. Doeven, et al. (2009). "The structural basis for peptide selection by the transport receptor OppA." EMBO J 28(9): 1332-1340.

Berntsson, R. P., S. H. Smits, et al. (2010). "A structural classification of substrate-binding proteins." FEBS Lett 584(12): 2606-2617.

Binda, C., M. Aldeco, et al. (2010). "Interactions of Monoamine Oxidases with the Antiepileptic Drug Zonisamide: Specificity of Inhibition and Structure of the Human Monoamine Oxidase B Complex." J Med Chem.

Blow, D. M., N. E. Chayen, et al. (1994). "Control of nucleation of protein crystals." Protein Sci 3(10): 1638-1643.

Blumenthal, D. K., K. Takio, et al. (1985). "Identification of the calmodulin-binding domain of skeletal muscle myosin light chain kinase." Proc Natl Acad Sci U S A 82(10): 3187-3191.

Bolotin, A., P. Wincker, et al. (2001). "The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. lactis IL1403." Genome research 11(5): 731-753.

Braig, K., Z. Otwinowski, et al. (1994). "The crystal structure of the bacterial chaperonin GroEL at 2.8 A." Nature 371(6498): 578-586.

Brzozowski, A. M. and J. Walton (2000). "Clear strategy screens for macromolecular crystallization." J Appl Cryst 34: 97-101.

Cayley, S., B. A. Lewis, et al. (1991). "Characterization of the cytoplasm of *Escherichia coli* K-12 as a function of external osmolarity. Implications for protein-DNA interactions in vivo." J Mol Biol 222(2): 281-300.

Chagnon, A. and M. Corbeil (1973). "Use of an organic buffer (HEPES) in human lymphocytoid cell line cultures." In Vitro 8(4): 283-287.

Chayen, N. E., J. W. Radcliffe, et al. (1993). "Control of nucleation in the crystallization of lysozyme." Protein Sci 2(1): 113-118.

Chayen, N. E., E. Saridakis, et al. (2001). "Porous silicon: an effective nucleation-inducing material for protein crystallization." J Mol Biol 312(4): 591-595.

Chayen, N. E., E. Saridakis, et al. (2006). "Experiment and theory for heterogeneous nucleation of protein crystals in a porous medium." Proc Natl Acad Sci U S A 103(3): 597-601.

Collins, B., R. C. Stevens, et al. (2005). "Crystallization Optimum Solubility Screening: using crystallization results to identify the optimal buffer for protein crystal formation." Acta Crystallogr Sect F Struct Biol Cryst Commun 61(Pt 12): 1035-1038.

Cregg, J. M. (2007). Pichia Protocols, Humana Press.

Cregg, J. M., J. L. Cereghino, et al. (2000). "Recombinant protein expression in Pichia pastoris." Molecular biotechnology 16(1): 23-52.

D., O. R., M. L.K., et al. (1992). Baculovirus Expression Vectors A Laboratory Manual. New York.

de Ruyter, P. G., O. P. Kuipers, et al. (1996). "Controlled gene expression systems for Lactococcus lactis with the food-grade inducer nisin." Applied and environmental microbiology 62(10): 3662-3667.

de Vos, W. M. and E. E. Vaughan (1994). "Genetics of lactose utilization in lactic acid bacteria." FEMS microbiology reviews 15(2-3): 217-237.

di Guan, C., P. Li, et al. (1988). "Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein." Gene 67(1): 21-30.

Drenth, J. (2006). "Principles of Protein X-Ray Crystallography." Springer 3.

Durst, R. A. and B. R. Staples (1972). "Tris-tris-HCl: a standard buffer for use in the physiologic pH range." Clin Chem 18(3): 206-208.

Eifler, N., M. Duckely, et al. (2007). "Functional expression of mammalian receptors and membrane channels in different cells." Journal of structural biology 159(2): 179-193.

Elowitz, M. B., M. G. Surette, et al. (1999). "Protein mobility in the cytoplasm of *Escherichia coli*." J Bacteriol 181(1): 197-203.

Ericsson, U. B., B. M. Hallberg, et al. (2006). "Thermofluor-based high-throughput stability optimization of proteins for structural studies." Anal Biochem 357(2): 289-298.

Erkens, G. B., R. P. Berntsson, et al. (2011). "The structural basis of modularity in ECF-type ABC transporters." Nat Struct Mol Biol 18(7): 755-760.

