

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

5,800

Open access books available

142,000

International authors and editors

180M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



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](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Structure- and Design-Based Difficulties in Recombinant Protein Purification in Bacterial Expression

*Kubra Acikalin Coskun, Nazlıcan Yurekli, Elif Cansu Abay, Merve Tutar, Mervenur Al and Yusuf Tutar*

## Abstract

Protein purification is not a simple task. Yet, overexpression at bacterial systems with recombinant modifications brings further difficulties. Adding a tag, an affinity label, and expressing particular domains of the whole protein, especially hydrophobic sections, make purification a challenging process. Protein folding pattern may perturb N- or C-terminal tag and this terminal preference may lead to poor purification yield. Codon optimization, solvent content and type, ionic conditions, resin types, and self-cleavage of recombinant proteins bring further difficulties to protein expression and purification steps. The chapter overviews problems of protein purification through a small peptide overexpression in bacteria (Recombinant anti-SARS Coronavirus 2 (SARS-Cov-2) Spike protein Receptor Binding Domain (RBD) antibody (Clone Sb#14)). The chapter also covers troubleshooting at distinct steps and highlights essential points to solve crucial issues of protein purification.

**Keywords:** protein, prokaryotic protein expression, and purification, protein modeling, protein aggregation, ionic strength

## 1. Introduction

Recombinant DNA technology involves genetic engineering; cutting DNA molecules from distinct biological species and then ligating them to a vector for expression [1, 2]. The technology helps to express the desired protein in large quantity rather than extracting from bulk amounts of tissues and animal fluids. Proteins are synthesized and modified depending on their functions in an organism. As an initial step, DNA encodes protein through transcription from mRNA synthesis. Then, mRNA is converted into protein. Transcription and translation occur simultaneously in prokaryotic organisms. The conversion of mRNA to protein begins before the synthesis of the mature mRNA transcript [3]. Protein expression involves the synthesis, modification, and regulation of a particular protein in a living organism. However, bacterial systems lack human protein modifications but overexpress recombinant proteins in bulk amounts [4]. Recombinant protein expression is useful to understand the structure and function of proteins. A network of protein complex functions can be distinguished by the characterization of individual

proteins function as well as interactions through recombinant protein techniques. Protein–protein and protein-ligand interactions may be highlighted by expressing interacting domains and by introducing key mutations to reveal key domains and residues, respectively. Considering the size and complexity of proteins, protein production is very efficient with vector templates using K12 bacterial systems [2, 4].

## **1.1 Bacterial protein synthesis**

Recombinant protein production in bacterial systems is fast, easy, and highly efficient [5]. The general strategy for recombinant protein production involves transforming the cell with a DNA vector containing the open reading frame of the gene, then after subcloning to the expression system, the protein is induced in the cells. After incubating the induced cells, harvested cells are lysed for further separation. The selection of the purification system depends on the type of protein, the affinity tag of the plasmid, the isoelectric point (pI) of the protein of interest, the molecular mass of the protein, the targeted yield, and the degree of functional activity. The lysed cells are purified through column chromatography with a proper resin(s) and a convenient buffer system [6, 7]. But in practice, several steps may cause problems. These include inadequate growth of the selected bacterial host cell, the formation of inclusion bodies, protein aggregation, structural alteration, recombinant protein nonspecific interaction with cellular proteins, problems in column systems used in purification [6, 8]. Further, as eukaryotic proteins expressed in *E. coli* may not perform post-translation modifications of other organism proteins, loss of function is observed. In addition, some of the proteins expressed are exposed to hard-denaturing agents or may cause collapses in structure and this often results in insolubility problems. Overexpression of recombinant proteins in bacterial systems leads to the formation of inclusion bodies. Re-folding of these proteins into their bioactive forms is cumbersome and requires a variety of agents and processes [1, 9].

### *1.1.1 Host cell selection*

First of all, the choice of the host cells to produce intact protein in the synthesis mechanism forms the mainline of the whole system. Microorganisms used in recombinant protein expression systems include bacteria and yeast. Each host has strengths and weaknesses. The organism to be selected varies depending on the particular protein, working conditions, desired yield. For example, if the desired protein has post-translational modifications, choosing a prokaryotic expression system would not be proper [10]. BL21 (DE3) and its derivatives are by far the most commonly used strains for recombinant protein synthesis. In addition, its genetics are characterized in more detail than any other microorganisms. Recent studies suggest that BL21 (DE3) gene-level research made this bacterium more important for the production of heterologous proteins. This host cell provides maximum efficiency in protein expression through inexpensive substrates, capable of rapid and high-yield growth. A modified form includes a pLysS plasmid that encodes T7 lysozyme. This lowers the background protein expression of recombinant protein but does not perturb IPTG induction. The plasmid is especially useful in toxic cases and provides an option for protein over-expression. Yeast is an alternative recombinant protein production host and provides eukaryotic post-translational modifications with high yield. Yeast growth temperature (30°C) is lower than that of bacteria (37°C) but the growth rate is much slower. Further, the transformation of plasmids to yeast is relatively difficult and the selection of transformed cells and growth conditions require special conditions [10, 11].

### 1.1.2 Plasmid selection

The expression plasmids consist of the replication origin, promoter, and multiple cloning sites. The most important issue to consider when choosing an appropriate vector is the copy number property. Because the number of copies is controlled by the replication. It is not always true to assume that the high amount of plasmid is proportional to the yield of recombinant protein expression. Because the high copy is inversely proportional to the rate of bacterial growth. In addition, this condition creates plasmid instability and creates a metabolic load. As a result protein production yield decreases [12, 13].

### 1.1.3 Promoter

Prokaryotes have to adapt to the environment by responding quickly to environmental changes. *E. coli* cells cannot use lactose directly as a source of carbon. But they use glucose, a component of lactose. For the bacterial cell to metabolize lactose, it is necessary to take lactose into the cell and break it down into a glucose monomer. For this, it is necessary to synthesize three different enzymes in the cell [6, 14]. As with *E. coli*, bacteria combine genes related to the same metabolic pathways to form clusters called operons. Transcription of the genes that make up the operon start from a single promoter. The resulting transcription product consists of an mRNA molecule containing information from multiple genes. Preserved DNA sequences in the promoter region help connect the enzyme to the DNA molecule. Induction is difficult in the presence of easily metabolized carbon sources. If lactose and glucose are present in the environment, expression from the *lac* promoter is not fully induced until all glucose is used up. In the absence of glucose, the promoter expresses the three enzymes to break down the lactose to obtain glucose. This property is used to induce prokaryotic expression vectors through IPTG (isopropyl 1-thio- $\beta$ -D-galactopyranoside); a lactose analog that binds *lac* repressor [14, 15]. In the commercial vectors, IPTG starts the transcription of the *lac* operon and eventually induces protein expression where the gene of interest is controlled by the *lac* operator.

### 1.1.4 Marker selection

A resistance marker is added to the plasmid to prevent the growth of cells that do not carry plasmids. This can be achieved by using a selection marker. For example, ampicillin resistance is conferred by the *bla* gene,  $\beta$ -lactamase, a periplasmic enzyme that inactivates the  $\beta$ -lactam ring of  $\beta$ -lactam antibiotics [16].

