

VILNIUS UNIVERSITY

Life Sciences Center



Molecular biology master studies 2nd year student

Laurynas KARPUS

Master's thesis

Development of high-throughput *in vivo* protein solubility screening system

Supervisor:


Dr. Linas Mažutis

Vilnius, 2020 m.

**Development of high-throughput *in vivo*
protein solubility screening system**

Thesis work was done in Sector of Microtechnologies, Institute of Biotechnology,
Vilnius University


Laurynas Karpus



.....

Supervisor of the work:

Allowed to defend:
Dr. Linas MAŽUTIS



.....

CONTENT

ABBREVIATIONS.....	5
INTRODUCTION.....	6
1. LITERATURE REVIEW.....	8
1.1. Solubility of proteins.....	8
1.1.1. Extrinsic properties affecting solubility.....	8
1.1.2. Intrinsic properties affecting solubility.....	12
1.2. Methods for measuring protein solubility.....	14
1.2.1. Addition of lyophilized protein.....	14
1.2.2. Concentration by ultrafiltration.....	14
1.2.3. Induction of amorphous precipitation.....	14
1.2.4. Antibody blot based screening.....	16
1.2.5. Split β -galactosidase based screening.....	17
1.2.6. Split GFP based screening.....	20
1.3. Droplet microfluidics.....	21
1.3.1. Droplet microfluidics in microbiology.....	21
1.3.2. Bacterial growth in agarose beads.....	23
2. MATERIALS AND METHODS.....	25
2.1. Materials and equipment.....	25
2.1.1. Reagents.....	25
2.1.2. Enzymes, solutions and kits.....	25
2.1.3. Bacterial strains and media.....	26
2.1.4. Oligonucleotides.....	26
2.1.5. Equipment.....	27
2.2. Methods.....	27
2.2.1. Production of PDMS chip.....	27
2.2.2. Competent cell preparation and transformation.....	29
2.2.3. Plasmid vector preparation.....	30
2.2.4. Insert preparation.....	31
2.2.5. Ligation and isolation of the constructed plasmids.....	33
2.2.6. Sequencing.....	33
2.2.7. Control protein solubility assessment.....	33
2.2.8. Preparation of testing samples.....	34
2.2.9. High-throughput system in bulk.....	34

2.2.10. High-throughput system in agarose beads	35
3. RESULTS.....	38
3.1. High-throughput solubility detection system design.....	38
3.2. Preparation of the insoluble and soluble protein variants	40
3.3. Validation and optimization of HT system in bulk	41
3.4. Validation of HT solubility system in agarose beads.....	44
3.5. Discussion	48
CONCLUSIONS.....	51
ABSTRACT	52
SANTRAUKA	53
ACKNOWLEDGEMENT	54
REFERENCES.....	55

ABBREVIATIONS

AHT – anhydrotetracycline

DNA – deoxyribonucleic acid

DTT – dithiothreitol

EA – polyethylenoxide-perfluoropolyether block-copolymer

eNDP-K – mutant of nucleoside diphosphate kinase

FACS – fluorescence activated cell sorting

GFP – green fluorescent protein

HFE-7500 – commercial name of fluorinated oil

HT – high-throughput

IPTG – isopropyl β -D-1-thiogalactopyranoside

LB – lysogeny broth

MBP – maltose binding protein

NGS – next generation sequencing

PAA – polyacrylamide

PCR – polymerase chain reaction

PDMS – polydimethylsiloxane

PFO – 1H,1H,2H,2H-perfluoro-1-octanol

pI – isoelectric point

RFU – relative fluorescence units

RNA – ribonucleic acid

v/v – volume per volume

w/v – weight per volume

wtNDP-K – wild type nucleoside diphosphate kinase

β -Gal – beta-galactosidase

INTRODUCTION

Heterologous expression of functional proteins is one of the cornerstones of modern biotechnology. Yet, in practice many important proteins are not effectively expressed in a soluble form (Endo & Kurusu, 2007). Sufficient solubility of proteins is immensely important to pharmaceutical industry (Caldwell et al., 2001; Fowler et al., 2005; Ricci & Brems, 2004) and structural biologists (Bagby et al., 2001; Riès-Kautt & Ducruix, 1997; C. H. Schein, 1993), wherein high concentration protein samples are often required. Generally, most scientists who work with proteins in a solution require them to be soluble, as there is a strong correlation between protein's solubility and the likelihood of that protein being functional (Vojdani, 1996).

A number of human diseases are also connected to protein solubility problems (Evans et al., 2004; W. Kim & Hecht, 2006; Pande et al., 2005), such as Alzheimer's (Dobson, 1999; Meyer et al., 2019; Thomas et al., 1995), Parkinson's (Enogieru et al., 2019) and mad cow disease (Anand et al., 2018). Most often these conditions are caused by either misfolding of proteins (Brown et al., 1997; Rao et al., 1994; Thomas et al., 1992) or by their aberrant processing, leading to the formation of aggregation-prone and protease-resistant products (Bruijn et al., 1998; Colon & Kelly, 1992; Davies et al., 1997; Galvin et al., 1999; Hind et al., 1983; Kawaguchi et al., 1994; Macdonald, 1993; McPhaul, n.d.; Spada et al., 1991). Therefore, a good understanding of the factors that influence protein solubility is an important area of research.

In biotechnology, protein solubility can often be improved by optimizing the amino acid sequence of the target protein (Huang et al., 1996) or genetic background of the expressing cells (Blackwell & Horgan, 1991; Bourot et al., 2000; Brown et al., 1997; Sugihara & Baldwin, 1988; Wynn et al., 1992). Yet, currently used protein solubility assays are slow and tedious, making the screening for improved protein variants and conditions yielding improved solubility highly inefficient (Knaust & Nordlund, 2001).

Collecting experimental data that describes how the protein amino acid sequences affects the protein's solubility *in vivo* is becoming extremely important. Such information can be used in the recently demonstrated, promising machine learning applications (Han et al., 2019.; Khurana et al., 2018), which aim to determine protein's solubility degree solely from their primary amino acid sequence. However, to perform well these algorithms require high amounts of currently unavailable data that connects protein sequence to its solubility.

Perhaps the most advanced system for solubility data collection is a method by Cabantous & Waldo (2006) based on split-GFP complementation *in vivo*. Nonetheless, it is limited by the

slow manual bacterial colony analysis for each unique protein variant of interest, making this method effectively low-throughput. Yet, successful developments of droplet microfluidic based high-throughput assays provide an example of how low-throughput approaches can be converted into effective high-throughput methods (Caen et al., 2018; Du et al., 2016; Lan et al., 2017; E. Y. Liu et al., 2018; Pham et al., 2017).

Combining low-throughput GFP complementation system and droplet microfluidics, this work describes the development and experimental testing of a high-throughput *in vivo* protein solubility determination system based on micrometer scale agarose beads and previously described GFP complementation system, that would enable efficient solubility screening and sorting for large protein libraries.

Therefore, the main **objective** of the work is:

To develop a proof of concept system for high-throughput *in vivo* protein solubility determination based on micron scale agarose beads.

To achieve this objective, the following **goals** were set:

1. Develop theoretical high-throughput solubility system design.
2. Select and characterize soluble and insoluble protein variants for the testing of the designed system.
3. Demonstrate the performance of the system experimentally in large volume (bulk) format.
4. Demonstrate the performance of the system experimentally using agarose beads.

1. LITERATURE REVIEW

1.1. Solubility of proteins

Protein solubility has several definitions, as the proteins, in an aqueous medium, can form true or colloidal solution or insoluble particles suspension (Arakawa & Timasheff, 1985). Thermodynamically, protein solubility is the protein concentration in a solvent in a simple or two-phase system (Vojdani, 1996). However, mathematically the protein solubility degree of a protein is the amount of protein present in liquid phase in relation to the total amount of protein in liquid and solid phases in balance. Finally, perhaps the common definition used in practice is the protein retention in the supernatant after the solution centrifugation for certain time period and under certain force centrifuge (Smith, 2000).

The solubility of a protein can be influenced by a number of extrinsic and intrinsic factors. Extrinsic factors that influence protein solubility include the presence of various solvent additives, ionic strength, temperature and pH (Arakawa & Timasheff, 1985; Riès-Kautt & Ducruix, 1997). However, solution condition alteration isn't always appropriate and/or sufficient to increase protein solubility to the required extent. On the other hand, the intrinsic factors that influence protein solubility are defined primarily by the protein's surface amino acids, yet a concrete understanding on how intrinsic properties of the protein can be altered in order to increase its solubility is still lacking (Sormanni & Vendruscolo, 2019; Touati et al., 1992).

In the following section, such parameters of protein solubility will be described in greater detail.

1.1.1. Extrinsic properties affecting solubility

Protein solubility is defined by interactions between the protein and solvent compounds, which are mostly solubilized in buffered water (Pace et al., 2004). In such solvent-solute system, water acts as the solvent via hydrogen bonds (Pace, Nick Pace, Fu, et al., 2014). However, in practice the solution is usually a ubiquitous buffer of biological solutions. Different compounds, such as the protein, water, buffer, crystallizing agents, interact with each other via multiple, often weak, types of interactions: H-Bonds, van der Waals, hydrophobic, dipole-dipole, monopole-monopole and monopole-dipole interactions (Dahal & Schmit, 2018). Monopole-monopole interactions, such as salt bridges, are affected by changes in pH and the presence of nonaqueous solvents and salts (Long & Labute, 2010). Monopole-dipole interactions, such as those between ions and water, are sensitive to the presence of compounds bearing polar groups (Lai et al., 2017).

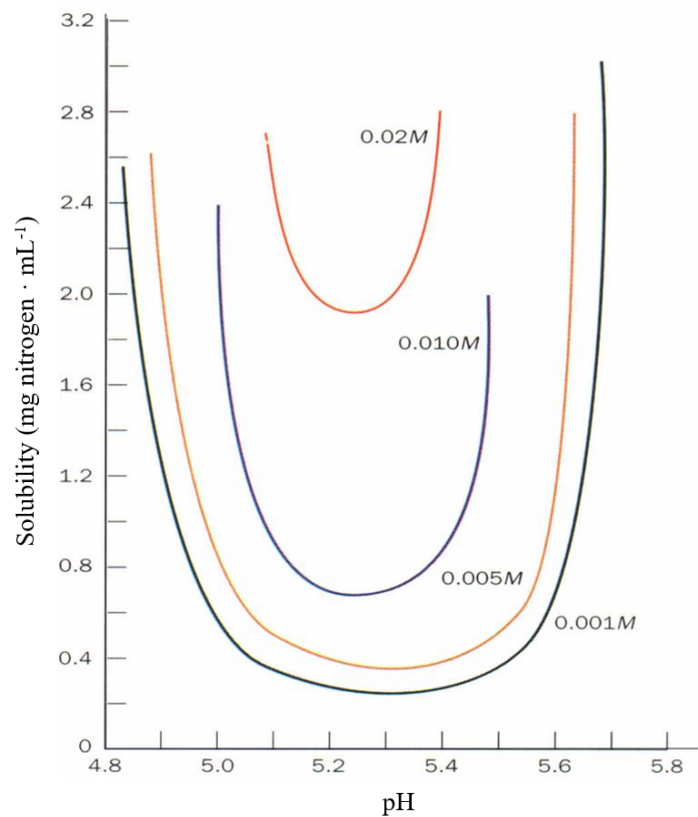


Figure 1.1 | Solubility of β -lactoglobulin as a function of pH at several different NaCl concentrations. Illustration adapted from Fox & Foster (1957).

Van der Waals hydrophobic interactions are of importance for intramolecular interactions promoting the tertiary structure of the protein (Pace, Nick Pace, Martin Scholtz, et al., 2014).

A. pH of the medium

The pH of the medium of one of the protein solubility determining factors. Generally, the degree of protein solubility is the result of electrostatic and hydrophobic interactions between proteins. The solubility is increased if electrostatic repulsion between the protein molecules is higher than the hydrophobic interactions. To become soluble, proteins should maximize the interaction with the solvent. In addition, at the isoelectric point (pI), proteins have a net zero charge, the attractive forces predominate and molecules tend to associate, resulting in a decrease of protein solubility (Kakalis & Regenstein, 1986; Singh & Ye, 2008; Vojdani, 1996). However, above the pI, the net charge is negative and the solubility is increased. Protein-water interactions increase at pH values higher or lower than the pI because protein carries positive molecules at pH values not far from pI.

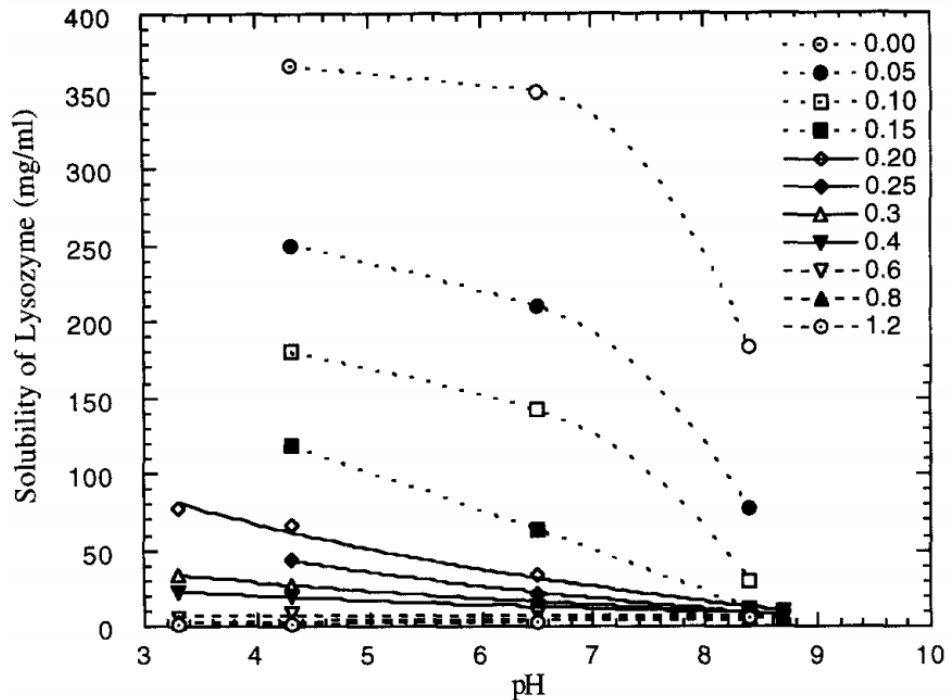


Figure 1.2 | Solubility of lysozyme as a function of pH, at 18°C and various NaCl concentrations indicated in the units of molarity. Illustration adapted from Guilloteau et al. (1992).

In such cases, minimal interactions with water are observed and protein aggregation occurs. Therefore, a U-shaped curve can be obtained experimentally if protein's solubility is plotted as a function of pH, wherein the minimum of the curve then corresponds to the pI (fig. 1.1).

For most of the naturally occurring proteins, their pI values are in the range of 3.5 and 6.5. For example, in the absence of a buffer solution, solubility of lysozyme was studied for a broad pH, ranging from 3 to 9 (Guilloteau et al., 1992). As shown in figure 1.2, the solubility of lysozyme has an inverse relationship with pH values, in other words, when approaching pI – at low ionic strength. At 0.6 M NaCl concentration, lysozyme solubility is nearly insensitive to pH variation, as seen from the flat solubility curve. However, at high NaCl concentrations (0.8-1.2 M), lysozyme solubility varies in the opposite direction, increasing when approaching pI value. This has been previously mentioned in the literature by Cacioppo & Pusey (1991). Evidently, pH is an important determinant of protein solubility, which also acts dynamically depending on the salt concentration of the solution.

B. Ionic strength of the solvent

The solubility of proteins is likely to increase in the presence of salt – such a process is called salting-in (Regenstein & Regenstein, 1984). This phenomenon is explained by the recently

generalized version of the Debye-Huckel theory (J.-L. Liu & Li, 2019). As salt concentration slightly increases, the proteins are being surrounded by salt counterions (ions of opposite net charge), which results in decreasing free energy of the protein and increasing activity of the solvent, which in turn leads to increased solubility (fig. 1.3). This theory also predicts that the logarithm of solubility is proportional to the square root of the ionic strength (Xiao & Song, 2017). However, at high concentrations of salt, the solubility of the proteins drop sharply and proteins can precipitate out - such a process is called salting-out (Foster et al., 1976).

For example, solubility data of carboxyhemoglobin showed that with increasing salt concentration, protein solubility also increases at first, but then decreases at high ionic strength (Catherine H. Schein, 1990).

C. Temperature

Temperature is another important factor that can influence the solubility of a protein. Generally, natural protein solubility is increased at temperatures between 40°C and 50°C. However, at high temperatures, protein solubility decreases (Kim & Lund, 1998; Langendorff et al., 1999; Mine, 1995).