Ewing, F., E. Forsythe, et al. (1994). "Orthorhombic lysozyme solubility." Acta Crystallogr D Biol Crystallogr 50(Pt 4): 424-428.

Fischer, M., A. P. Leech, et al. (2011). "Comparative Assessment of Different Histidine-Tags for Immobilization of Protein onto Surface Plasmon Resonance Sensorchips." Analytical chemistry.

Gaberc-Porekar, V. and V. Menart (2001). "Perspectives of immobilized-metal affinity chromatography." Journal of biochemical and biophysical methods 49(1-3): 335-360.

Garman, E. (1999). "Cool data: quantity AND quality." Acta Crystallogr D Biol Crystallogr 55(Pt 10): 1641-1653.

Garman, E. F. and R. L. Owen (2006). "Cryocooling and radiation damage in macromolecular crystallography." Acta Crystallogr D Biol Crystallogr 62(Pt 1): 32-47.

Gasson, M. J. (1983). "Plasmid complements of *Streptococcus lactis* NCDO 712 and other lactic streptococci after protoplast-induced curing." Journal of bacteriology 154(1): 1-9.

Gingras, A. C., R. Aebersold, et al. (2005). "Advances in protein complex analysis using mass spectrometry." The Journal of physiology 563(Pt 1): 11-21.

Goldschmidt, L., D. R. Cooper, et al. (2007). "Toward rational protein crystallization: A Web server for the design of crystallizable protein variants." Protein Sci 16(8): 1569-1576.

Groß, R., R. Pisa, et al. (2002). "Isolierung der trimeren Hydrgenase aus Wollinella succinogenes durch StrepTactin-Affinitätschromatographie." Biospektrum 1: 101.

Guillot, A., C. Gitton, et al. (2003). "Proteomic analysis of *Lactococcus lactis*, a lactic acid bacterium." Proteomics 3(3): 337-354.

Haas, C. and J. Drenth (1999). "Understanding protein crystallization on the basis of the phase diagram." Journal of Crystal Growth 196(2-4): 388-394.

Hampton, E. N., M. W. Knuth, et al. (2007). "The self-inhibited structure of full-length PCSK9 at 1.9 A reveals structural homology with resistin within the C-terminal domain." Proc Natl Acad Sci U S A 104(37): 14604-14609.

Hansen, S. B., G. Sulzenbacher, et al. (2006). "Structural characterization of agonist and antagonist-bound acetylcholine-binding protein from *Aplysia californica*." J Mol Neurosci 30(1-2): 101-102.

Hefti, M. H., C. J. Van Vugt-Van der Toorn, et al. (2001). "A novel purification method for histidine-tagged proteins containing a thrombin cleavage site." Analytical biochemistry 295(2): 180-185.

Hochuli, E., H. Dobeli, et al. (1987). "New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues." Journal of chromatography 411: 177-184.

Holm, J. and S. I. Hansen (2001). "Effect of hydrogen ion concentration and buffer composition on ligand binding characteristics and polymerization of cow's milk folate binding protein." Biosci Rep 21(6): 745-753.

Hope, H. (1988). "Cryocrystallography of biological macromolecules: a generally applicable method." Acta Crystallogr B 44 (Pt 1): 22-26.

Horwich, A. (2000). "Working with Paul Sigler." Nat Struct Biol 7(4): 269-270.

Huber, T., D. Steiner, et al. (2007). "In vitro selection and characterization of DARPins and Fab fragments for the co-crystallization of membrane proteins: The Na(+)-citrate symporter CitS as an example." J Struct Biol 159(2): 206-221.

Hunt, I. (2005). "From gene to protein: a review of new and enabling technologies for multi-parallel protein expression." Protein expression and purification 40(1): 1-22.

Hunte, C. and H. Michel (2002). "Crystallisation of membrane proteins mediated by antibody fragments." Curr Opin Struct Biol 12(4): 503-508.

Hunte, C., E. Screpanti, et al. (2005). "Structure of a Na+/H+ antiporter and insights into mechanism of action and regulation by pH." Nature 435(7046): 1197-1202.

Jana, S. and J. K. Deb (2005). "Strategies for efficient production of heterologous proteins in Escherichia coli." Applied microbiology and biotechnology 67(3): 289-298.

Jancarik, J. and S. H. Kim (1991). "Sparse matrix sampling: a screening method for crystallization of proteins." J Appl Cryst 24: 409-411.