### 1.1.5 Affinity tags and its contribution to protein solubility

The addition of affinity tags to the plasmid (such as His Tag, glutathione-S-transferase, and cellulose-binding domain) is employed to separate a particular protein from the heterogeneous protein mixture during purification, forming disulfide bonds, increasing the solubility of the recombinant proteins and transferring them to the periplasm region. Affinity tags have a great role in separating the desired protein from cell lysate in recombinant protein purification. Affinity tags are divided into small peptide tags (amino acids) and large polypeptide tags (fusion partners) [17]. Small peptide tags are less likely to interfere when fused to the protein. In some cases, this may have negative consequences on the tertiary conformation and biological activity of the fused chimeric protein. Vectors are available that allow tags to be placed optionally at the N-terminal or C-terminal

end. It is more advantageous to position a signal peptide at the N-terminal end for better secretion of the recombinant protein. At this point, it is important to know which end of the protein is embedded in the folding pattern by examining the three-dimensional structure of a particular protein, and it is necessary to place the label on the solvent-exposed end. Examples of small peptide tags are poly-His, c-Myc, and FLAG [18]. His-tagged proteins can be purified by affinity chromatography in resins containing positively charged metal ion nickel. In addition, at the end of purification, with commercial antibodies, labeled recombinant protein can be detected by western blot [17–21]. On the other hand, it increases the solubility of the recombinant protein produced by the addition of a non-peptide fusion partner (large polypeptide label). The most commonly used fusion labels include Thioredoxin (Trx), Ubiquitin, SUMO, Maltose binding protein (MBP), Glutathione S-transferase (GST) [17, 22]. The reason why fusion partners show properties that increase the solubility of the protein is still not fully explained. Though, MBP label has been shown to carry a small chaperone activity. The GST label has been shown to have the weakest solubility-enhancing effect among fusion partners. Trx has the most solubility-enhancing properties, but due to its size, it may cause adverse effects. In recent years, studies have shown that “Calcium-Binding Protein Fh8” tag derived from a parasite called “*Fasciola hepatica*” recombinantly added to proteins increases protein solubility [6, 17, 20, 21]. Studies are underway for better solubility enhancing effect of recombinant protein tags.

## 1.2 Troubleshooting strategies for recombinant protein expression

Even if the effective parameters are provided in the production of recombinant protein, it may not be determined exactly whether the desired protein will be eluted excessively and in active soluble form. Therefore, there are additional strategies for optimizing protein expression [7].

### 1.2.1 Low or no protein production

If the desired protein cannot be detected using sensitive techniques or is detected at a low expression rate, the problem is usually caused by a toxic effect of the heterologous protein in the cell. As a result of protein toxicity in the host cell, cells cannot proliferate at a sufficient level and show a low growth rate [7, 23]. The first measure to solve this problem should be followed before proliferating cells are induced. If the growth rate of the recombinant cell is slower than that of the strain with empty vectors, it is related to either gene toxicity or the basal expression of toxic mRNA and protein. Control of basal production is associated with the operon system. *LacI* or *LacIQ* expression blocks transcription in *Lac*-based promoters. High-copy plasmids must be cloned in the *LacI* Q expression vector. Since the presence of tryptone or peptone in the growth medium contains inducing lactose, a more controlled expression is provided with the addition of glucose at 0.2–1 w/v. Plasmids containing T7-based promoters prevent leaky production, such as BL21DE3-pLYS (S) [8, 24].

### 1.2.2 Limiting factors in the medium

Luria Bertani (LB), the most commonly used growth medium environment for *E. coli* culture, is an ideal environment for high-nutrient cell growth. When recombinant protein production cannot be replicated with the recommended mechanisms, production efficiency can be increased by increasing the volume of the targeted protein. A successful result can be achieved with adequate ventilation

with rigorous shaking of the growth medium. Although LB has a high protein content, cell proliferation is partially reduced. This is due to the low carbohydrate content of LB. As a solution to this situation, increasing peptone and yeast extract provides higher cell proliferation with the addition of  $MgSO_4$ , which contributes to the ionic intensity of the environment. In addition, the amount of acid released as a result of increased glucose metabolism over time exceeds the buffering capacity of LB. In case of acidification of the growth medium, 50 mM phosphate salts can be added to the environment and buffered [7, 11]. In the broth culture, as the number of cells per unit media increases, oxygen limitation occurs and changes the metabolic capacity of the cell. This prevents optimal growth and the easiest way to increase the amount of oxygen in the growth medium is to increase the speed of the shaking containers. The optimum shaking speed range is 300–400 rpm. Several anti-foaming agents can be added to the broth culture to prevent the negative effect of the foams formed by strong shaking on oxygen circulation [24].

### *1.2.3 Formation of inclusion bodies*

The inclusion bodies formed in *E. coli* are denatured protein molecules that do not display biological activity. Dissolving, refolding, and purification protocols should be applied, respectively, to make inclusion objects functionally active and soluble. In the transfer of a foreign gene to *E. coli*, control of gene expression is lost. The nascent polypeptide expression depends on several factors such as osmosis, folding pattern, and pH. If expression increases, the number of unspecific hydrophobic interactions in the polypeptide chain increases. This causes instability and clustering in polypeptization. The resulting protein aggregation is called “inclusion bodies.” The main reason for the formation of clustering is due to the deterioration of the balance between protein aggregation and protein resolution [1, 25]. Therefore, a soluble recombinant protein can be obtained through strategies that eliminate the factors causing the formation of inclusion bodies. As mentioned in the “Affinity tags” section, one way to prevent the solubility problem that may occur in the expressed protein is; combining the desired protein with a fusion partner (large polypeptide tag) that acts as a solubilizer [17].

### *1.2.4 Disulfide bond formation*

To obtain the biologically active three-dimensional structure of recombinant proteins, it is important to establish the right disulfide bonds. The formation of improper disulfide bonds causes the protein to fold incorrectly and the formation of inclusion bodies. Disulfide change reactions catalyzed by many enzymes in the Dsb family, where cysteine oxidation occurs in *E. coli* periplasm, form disulfide bonds in the polypeptide chain [26]. In the cytoplasm, the formation of disulfide bonds is rare because the remnants of cysteine are catalytic regions for many enzymes in the cytoplasm. The wrong disulfide bonds in these regions can cause protein inactivation, clustering, and incorrect folding. However, some strains of *E. coli* have conditions that trigger the formation of a disulfide bond [5].

### *1.2.5 Addition of chemical chaperones and co-factors*

Molecular chaperones form the heart of protein synthesis and help nascent polypeptides fold into their active structures. Some specific types of chaperones, such as ClpB, can cleave unfolded polypeptides contained in inclusion bodies. However, high levels of recombinant protein production may result in increased molecular traffic in the cytoplasm, resulting in uncontrolled protein folding control.