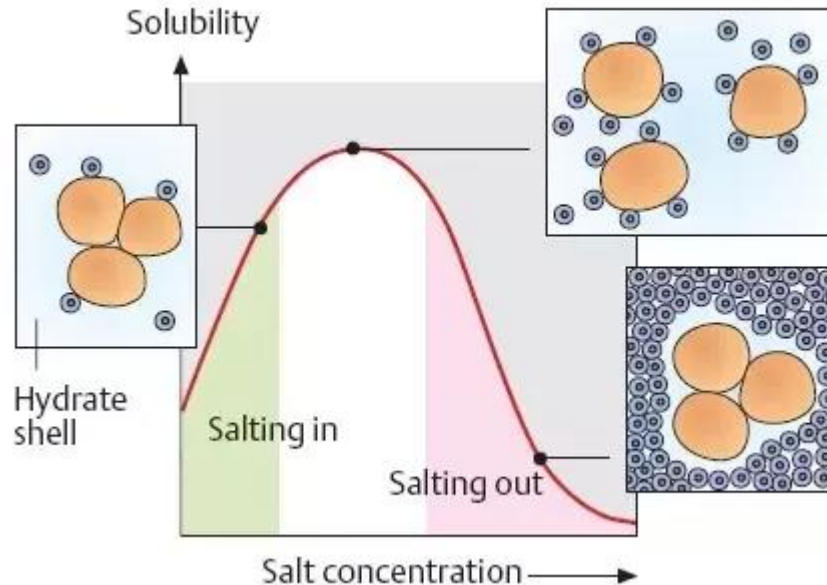


Figure 1.3 | Salting-in and -out processes described by the Debye-Huckel theory. Initially, increasing salt concentration increases protein solubility due to the counterion screening of the protein and increased activity of the solvent (salting-in). At high salt concentrations the solubility of proteins drops sharply and they tend to precipitate out (salting-out). Illustration adapted from Kalra et al. (2001).

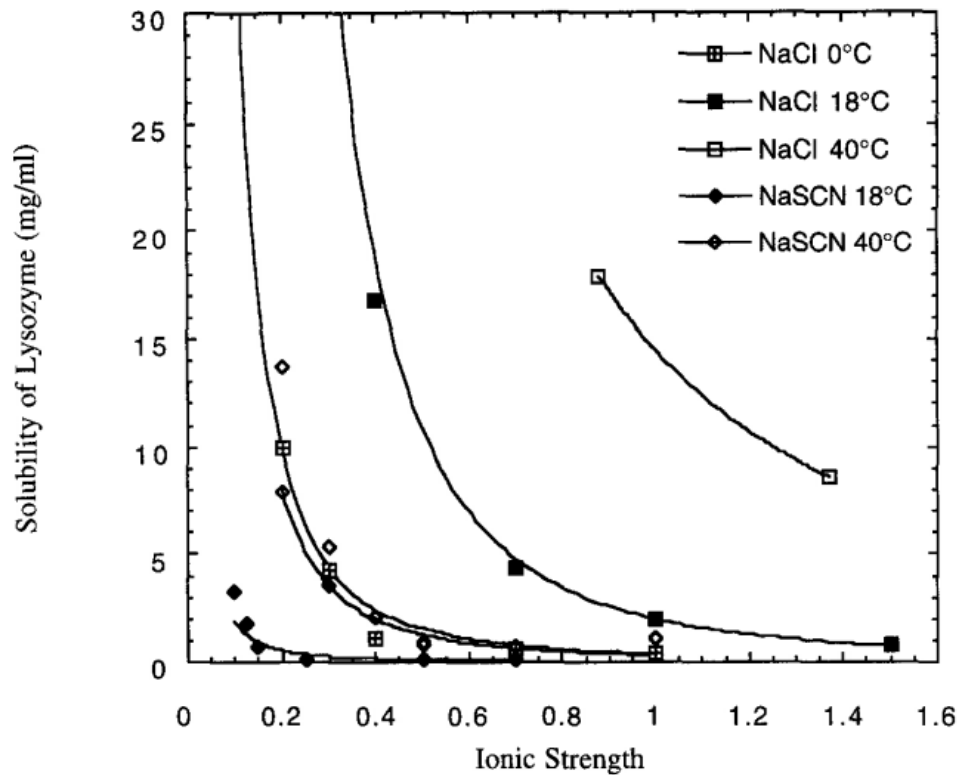


Figure 1.4 | Solubility of lysozyme as a function of ionic strength for various temperatures in the presence of NaCl and NaSCN, at pH 4.5. Illustration adapted from Guilloteau et al. (1992).

This is explained by the fact that at high temperatures proteins are denatured by the effect of temperature on the non-covalent bonds (hydrogen, hydrophobic, electrostatic). As the secondary and tertiary structures become unfolded, the hydrophobic groups of protein molecules interact with each other and, in turn, reduce the binding of the water. Such hydrophobic interactions eventually lead to aggregation and precipitation. In other words, the solubility decreases. By analyzing the combined effects of temperature and ionic strength, it was determined that the solubility is more sensitive to a change of temperature at low ionic strength (Howard et al., 1988). In the case of lysozyme, at pH 4.5 (Guilloteau et al., 1992), solubility strongly increases with temperature for NaCl concentrations lower than 0.3-0.5 M (fig. 1.4).

1.1.2. Intrinsic properties affecting solubility

Multiple researchers have described the significant relationship between protein's amino acid sequence derived features and its solubility (Doi et al., 2005; Trainor et al., 2017). The first simple method for calculating the protein's solubility from its sequence was proposed by Wilkinson & Harrison (1991) and then improved by Davis et al. (1999). The model by Wilkinson and Harrison was based on a pair of parameters. First, the average charge determined by the relative numbers of

aspartate, glutamine, lysine and arginine amino acid residues. Second, the amount of the turn-forming asparagine, glycine, proline and serine residues.

Furthermore, Christendat et al. (2000) have examined experimental success (soluble) and failure (insoluble) data accumulated in a high-throughput structural genomics project on multiple non-membrane proteins from *Methanobacterium thermoautotrophicum*. This research has demonstrated that insoluble proteins tended to have more hydrophobic stretches (20 amino acids or longer), fewer negatively charged residues and higher percentage of aromatic amino acids than soluble ones and a lower glutamine content. Following these observations, a set of simple rules were derived that allowed to predict protein solubility from its primary sequence with a 65% accuracy (for the same set of proteins). Afterwards, Bertone et al. (2001) have reanalyzed the same proteins from *Methanobacterium thermoautotrophicum* and confirmed that the absence of hydrophobic patches and high content of negative residues are associated with improved solubility of proteins. Additionally, the researchers have found that low percentage of aspartic acid, glutamic acid, asparagine and glutamine residues increases the probability of a protein to be insoluble.

Computational analyses of solubility data have also yielded promising results. Goh et al. (2004) have applied random forest and decision tree analysis to various attributes of more than 27000 proteins from multiple organisms and found that protein solubility is influenced by (in increasing order of importance): length (< 516 amino acids), percentage of serine, cysteine, threonine and methionine amino acids, fraction of negatively charged residues and percentage of serine. However, in a high-throughput study describing overexpression in *E. coli* of 10167 proteins of *Caenorhabditis elegans*, no statistically significant correlation between the presence of rare codons, molecular weight, protein's pI and overall sequence hydrophobicity and protein solubility was observed (Luan & -H. Luan, 2004). The authors have indicated that proteins highly homologous to those with known soluble structures have higher chances to be soluble (Luan & -H. Luan, 2004).

Finally, Chan et al. (2013) have demonstrated that one of the most important factors that contribute to the protein solubility are positively charged and non-polar surface amino acid patches, i.e. structural stretches of amino acids on the protein's surface. The determined correlation between insolubility and patches of positively charged amino acids implies that this property, or similar feature to which it is strongly related, promotes protein-protein interactions (Chan et al., 2013). According to the authors, it could be argued that the increase of positive charge patches may lower the protein's fold stability through unfavorable charge interactions, and thus influence its solubility via partial unfolding.

1.2. Methods for measuring protein solubility

1.2.1. Addition of lyophilized protein

Protein solubility can be measured by adding lyophilized protein to a solution, up until the solution becomes saturated and the limit of solubility is reached (Evans et al., 2004; Catherine H. Schein, 1990). In this way, proteins with different degrees of solubility will reach the limit sooner or later, if the protein is less or more soluble respectively. However, such an approach is problematic for proteins that are highly soluble, because large amounts of purified protein are needed to reach the limit of solubility (Ahern & Manning, 1992). Additionally, when adding large amounts of protein to the solution, a gel may form consisting of mixed aqueous and solid phases. This makes the protein concentration determination, as there is no clear supernatant separated in the solution. Furthermore, the water and buffer content of the lyophilized protein is a critical and difficult variable to control (Ahern & Manning, 1992), and efforts to completely remove water, buffer, and cryo-protectants from lyophilized protein samples tend to make the protein harder to dissolve. Finally, denaturation and structural damage may occur in cases of extensive lyophilization (Prestrelski et al., 1993).

1.2.2. Concentration by ultrafiltration

Another method of protein solubility determination involves concentrating protein solution by centrifuging the sample in micro concentrator with a suitable molecular weight cut-off (Evans et al., 2004). Using this approach, the soluble and insoluble fractions are separated and can then be quantified using standard protein gel electrophoresis. However, this method also tends to produce gel-like solutions (Ahern & Manning, 1992) and would consume large quantities of protein in cases of highly soluble protein variants. Shire et al. (2010) reviewed several other concentration methods for determining the highest achievable protein concentration, yet have concluded that most of the concentrating methods have previously mentioned drawbacks.

1.2.3. Induction of amorphous precipitation

Described my multiple articles, determination of protein solubility by precipitating proteins with ammonium sulfate has proved to be a reliable method (Qamar et al., 1993; Trevino et al., 2007). Even though the method does not provide pharmaceutically relevant conditions, the ammonium sulfate precipitation can give a rapid and reliable information on the solubility of a specific protein.

This precipitation technique is experimentally reliable because factors which are difficult to control such as the water/buffer content of lyophilized protein or incidental ions that might get introduced during the course of the experiment become masked by the high concentration of salt. Advantageously, ammonium sulfate precipitation requires relatively small amounts of protein (10 mg or less) — even when studying a highly soluble protein variant.

At the beginning, as ammonium sulfate concentration increases, the protein solubility also increases, and this is attributed to electrostatic effects that can be described by the Debye–Huckel theory described in previous sections (Tanford, 1961). Collins (1997) pointed out that in the salting-in region, intramolecular ion pairs of the protein are replaced by protein–salt ion pairs and unpaired protein charges are paired with salt ions. The newly formed ion pairs enhance water’s ability to bind the surface residues of the protein and therefore increases protein stability.

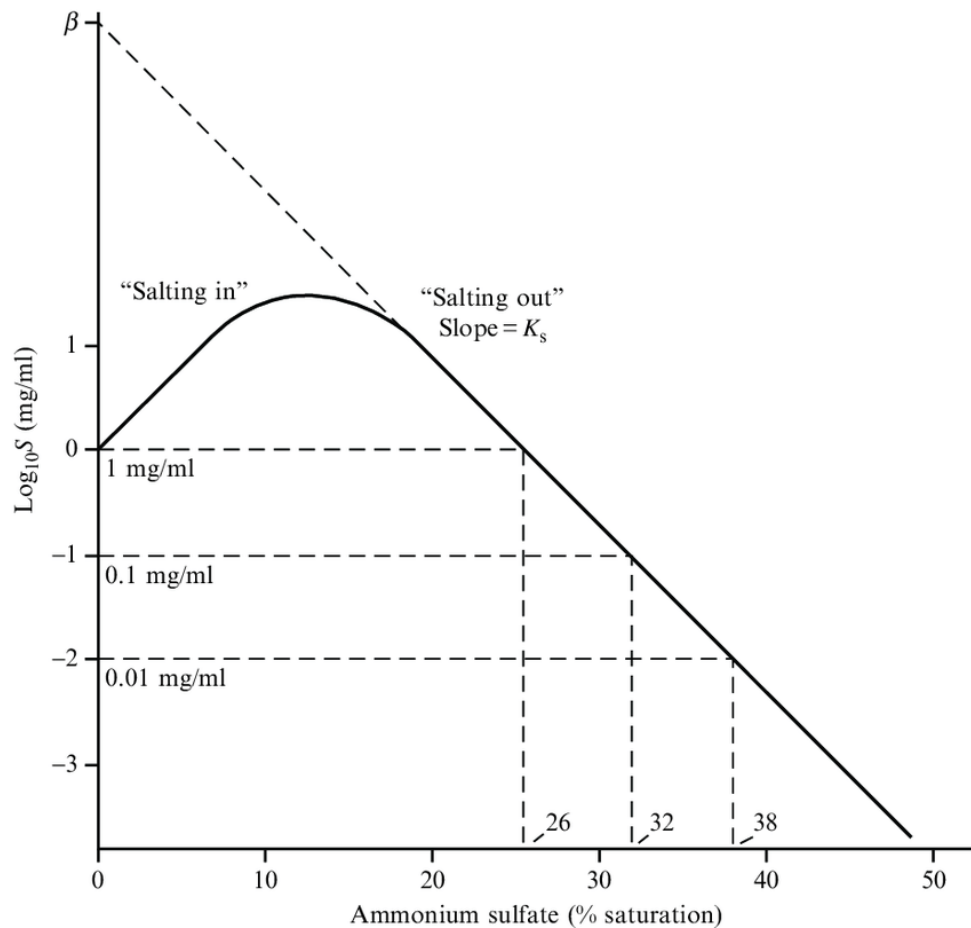


Figure 1.5 | Effect of increasing ammonium sulfate concentration on protein solubility. Illustration adapted from Burgess (2009).

Generally, the salting-in region occurs in salt concentrations between 0 and 0.3 M (Collins, 1997). Hence, as figure 1.5 illustrates, at around 0.3 M salt concentration, the curve can reach the maximum value. In contrast, as the salting-out occurs and salt concentration increases, protein solubility decreases. This can be described by the linear relationship described in Equation 1.

$$\text{Log}_{10}S = \beta - K_S C_S \text{ (Eq. 1)}$$

In this equation, the S is the solubility of the protein, β is the theoretical solubility value at zero salt concentration, K_S is the salting-out constant, and C_S is the molar concentration of the salt. K_S is expected to be dependent on the nature of the salt and protein (Dixon & Webb, 1961), yet independent from pH and temperature (McMeekin et al., 1937; Sandstrom, 1930). However, the dependence of protein on the K_S was determined to be insignificant. For example, Dixon & Webb (1961) have demonstrated that the variation of K_S for different proteins was not larger than two-fold.

1.2.4. Antibody blot based screening

Solubility detection methods described in the previous sections do not allow the efficient measurement of multiple different proteins simultaneously. For this reason, in order to overcome the drawbacks of the previous methods, a system for solubility detection based on antibody blot was developed (Knaust & Nordlund, 2001). The method is based on the efficient separation of insoluble and soluble cellular lysate fractions using filter plates. Once the fractions are separated, staining antibodies bind only to the soluble fraction of the lysate as the insoluble fraction is shielded within the inclusion bodies formed in bacteria.

In the described method, single *E. coli* clones containing a recombinant protein library, are picked and distributed randomly in a 96-well microtiter plate. After the target protein induction in each well, the cultures are lysed and harvested by filtration using 96-well filter plates. The filter plates separate soluble protein from inclusion bodies (which contain the insoluble proteins). Two separate dot blots are then prepared from the lysates and probed with RGS-His₄ antibody. The first blot containing the samples taken from the lysates before filtering corresponds to total protein (fig. 1.6 A). The second blot prepared after the filtering step represents the filtrates containing soluble protein (fig. 1.6 B). By comparing western blot band intensities of total and soluble proteins, protein solubility degree can be approximated for each of the 96 wells.

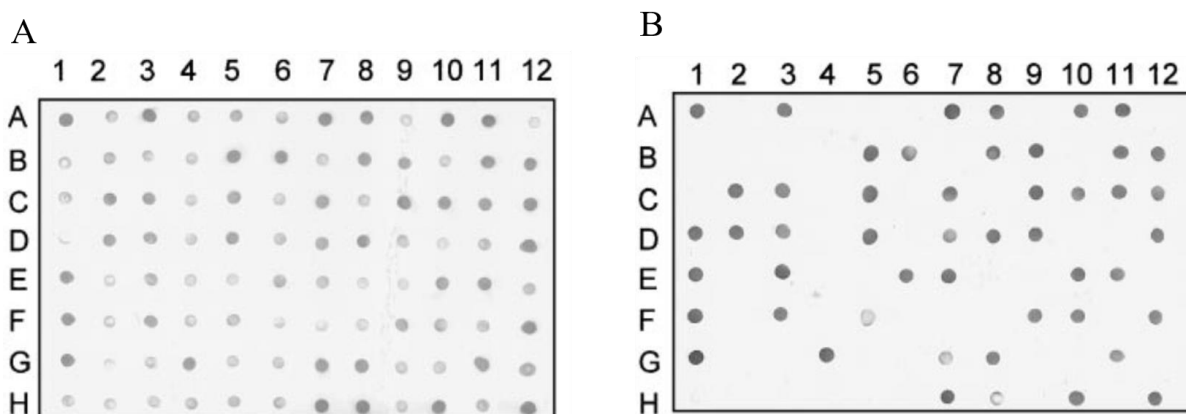


Figure 1.6 | Dot blots of protein lysates from 96-well filter plates. Protein samples were dotted onto nitrocellulose membrane and probed for His-tagged protein with anti-RGS-His Antibody. **(A)** Total Protein from cells lysed under non denaturing conditions taken before the filtering step. **(B)** Filtrates containing the soluble protein from cells lysed under non denaturing conditions. Illustration adapted from Knaust & Nordlund (2001).