Jancarik, J., R. Pufan, et al. (2004). "Optimum solubility (OS) screening: an efficient method to optimize buffer conditions for homogeneity and crystallization of proteins." Acta Crystallogr D Biol Crystallogr 60(Pt 9): 1670-1673.

Johari, G. P., A. Hallbrucker, et al. (1987). "The glass-liquid transition of hyperquenched water." Nature 330: 552-553.

Jonasson, P., S. Liljeqvist, et al. (2002). "Genetic design for facilitated production and recovery of recombinant proteins in Escherichia coli." Biotechnology and applied biochemistry 35(Pt 2): 91-105.

Juers, D. H. and B. W. Matthews (2004). "Cryo-cooling in macromolecular crystallography: advantages, disadvantages and optimization." Q Rev Biophys 37(2): 105-119.

Kean, J., R. M. Cleverley, et al. (2008). "Characterization of a CorA Mg2+ transport channel from Methanococcus jannaschii using a Thermofluor-based stability assay." Mol Membr Biol 25(8): 653-663.

Konings, W. N., J. Kok, et al. (2000). "Lactic acid bacteria: the bugs of the new millennium." Curr Opin Microbiol 3(3): 276-282.

Kopecky, D. J., X. Hao, et al. (2008). "Identification and optimization of N3,N6-diaryl-1H-pyrazolo[3,4-d]pyrimidine-3,6-diamines as a novel class of ACK1 inhibitors." Bioorg Med Chem Lett 18(24): 6352-6356.

Kuipers, O. P., M. M. Beerthuyzen, et al. (1995). "Autoregulation of nisin biosynthesis in *Lactococcus lactis* by signal transduction." The Journal of biological chemistry 270(45): 27299-27304.

Kuipers, O. P., P. G. de Ruyter, et al. (1998). "Quorum sensing-controlled gene expression in lactic acid bacteria." J Biotechnol 64: 15-21.

Kulahin, N., O. Kristensen, et al. (2011). "Structural model and trans-interaction of the entire ectodomain of the olfactory cell adhesion molecule." Structure 19(2): 203-211.

Kunji, E. R., D. J. Slotboom, et al. (2003). "*Lactococcus lactis* as host for overproduction of functional membrane proteins." Biochimica et biophysica acta 1610(1): 97-108.

Lawson, D. M., P. J. Artymiuk, et al. (1991). "Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts." Nature 349(6309): 541-544.

Li, M. Z. and S. J. Elledge (2007). "Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC." Nat Methods 4(3): 251-256.

Linding, R., L. J. Jensen, et al. (2003). "Protein disorder prediction: implications for structural proteomics." Structure 11(11): 1453-1459.

Lo Conte, L., C. Chothia, et al. (1999). "The atomic structure of protein-protein recognition sites." J Mol Biol 285(5): 2177-2198.

Low, B. W., C. C. Chen, et al. (1966). "Studies of insulin crystals at low temperatures: effects on mosaic character and radiation sensitivity." Proc Natl Acad Sci U S A 56(6): 1746-1750.

Lundstrom, K. (2006a). "Structural genomics for membrane proteins." Cellular and molecular life sciences : CMLS 63(22): 2597-2607.

Lundstrom, K., R. Wagner, et al. (2006b). "Structural genomics on membrane proteins: comparison of more than 100 GPCRs in 3 expression systems." Journal of structural and functional genomics 7(2): 77-91.

Madoori, P. K., H. Agustiandari, et al. (2009). "Structure of the transcriptional regulator LmrR and its mechanism of multidrug recognition." EMBO J 28(2): 156-166.

Mavaro, A., A. Abts, et al. (2011). "Substrate recognition and specificity of NISB, the lantibiotic dehydratase involved in nisin biosynthesis." J Biol Chem.

McElroy, H. H., G. W. Sissom, et al. (1992). "Studies on engineering crystallizability by muttion of surface residues of human thymidylate synthase." J. Cryst. Growth 122: 265-272.

McPherson, A. and P. Shlichta (1988). "Heterogeneous and epitaxial nucleation of protein crystals on mineral surfaces." Science 239(4838): 385-387.

Menetrey, J., A. Bahloul, et al. (2005). "The structure of the myosin VI motor reveals the mechanism of directionality reversal." Nature 435(7043): 779-785.

Midgett, C. R. and D. R. Madden (2007). "Breaking the bottleneck: eukaryotic membrane protein expression for high-resolution structural studies." Journal of structural biology 160(3): 265-274.

Mierau, I. and M. Kleerebezem (2005). "10 years of the nisin-controlled gene expression system (NICE) in Lactococcus lactis." Applied microbiology and biotechnology 68(6): 705-717.