One strategy used to solve this problem is to arrest protein expression by removing the inducer after a centrifugation step and adding a fresh medium containing chloramphenicol, the protein expression inhibitor. Thus, it allows the recruitment of molecular chaperones to enable the folding of newly synthesized recombinant polypeptides [27, 28]. One of the systems used commercially for protein folding is chaperone plasmids. This system consists of plasmids that allow overexpression of different chaperones or their combinations. Examples of these are GroES-GroEL, DnaK/DnaJ/GrpE [27]. When proteins are released from inclusion bodies, denatured with urea, and subsequently folded *in vitro*, the addition of osmolytes (proline, trehalose) at a concentration ratio of 0.1–1M increases the yield of soluble protein. In addition, the correctly folded protein may require special cofactors such as metal ions (such as magnesium, iron/sulfur) or polypeptide cofactors in the media medium to reach its final conformation. The addition of these compounds to the culture increases the yield and the folding rate of soluble proteins [8, 28].

### 1.2.6 Slowing down the production rate

Slowing the production rate of the recombinant protein reduces cellular protein concentration and protein trafficking, allowing the synthesized polypeptides to fold more smoothly. The most common method of reducing the rate of protein synthesis is to lower the incubation temperature [29]. Decreased temperature prevents the formation of aggregation due to its reduction of hydrophobic interactions. Recombinant protein synthesis occurs in the temperature range of 15–25°C. However, when working at the lower temperature range, this causes slower growth and therefore lower protein synthesis. This obstacle can be overcome with commercial products. The ArticExpress™ (Agilent Technologies) competent cells improve recombinant protein expression at low temperatures through co-expressing ortholog genes of *E. coli* GroEL and GroES from *Oleispira Antarctica*, namely Cpn60 and co-chaperone Cpn10. These chaperones work together to fold a substrate protein, and usually carry re-folding activity at 4–12°C temperature range, increasing recombinant protein yield and solubility at lower temperatures [30].

## 2. Techniques used in recombinant protein purification and detection

Selection of the purification methods generally uses distinct characteristics of the proteins. The distinct properties of recombinant proteins may include chemical, biological, and physical features due to differences in spatial structure and amino acid sequences. Usually, to benefit from these differences, multiple steps are required in the optimal purification process but it should be noted that each step may cause loss of product stability and/or yield, therefore the lowest number of steps are recommended overall for maximum yield. So, method selection determines the ratio of better yield to the better-purified product. Key factors that can affect the purification selection steps include the solubility of the lysate, sample size, and physicochemical properties of the target protein. The first step for purification is to analyze the protein characteristics and match them with literature reports-protocols. For example, a useful parameter in the purification process is amino acid composition. pKa and pI values can be calculated using the amino acid composition. Determination of the values helps to select column type, buffer, pH, or resin type. Once optimization of the purification is established, the method may be employed for a protein with similar sequences or motifs at least for orthologs or isoforms [31]. The characteristic features of proteins that are mainly used for purification type selection are solubility, size, charge, and specific binding

affinity [32]. By using these properties, numerous techniques may be employed in protein purification. Solubility parameter can be used with “salting out” through the knowledge of proteins mostly being less soluble in high salt concentrations. And hence, this strategy can be used to separate the protein of interest. Further, dialysis can be used after salting out to remove the salt molecules [33]. Another technique that uses size difference is gel-filtration or size exclusion chromatography (SEC). A column with porous beads resin is used for this and as the sample goes through the column, beads help to separate molecules. The beads are usually 0.1 mm in diameter, so bigger molecules cannot permeate the pores but small molecules penetrate into the pores and are trapped there for a while until the molecules exit again and return to solvent. This action retards small molecules but bigger molecules travel rapidly through a void volume with buffer flow. Small molecules shielding and bigger molecules faster flow separate molecules from each other in fractions depending on their sizes. And as the molecules exit the column, bigger molecules elute first and then, smaller molecules come after.

If the net charge is criteria to be used as a separating feature, ion-exchange chromatography can be used. If the target protein is positively charged as a cationic protein, then, a negatively charged carboxymethyl-cellulose (CM-cellulose) pre-packed column/resin can be used. But if the protein is negatively charged as in anionic proteins, then positively charged diethylaminoethyl-cellulose (DEAE-cellulose) pre-packed columns/resin can be used [33]. It is also known that proteins can have high affinities for certain chemical groups. Affinity chromatography can use this feature to purify proteins and its effect is the best on proteins with affinities to highly specific molecules.

Distinct separation techniques, that is, ion exchange and gel filtration may be employed at high-pressure liquid chromatography (HPLC) with proper column selection. This technique differs from the others because the applied pressure is significantly higher and it does not rely on gravity for sample flow. However, high-pressure limits the purification of higher molecular weight proteins as the pressure denatures protein structure. For higher-molecular-weight proteins FPLC (fast-pressure liquid chromatography) is preferred to prevent pressure-dependent denaturation. FPLC is the preferred technique for protein chemists since any target protein can be separated from cellular lysate readily. And the technique provides a wide range of column options. The flexibility of this technique provides the purification of stable proteins with a high yield. Lower pressure provides advantages as well. Clogging due to lysate content and backpressure problems are less likely encountered compared to that of HPLC. The techniques may also be used with tagged proteins. Histidine tag is one of the most common ones that are used with recombinant proteins and it has a high affinity of metal ions like  $\text{Ni}^{2+}$  [17]. To screen if the purification steps are working, gel electrophoresis can be used. In-gel electrophoresis, proteins are separated by their mass as they go through the gel and the smaller ones move faster. As they get separated by their masses, proteins can be visualized in the gel and the gel show protein of interest among others. Another feature to separate proteins is their isoelectric point. This point represents the pH level where the protein has zero net charges. The technique that uses this property to separate proteins is called isoelectric focusing. When proteins go through a pH gradient gel, they will stop at the point where they have no net charge and get separated from the proteins with distinct isoelectric points. To get more specific results, isoelectric focusing can be coupled with SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) in a technique called two-dimensional electrophoresis. In this technique, first isoelectric focusing is done as the sample goes through the gel horizontally and after proteins stop at their respective pH levels, vertical electrophoresis starts. So, the sample



is separated according to the isoelectric points horizontally and by their masses vertically. The two-dimensional separation technique is employed to distinguish differences in two different states. Actually, distinct spots may be characterized by MALDI-TOF mass spectrometry.