This antibody blot detection method was the first step towards higher throughput solubility measurement systems. However, the provided throughput of the method is insufficient for large scale studies as the antibody blot based method is rather complex, expensive (due to the usage of antibodies) and therefore difficult to scale.

1.2.5. Split β -galactosidase based screening

In 2001, Wigley et al. (2001) presented a general method to assess the solubility and folding of proteins *in vivo*. The basis of the assay was structural complementation between the α - and ω -fragments of β -galactosidase (β -gal). Small α -fragment is fused to the target proteins or protein libraries (fig 1.7).

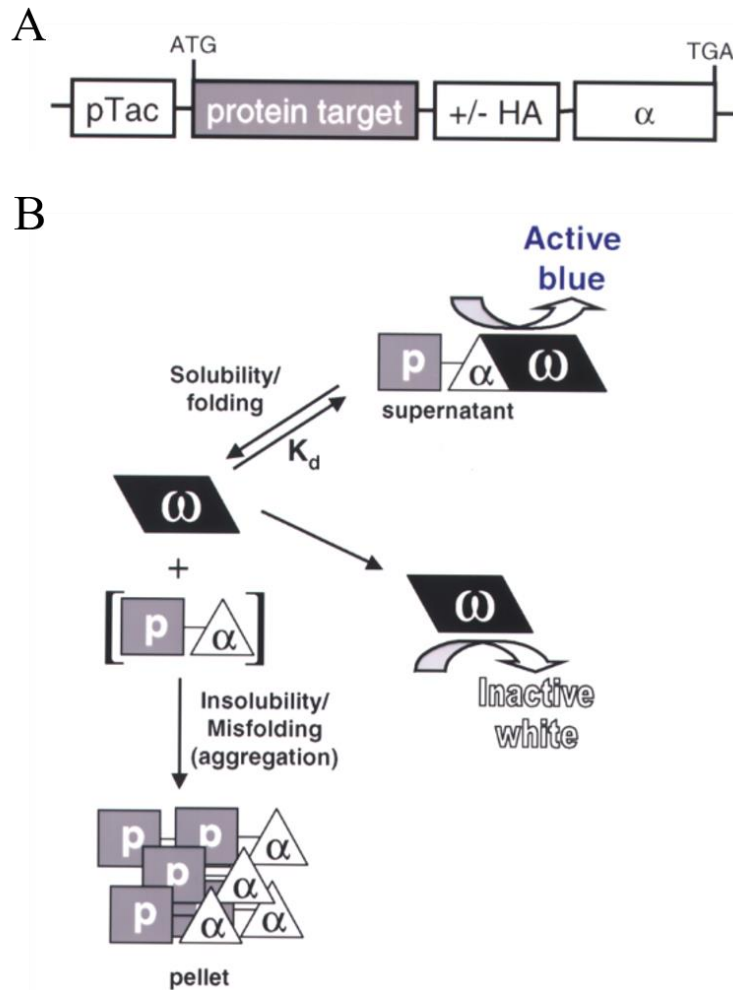


Figure 1.7 | An *in vivo* solubility assay based on structural complementation of β -galactosidase. (A) The target protein/ α -fragment C-terminal fusion expression construct (α -fragment, residues 7–58 from full-length β -gal). (B) Schematic depicting the complementation solubility assay. p represents the target protein, and α and ω represent each of the complementing fragments of the tetrameric β -galactosidase (β -gal). Illustration adapted from Wigley et al. (2001).

Expressed target proteins contain the small structural α -fragments. If the protein is soluble the fragment is exposed and can complement the ω fragment. This can be seen as blue color in the classical blue/white screening (fig. 1.7 B). However, if the protein is insoluble, it forms inclusion bodies and the α fragment is shielded from binding, resulting in a white color in the blue/white screening (fig. 1.7 B).

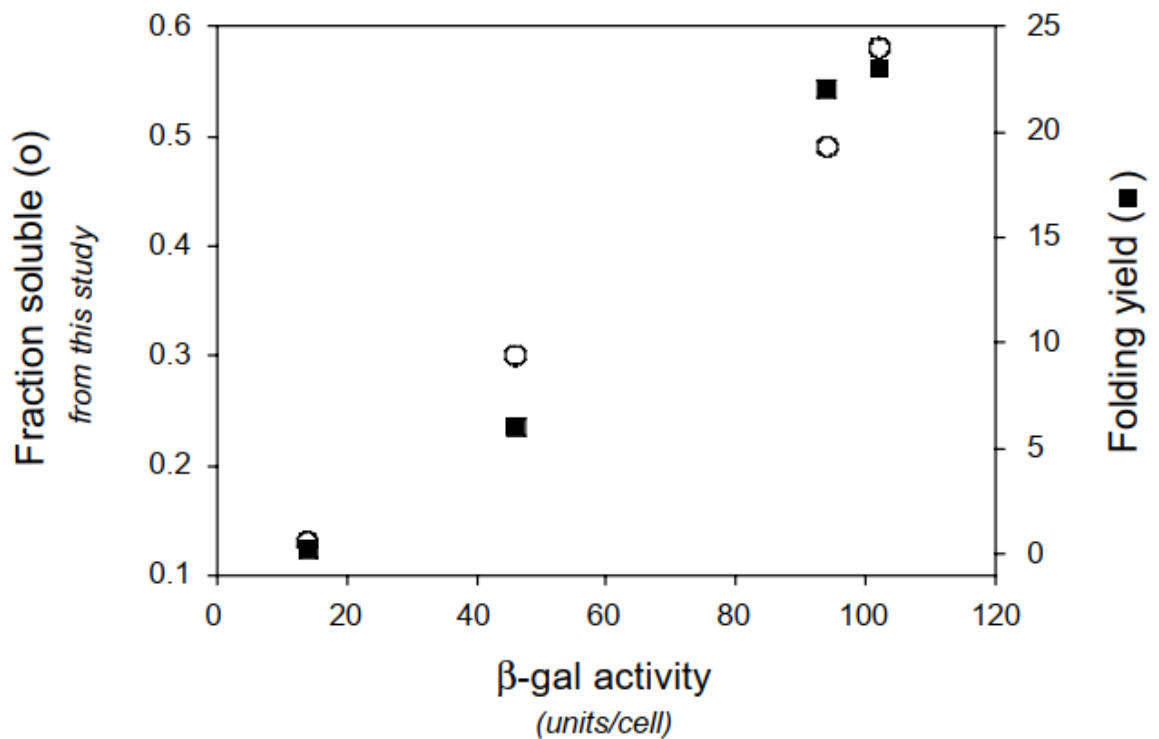


Figure 1.8 | Correlation of β -gal activity with fusion protein solubility and folding. The *in vitro* β -gal activity, measured in cell lysates, exhibits a linear correlation with the fraction soluble for each of the MBP and α fusion proteins. Illustration adapted from Betton & Hofnung (1996).

Fusions of the α -fragment to the C terminus of target proteins with widely varying *in vivo* solubility levels revealed a strong correlation between β -gal activity and the solubility of the target. Figure 1.8 shows the linear relationship between the β -gal activity and both the soluble fraction of the maltose binding protein (MBP) and α fusions as assessed by densitometry of Coomassie stained gels and the reported periplasmic folding yields for the unfused MBPs (Betton & Hofnung, 1996).

Thus, the described structural complementation provided a relatively simple and straightforward means of monitoring protein solubility *in vivo*. This method was the first step into a more affordable and non-complex determination of *in vivo* protein solubility. However, the main drawback of the described method is the attachment of large α fragment of the β -galactosidase which, according to the method authors, may significantly perturb the protein folding and solubility, leading to error-prone measurements.

1.2.6. Split GFP based screening

A method for protein solubility screening, developed by Cabantous & Waldo (2006), was created to overcome the shortcomings of the previously described split β -galactosidase based screening. While the main principle of soluble screening is similar, the main difference is that the new method uses green fluorescent protein (GFP) instead of β -galactosidase. An exceptionally well folded variant of GFP termed “superfolder GFP” (Pédelacq et al., 2006), which was split into two self-complementing parts (Cabantous et al., 2005). Similar to the case of β -galactosidase, split GFP fragment, consisting of 15 amino acids (GFP11), is fused to a target protein. Concurrently, the rest of the GFP sequence (GFP1-10) is expressed separately in a different plasmid. Such a system consisting of two plasmids – GFP11 and GFP1-10 – will be termed “solubility system” or “solubility assay” in the results and discussion sections of this work.

These fragments associate in *E. coli* spontaneously to form fluorescent GFP. Much like in the case of β -galactosidase based screening, the complementation degree correlates with target protein solubility in *E. coli*. This is, again, the result of shielding of the fused GFP fragment in the inclusion bodies in cases of insoluble proteins. However, compared to the β -galactosidase fragment complementation method, the GFP11 tag has minimal effect on protein solubility and folding (Cabantous et al., 2005).

The *in vivo* protein solubility detection is performed on agarized LB plates. Using the transfer membranes, colonies can be moved to different plates containing different inducers (fig. 1.9 A).

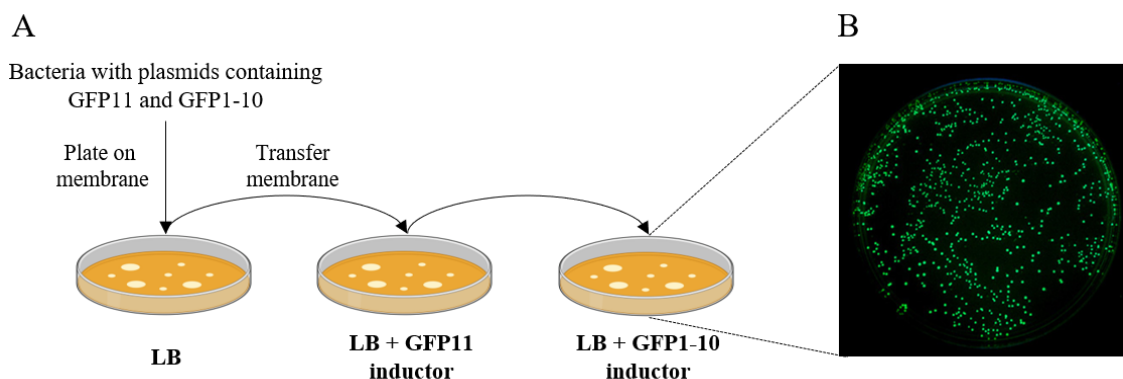


Figure 1.9 | In vivo solubility screening using sequential induction. (A) After the sequential induction, soluble GFP11 fused proteins spontaneously bind GFP1–10, and the resulting fluorescence is proportional to the amount of soluble, nonaggregated GFP11 tagged protein. Desired clones are then picked for propagation. (B) Fluorescent colonies after the sequential induction of plasmids. Illustration adapted from Cabantous & Waldo (2006).

First, the target protein fused with GFP11 is induced and expressed. After the target protein is expressed, the GFP1-10 is induced. In cases where target protein was soluble, the GFP11 binds to the GFP1-10 and green fluorescence can be detected in grown colonies on the agarized LB (fig. 1.9 B). On the other hand, insoluble proteins will shield the GFP11 from complementing, and the fluorescence would not be detected.

The described split GFP method have made a large step forward and improved upon the β -galactosidase method by improving the fused part to the target protein. The authors have demonstrated that GFP11 fused to target protein does not perturb protein folding or solubility, as it is in the case of alpha. However, the main drawback of this method is that it is difficult to analyze large number (e.g. thousands) of soluble protein variants, as each fluorescent colony must be picked manually from the plate.

1.3. Droplet microfluidics

Droplet based microfluidic systems manipulate discrete volumes of fluids in immiscible phases (Shang et al., 2017). Such generated microdroplets offer the possibility of handling very low volumes (μL to fL) of fluids (Suea-Ngam et al., 2019). Each rapidly generated droplet acts as a microreactor, thus enabling high-throughput processes (Zhu & Wang, 2016). Introduction of microfluidic technologies into the field of microbiology enabled new possibilities in, for example, biotechnological applications (Basova & Foret, 2015; Zhu & Wang, 2016), studies of microbial physiology (J. Chen et al., 2018) and detection of disease biomarkers (Kaushik et al., 2018).

1.3.1. Droplet microfluidics in microbiology

The application of droplets to study microorganisms have several advantages over classical methods such as multi-well plates, petri dishes or large volume flasks. First advantage of microfluidic systems is the possibility to isolate single cells from bulk in their own small liquid compartment (R. Chen et al., 2019). This allows to separate cells and analyze them individually, as well as use minimal amounts of reagents (μL to fL), making such an approach cost effective. Secondly, droplet microfluidics provide the ability to generate and analyze extremely large numbers (millions to billions) of individual droplets (Headen et al., 2018). For example, if bacteria would contain a large library of plasmids, each library variant could be analyzed (e.g. detection of color or fluorescence) separately in a high throughput fashion. The third advantage of droplet microfluidics is the possibility to perform complex operations on droplets, allowing unique

experimental protocols to be executed at a fast pace. Current state of the art microfluidic chips allow automated and controlled droplet formation (Chong et al., 2016; Churski et al., 2010), merging of different droplets or additional reagents (Varma et al., 2016), incubation (Berry et al., 2019), splitting (Raveshi et al., 2019) and physical sorting (Caen et al., 2018). As an example, this enables researchers to conduct multiple measurements on the same droplets or to track population dynamics in various chemical settings (Jakiela et al., 2013).

In microbiology, the identification of bacteria and their activity is one of the most important experimental protocols. Martin et al. (2003) have developed a system based on droplet microfluidics that was able to detect fluorescent proteins synthesized by bacteria, that were encapsulated in 60 nL droplets.

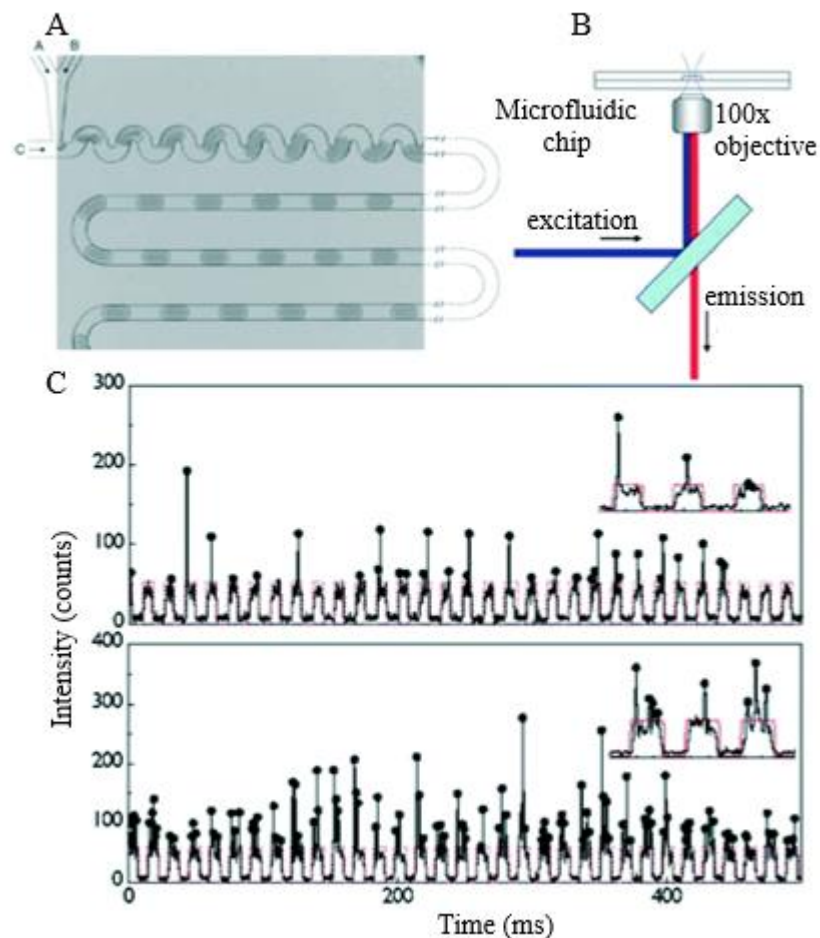


Figure 1.10 | Detection of bacteria expressing fluorescent protein. (A) Microfluidic chip for generation of the 60 nL droplets. (B) Laser setup for fluorescence reading from the microfluidic chip. (C) Example graphs of the traces of the fluorescence signal readout. Illustration adapted from Huebner et al. (2007).

This was one of the first times that microbial growth was detected directly in droplets by measuring the fluorescence intensity produced by *Escherichia coli* (fig. 1.10 B). Afterwards, similar principle of growth measurement using fluorescence was done for the analysis of other species, e.g. *Salmonella typhimurium* (Leung et al., 2012) and *Bacillus subtilis* (Bjork et al., 2015).