Muller, K. M., K. M. Arndt, et al. (1998). "Tandem immobilized metal-ion affinity chromatography/immunoaffinity purification of His-tagged proteins--evaluation of two anti-His-tag monoclonal antibodies." Anal Biochem 259(1): 54-61.

Murphy, K. P., P. L. Privalov, et al. (1990). "Common features of protein unfolding and dissolution of hydrophobic compounds." Science 247(4942): 559-561.

Ostermeier, C., S. Iwata, et al. (1995). "Fv fragment-mediated crystallization of the membrane protein bacterial cytochrome c oxidase." Nat Struct Biol 2(10): 842-846.

Oswald, C., S. H. Smits, et al. (2008). "Microseeding - a powerful tool for crystallizing proteins complexed with hydrolyzable substrates." Int J Mol Sci 9(7): 1131-1141.

Pittelkow, M., B. Tschapek, et al. (2011). "The Crystal Structure of the Substrate-Binding Protein OpuBC from *Bacillus subtilis* in Complex with Choline." Journal of molecular biology 411(1): 53-67.

Porath, J., J. Carlsson, et al. (1975). "Metal chelate affinity chromatography, a new approach to protein fractionation." Nature 258(5536): 598-599.

Regnier, F. E. (1983a). "High-performance liquid chromatography of biopolymers." Science 222(4621): 245-252.

Regnier, F. E. (1983b). "High-performance liquid chromatography of proteins." Methods Enzymol 91: 137-190.

Ressl, S., A. C. Terwisscha van Scheltinga, et al. (2009). "Molecular basis of transport and regulation in the Na(+)/betaine symporter BetP." Nature 458(7234): 47-52.

Rodgers, D. W. (1994). "Cryocrystallography." Structure 2(12): 1135-1140.

Rosenbaum, D. F. and C. F. Zukoski (1996). "Protein interactions and crystallization." Journal of Crystal Growth 169(4): 752-758.

Rubio, V., Y. Shen, et al. (2005). "An alternative tandem affinity purification strategy applied to Arabidopsis protein complex isolation." The Plant journal : for cell and molecular biology 41(5): 767-778.

Rupp, B. (2007). "Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology." Taylor & Francis Ltd 1.

Sachdev, D. and J. M. Chirgwin (1999). "Properties of soluble fusions between mammalian aspartic proteinases and bacterial maltose-binding protein." Journal of protein chemistry 18(1): 127-136.

Sala, E. and A. de Marco (2010). "Screening optimized protein purification protocols by coupling small-scale expression and mini-size exclusion chromatography." Protein Expr Purif 74(2): 231-235.

Sauter, C., B. Lorber, et al. (1999). "Crystallogenesis studies on yeast aspartyl-tRNA synthetase: use of phase diagram to improve crystal quality." Acta Crystallogr D Biol Crystallogr 55(Pt 1): 149-156.

Schmidt, T. G., J. Koepke, et al. (1996). "Molecular interaction between the Strep-tag affinity peptide and its cognate target, streptavidin." Journal of molecular biology 255(5): 753-766.

Schwarz, C. K., B. Tschapek, et al. (2011). "Crystallization and preliminary X-ray crystallographic studies of an oligomeric species of a refolded C39 peptidase-like domain of the *Escherichia coli* ABC transporter haemolysin B." Acta Crystallogr Sect F Struct Biol Cryst Commun 67(Pt 5): 630-633.

Senkovich, O., W. J. Cook, et al. (2007). "Structure of a complex of human lactoferrin N-lobe with pneumococcal surface protein a provides insight into microbial defense mechanism." Journal of molecular biology 370(4): 701-713.

Smith, D. B. and K. S. Johnson (1988). "Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase." Gene 67(1): 31-40.

Smith, G. E., M. D. Summers, et al. (1983). "Production of human beta interferon in insect cells infected with a baculovirus expression vector." Molecular and cellular biology 3(12): 2156-2165.

Smits, S. H., A. Mueller, et al. (2008a). "Coenzyme- and His-tag-induced crystallization of octopine dehydrogenase." Acta Crystallogr Sect F Struct Biol Cryst Commun 64(Pt 9): 836-839.

Smits, S. H., A. Mueller, et al. (2008b). "A structural basis for substrate selectivity and stereoselectivity in octopine dehydrogenase from *Pecten maximus*." J Mol Biol 381(1): 200-211.