## 2.1 Protein structure determination and screening databases

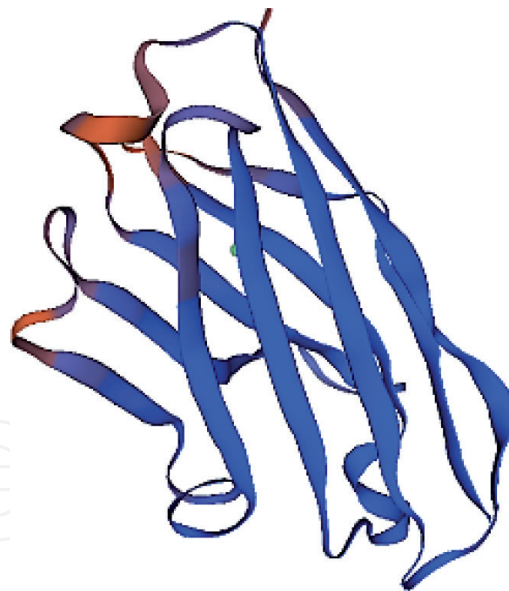
The structure and function of a protein are essential to characterize the linkage associated between common motifs and biochemical activity [34]. These features are normally determined by NMR or X-ray crystallography techniques [35]. NMR determines dynamic structure but the technique is limited to protein molecular mass. However, an instant picture of the protein structure with a relatively higher mass can be taken by X-ray crystallography. After elucidation of information from these structure determination techniques, scientists concluded that similar sequences show similar structural patterns [36]. Then, different databases which can display protein 2 and 3-dimensional structures have been developed. Determination of protein structure is important for not only understanding the function but also important for protein experiments, such as protein purification. Proteins have alpha helices, turns, and beta sheets as secondary structures. Beta sheet structures and the outer surface of alpha helices of proteins can accumulate within the cellular medium or stick to each other and other proteins during aggregation. *In vitro* experiments of proteins showed that proteins do not act like they are within the cell in terms of their stability, charge, and interactions properties. To understand the optimum conditions for protein purification, structure determination databases and pI calculation play crucial roles. In the case of His-tagged proteins, pI is one of the most important parameters for purification. pI is determined with a special calculation by considering amino acid sequences. NCBI Protein Data Bank provides the amino acid sequence of desired proteins (<https://www.rcsb.org/>). Several properties of the selected protein can be calculated by Expasy (Expert Protein Analysis System) Tool which is one of the most convenient tools on the web ([https://web.expasy.org/compute\\_pi/](https://web.expasy.org/compute_pi/)) [36].

## 2.2 Swiss modeling and I-Tasser

Protein structure can be determined by bioinformatic tools such as Swiss Modeling or I-Tasser. Swiss model and I-Tasser are Protein Data Bank (PDB) dependent protein homology modeling databases. They use the known templates from PDB. Swiss Modeling is quicker than I-Tasser, however, I-Tasser produces better and more stable results. Swiss Modeling uses known sequences on the internet and generates data by comparing known structures. Both Swiss Modeling and I-Tasser have the advantage of understanding the main structure of the protein. Protein structure screening is a key factor to understanding interactions of proteins within themselves and with the environment. 3D modeled protein structures can be screened with commercially available tools like YASARA and Discovery Studio Tool. In these tools, not only protein structure is screened, but also domains of proteins can be separated, deleted and water molecules can be removed. UCSF Chimera and Autodock Tools also can be used for screening. After modeling, structures may be downloaded as PDB format and can be visualized by several programs: Discovery Studio, Autodock Vina, UCSF Chimera, etc. Interactions within the protein and secondary-tertiary structures can be obtained from these tools. All these tools work with a PDB file. There are other tools for protein structure determination and can be found at <https://www.click2drug.org/>. This website provides database links for distinct applications.

### 3. Importance of databases and protein structure determination in recombinant protein purification

Protein databases play a crucial role in bioinformatics and help to find information related to their research. In this way, all biological information becomes accessible through data mining tools saving time and resources. The first step in the study of a new protein is searching databases. Without the prior knowledge from such searches, previously known protein information could be missed, or an experiment could be repeated unnecessarily. There are hundreds of useful databases that can be used in protein research. However, in this study, the pI and 3D structure of the peptide were obtained using the Swiss database and ExPASy Database. The purpose of the Swiss Database is to make protein structure modeling accessible. Therefore, with this database, the 3D shape of the protein provided us with information to understand the structure of the peptide and its interactions. Additionally, ExPASy Database provides information about proteomics, post-translational modification prediction, primary, secondary, and tertiary structure analysis, sequence alignment, and pI of the protein [37]. The pI of the peptide provides the pH range where that peptide has a negative charge and prepares the buffer solution accordingly. Consequently, databases have key roles in biological research, and enormous data for protein structures, functions, and sequences can be generated by these available databases. These data offer essential information about our protein research as well. **Figure 1** provides the predicted structure for our research. SB#14 model indicates that the protein is formed from  $\beta$ -sheet structures.



**Figure 1.**

*Predicted structure of Sb#14. Sb#14 is a recombinant synthetic monoclonal antibody 14 used to detect spike protein of COVID-19 and used for immunodetection of the virus. SB#14 is modeled by the Swiss model to design purification steps for the TUSEB project.*

### 4. Structure- and design-based difficulties in recombinant protein purification

#### 4.1 Protein insolubility

Protein solubility is one of the most important protein properties and it can be defined as the protein concentration in a saturated solution that is in equilibrium with

a solid phase [38]. Not only some extrinsic factors, including pH, ionic strength, temperature, and some solvent additives, can affect the protein solubility but also several intrinsic factors influence protein solubility. Moreover, the amino acids on the protein surface are the primary intrinsic factors that impact protein solubility [39]. Several studies have revealed the relationship between protein solubility and sequence-derived characteristics. Wilkinson and Harrison et al. provided a simple approach for predicting protein solubility from the sequence, which was further refined by Davis et al. [40]. The average charge, which is derived by the relative quantities of Asp, Glu, Lys, and Arg residues, and the concentration of turn-forming residues are the two parameters used in their solubility model (Asn, Gly, Pro, and Ser). In addition, Christendat et al. have demonstrated that insoluble proteins had more hydrophobic stretches (more than 20 amino acids), less glutamine (Q 4%), fewer negatively charged residues (DE 17%), and a higher percentage of aromatic amino acids (FYW >7.5%) than soluble proteins [41]. The affinity tag (His/GST) in recombinant protein purified by affinity chromatography allows the protein to be purified. However, affinity tag has been observed to alter the biological activity of the protein. Because a minor difference affects protein solubility, the choice of affinity tag at the N- or C-terminus is important when expressing a protein domain. Klock and colleagues investigated a nested collection of 2143N- and C-terminal truncations from 96 targets and found significant variance in both solubility and aggregation processes by changing just a few amino acids in a protein length [42]. Therefore, it is essential to analyze which end of the protein is hidden. Furthermore, if the three-dimensional structure is known, the tag should be kept in a solvent-accessible end. In this way, the solubility of the protein can be increased. Insoluble proteins can aggregate during the expression process. That is why the different parameters should be optimized. On the other hand, during downstream purification steps, protein aggregation can occur. In these cases, developing a suitable and optimized purification procedure for each protein is critical. Sb#14 hydrophobic nature (**Figure 1**) leads to solubility problems as well as aggregation. The protein sticks to larger proteins and this led to difficulties in purification.