1.3.2. Bacterial growth in agarose beads

Microbiologists often grow bacterial cells on agar media plates in order to isolate, screen and select clonal cell populations. However, for multiple applications, the isolation of mutants grown by bacteria on agar often requires large quantities of chemicals, multiple iterations of screening and tedious manual work (Alain & Querellou, 2009). This can be problematic if the screening reagents are difficult to obtain (structurally complex) and only available in low amounts. Additionally, colony selection on agar media plates screening time is restricted by the growth of individual cells into colonies, as they must grow large enough to be visualized to pick. To overcome the limitations of agar media plate screening, researchers have developed a microfluidic based system for monodisperse agarose microparticle production (McDonald et al., 2000). In this system, single bacteria can be rapidly encapsulated and grown in separate agarose beads, allowing parallel and fast colony analysis (Katsuragi et al., 2000; Luo et al., 2007).

In 2011, Eun et al. (2011) have isolated single bacteria in agarose microparticles and demonstrated that they can rapidly determine the minimum inhibitory concentration of rifampicin and isolate mutant MG1655 *E. coli* that are resistant to the antibiotic. The authors have used flow focusing devices in order to produce the microparticles. Bacterial cells mixed with 1.5% agarose are pumped into the junction, through the inlet and is met by the mineral oil, flowing from two orthogonally oriented channels (fig. 1.11 A).

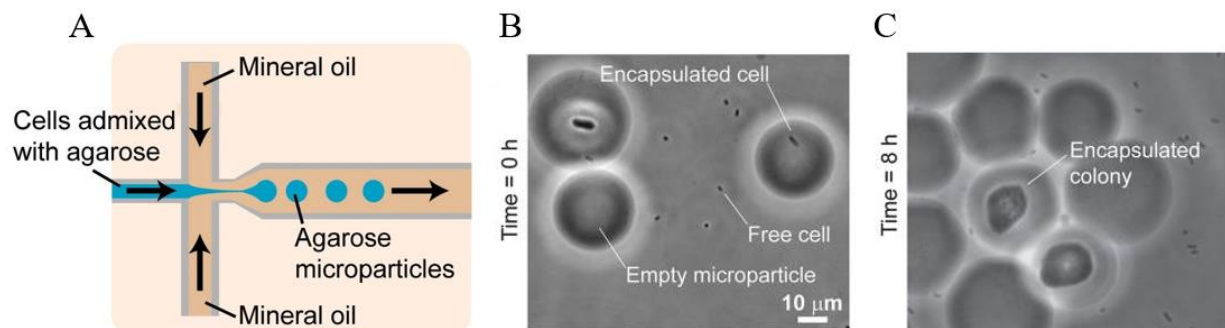


Figure 1.11 | Agarose microparticle system. (A) Schematic diagram of PDMS device used to generate microfluidic droplets containing agarose mix and bacterial cells. (B) Agarose bead microparticles right after bacteria encapsulation. (C) Growth of bacteria inside agarose microparticles after 8 hours. Illustration adapted from Eun et al. (2011).

In the same article, the authors have demonstrated bright field microscope images of droplets immediately after the agarose mix and bacterial cells encapsulation (fig. 1.11 B). Afterwards, the microparticles were incubated at 37°C for 8 hours, which resulted in the formation of 5–20 µm diameter microcolonies (fig. 1.11 C). The authors have then prepared bacteria containing GFP plasmids and encapsulated them into the agarose beads. After bacterial growth and during FACS analysis, it was determined that 24% of microparticles emit detectable green fluorescence, demonstrating that agarose based microparticles can be used for successful colony growth and protein expression. Therefore, it is evident that agarose based microparticles can be successfully used in order to overcome the obstacles of standard agarized plate screening.

To summarize, in the literature analysis part of this work main determinants of protein solubility were reviewed. In accordance to published data, both extrinsic (e.g. pH, temperature, ionic strength of the solvent) and intrinsic (e.g. primary amino acid sequence of the protein) properties may highly affect the solubility of protein. Interestingly, for a set of external conditions, a protein sequence variation (of the same function and/or structure) is likely to exist that would result in a soluble protein. This makes the understanding of the delicate relationship between protein's primary sequence and its solubility extremely important. On that note, several important methods – especially the split β -galactosidase (Wigley et al., 2001) and split-GFP (Cabantous & Waldo, 2006) – for protein solubility determination were developed in the past decades. These systems made the first steps towards acquisition of data that describes how proteins sequence influences the solubility. Yet, these systems have significant drawbacks due to the usage of agarized plates, which result in low-throughput and highly tedious workflow. This limits the amount of protein variants that can be analyzed in a given amount of time. On another note, droplet microfluidic systems developed for microbiology offer a way to increase throughput of the experiments (J. Chen et al., 2018). Important work was done by Eun et al. (2011), which demonstrated that colony growth and analysis on agar plates can be performed using droplet microfluidics by producing agarose microparticles. Using the same principles, in this work, the split-GFP system (Cabantous & Waldo, 2006) is converted into high-throughput (HT) solubility system using agarose beads produced with droplet microfluidic techniques.

2. MATERIALS AND METHODS

2.1. Materials and equipment

2.1.1. Reagents

- AB „Vilniaus degtinė“: 96% ethanol;
- „Aquapel Glass Treatment“: commercial silane mix;
- „Carl Roth“: 10 mg/mL ethidium bromide;
- „Dow Corning“: polydimethylsiloxane (PDMS), PDMS polymer fixer;
- „Gibco“: 10 × PBS (pH 7.2);
- „Lonza“: agarose;
- „RAN Biotechnologie“: RAN surfactant;
- „Sigma-Aldrich“: ≥ 99% hexane, ≥ 99% potassium chloride (KCl), ≥ 99% sodium chloride (NaCl), ≥ 99.99 %, 1M Tris-HCl (pH 8.0), 1H,1H,2H,2H-Perfluoro-1-octanol (PFO), fluorinated oil HFE-7500, light mineral oil, Propylene glycol methyl ether acetate (PGMEA);
- „Thermo Fisher Scientific“: dNTP mix (10 mM), 50 × TAE buffer, MgCl (50 mM).

2.1.2. Enzymes, solutions and kits

Enzymes

- „Thermo Fisher Scientific“:
 - Phusion Master Mix;
 - Platinum Taq polymerase;
 - FastDigest restrictases.

Solutions

- 0.1 M CaCl₂ with 15% glycerol;
- 0.1 M CaCl₂ solution;
- TBE buffer: 89 mM Tris-HCl, 89 mM H₃BO₃, 2 mM EDTA, pH 8.3 (25 °C);
- TAE buffer: 40 mM Tris-acetate, 1 mM EDTA, pH 8.3 (25 °C);

Kits

- “Invitrogen”: Agarose Gel DNA Extraction kit (Invitrogen);
- „Thermo Fisher Scientific“: „PCR PurificationKit“, „PCR PurificationKit“, „Rapid DNA LigationKit“.

DNA molecular mass standards and dyes

- „100 bp DNA Ladder“ (Thermo Fisher Scientific)
- „DNA ladder mix“(Thermo Fisher Scientific)
- „DNA loading dye 6x“ (Thermo Fisher Scientific)

2.1.3. Bacterial strains and media

Bacteria media

- LB Media: 1 % peptone, 0.5 % yeast extract, 0.5 % NaCl solution, pH 7.0 (25°C);
- Agarized LB media: 1 % peptone, 0.5 % yeast extract, 0.5 % NaCl solution, 1.5 % agar, pH 7.0 (25°C).

Bacteria strains

- *E. coli*: DH5 α ;
- *E. coli*: BL21 (DE3).

2.1.4. Oligonucleotides

Table 2.1. Oligonucleotides.

Name	Sequence (5' ->3')	Comment
Bba_Fw	GAATTCGCGGCCGCTTCTAG	Forward primer for the amplification of NDP-K cassettes
Bba_Rev	CTGCAGCGGCCGCTACTAGTA	Reverse primer for the amplification of NDP-K cassettes

2.1.5. Equipment

- Centrifuge “Eppendorf 5840R”
- Fluorometer “FlexStation II 384”
- Laminar „Telstar AV-100”;
- Thermoblock „Biosan“;
- Syringe pump „Harvard Apparatus PHD 2000”;
- Incubator „Mettler“;
- Table centrifuge „Grant-bio PVC-2400“;
- Thermostat „Corning“;
- Electrophoresis machine “Bio-Rad”;
- Electric scale “Ohaus Scout”;
- Inverted optical microscope “Nikon eclipse Ti”;
- Water bath “SBB Aqua 18 plus”;
- Gel imaging system “MiniBIS Pro”;
- Shaker “Vortex-Genie 2”;
- Thermocycler “Eppendorf Mastercycler gradient”;
- Spectrophotometer“Nanodrop”;
- Needles – „Terumo Neolus“ Neolus“ 0.4 × 16 mm, 27 G × 5/8 ir 0.6 × 16 mm;
- Plate reader “Biotek PowerWave XS”
- Syringes – „Braun“ 1ml OmniFix;
- Tubing – „Adtech“ PTFE 0.56 mm;
- Microfluidic chip– single water phase;
- Tubes – „Eppendorf“ 1.5 mL, 15mL, 50 mL DNA LoBind.

2.2. Methods

2.2.1. Production of PDMS chip

Microfluidic chips are produced using soft lithography techniques. Photomasks that are used for the formation of PDMS are manufactured in U.S., while the rest of the procedures are executed locally. The preparation of wafer and PDMS mold is illustrated below (fig. 2.1).

Preparation of master wafer

1. Light sensitive material (photoresist) is placed on a silicon wafer. The wafer is then spin-coated (500 rounds/minute) in order to distribute the photoresist evenly throughout the silicon wafer, i.e. the height of the material is the same in the whole area.
2. The silicon wafer with photoresist is placed on a 65°C heating block for 1 minute. Afterwards, it is transferred onto 95°C heating block for 3 minutes.
3. Photomask is placed on the heated substrate and the wafer is subjected to UV light for 20–30 seconds.
4. After irradiation, uncured substrate is washed from the wafer using PGMEA.
5. Wafer is carefully washed with isopropanol.

Production of PDMS mold and substrate binding

1. PDMS mix is prepared by mixing PDMS polymer and fixer with a ratio of 10:1 (w/w) until the mix becomes foggy and full of small air bubbles.
2. Using a vacuum pump, the air bubbles are removed from the mix.
3. Previously prepared wafer is transferred into a plastic petri dish and inserted into a vacuum pump. The prepared PDMS mix is poured into the petri dish, over the wafer.
4. Petri dish is kept in the vacuum until all of the air bubbles are removed.
5. The petri dish with poured PDMS mix is transferred into thermostat and incubated for 2 hours at 65°C.
6. After the heating step, PDMS slab is cut using a scalpel and carefully separated from the silicon wafer.
7. Using a biopsy needle (puncher) inlet and outlet holes are punched (entry and exit point for tubes containing fluids).
8. In order to remove PDMS leftovers and dust in the holes, small pressure nitrogen gas is used.
9. The PDMS slab and glass substrate are placed upside down in the plasma generating equipment.
10. Plasma-activated PDMS slab is then placed on the plasma activated glass surface and they are incubated at 65°C for 10 minutes.
11. PDMS chip channels are then filled with silane mix solution by a syringe.

12. The silane mix is kept in the channels for 10 to 30 seconds and then removed by pressurized air.
13. After the silane mix coating, microfluidic chip is placed on a 65°C heating block for 20 minutes.

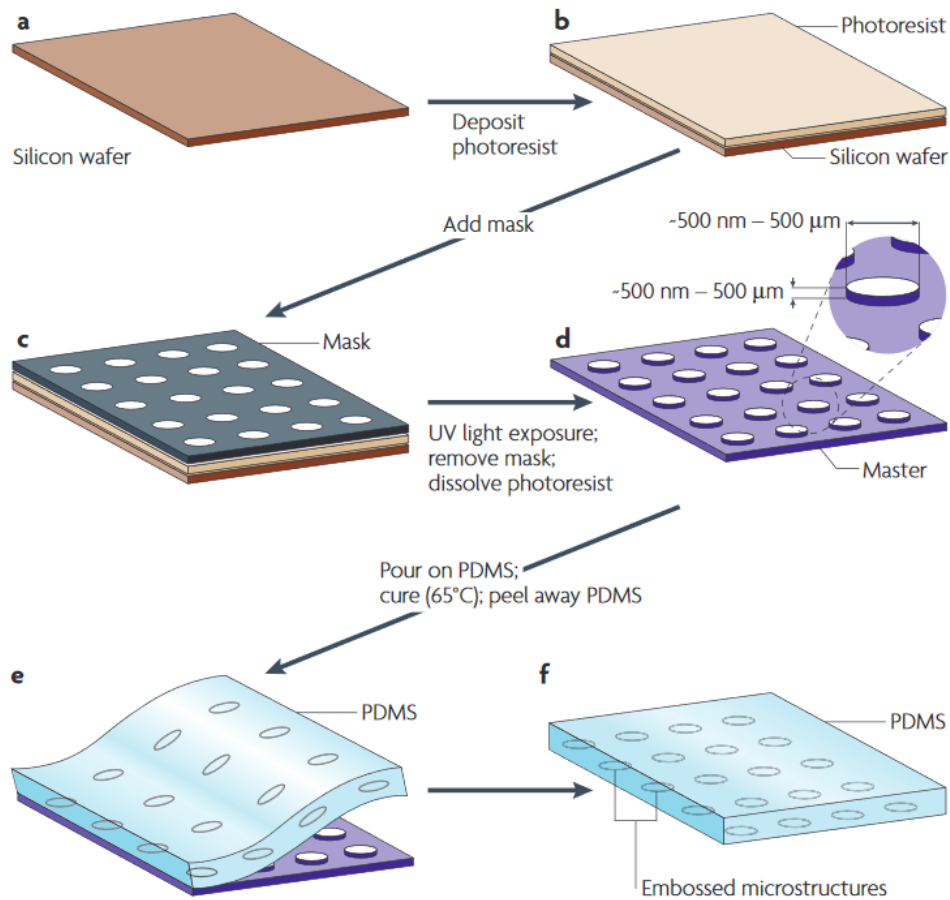


Figure 2.1 | The fabrication of micropatterned slabs of PDMS. (A, B) Photoresist is spin-coated on a silicon wafer. (C) A mask is placed in contact with the layer of photoresist. (D) the photoresist is illuminated with ultraviolet (UV) light through the mask. (E) PDMS is poured onto the master, cured thermally and peeled away. (F) The resulting layer of PDMS has microstructures embossed in its surface. Adapted from Weibel et al. (2007).

2.2.2. Competent cell preparation and transformation

Preparation of competent cells

1. 2 mL of overnight *E. coli* cultures are inoculated into 200 mL of fresh LB medium (1:100 dilution).
2. The bacteria are grown in 37°C until optical density reaches 0.6.
3. Bacteria are harvested by centrifugation at 4000 rpm for 10 minutes.

4. The pellet is washed twice with 0.1 M CaCl₂ solution (ice cold) and resuspended in 2 mL of 0.1 M CaCl₂ solution containing 15 % glycerol (ice cold).

Cell transformation

1. 1 μL of plasmid DNA (~1 ng) is carefully mixed with previously prepared and thawed DH5a or BL21 (DE3) competent cells.
2. The solution is incubated for 30 minutes on ice.
3. Heat shock is then applied to cells by placing them in 42°C for 45 seconds.
4. Cells are transferred on ice for 2 minutes.
5. 500 μL of SOC medium is added to the cell mix and incubated at 37°C on a 300 rpm shaking plate for 1 hour.
6. Resulting cells are spread onto LB agar plates supplemented with appropriate antibiotics (spectinomycin 75 μg/mL or kanamycin 35 μg/mL) and incubated at 37°C overnight.

2.2.3. Plasmid vector preparation

Plasmids pTET_GFP11 and pET_GFP1-10 (provided by Cabantous, fig. 2.2) were transformed into DH5α competent cells. 5 mL of culture were centrifuged and the plasmid DNA was purified with “Zymo Research” plasmid miniprep kit, following the provided manual.

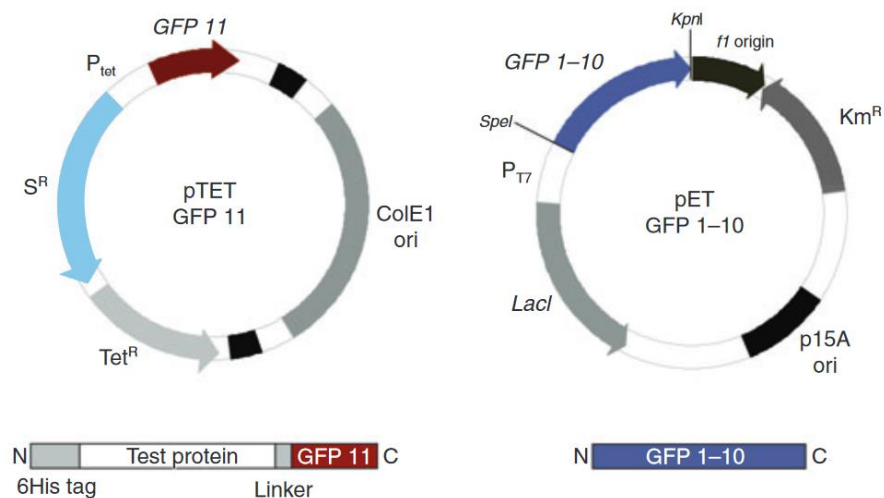


Figure 2.2 | Plasmid vectors used in this work. (Left) pTET vector containing the GFP11 and the test protein, carrying spectinomycin resistance. (Right) pET vector containing the GFP1-10 and carrying kanamycin resistance.