Sorensen, H. P. and K. K. Mortensen (2005). "Advanced genetic strategies for recombinant protein expression in *Escherichia coli*." Journal of biotechnology 115(2): 113-128.

Stofko-Hahn, R. E., D. W. Carr, et al. (1992). "A single step purification for recombinant proteins. Characterization of a microtubule associated protein (MAP 2) fragment which associates with the type II cAMP-dependent protein kinase." FEBS letters 302(3): 274-278.

Taha, M. (2005). "Buffers for the physiological pH range: acidic dissociation constants of zwitterionic compounds in various hydroorganic media." Ann Chim 95(1-2): 105-109.

Tang, C., C. D. Schwieters, et al. (2007). "Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR." Nature 449(7165): 1078-1082.

Tate, C. G., J. Haase, et al. (2003). "Comparison of seven different heterologous protein expression systems for the production of the serotonin transporter." Biochimica et biophysica acta 1610(1): 141-153.

Terpe, K. (2003). "Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems." Applied microbiology and biotechnology 60(5): 523-533.

Teuber, M. (1995). "The genus Lactococcus." The genera of lactic acid bacteria 2.

Tornquist, K., J. Paallysaho, et al. (1995). "Influence of Hepes- and CO2/HCO(3-)-buffer on Ca2+ transients induced by TRH and elevated K+ in rat pituitary GH4C1 cells." Mol Cell Endocrinol 112(1): 77-82.

Tschapek, B., M. Pittelkow, et al. (2011). "Arg149 Is Involved in Switching the Low Affinity, Open State of the Binding Protein AfProX into Its High Affinity, Closed State." Journal of molecular biology 411(1): 36-52.

Urh, M., D. York, et al. (1995). "Buffer composition mediates a switch between cooperative and independent binding of an initiator protein to DNA." Gene 164(1): 1-7.

van de Guchte, M., J. Kok, et al. (1992). "Gene expression in *Lactococcus lactis*." FEMS microbiology reviews 8(2): 73-92.

Voss, S. and A. Skerra (1997). "Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification." Protein engineering 10(8): 975-982.

Waugh, D. S. (2005). "Making the most of affinity tags." Trends in biotechnology 23(6): 316-320.

Wegner, G. H. (1990). "Emerging applications of the methylotrophic yeasts." FEMS microbiology reviews 7(3-4): 279-283.

Weik, M., G. Kryger, et al. (2001). "Solvent behaviour in flash-cooled protein crystals at cryogenic temperatures." Acta Crystallogr D Biol Crystallogr 57(Pt 4): 566-573.

Zaitseva, J., I. B. Holland, et al. (2004). "The role of CAPS buffer in expanding the crystallization space of the nucleotide-binding domain of the ABC transporter haemolysin B from *Escherichia coli*." Acta Crystallogr D Biol Crystallogr 60(Pt 6): 1076-1084.

Zaitseva, J., S. Jenewein, et al. (2005). "Functional characterization and ATP-induced dimerization of the isolated ABC-domain of the haemolysin B transporter." Biochemistry 44(28): 9680-9690.

Zheng, C. F., T. Simcox, et al. (1997). "A new expression vector for high level protein production, one step purification and direct isotopic labeling of calmodulin-binding peptide fusion proteins." Gene 186(1): 55-60.

Zimmerman, S. B. and S. O. Trach (1991). "Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*." J Mol Biol 222(3): 599-620.
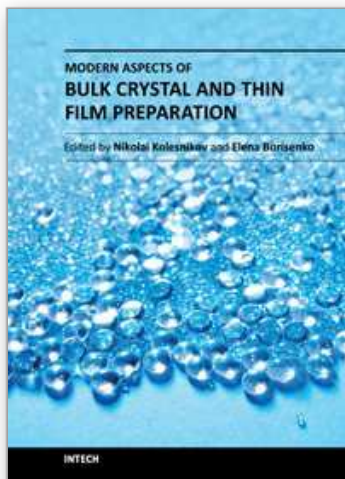
Modern Aspects of Bulk Crystal and Thin Film Preparation

Edited by Dr. Nikolai Kolesnikov

In modern research and development, materials manufacturing crystal growth is known as a way to solve a wide range of technological tasks in the fabrication of materials with preset properties. This book allows a reader to gain insight into selected aspects of the field, including growth of bulk inorganic crystals, preparation of thin films, low-dimensional structures, crystallization of proteins, and other organic compounds.

INTECH

open science | open minds