#### 4.1.1 Effects of imidazole on protein solubility

Imidazole is one of the most widely used organic compounds in protein affinity purification processes. It is used as a competitive agent to elute the histidine-tagged proteins. High concentrated imidazole that includes protein samples should be eliminated after eluting from the nickel column by dialysis [43]. In spite of all precautions, his tagged Sb#14 sticks to other proteins and has low solubility, therefore, SEC is used for protein purification rather than affinity purification. As mentioned, SEC separates proteins according to the molecular weight of the molecule. SEC performed with Superdex 75 size-exclusion column 10/30 (GE Healthcare, Princeton, NJ, USA). Moreover, SEC (15 cm length with r, 3 cm column) used in this study separates aggregates readily. Lower molecular weight of Sb#14 provides an advantage in the purification process, recalling that larger proteins elute first. This custom SEC column was unique as the resin resolution is high while the column length is relatively lower. The choice of resin and column size helped to resolve Sb#14 from bacterial lysate in a single purification step. The peptide (MW: 12.468 g/mol, pI: 8.91) is small and prone to aggregate. Therefore, the single-step purification blocks the self-cleavage of protein domains.

#### 4.1.2 Protein folding

The stability of the protein in various buffer compositions and pH levels with and without ligands should be determined. There are some useful websites for fold

recognition that can be used to predict the protein fold (PSI-BLAST and SEARCH). Some proteins are misfolded and require the addition of a cofactor, or ligand to restore proper folding and increase stability. For instance, beta-sheets are more prone to form amyloid-like aggregates if there are other binding partners that support protein stabilization and folding [44]. If the protein has a large number of beta-sheets, aggregation may be observed. This can be explained by the tendency of sticking together at Sb#14 and leading to the formation of insoluble aggregates. Tris-HCl buffer is used to stabilize Sb# 14.

#### 4.1.3 Reducing agents

To reduce aggregation, reducing agents such as dithiothreitol (DTT) may be used and added to the buffer. DTT is called Cleland's reagent and is used for protein reduction. However, a high concentration of DTT can reduce the nickel ion in the resin of the column. That is why the determination of the optimal concentration of DTT is essential.  $\beta$ -ME (Beta-mercaptoethanol) cleaves protein disulfide bonds (cysteine), and TCEP (Tris phosphine hydrochloride) can also be used as reducing agents, considering longer half-life  $\beta$ -ME. DTT reacts easily with nickel ions whereas  $\beta$ -ME reacts easily with cobalt, copper ions, and other phosphate buffers [44]. A precaution is required to obtain optimal conditions.

#### 4.1.4 Isoelectric point (pI) and pH

Each protein has a pI, where the protein's net charge is zero. Protein does not migrate at that point, and aggregation occurs [45]. On one hand, acidic proteins are likely to crystallize 0–2.5 pH units above their isoelectric point. On the other hand, basic proteins are more likely to crystallize 1.5–3 pH units below their pI. Hence, different pH values affect the protein's stability and solubility [46]. The pI of the peptide is important for us to know the pH range where that peptide has a negative charge and to prepare the buffer solution accordingly. That is why the pH of the buffer component is one of the most critical parameters. Sb #14 has a pI value of 8.91. This value set the pH parameter (pH:7.91) of the buffer used in the purification process.

### 4.2 Importance of protein isoelectric point in tagged protein purification

pI represents the pH level of a molecule where the net charge is zero. Amino acid composition of the protein can be used to calculate an estimated value with the help of databases [47]. If the pI is lower than the pH of a solution, protein will have a negative charge but if it is the opposite then the protein will have a positive charge. This feature can be used for purification purposes since it is a specific physicochemical parameter to distinguish between amphoteric molecules [39]. Also, it can be used to understand how solution pH can affect the protein stability in the pH range. So, buffers are used to keep proteins stable. To create an environment for protein to be stable, generally, the buffer is selected to have a pH level around the pI of the protein. If this difference between the pI of protein and pH of the solution gets larger then protein gets a greater net charge too. With this greater net charge, ionic compounds will be able to bind residues [48]. To avoid this unspecific interaction, the buffer's pH range should be selected accordingly to the protein's pI. And this knowledge of pH values with their effects on the proteins can be useful in the purification process. In tagged protein purification, affinity chromatography is a commonly used technique. The pH levels also affect this technique since affinity resins have their pH ranges to provide more stable links for not only the ligand

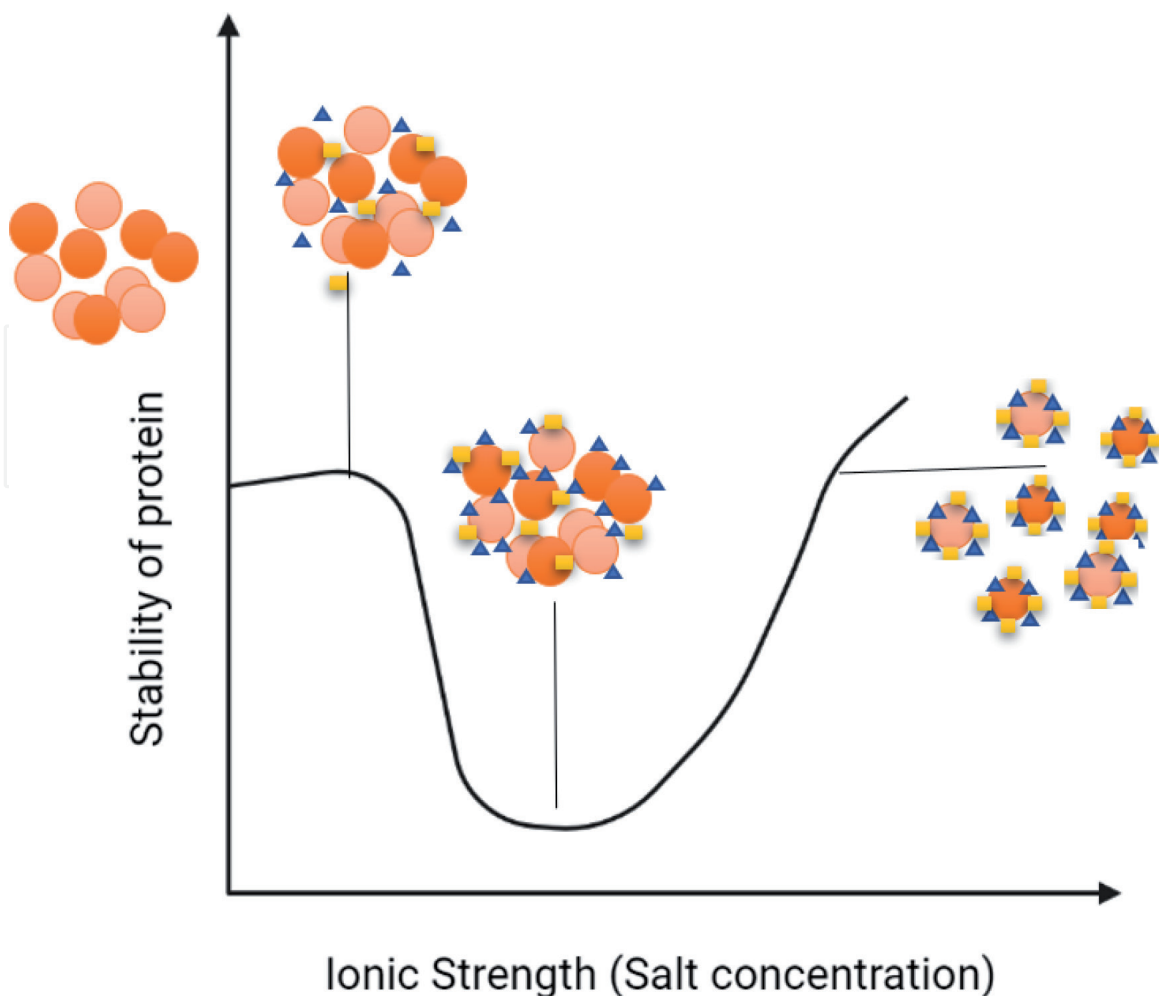
and the bead but also for the tag and the ligand. While making decisions about the purification protocol, the affinity resins' and the tags' working pH ranges should be kept in mind to create a better environment and more stable interaction. Also choosing affinity resins and tags that have a wider range of pH that they can work may be useful for the purification of proteins.