2.2.4. Insert preparation

The genes for soluble (eNDP-K) and insoluble (wtNDP-K) proteins were commercially synthesized (IDT). The sequences of the cassettes are provided below (table 2.2).

Table 2.2 DNA sequences of eNDP-K and wtNDP-K cassettes.

eNDP-K	<p>gaattcgcggccgcttctagggctctccatgcacgcgattaacattgcgtttttgcgctgattatgcctgttg agaaaactctactcactgaagccagacgcagtgggcgcgggggcttgcggcgagattattctaggttgaanaag ctggcctaaagatagtagccctcaaaatggtaaggcatctccagaggaaatagagagattttaccctcatcagagg aatggctccggctcggcggggcagaagcgtttaaaggcgtatcaagagcttggcatagatccaaggcgaagattgg cactgacgatcccgtggaggtaggtcggattattaacgtagtttagttaagtacatgacatcggggcctatcgttgtaa tgggttaaaggggaatagggtgtgaaatcgcagaaagctggtgggccccacgtcgcctactcggcggccgc ggggacaataagggcgactactcaattgactcgcctgacttagcggctgaggaggggagggtggttttaacttgg ccacgcgtcggatagtcctccgaagccgagagagaaataagattttggttcgagaagaggaggttttagaggccg gctccgatggagggtctggtggcgatcaacaagtcgtgaccacatggtcctcatgagtacgtaaatgctgctggga ttacataaggtacttaactcagcaccaccaccaccactgagatccggctgtaacaaagcccgaaggaagct gagttggctgctccaccgctgagcaataactagcataaccggtacctaggtactagtagcggccgctgcag</p>
wtNDP-K	<p>gaattcgcggccgcttctagggctctccatgcacgcgattaacattgcgtttttgacctgattatgcctgttg agaaaactctactcactgaagccagacgcagtgggcgcgggggcttgcgacgagattattctaggttgaanaag ctggcctaaagatagtagccctcaaaatggtaaggcatctccagaggaaatagagagattttaccctcatcagagg aatggctccagtcggcggggcagaagcgtttaaaggcgtatcaagagcttggcatagatccaaggcgaagattgg actgacgatcccgtggaggtaggtcggattattaacgtaacttagttaagtacatgacatcggggcctaactgtgta ggtgtaaaggggaatagggtgtgaaatcgcagaaagctggtgggccccacgtcgcctactcggcggccg gggacaataagggcgactactcaattgactcgcctgacttagcggctgaggaggggagggtggttttaacttgg caccgcgtcggatagtcctccgaagccgagagagaaataagattttggttcgagaagaggaggttttagaggccg ctccgatggagggtctggtggcgatcaacaagtcgtgaccacatggtcctcatgagtacgtaaatgctgctggg acataaggtacttaactcagcaccaccaccaccactgagatccggctgtaacaaagcccgaaggaagct agttggctgctccaccgctgagcaataactagcataaccggtacctaggtactagtagcggccgctgcag</p>

Amplification of inserts

The cassettes were amplified by PCR (Phusion master mix) using bba_Fw and bba_Rv as forward and reverse primers respectively. PCR mix and cycling conditions can be found in the tables below (tables 2.3 and 2.4).

Table 2.3 PCR mix for cassette amplification.

Component	Volume (μL)	Final concentration
Phusion master mix (2X)	25	1x
Forward primer	2.5	0.5 μM
Reverse primer	2.5	0.5 μM
DNA template	1	0.2 ng/ μl
H ₂ O	Up to 50	-
Final volume	50	-

Table 2.4 PCR cycling conditions

	Initial denaturation	30 cycles			Final elongation
		Denaturation	Annealing	Elongation	
Temperature	98°C	98°C	60°C	72°C	72°C
Time	30 sec	10 sec	15 sec	15 sec	5 min

PCR product purification

Amplified PCR products were purified from agarose gel to avoid additional PCR products with Agarose Gel DNA Extraction kit (Invitrogen).

Purified product digestion

The purified PCR products were digested by restriction enzymes in order to prepare them for ligation into vector. The digestion mix components are provided below (table 2.5). The plasmid vectors (pTET and pET) were digested accordingly.

Table 2.5 Restriction enzyme mix for DNA digestion.

Component	Volume μL)	Final concentration
PCR product (eNDP-K or wtNDP-K)	1	16 ng/ μ l (480 ng)
FastDigest buffer (10x)	3	1x
FD Eco31I	1	-
FD KpnI	1	-
H ₂ O	24	-
Final volume	30	-

2.2.5. Ligation and isolation of the constructed plasmids

The ligation was carried out with 1 U T4 ligase using a molar vector: insert ratio 1:5 (~100 ng of vector and ~ 500 ng of insert) with addition of T4 ligase buffer. Ligation was carried out at 22 °C for 1 hour. The ligation mixture was incubated at 72 °C for 10 minutes to inhibit the T4 ligase. New constructed vectors were transformed into competent DH5a cells and isolated (see “Competent cell preparation and transformation”). The DNA sequence of plasmids was confirmed by sequencing.

2.2.6. Sequencing

Amplified and plasmid DNA sequencing was performed in the VU Institute of biotechnology using standard sanger sequencing protocols.

2.2.7. Control protein solubility assessment

In order to test control protein - wtNDP-K and eNDP-K - solubility, after initial protein induction proteins are collected by centrifugation (2 mL of culture centrifuged for 10 minutes at 4000 g). Afterwards, the cells are washed with 500 μ L of PBS, resuspended in 800 μ L of PBS and lysed by sonication. The sonication was performed for 4 minutes (cycle 9 \times 10%, power 30%). Then, the cells were centrifuged for 20 minutes at 14 000 g (4°C) and soluble fraction was collected. The insoluble fraction was collected by resuspending the pellet in 800 μ L of PBS. For each fraction,

10 μ L was mixed with 2 \times SDS dye containing DTT, heated at 95°C for 5 minutes and subjected to SDS-PAGE electrophoresis (stacking gel 4 %, resolving gel 15 % PAA).

2.2.8. Preparation of testing samples

Different plasmid combinations were transformed into *E. coli* BL21 (DE3) cells to be tested throughout the experiments. After the transformation, the bacteria were plated on LB agar medium with appropriate antibiotics (spectinomycin and/or kanamycin). Samples and antibiotics are listed below (table 2.6).

Table 2.6 Antibiotics used in different samples.

Sample	Antibiotics
eNDP-K (pTET) + GFP1-10 (pET)	75 μ g/mL Spectinomycin + 35 μ g /mL Kanamycin
wtNDP-K (pTET) + GFP1-10 (pET)	75 μ g/mL Spectinomycin + 35 μ g /mL Kanamycin
eNDP-K (pTET) only	75 μ g/mL Spectinomycin
wtNDP-K (pTET) only	75 μ g /mL Spectinomycin
GFP1-10 only (pET)	35 μ g/mL Kanamycin (Kan)

2.2.9. High-throughput system in bulk

The following step by step protocol was used for the in-bulk experiments of the HT solubility system:

1. *E. coli* BL21 (DE3) cells are transformed with plasmids and plated on LB agar medium with appropriate antibiotics.
2. Single colonies from freshly transformed plates are inoculated in overnight LB medium supplemented with appropriate antibiotics.
3. 1:100 of overnight culture is transferred to fresh LB medium supplemented with appropriate antibiotics and cultivated at 37°C for 2 hours or until OD₆₀₀ reaches 0.5-0.6.

4. Protein expression of pTET plasmids is induced using 0.3 $\mu\text{g/mL}$ AHT (anhydrotetracycline; stock solution prepared in ethanol with 0.3 mg/mL concentration) at 37°C for 2 hours.
5. The cells are collected by centrifugation for 10 minutes at 4000 rpm.
6. The cell medium is discarded and cells are resuspended in 0.9 % NaCl solution.
7. Washing steps 5-6 are repeated 3 times.
8. After an additional spin and discard step, cells are resuspended in a fresh LB medium with appropriate antibiotics.
9. Cells are cultivated for 1 hour at 37°C (inductorless growth).
10. Protein expression of pET plasmid is induced with 0.3 mM IPTG at 37°C for 1 or 2 hours.
11. The cells are collected by centrifugation for 10 minutes at 4000 rpm and resuspended with 0.9 % NaCl solution.
12. OD₆₀₀ is estimated and the bacteria amount is then normalized throughout the samples.
13. The fluorescence is then measured (see equipment section) using excitation and emission wavelengths of 450 nm and 510 nm respectively.

2.2.10. High-throughput system in agarose beads

A. Preparation of bacteria

1. *E. coli* BL21 (DE3) cells are transformed with plasmids and plated on LB agar medium with appropriate antibiotics.
2. Single colonies from freshly transformed plates are picked into 1.5 mL eppendorf tube which contains 1 mL of LB media.
3. Samples are transferred to cuvettes and OD₆₀₀ is measured for the samples.
4. Optical density of different samples is normalized for a more convenient sample preparation (e.g. 0.5).

B. Encapsulation of bacteria

1. 2 mL of HFE-7500 oil is loaded into a 3 mL syringe (oil is always supplemented with 1 % (w/v) EA surfactant).
2. Needle is attached to the syringe and using tweezers tubing is applied.
3. Using microfluidic syringe pumps, the solution is set to flow at 2000–5000 $\mu\text{l/h}$ until it reaches the end of the tubing.
4. LB and 2 % agarose solution is prepared by melting the agarose.

5. Final agarose mix is prepared by adding varying amounts of previously prepared bacteria to the 2 % agarose mix, according to the Poisson distribution (λ values from 0.1 to 0.8).
6. 1 mL of final agarose mix is loaded into a 3 mL syringe.
7. After attaching a needle and the tubing to the syringe, the solution is set to flow at 2000–5000 $\mu\text{l/h}$ until it reaches the end of the tubing.
8. A separate 10 cm tubing is cut and one end is inserted into PDMS chip outlet (fig. 2.3 C). The other end is inserted into a 2 mL collection tube filled with 300 μl of mineral oil.
9. Rest of the tubing are inserted in the appropriate inlets in the “simple drop maker” PDMS chip (fig. 2.3).
10. The pump flow speeds are 100 $\mu\text{l/h}$ and 300 $\mu\text{l/h}$ for agarose mix and oil mix respectively.

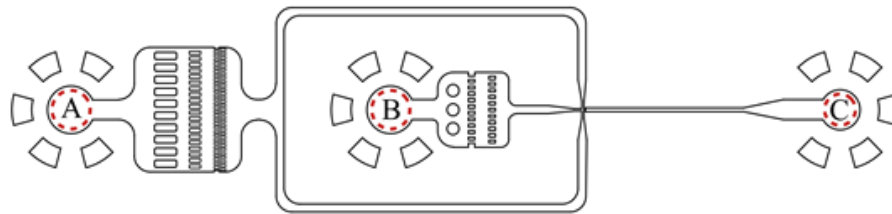


Figure 2.3 | Single water phase microfluidic chip. Letters denote oil inlet (A), agarose solution inlet (B) and droplet outlet (C) structures. Illustration adapted from Mazutis et al. (2013).

C. Preparation of agarose beads

1. After the encapsulation, collected emulsion is placed in a 37°C thermostat for 2 hours for the bacteria to grow.
2. After the incubation, 2 μL of emulsion is mixed with 8 μL of HFE-7500 oil inspected (for bacterial growth) under a brightfield microscope using a hemocytometer.
3. Emulsion is incubated at 4°C for 1 hour.
4. After the incubation, 500 μL of 20 % (v/v) PFO and 500 μL of LB media is added to the emulsion.
5. The solution is mixed by pipetting and gentle vortexing.
6. Sample is then centrifuged for 5 minutes at 5000 g.

7. Bottom phase of solution (PFO) is carefully removed from the tube.
8. The washing steps 4-7 are repeated 3 times and after the last centrifugation step, agarose beads are resuspended in DPBS.
9. Additionally, the agarose beads are washed 2 times with DPBS and finally 2 times with LB media containing appropriate antibiotics.

D. Induction of target proteins

1. Agarose bead solution is incubated at 37°C for 1 hour.
2. pTET plasmid is induced using 0.3 µg/mL AHT at 37°C for 2 hours.
3. After the incubation, agarose bead solution is centrifuged at 2000g for 2 minutes.
4. Supernatant is removed and beads are resuspended in 1mL of DPBS.
5. Washing described in steps 3 and 4 is repeated two times by resuspending with DPBS and finally two more times by resuspending the beads in LB media containing appropriate antibiotics.
6. Solution is incubated is incubated at 37°C for 1 hour.
7. Protein expression of pET plasmid is induced with 0.3 mM IPTG at 37°C for 1 hour.

E. Visualization of agarose beads

1. After the induction steps, agarose beads are washed 2 times with DPBS as described in steps D3 and D4.
2. 10 µL of agarose bead solution is inserted into hemocytometer and placed under fluorescent microscope
3. Brightfield images are captured using 10×, 20× or 40× magnification.
4. Without moving the samples, fluorescence images are captured, tracking excitation and emission wavelengths of 450 nm and 510 nm respectively.

3. RESULTS

3.1. High-throughput solubility detection system design

The aim of this work is to create a proof of concept assay that would allow high-throughput protein solubility screening. The *in vivo* protein solubility system based on split-GFP complementation by Cabantous & Waldo (2006) is an effective method that enables rapid identification of soluble protein variants from protein libraries. However, in order to analyze the amino acid sequences of the detected soluble proteins, it is required to manually pick and analyse each protein variant, making the whole system effectively low throughput (fig. 3.1). This can be seen as a major drawback when working with large protein libraries and/or the goal of the analysis is to relate protein sequence to its solubility level.

In this work, the split-GFP principle developed by Cabantous & Waldo (2006) was used as a starting point for the droplet microfluidic based high-throughput (HT) solubility detection system, which will be referred to as the HT system or HT solubility system from here on out. In this assay, single bacteria are grown inside monodisperse micrometer-scale agarose beads instead of agarose plates. The bacteria separation into unique compartments allows flexible downstream analysis of protein libraries – fluorescence detection and sorting of beads.

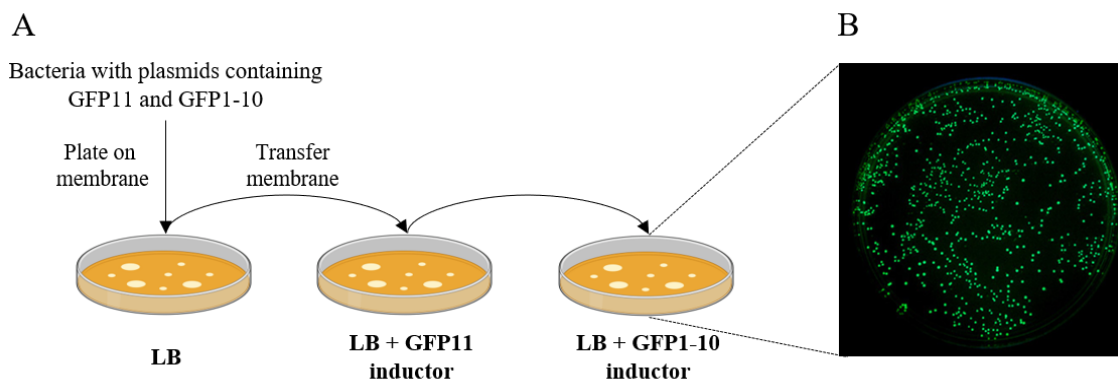


Figure 3.1 | In vivo solubility screening using sequential induction. (A) After the sequential induction, soluble GFP11 fused proteins spontaneously bind GFP1-10, and the resulting fluorescence is proportional to the amount of soluble, non-aggregated GFP11 tagged protein. Desired clones are then picked for propagation. (B) Fluorescent colonies after the sequential induction of plasmids. Illustration adapted from Cabantous & Waldo (2006).