### 4.3 Protein aggregation and importance of ionic strength

Proteins are special structures that work with covalent and non-covalent interactions. They have cellular wide roles, including signaling, structural and metabolic processes. Their special structural features and 3D architecture determine their roles and interactions. These forms of proteins are determined by the amino acid sequences [49]. Proteins are not synthesized in their functional form. When their translation process is finished, the primary protein structure is formed. After that, they form alpha helices and beta sheets by hydrogen bonds. Alpha helices and beta sheets interact with each other with weak interactions and disulfide bonds and tertiary structure is formed. In some instances, the quaternary structure may be formed when tertiary structures interact and eventually in all cases functional protein forms [55]. However, in some cases, proteins can accumulate and form aggregates which may cause failure in protein purification experiments. Mostly, beta-sheets tend to interact with each other and accumulate. This event can be exemplified by amyloid aggregates in Alzheimer's disease. Recombinant protein aggregates resulted in the prevention of exposure of tags in tagged protein that causes failure in purification. Also, solubility prevents aggregate formation in proteins [50]. Protein aggregation can be prevented by adding salt to the proteins. Salt ions interact with the charged protein surface areas and prevent non-specific interactions, aggregation, and lower protein-protein interaction. However, a precaution is a must when preparing protein for binding experiments. Please note that high ionic concentration blocks ligand/protein binding experiments. As shown in **Figure 2**, Sb #14 is prone to aggregation and the process may be prevented/decreased through proper conditions.

### 4.4 Usage of additional agents to prevent protein accumulation, attachment, and insolubility

Urea dissolves the aggregated protein solutions. The efficiency of the process is increased by taking the necessary purification steps [1, 51]. Among these processes, protein dissolving and refolding steps constitute the most important steps for optimal protein activity and higher recovery. The protein precipitate is generally separated from other cellular components by low-speed centrifugation after cell lysis. Because protein aggregates are denser than cellular components, the lysate proteins are precipitated by centrifugation and then dissolved using detergents such as urea, guanidine-HCl, high concentrations of chaotropic denaturants, sodium N-lauroyl sarcosine, SDS, N-acetyl trimethyl ammonium chloride [52]. Further, additional reducing agents such as DTT, cysteine, Triton X-100,  $\beta$ -ME are used to dissolve inclusion bodies. These agents retain cysteine residues, minimizing the formation of false and unnatural disulfide bonds in the protein solution. Metal-containing oxidation of cysteine is prevented by using chelating agents such as EDTA in dissolution buffers [44, 52]. By removing the soluble protein content, removing the chaotropic reagents, and diluting them directly into the renaturation buffer, the recombinant proteins are folded back into their native form [44]. Protein collapse is a higher-order reaction while protein folding is a lower-order reaction. Therefore, the aggregation rate is higher than the folding rate. Due to the kinetic



**Figure 2.**

*Ionic strength is important for the separation of proteins from each other that can be aggregated. The strength of ions resists accumulation by preventing protein-protein interaction.*

competition that occurs, the increase in protein concentration decreases the folding efficiency of the protein. For accurate and efficient folding kinetics, the preferred protein concentration is used in the range of  $10\text{--}50\ \mu\text{g}\cdot\text{ml}^{-1}$  [1, 53]. As explained in the section of ‘Disulfide bond formation’, recombinant proteins with multiple disulfide bonds in their structure tend to be in a correct folding process in the presence of both oxidizing and reducing agents for the formation of these bonds. The simplest way for oxidation is to oxidize the protein with air in the presence of a metal catalyst. Another common oxidation option is the addition of thiol agents containing compounds such as glutathione, cysteine, cysteamine to the protein mixture. The most commonly used thiol reagents are reduced/oxidized glutathione (GSH-GSSH), cysteine/cystine, DTT/GSSH, cysteamine compounds [1, 36]. There are also low-molecular-weight additives that help refolding process. There are studies on the use of additives such as acetone, DMSO, short-chain alcohols, PEG in the bioactive protein process. In addition, it has been observed that L-arginine/HCl reduces aggregation on protein. The  $0.4\text{--}1\ \text{M}$  arginine used in the studies also increases the protein folding efficiency by reducing the aggregation in the recombinant protein solution. This feature of arginine has been attributed to the interaction of the guanidino structure in its structure with tryptophan residues in proteins [44, 52–55]. Sb #14 was also treated with DTT but when overexpressed, the protein has solubility problems. The structure is highly prone to aggregation and solubility may be increased upon co-expressing chaperones/Heat Shock Proteins or yeast systems seem proper for preventing aggregation. Yet, this may not solve the problem but mutational studies may provide more soluble and stable structures.

## 5. Conclusion

Protein purification depends on several factors: resin type, solvent, ionic strength, pH, protein structural tendency to aggregation, buffer systems, protein structure, ligand if any, column dimension. For each factor, problems may be encountered. To eliminate these problems and decide on protein purification protocol, protein structural properties must be examined initially. Tandem purification steps may also increase the purification yield. However, self-cleavage of certain proteins or oxidation that may distort the protein function leads to problems. Therefore, several distinct protocols may be tested before purifying the targeted protein with high efficiency and functionality. Sb#14 structure mainly consists of  $\beta$ -sheets and overexpressing this petit protein lead aggregation. Solubility is another problem in the cellular milieu as Sb#14 hydrophobic nature interacts with other proteins in the lysate. Therefore, proper solvent selection (phosphate buffer) and adjusting the pH (1 unit lower than pI, 7.91) provide soluble protein. Further, we take advantage of the protein's lower molecular weight and employed a convenient resin (Superdex 75-separates 3000–70,000 molecular weights, most of the lysate elutes before Sb#14) and custom size column (15 cm length with r: 3 cm-lower pressure yet increase resolution). The purification was performed with high yield by AKTA go FPLC system. Additionally, co-expressing heat shock proteins with this type of protein may help in folding and dissolving aggregates. All these conditions must be tested for individual proteins for optimum purification yield.

## Acknowledgements

Prof. Dr. Yusuf TUTAR acknowledges grant from TUSEB (Project # 8970-220-CV-01) and infrastructure grant from the University of Health Sciences-Turkey (Project #2017-041). Sb#14 is from Addgene (#153522).

IntechOpen

IntechOpen

## Author details

Kubra Acikalin Coskun<sup>1</sup>, Nazlıcan Yurekli<sup>2</sup>, Elif Cansu Abay<sup>2</sup>, Merve Tutar<sup>3</sup>,  
Mervener Al<sup>4</sup> and Yusuf Tutar<sup>5,6,7\*</sup>

1 Faculty of Medicine, Division of Medicinal Biology, Department of Basic Sciences,  
Istanbul Aydin University, Istanbul, Turkey

2 Department of Molecular Biology and Genetics, Arel University, Istanbul, Turkey

3 Department of Molecular Biology and Genetics, Bilkent University, Ankara,  
Turkey

4 Hamidiye Faculty of Medicine, Division of Medicinal Biochemistry, Department  
of Basic Sciences, University of Health Sciences-Turkey, Istanbul, Turkey

5 Hamidiye Faculty of Pharmacy, Division of Biochemistry, Department of Basic  
Pharmaceutical Sciences, University of Health Sciences-Turkey, Istanbul, Turkey

6 Division of Molecular Oncology, Hamidiye Health Sciences Institutes, University  
of Health Sciences-Turkey, Istanbul, Turkey

7 Validebağ Research Center, University of Health Sciences-Turkey, Istanbul,  
Turkey

\*Address all correspondence to: [yusuf.tutar@sbu.edu.tr](mailto:yusuf.tutar@sbu.edu.tr)

## IntechOpen

© 2022 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms  
of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium,  
provided the original work is properly cited. 



## References

- [1] Singh SM, Panda AK. Solubilization and refolding of bacterial inclusion body proteins. *Journal of Bioscience and Bioengineering*. 2005;**99**(4):303-310. DOI: 10.1263/jbb.99.303
- [2] Stollar EJ, Smith DP. Uncovering protein structure. *Essays in Biochemistry*. 2020;**64**(4):649-680. DOI: 10.1042/EBC20190042
- [3] Buccitelli C, Selbach M. mRNAs, proteins and the emerging principles of gene expression control. *Nature Reviews Genetics*. 2020;**21**(10):630-644. DOI: 10.1038/s41576-020-0258-4
- [4] Baneyx F. Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology*. 1999;**10**(5):411-421. DOI: 10.1016/S0958-1669(99)00003-8
- [5] Andersen DC, Andersen DC, and Krummen L. Recombinant protein expression for therapeutic applications. *Current Opinion in Biotechnology*. May 2002;**1669**:117-123
- [6] Rosano GL and Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Frontiers in Microbiology*. 2014;**5**:1-17. DOI: 10.3389/fmicb.2014.00172
- [7] Peleg Y, Prabakar V, Bednarczyk D, Unger T. Heterologous Gene Expression in *E. coli*. *Current Opinion in Biotechnology*. 2017;**1586**:33-43. DOI: 10.1007/978-1-4939-6887-9
- [8] Schumann W, Ferreira LCS. Production of recombinant proteins in *Escherichia coli*. *Genetics and Molecular Biology*. 2004;**27**(3): 442-453. DOI: 10.1590/S1415-47572004000300022
- [9] de Marco A. Recombinant expression of nanobodies and nanobody-derived immunoreagents. *Protein Expression and Purification*. 2020;**172**:105645. DOI: 10.1016/j.pep.2020.105645
- [10] Demain AL, Vaishnav P. Production of recombinant proteins by microbes and higher organisms. *Biotechnology Advances*. 2009;**27**(3):297-306. DOI: 10.1016/j.biotechadv.2009.01.008
- [11] Sezonov G, Joseleau-Petit D, D'Ari R. *Escherichia coli* physiology in Luria-Bertani broth. *Journal of Bacteriology*. 2007;**189**(23):8746-8749. DOI: 10.1128/JB.01368-07
- [12] Lee C, Kim J, Shin SG, Hwang S. Absolute and relative QPCR quantification of plasmid copy number in *Escherichia coli*. *Journal of Biotechnology*. 2006;**123**(3):273-280. DOI: 10.1016/j.jbiotec.2005.11.014
- [13] Del Solar G, Espinosa M. Plasmid copy number control: An ever-growing story. *Molecular Microbiology*. 2002;**37**(3):492-500. DOI: 10.1046/j.1365-2958.2000.02005.x
- [14] Browning DF, Godfrey RE, Richards KL, Robinson C, Busby SJW. Exploitation of the *Escherichia coli* lac operon promoter for controlled recombinant protein production. *Biochemical Society Transactions*. 2019;**47**(2):755-763. DOI: 10.1042/BST20190059
- [15] Lalwani MA et al. Optogenetic control of the lac operon for bacterial chemical and protein production. *Nature Chemical Biology*. 2021;**17**(1): 71-79. DOI: 10.1038/s41589-020-0639-1
- [16] Peubez I et al. Esters of dicarboxylic acids as additives for lubricating oils. *Tribology International*. 2006;**39**(6):560-564. DOI: 10.1016\_j.triboint.2005.06.001

- [17] Kimple ME, Brill AL, and Pasker RL. Overview of affinity tags for protein purification. *Current Protocols in Protein Science*. 2013;**73**:608-616. DOI: 10.1002/0471140864.ps0909s73
- [18] Terpe K. Overview of tag protein fusions: From molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology*. 2003;**60**(5):523-533. DOI: 10.1007/s00253-002-1158-6
- [19] Walls D, Walker JM. Protein Chromatography. *Protein Chromatography*. 2017;**1485**:423. DOI: 10.1007/978-1-4939-6412-3
- [20] Costa SJ, Coelho E, Franco L, Almeida A, Castro A, Domingues L. The Fh8 tag: A fusion partner for simple and cost-effective protein purification in *Escherichia coli*. *Protein Expression and Purification*. 2013;**92**(2):163-170. DOI: 10.1016/j.pep.2013.09.013
- [21] Costa S, Almeida A, Castro A, and Domingues L. Fusion tags for protein solubility, purification, and immunogenicity in *Escherichia coli*: The novel Fh8 system. *Frontiers in Microbiology*. 2014;**5**(FEB):63. DOI: 10.3389/fmicb.2014.00063
- [22] Chang HM, Yeh ETH. Sumo: From bench to bedside. *Physiological Reviews*. 2020;**100**(4):1599-1619. DOI: 10.1152/physrev.00025.2019
- [23] Dumon-Seignovert L, Cariot G, Vuillard L. The toxicity of recombinant proteins in *Escherichia coli*: A comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expression and Purification*. 2004;**37**(1):203-206. DOI: 10.1016/j.pep.2004.04.025
- [24] Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification*. 2005;**41**(1):207-234. DOI: 10.1016/j.pep.2005.01.016
- [25] Jäger VD et al. Catalytically-active inclusion bodies for biotechnology—General concepts, optimization, and application. *Applied Microbiology and Biotechnology*. 2020;**104**(17):7313-7329. DOI: 10.1007/s00253-020-10760-3
- [26] Messens J, Collet JF. Pathways of disulfide bond formation in *Escherichia coli*. *International Journal of Biochemistry and Cell Biology*. 2006;**38**(7):1050-1062. DOI: 10.1016/j.biocel.2005.12.011
- [27] Thomas JG, Baneyx F. ClpB and HtpG facilitate de novo protein folding in stressed *Escherichia coli* cells. *Molecular Microbiology*. 2000;**36**(6):1360-1370. DOI: 10.1046/j.1365-2958.2000.01951.x
- [28] Sørensen HP, Mortensen KK. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *Journal of Biotechnology*. 2005;**115**(2):113-128. DOI: 10.1016/j.jbiotec.2004.08.004
- [29] Campani G, Gonçalves da Silva G, Zangirolami TC, Perencin de Arruda Ribeiro M. Recombinant *Escherichia coli* cultivation in a pressurized airlift bioreactor: Assessment of the influence of temperature on oxygen transfer and uptake rates. *Bioprocess and Biosystems Engineering*. 2017;**40**(11):1621-1633. DOI: 10.1007/s00449-017-1818-7
- [30] Ferrer M, Lünsdorf H, Chernikova TN, Yakimov M, Timmis KN, Golyshin PN. Functional consequences of single:Double ring transitions in chaperonins: Life in the cold. *Molecular Microbiology*. 2004;**53**(1):167-182. DOI: 10.1111/j.1365-2958.2004.04077.x
- [31] Wingfield PT. Overview of the purification of recombinant proteins. *Current Protocols in Protein Science*. 2015;**80**:6.1.1-6.1.35. DOI: 10.1002/0471140864.ps0601s80