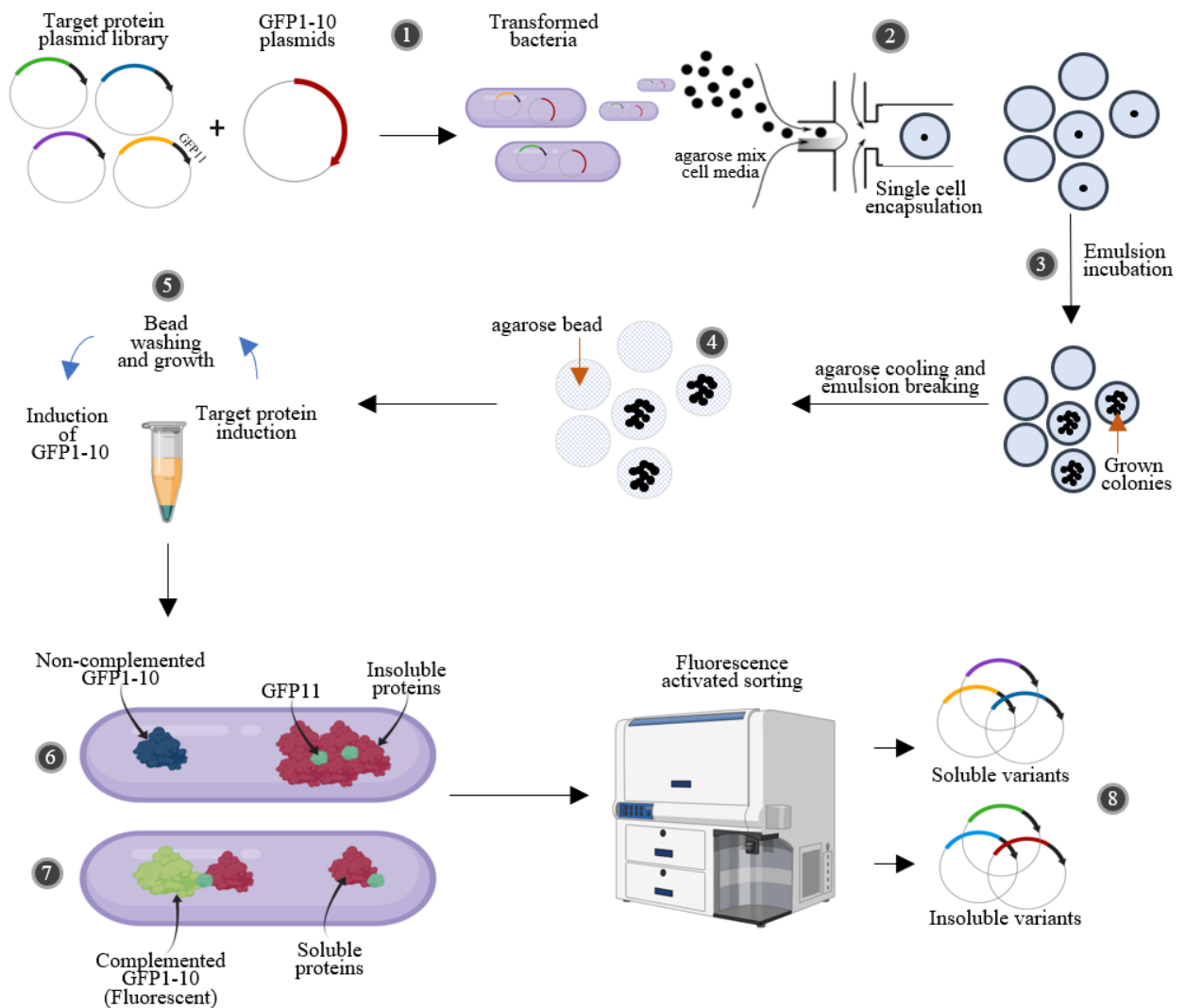


Figure 3.2 | Conceptual design of the high-throughput *in vivo* protein solubility system.

In the conceptual HT system design, protein library fused with the small GFP11 is transformed into bacteria in the form of plasmid DNA, together with a secondary plasmid containing GFP1-10 (fig. 3.2-1). The resulting library is then encapsulated into water droplets together with agarose mix and LB growth media, in such a way that on average a single droplet contains only a single bacterial cell (fig. 3.2-2). The formed emulsion is incubated in order to grow the bacteria inside the droplets (fig. 3.2-3). Afterwards, the emulsion is cooled in order to solidify agarose inside the droplets, which results in the formation of agarose beads (fig. 3.2-4).

The agarose beads are impermeable to bacterial cells, yet permeable to small molecules. This allows to change the growth medium, wash the beads and add expression inductors multiple times during the experiment (see methods). This is the main reason why the described assay is

designed with agarose beads instead of water droplets, which generally do not allow simple washing or molecule exchange steps.

The formed agarose beads are then subjected to subsequent induction and washing steps. First, the target protein with GFP11 fragment is induced. After the induction, the beads are washed and secondary inductor is added for the induction of GFP1-10 fragment (fig. 3.2-5). In case of an insoluble target protein, the fused GFP11 tail is hidden inside inclusion bodies and does not complement the GFP1-10 (fig. 3.2-6). However, in case of soluble protein, the small GFP fragment attaches to GFP1-10 which results in a complete green fluorescent protein formation (fig. 3.2-7). The beads may then be subjected to fluorescence activated physical sorting and sequencing of the sorted and amplified soluble and insoluble protein DNA libraries (fig. 3.2-8).

3.2. Preparation of the insoluble and soluble protein variants

In order to show the proof of concept of the HT solubility system, well characterized control proteins of known solubility level were required. Ideally, these proteins would have a similar amino acid sequence, in order not to perturb the expression level. Also, they should have a drastically different solubility degree, e.g. one soluble and one insoluble variant.

Proteins with such requirements were described by Pédelacq et al. (2002). The authors of the article have demonstrated that a ~25 kDA nucleoside diphosphate kinase (wtNDP-K) from *Pyrobaculum aerophilum* is completely insoluble and directed into inclusion bodies when expressed in *E. coli*. The authors have also performed several rounds of directed evolution, which resulted in a selection of a soluble NDP-K mutant (eNDP-K) containing 6 amino acid substitutions when compared to the wild type variant.

These proteins were chosen to evaluate the positive (soluble) to negative (insoluble) signal ratio of the HT system. The genes of both variants were commercially synthesized and cloned into pTET vectors, that will be further used in combination with pET vectors containing the GFP1-10. Bacteria, transformed with the prepared pTET vectors were grown and induced with AHT (for a detailed protocol, see the method section). After the protein expression, the cells were sonicated and soluble and insoluble fractions were separated by centrifugation. Polyacrylamide gel electrophoresis was used to analyze the fractions of both wtNDP-K and eNDP-K variants.

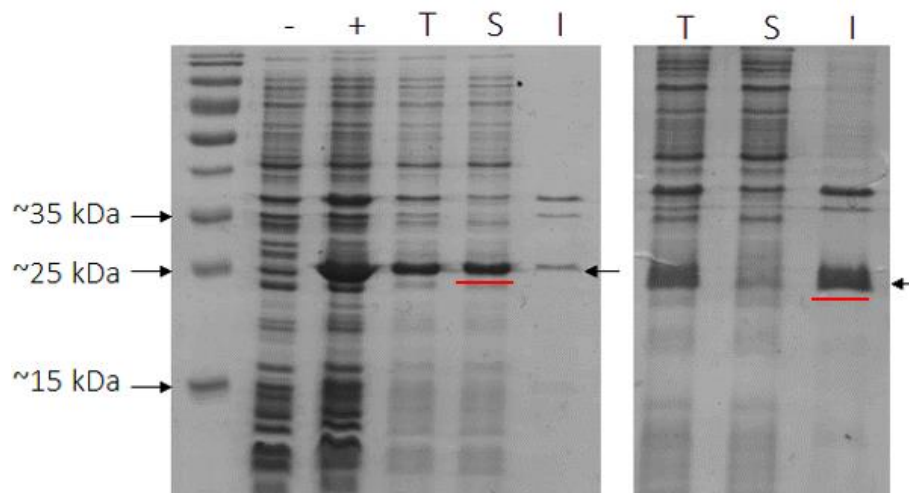


Figure 3.3 | Polyacrylamide gel electrophoresis gel image visualizing the eNDP-K (left side) and wtNDP-K (right side) protein synthesis. T – total lysate, S – soluble fraction, I – insoluble fraction, “+” – induced sample, “-” – uninduced sample. Red lines mark soluble and insoluble for eNDP-K and wtNDP-K fractions respectively.

For the eNDP-K, gel results show a dense dark band of soluble fraction and a dim band of insoluble fraction at 25 kDa (fig. 3.3). For the wtNDP-K, the soluble and insoluble band intensities are reversed - dense and dim bands for the insoluble and soluble fractions respectively (fig. 3.3). Agreeing with the literature, these protein variants are polar opposites in terms of solubility and can be further used to characterize the HT system.

3.3. Validation and optimization of HT system in bulk

After the validation of the soluble (eNDP-K) and insoluble (wtNDP-K) control proteins, the HT system concept was initially tested in bulk, low throughput format.

Split GFP method by Cabantous & Waldo (2006) was performed on agarized LB media plates using transfer membranes. However, usage of liquid growth media is a prerequisite for the conceptual HT solubility assay (see HT system overview section) in order to be able to use droplet microfluidic techniques. The original authors have not shown their method to work in liquid setting, therefore this had to be validated in this work.

The experiment was performed in parallel for either eNDP-K or wtNDP-K plasmid containing bacteria together with GFP1-10 plasmid. Also, additional controls were used. The full list of samples used is described below (table 3.1).

Table 3.1 Samples used in the bulk system validation.

Sample name	Sample description
eNDP-K + GFP1-10	Bacteria with both (i) eNDP-K gene (soluble protein) fused with GFP11 in pTET plasmid and (ii) GFP1-10 gene in pET plasmid.
wtNDP-K + GFP1-10	Bacteria with both (i) wtNDP-K (insoluble protein) gene fused with GFP11 in pTET plasmid and (ii) GFP1-10 gene in pET plasmid.
eNDP-K only	Bacteria with eNDP-K (soluble protein) gene fused with GFP11 in pTET plasmid only.
wtNDP-k only	Bacteria with wtNDP-K gene fused with GFP11 in pTET plasmid only.
GFP1-10 only	Bacteria with GFP1-10 gene in pET plasmid only.

The system validation consisted of the following steps: growth of overnight culture (fig. 3.4-1), dilution and growth (fig. 3.4-2), induction of pTET promoter using AHT (fig. 3.4-3), induction of pET promoter using IPTG (fig. 3.4-4), washing away of growth media (fig. 3.4-5), measurement of fluorescence (fig. 3.4-6).

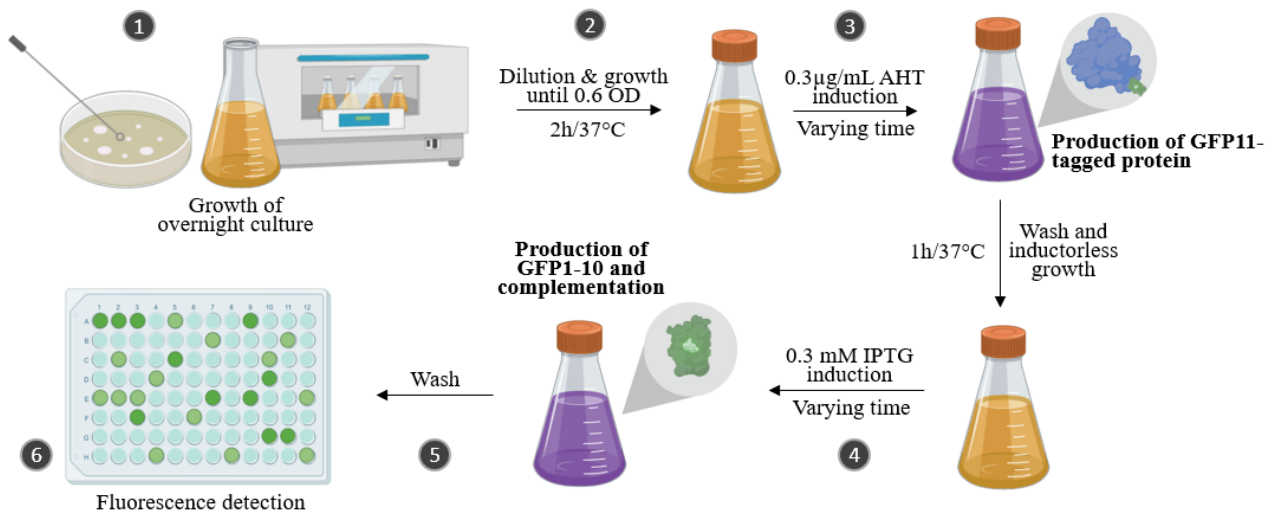


Figure 3.4 | Scheme of experimental for evaluation of the solubility system in bulk.

After the experiment, the fluorescence was measured for each of the sample (table 3.1) with 5 technical and 3 biological repeats. In cases where only pTET promoter was induced, background level fluorescence was detected for all of the samples (fig. 3.5, AHT induction only). Furthermore, samples containing only one plasmid also resulted in background level fluorescence (fig. 3.5, eNDP-K, wtNDP-K, GFP1-10 only). Samples containing both plasmids and soluble protein demonstrated strong average fluorescence intensity of 13.42×10^3 and 14.36×10^3 for 1-hour and 2-hour inductions of pET plasmid respectively (fig. 3.5, eNDP-K+GFP1-10). For insoluble protein, the fluorescence intensities were 6.87×10^3 and 6.23×10^3 for 1- and 2-hour induction of pET (fig. 3.5, wtNDP-K+GFP1-10). Because both 1- and 2-hour IPTG induction times yielded similar results fluorescence signal intensities, 1-hour induction was chosen for further system development.

For a solubility detection method to be reliable, the signal ratio between soluble and insoluble protein must be high in order to reliably separate the two signals, e.g. during the physical sorting of samples. In the case of 1-hour of pET plasmid induction, the average fluorescence signal intensity ratio between eNDP-K and wtNDP-K samples was 1.95.

We have hypothesized that the origin of relatively high fluorescence level of insoluble protein sample may come from the undepleted pTET inducer (AHT) activity during the pET induction. It may take time for the insoluble protein to aggregate after translation - which makes the immediate complementation of GFP1-10 and GFP11 possible for insoluble proteins.

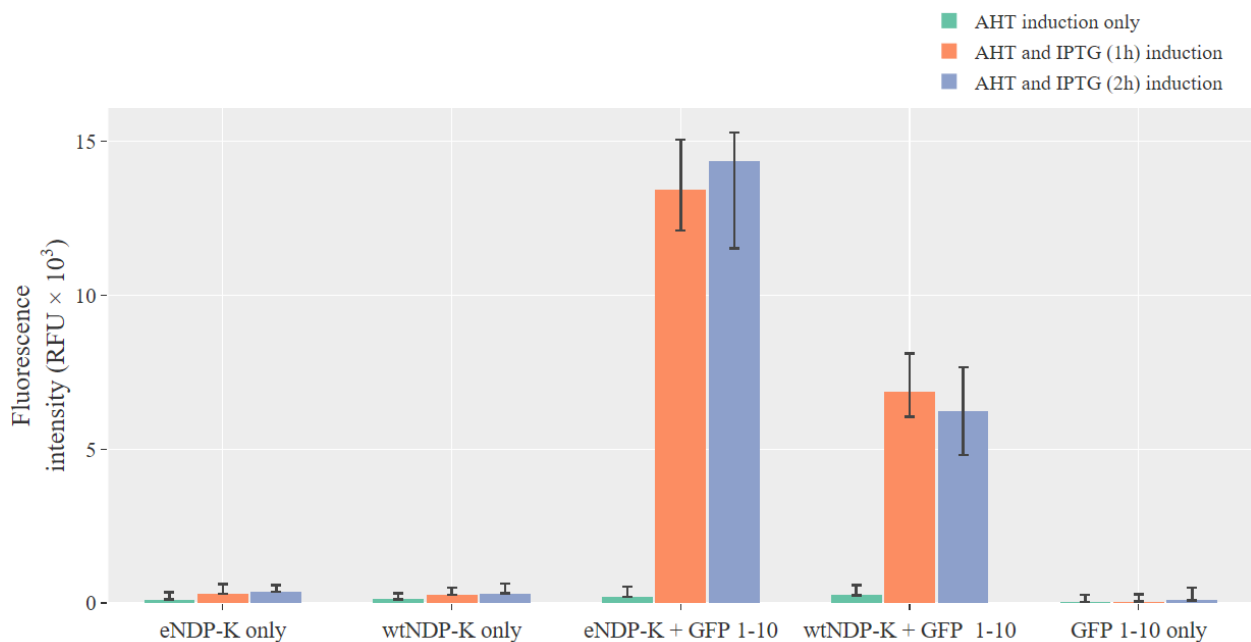


Figure 3.5 | Fluorescence intensity results of the samples after the in bulk experimental steps. Samples induced with AHT only are in green. Samples induced with AHT and either 1 or 2 hours of IPTG are in orange and purple respectively.