- [32] Berg JM, Tymoczko JL, Stryer L. Section 4.1, The purification of proteins is an essential first step in understanding their function. In: *Biochemistry*. 5th ed. New York: W H freeman; 2002. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK22410/>
- [33] Harcum S. Purification of protein solutions. *Biologically Inspired Textiles*. 2008;**1**:26-43. DOI: 10.1533/9781845695088.1.26
- [34] Righetti PG. Electrophoresis| Isoelectric focusing. *Encyclopedia of Analytical Science*. 2005;**1079**:382-392. DOI: 10.1016/b0-12-369397-7/00124-2
- [35] Tramontano A. *Function Prediction*. Oxfordshire, United Kingdom: Taylor & Francis Group; 2005. pp. 45-67. DOI: 10.1201/9781420035001.ch3
- [36] Factors R. Chapter 11: Orthodontic Treatment of Class III Malocclusion. Vol. 1990. Amsterdam, Netherlands: Elsevier; 1999. p. 306. DOI: 10.1007/978-1-4939-7315-6
- [37] Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*. 2018;**46**:W296-W303
- [38] Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*. 2003;**31**(13):3784-3788. DOI: 10.1093/nar/gkg563. PMID: 12824418; PMCID: PMC168970.
- [39] Kramer RM, Shende VR, Motl N, Pace CN, Scholtz JM. Toward a molecular understanding of protein solubility: Increased negative surface charge correlates with increased solubility. *Biophysical Journal*. 2012;**102**(8):1907-1915. DOI: 10.1016/j.bpj.2012.01.060
- [40] Trevino SR, Scholtz JM, Pace CN. Amino acid contribution to protein solubility: Asp, Glu, and ser contribute more favorably than the other hydrophilic amino acids in RNase Sa. *Journal of Molecular Biology*. 2007;**366**(2):449-460. DOI: 10.1016/j.jmb.2006.10.026
- [41] Davis GD, Elisee C, Mewham DM, Harrison RG. New fusion protein systems designed to give soluble expression in *Escherichia coli*. *Biotechnology and Bioengineering*. 1999;**65**(4):382-388. DOI: 10.1002/(SICI)1097-0290(19991120)65:4<382::AID-BIT2>3.0.CO;2-I
- [42] Smialowski P, Martin-Galiano AJ, Mikolajka A, Girschick T, Holak TA, Frishman D. Protein solubility: Sequence based prediction and experimental verification. *Bioinformatics*. 2007;**23**(19):2536-2542. DOI: 10.1093/bioinformatics/btl623
- [43] European Commission EUROSTAT. Glossary:Special-purpose entity (SPE). *Nature Methods*. 2008;**5**(2):135-146. DOI: 10.1038/nmeth.f.202.Protein
- [44] Bhat E, Abdalla M, Rather I. Key Factors for Successful Protein Purification and Crystallization. *Global Journal of Biotechnology and Biomaterial Science*. 2018;**4**(1):001-007. DOI: 10.17352/gjbbs.000010
- [45] Novák P, Havlíček V. Protein extraction and precipitation. In: *Proteomic Profiling and Analytical Chemistry: The Crossroads*. Second ed. Amsterdam, Netherlands: Elsevier; 2016. pp. 52-62. DOI: 10.1016/B978-0-444-63688-1.00004-5
- [46] Friedman DB, Hoving S, Westermeier R. Chapter 30 isoelectric focusing and two-dimensional gel electrophoresis. *Methods in Enzymology*. 2009:515-540. DOI: 10.1016/s0076-6879(09)63030-5
- [47] Lee K. Protein Folding. *Genome Biology*. 2001;**2**(1):1-20. DOI: 10.1186/gb-spotlight-20010313-01

[48] Dobson C. Protein folding and misfolding. *Nature*. 2003;**426**:884-890.  
DOI: 10.1038/nature02261

[49] Roberts CJ. Protein aggregation and its impact on product quality. *Current Opinion in Biotechnology*. Dec 2014;**30**:211-217. DOI: 10.1016/j.copbio.2014.08.001. Epub 2014 Aug 28. PMID: 25173826; PMCID: PMC4266928

[50] Bao RM, Yang HM, Yu CM, Zhang WF, Tang JB. An efficient protocol to enhance the extracellular production of recombinant protein from *Escherichia coli* by the synergistic effects of sucrose, glycine, and triton X-100. *Protein Expression and Purification*. 2016;**126**:9-15.  
DOI: 10.1016/j.pep.2016.05.007

[51] Fahnert B. Using folding promoting agents in recombinant protein production: A review. *Recombinant Gene Expression*. 2012;**824**:3-36

[52] Cline DJ, Redding SE, Brohawn SG, Psathas JN, Schneider JP, Thorpe C. New water-soluble phosphines as reductants of peptide and protein disulfide bonds: Reactivity and membrane permeability. *Biochemistry*. 2004;**43**(48):15195-15203.  
DOI: 10.1021/bi048329a

[53] Kumar A et al. Optimization and efficient purification of recombinant Omp28 protein of *brucella melitensis* using triton X-100 and  $\beta$ -mercaptoethanol. *Protein Expression and Purification*. 2012;**83**(2):226-232.  
DOI: 10.1016/j.pep.2012.04.002

[54] Mayer M, Buchner J. Refolding of inclusion body proteins. *Methods in Molecular Medicine*. 2004;**94**:239-254.  
DOI: 10.1385/1-59259-679-7:239

[55] Gunner MR, Mao J, Song Y, Kim J. Factors influencing the energetics of electron and proton transfers in proteins. What can be learned from calculations. *Biochimica Et Biophysica Acta (BBA) – Bioenergetics*. 2006;**1757**(8):942-968.  
DOI: 10.1016/j.bbabi.2006.06.005