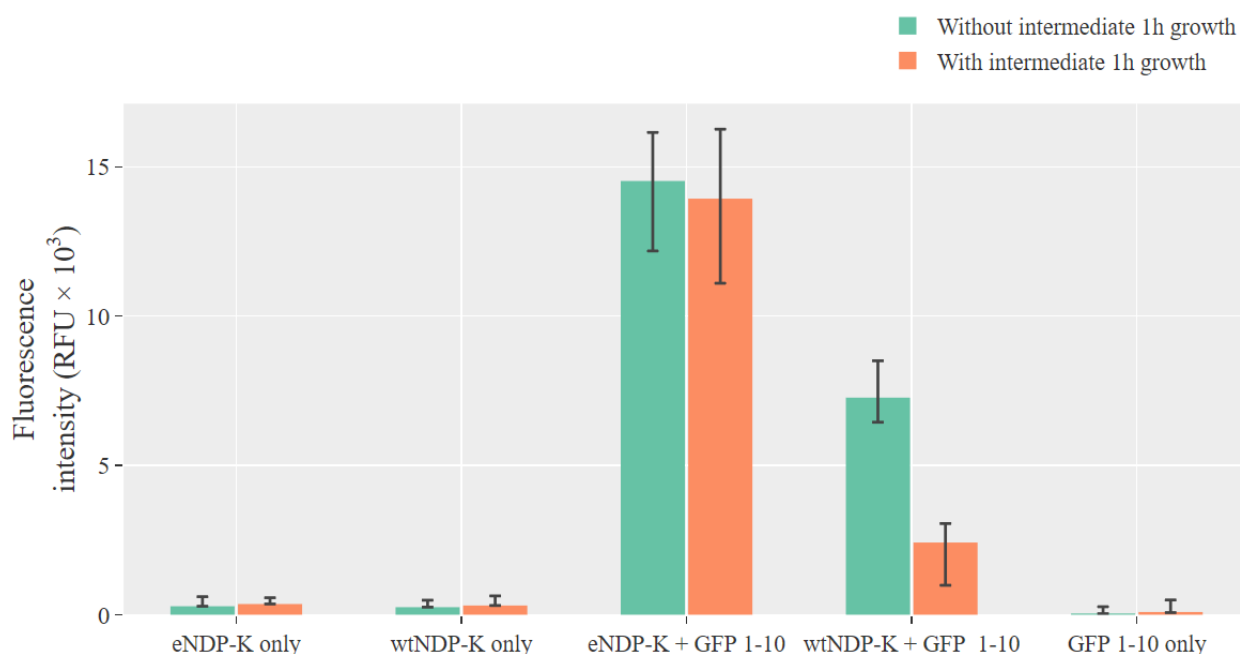


Figure 3.6 | Fluorescence results with or without intermediate growth steps. Samples with and without intermediate growth step are in green and orange colors respectively.

In other words, even for insoluble proteins, the GFP11 tail may be accessible for a short time after synthesis. Therefore, when inducing GFP1-10 fragment (pET), the expression of target protein with GFP11 (pTET) should be as minimal as possible. For this reason, an additional inductorless growth step was included in the experiment, during which the pTET plasmid inductor would be completely depleted from the bacteria before the GFP1-10 induction.

The addition of inductorless growth yielded fluorescence signal intensity of eNDP-K and wtNDP-K of 13.93×10^3 and 2.42×10^3 respectively. Therefore, the addition of growth step in between AHT and IPTG inductions increased the soluble to insoluble fluorescence signal ratio from 1.95-fold to 5.7-fold (fig. 3.6, eNDP-K and wtNDP-K+GFP1-10). Changes in control samples were not detected compared to results gathered without the additional step (fig. 3.5, eNDP-K, wtNDP and GFP1-10 only).

3.4. Validation of HT solubility system in agarose beads

After the bulk validation of the solubility system using liquid LB media, the final assay incorporating agarose beads was tested.

The proof of concept of the method was demonstrated using the following experimental steps: encapsulation of single bacteria into water droplets together with LB growth media and 2 % agarose mix (fig. 3.7-1), incubation of emulsion for bacterial growth (fig. 3.7-2), agarose cooling

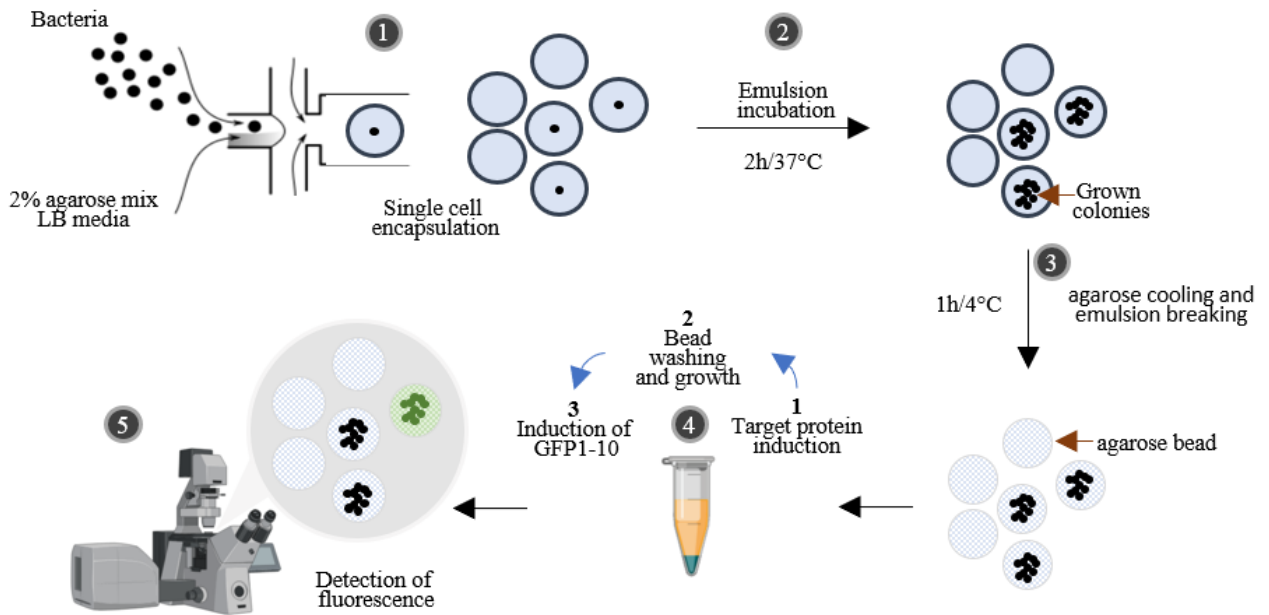


Figure 3.7 | Scheme of experimental for evaluation of the *in vivo* high-throughput protein solubility system in agarose beads.

and emulsion breaking (fig. 3.7-3), subsequent induction and washing of agarose beads (fig. 3.7-4), visualization of agarose bead fluorescence under fluorescence microscope (fig. 3.7-5). A detailed description of each step, as well as additional ones, can be found in the method section.

After the encapsulation of different samples into droplets (table 3.2), the collected emulsion is visualized under a brightfield microscope (fig. 3.8). Average measured droplet diameter was $31.1 \mu\text{m} \pm 2.6 \mu\text{m}$. Afterwards, the emulsion was incubated for bacteria to grow inside the droplets. Once the agarose beads containing grown bacteria colonies were prepared, the subsequent induction was executed. Agarose beads were visualized using both brightfield and fluorescence microscopy. First, in the brightfield images empty agarose beads were noticed (fig. 3.9, white arrows). Likely, these agarose beads, by chance, did not receive a bacterial cell during the encapsulation of samples. It is also possible that those beads did receive a cell, yet failed to grow over 2 hours during the bacterial growth period. In all of the samples grown bacterial colonies are present inside the agarose beads (fig. 3.9, green arrows).

Table 3.2 Samples used in the HT system validation.

Sample name	Sample description
eNDP-K + GFP1-10	Bacteria with both (i) eNDP-K gene (soluble protein) fused with GFP11 in pTET plasmid and (ii) GFP1-10 gene in pET plasmid.
wtNDP-K + GFP1-10	Bacteria with both (i) wtNDP-K (insoluble protein) gene fused with GFP11 in pTET plasmid and (ii) GFP1-10 gene in pET plasmid.
eNDP-K only	Bacteria with eNDP-K (soluble protein) gene fused with GFP11 in pTET plasmid only.
GFP1-10 only	Bacteria with GFP1-10 gene in pET plasmid only.

Lose bacterial cells can be noticed in the background. These bacterial cells may have escaped the agarose beads during the whole procedure (6 to 8 hours).

Sample containing insoluble protein variant, wtNDP-K+GFP1-10, emitted a barely detectable fluorescence signal at emission wavelength of 510 nm (fig. 3.9, A). However, the soluble protein variant, eNDP-K+GFP1-10, produced a strong fluorescence signal at the same wavelength (fig. 3.9, B).

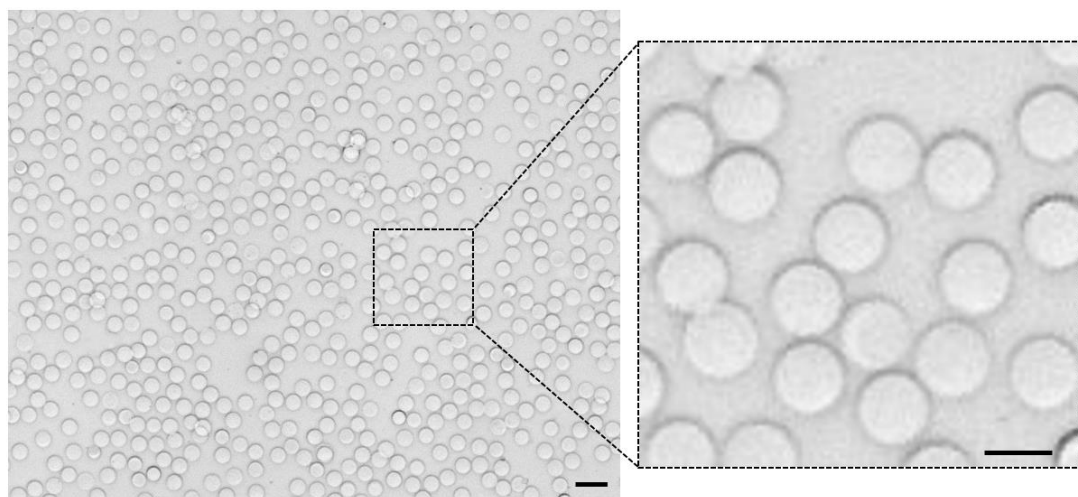


Figure 3.8 | Bright field image of emulsion containing agarose and bacteria minutes after encapsulation. Nikon 10x objective used for imaging. Scale bars 75 µm and 30 µm (zoomed).

For both wtNDP-K+GFP1-10 and eNDP-K+GFP1-10, the fluorescence only occurred in the same locations wherein bacterial colonies can be found. Identical locations are visualized by two red arrows of brightfield and fluorescence images (fig. 3.9). Control sample images containing only one plasmid (eNDP-K and GFP1-10 only) are not shown, as these samples did not exhibit any fluorescence.

The fluorescence strength of agarose beads in different samples was then quantified using imaging editing software (see methods). The fluorescence intensity was calculated for each agarose bead separately and averaged afterwards. In other words, each bead in the same sample acted as a technical replicate when quantifying the fluorescence intensity. In order to account for different amounts of grown bacteria in a bead, the fluorescence signal intensity integral was normalized to the calculated integral of the relative occupancy of the agarose bead by bacteria.

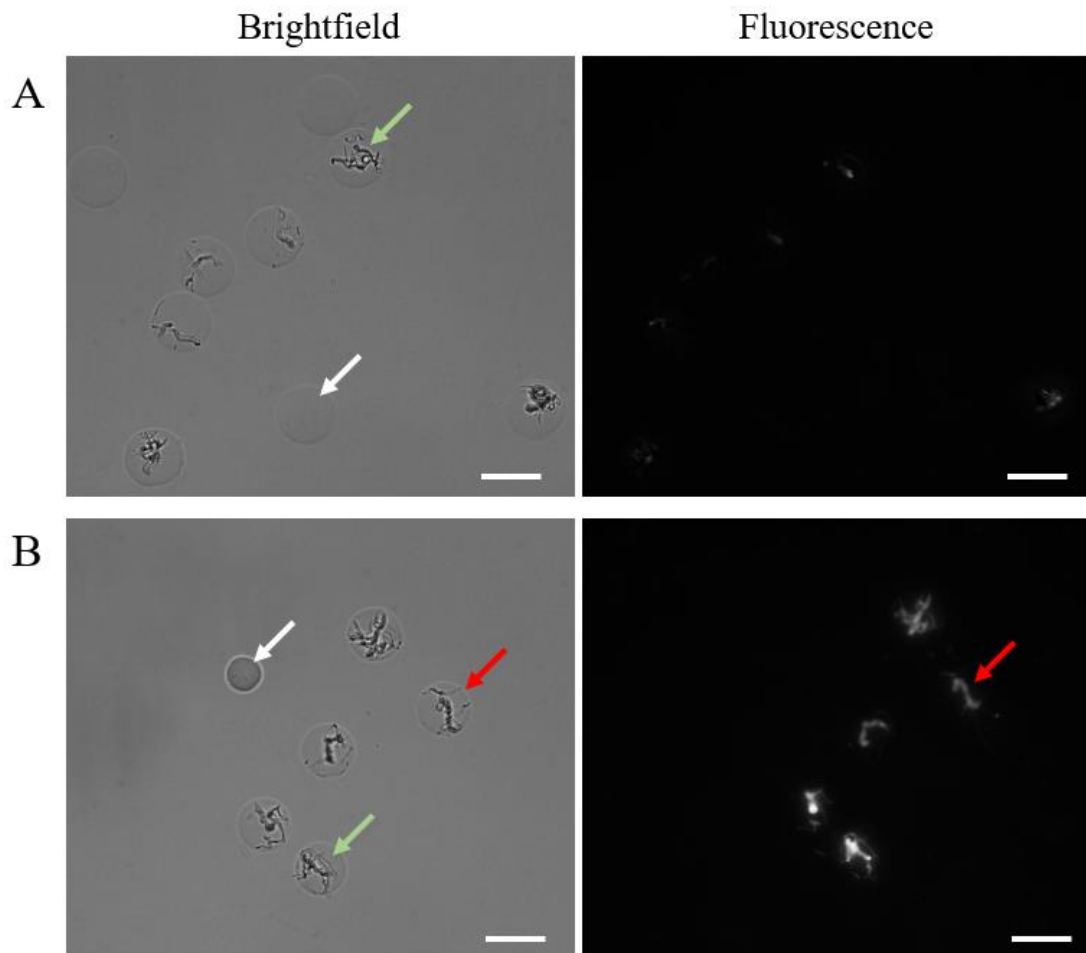


Figure 3.9 | Brightfield (left section) and fluorescent (right section) images of samples. (A) Images of wtNDP-K (insoluble protein) sample. **(B)** Images of eNDP-K (soluble protein) sample. Green arrows indicate bacterial colony growth inside agarose beads. White arrows indicate empty agarose beads. Red arrows indicate the same location in both brightfield and fluorescent images. Nikon 20x objective used for imaging. Scale bar 75 μm .

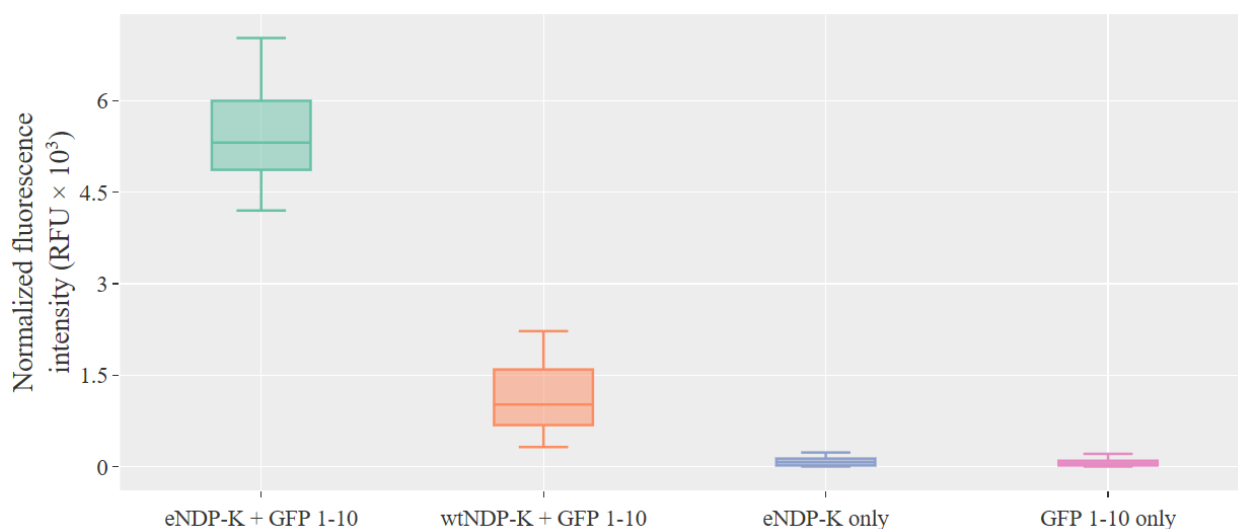


Figure 3.10 | Estimation of normalized fluorescence intensity for the samples. Fluorescence was normalized by integrating the occupancy of bacterial colonies inside the agarose beads. For each sample, each separate agarose bead acted as a technical replicate.

The fluorescence signals of single plasmid control samples (eNDP-K and GFP1-10) were negligible, with values scattered around zero. The median values were 0.07×10^3 and 0.05×10^3 for eNDP-K and GFP1-10 samples respectively (fig. 3.10). For the insoluble wNDP-K+GFP1-10 sample the median of calculated and normalized fluorescence was 1.02×10^3 . Finally, for the soluble protein sample, eNDP-K+GFP1-10, the calculated intensity was 5.32×10^3 . Using this information, the fluorescence intensity ratio of soluble to insoluble protein sample was estimated to be 5.21, by dividing the soluble to insoluble sample fluorescence intensities.

3.5. Discussion

At the start of this work, a proof of concept of high-throughput solubility system was designed. The assay design combines the main ideas of the previously described *in vivo* split GFP complementation system for protein solubility determination (Cabantous & Waldo, 2006) and droplet microfluidics.

In essence, the low-throughput split-GFP method based on agarose plates is transformed into a microscale agarose bead method, wherein each agarose bead grows a colony from a single bacterial cell. Before the experimental testing of the system, soluble and insoluble protein variants - mutant and wild type variants of nucleoside diphosphate kinase - are selected from the literature. The *in vivo* protein solubility was confirmed using a standard polyacrylamide gel electrophoresis - the soluble variant eNDP-K and insoluble variant wtNDP-K, had high intensity bands at the correct visualized soluble and insoluble fractions respectively (fig. 3.3, red lines). Confirmation of

the different solubility degrees of protein variants allowed to further use them for the testing of designed solubility system.

First, the solubility system was tested in large volumes (bulk format). After the initial experiments, fluorescence signal intensities of 13.42×10^3 and 6.87×10^3 (ratio of 1.95) were recorded for the soluble and insoluble samples respectively (fig. 3.5). We have hypothesized that large fluorescence signal from the insoluble protein variant may arise from the undepleted pTET inductor (AHT) activity during the pET induction. If the target protein expression is not stopped during the GFP1-10 induction, the GFP1-10 can bind small GFP11 fragment right after its expression, even if the protein is insoluble. In other words, it may take time for the protein to aggregate and hide its GFP11 tail. This hypothesis was tested by incorporating a 1-hour inductorless growth step between the inductions of pTET and pET plasmids. It was expected that target protein expression from pTET would stop completely during that hour. As a result of the additional step, fluorescence ratio between soluble and insoluble variants have increased from 1.95-fold to 5.7-fold (fig. 3.6). The ratio of soluble to insoluble sample fluorescence intensity was large enough to confidently separate them, thus we have decided to test the system in the high-throughput (agarose bead) setting.

The bacteria samples were encapsulated into 30 μm diameter agarose beads (fig. 3.8) and the whole designed HT solubility system was executed. The soluble and insoluble samples yielded GFP fluorescence emission signal strength of 5.32×10^3 and 1.02×10^3 (5.21-fold difference) respectively (fig. 3.10). The results of the proof of concept assay for HT solubility detection described demonstrate that the method successfully discriminate between soluble and insoluble proteins by a 5.21-fold fluorescence signal intensity.

In theory, this soluble to insoluble signal ratio may be enough for a binary type sorting (either soluble or insoluble), as it is described in literature that fluorescence changes as low as 2-fold can be successfully detected and samples may be sorted into two groups (You et al., 2014). However, further increase in positive to negative signal ratio would be beneficial. For example, with further ratio increase it may be possible to accurately discriminate solubility in a non-binary fashion, i.e. discriminate not only between fully insoluble and fully soluble proteins but also partly soluble. This may be important when collecting information on how changes in protein's amino acid sequence are influencing its solubility. It has been shown that sequence changes may influence the solubility in both minor and major ways, making the non-binary discrimination relevant (Trainor et al., 2017).

It is also important to acknowledge the multiple possible shortcomings of this HT solubility system. First, when analyzing large protein libraries, it may be difficult to determine what factor(s) influences the change in solubility. For example, it is shown that protein expression levels can vary depending on its amino acid sequence (Bivona et al., 2010). Furthermore, it is known that higher protein expression level reduces solubility significantly (Trevino et al., 2008). Therefore, it may be hard to tell whether the solubility changes are “direct” (sequence related) or “indirect” (expression related). In theory, this can be mitigated by expanding the system - adding a way to track the amount of protein produced in each bacterial cell or ensuring sequence independent expression level. Secondly, split GFP is not a dynamic reporter of solubility. Once the full GFP construct is complemented, the fluorescence is stable under physiological conditions. Third, it has been shown that even functional and/or soluble proteins may nonetheless be aggregated and directed to inclusion bodies (Armstrong et al., 1999; Davis et al., 1999; Makrides, 1996).

Finally, the proof of concept of the high-throughput system based on agarose beads for solubility determination may provide a way to quickly analyze large protein libraries. In the performed experiments, 7.09×10^6 of agarose beads were generated per hour. This speed can theoretically be further increased.

Future work should be focused on increasing the soluble to insoluble protein variant fluorescence intensity signal ratio, demonstrating the fluorescence activated sorting of the protein samples and screening of larger protein variant libraries.

CONCLUSIONS

1. Combination of GFP complementation system and agarose microparticles allowed a design of a theoretical system for high-throughput solubility.
2. The developed system permitted a correct assessment of solubility of the selected model both soluble and insoluble proteins – mutant and wild type variants of nucleoside diphosphate kinase, respectively.
3. Soluble and insoluble protein variants tested in bulk format yielded the complemented GFP fluorescence emission signal strength of 13.93×10^3 and 2.42×10^3 , respectively, resulting in a 5.7-fold difference, which is appropriate for evaluation of solubility.
4. Soluble and insoluble protein variants tested in 30 μm diameter agarose beads yielded the complemented GFP fluorescence emission signal strength of 5.32×10^3 and 1.02×10^3 , respectively, resulting in a 5.21-fold difference, which is appropriate for a high throughput evaluation of solubility of the target proteins.

ABSTRACT

Laurynas Karpus

Development of high-throughput *in vivo* protein solubility screening system

Solubility of proteins is an immensely important parameter for structural biologists, pharmaceutical industry and general scientist who work with proteins. In biotechnology, protein solubility can often be improved by optimizing the amino acid sequence of the target protein. Yet, currently used protein solubility assays are slow and tedious, making the screening for improved protein variants and conditions yielding improved solubility highly inefficient. Therefore, in this work, a high-throughput protein solubility assay was designed and experimentally validated. The developed assay is based on droplet microfluidics and a previously described low-throughput *in vivo* GFP complementation method. Proof of concept results of the assay showed that soluble and insoluble proteins can be discriminated with a 5.2-fold fluorescence signal intensity difference. The developed system provides a way to quickly analyze the solubility of large libraries of proteins.

SANTRAUKA

Laurynas Karpus

Didelio našumo *in vivo* baltymų tirpumo nustatymo sistemos kūrimas

Tirpumas yra itin svarbi baltymų savybė struktūriniais biologams, farmacijos industrijai ir visiems mokslininkams dirbantiems su rekombinantiniais baltymais. Biotechnologijoje, baltymų tirpumas dažnai gali būti pagerinamas optimizuojant tam tikro baltymo aminorūgščių seką. Tačiau, šiuo metu naudojami baltymų tirpumui nustatyti skirti metodai yra itin neefektyvūs dėl savo mažo našumo. Dėl šios priežasties, šiame darbe aprašoma ir eksperimentiškai patikrinama nauja, didelio našumo sistema baltymų tirpumo nustatymui. Ši sistema yra paremta lašelių mikroskopsčių technologijomis bei prieš tai aprašytu mažo našumo *in vivo* GFP komplementacijos metodu. Gauti eksperimentiniai sistemos rezultatai rodo, jog tirpūs ir netirpūs baltymų variantai gali būti atskiriami su 5,2 karto fluorescencijos signalo intensyvumo skirtumu. Sukurta sistema suteikia galimybę greitai analizuoti baltymų tirpumą didelėms baltymų bibliotekoms.

ACKNOWLEDGEMENT

I would like to thank my supervisor Linas Mažutis for allowing me to complete my thesis in the sector of microtechnologies, advising and guiding me on the whole system concept.

I also want to thank Rolandas Meškys, Greta Stonytė, Diana Iksalaitė and Simona Povilonienė for technical and intellectual support.

REFERENCES

1. Ahern, T. J., & Manning, M. C. (1992). *Stability of Protein Pharmaceuticals: Part B: In Vivo Pathways of Degradation and Strategies for Protein Stabilization*. Springer.
2. Alain, K., & Querellou, J. (2009). Cultivating the uncultured: limits, advances and future challenges. *Extremophiles: Life under Extreme Conditions*, 13(4), 583–594.
3. Anand, U., Garen, C., & Woodside, M. T. (2018). Probing the Single Molecule Folding Dynamics of Mammalian Prion Proteins from Species with Different Disease Susceptibility. In *Biophysical Journal* (Vol. 114, Issue 3, p. 355a). <https://doi.org/10.1016/j.bpj.2017.11.1976>
4. Arakawa, T., & Timasheff, S. N. (1985). [3]Theory of protein solubility. In *Methods in Enzymology* (pp. 49–77). [https://doi.org/10.1016/0076-6879\(85\)14005-x](https://doi.org/10.1016/0076-6879(85)14005-x)
5. Armstrong, N., de Lencastre, A., & Gouaux, E. (1999). A new protein folding screen: application to the ligand binding domains of a glutamate and kainate receptor and to lysozyme and carbonic anhydrase. *Protein Science: A Publication of the Protein Society*, 8(7), 1475–1483.
6. Bagby, S., Tong, K. I., & Ikura, M. (2001). Optimization of Protein Solubility and Stability for Protein Nuclear Magnetic Resonance. In *Methods in Enzymology* (pp. 20–41). [https://doi.org/10.1016/s0076-6879\(01\)39307-2](https://doi.org/10.1016/s0076-6879(01)39307-2)
7. Basova, E. Y., & Foret, F. (2015). Droplet microfluidics in (bio)chemical analysis. In *The Analyst* (Vol. 140, Issue 1, pp. 22–38). <https://doi.org/10.1039/c4an01209g>
8. Berry, S. B., Lee, J. J., Berthier, J., Berthier, E., & Theberge, A. B. (2019). Droplet incubation and splitting in open microfluidic channels. In *Analytical Methods* (Vol. 11, Issue 35, pp. 4528–4536). <https://doi.org/10.1039/c9ay00758j>
9. Bertone, P., Kluger, Y., Lan, N., Zheng, D., Christendat, D., Yee, A., Edwards, A. M., Arrowsmith, C. H., Montelione, G. T., & Gerstein, M. (2001). SPINE: an integrated tracking database and data mining approach for identifying feasible targets in high-throughput structural proteomics. *Nucleic Acids Research*, 29(13), 2884–2898.
10. Betton, J. M., & Hofnung, M. (1996). Folding of a mutant maltose-binding protein of *Escherichia coli* which forms inclusion bodies. *The Journal of Biological Chemistry*, 271(14), 8046–8052.
11. Bivona, L., Zou, Z., Stutzman, N., & Sun, P. D. (2010). Influence of the second amino acid on recombinant protein expression. *Protein Expression and Purification*, 74(2), 248–256.
12. Bjork, S. M., Sjostrom, S. L., Andersson-Svahn, H., & Joensson, H. N. (2015). Metabolite profiling of microfluidic cell culture conditions for droplet based screening. *Biomicrofluidics*, 9(4), 044128.
13. Blackwell, J. R., & Horgan, R. (1991). A novel strategy for production of a highly expressed recombinant protein in an active form. In *FEBS Letters* (Vol. 295, Issues 1-3, pp. 10–12). [https://doi.org/10.1016/0014-5793\(91\)81372-f](https://doi.org/10.1016/0014-5793(91)81372-f)
14. Bourot, S., Sire, O., Trautwetter, A., Touze, T., Wu, L. F., Blanco, C., & Bernard, T. (2000). Glycine Betaine-assisted Protein Folding in a lysA Mutant of *Escherichia coli*. In *Journal of Biological Chemistry* (Vol. 275, Issue 2, pp. 1050–1056). <https://doi.org/10.1074/jbc.275.2.1050>
15. Brown, C. R., Hong-Brown, L. Q., & Welch, W. J. (1997). Correcting temperature-sensitive protein folding defects. *The Journal of Clinical Investigation*, 99(6), 1432–1444.
16. Bruijn, L. I., Houseweart, M. K., Kato, S., Anderson, K. L., Anderson, S. D., Ohama, E., Reame, A. G., Scott, R. W., & Cleveland, D. W. (1998). Aggregation and motor neuron toxicity of an ALS-linked SOD1 mutant independent from wild-type SOD1. *Science*, 281(5384), 1851–1854.
17. Burgess, R. R. (2009). Chapter 20 Protein Precipitation Techniques. In *Methods in Enzymology* (pp. 331–342). [https://doi.org/10.1016/s0076-6879\(09\)63020-2](https://doi.org/10.1016/s0076-6879(09)63020-2)

18. Cabantous, S., Terwilliger, T. C., & Waldo, G. S. (2005). Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nature Biotechnology*, 23(1), 102–107.
19. Cabantous, S., & Waldo, G. S. (2006). In vivo and in vitro protein solubility assays using split GFP. *Nature Methods*, 3(10), 845–854.
20. Cacioppo, E., & Pusey, M. L. (1991). The solubility of the tetragonal form of hen egg white lysozyme from pH 4.0 to 5.4. In *Journal of Crystal Growth* (Vol. 114, Issue 3, pp. 286–292). [https://doi.org/10.1016/0022-0248\(91\)90043-5](https://doi.org/10.1016/0022-0248(91)90043-5)
21. Caen, O., Schütz, S., Jammalamadaka, M. S. S., Vrignon, J., Nizard, P., Schneider, T. M., Baret, J.-C., & Taly, V. (2018). High-throughput multiplexed fluorescence-activated droplet sorting. *Microsystems & Nanoengineering*, 4, 33.
22. Caldwell, G. W., Ritchie, D. M., Masucci, J. A., Hageman, W., & Yan, Z. (2001). The New Pre-Preclinical Paradigm: Compound Optimization in Early and Late Phase Drug Discovery. In *Current Topics in Medicinal Chemistry* (Vol. 1, Issue 5, pp. 353–366). <https://doi.org/10.2174/1568026013394949>
23. Chan, P., Curtis, R. A., & Warwicker, J. (2013). Soluble expression of proteins correlates with a lack of positively-charged surface. *Scientific Reports*, 3, 3333.
24. Chen, J., Vestergaard, M., Shen, J., Solem, C., Dufva, M., & Jensen, P. R. (2018). Droplet-based microfluidics as a future tool for strain improvement in lactic acid bacteria. *FEMS Microbiology Letters*, 365(23). <https://doi.org/10.1093/femsle/fny258>
25. Chen, R., Sun, Z., & Chen, D. (2019). Droplet-based microfluidics for cell encapsulation and delivery. In *Microfluidics for Pharmaceutical Applications* (pp. 307–335). <https://doi.org/10.1016/b978-0-12-812659-2.00011-9>
26. Chong, Z. Z., Tan, S. H., Gañán-Calvo, A. M., Tor, S. B., Loh, N. H., & Nguyen, N.-T. (2016). Active droplet generation in microfluidics. *Lab on a Chip*, 16(1), 35–58.
27. Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Savchenko, A., Cort, J. R., Booth, V., Mackereth, C. D., Saridakis, V., Ekiel, I., Kozlov, G., Maxwell, K. L., Wu, N., McIntosh, L. P., Gehring, K., Kennedy, M. A., Davidson, A. R., Pai, E. F., Gerstein, M., ... Arrowsmith, C. H. (2000). Structural proteomics of an archaeon. *Nature Structural Biology*, 7(10), 903–909.
28. Churski, K., Korczyk, P., & Garstecki, P. (2010). High-throughput automated droplet microfluidic system for screening of reaction conditions. *Lab on a Chip*, 10(7), 816–818.
29. Collins, K. D. (1997). Charge density-dependent strength of hydration and biological structure. In *Biophysical Journal* (Vol. 72, Issue 1, pp. 65–76). [https://doi.org/10.1016/s0006-3495\(97\)78647-8](https://doi.org/10.1016/s0006-3495(97)78647-8)
30. Colon, W., & Kelly, J. W. (1992). Partial denaturation of transthyretin is sufficient for amyloid fibril formation in vitro. *Biochemistry*, 31(36), 8654–8660.
31. Dahal, Y. R., & Schmit, J. D. (2018). Ion Specificity and Nonmonotonic Protein Solubility from Salt Entropy. *Biophysical Journal*, 114(1), 76–87.
32. Davies, S. W., Turmaine, M., Cozens, B. A., DiFiglia, M., Sharp, A. H., Ross, C. A., Scherzinger, E., Wanker, E. E., Mangiarini, L., & Bates, G. P. (1997). Formation of Neuronal Intranuclear Inclusions Underlies the Neurological Dysfunction in Mice Transgenic for the HD Mutation. In *Cell* (Vol. 90, Issue 3, pp. 537–548). [https://doi.org/10.1016/s0092-8674\(00\)80513-9](https://doi.org/10.1016/s0092-8674(00)80513-9)
33. Davis, G. D., Elisee, C., Newham, D. M., & Harrison, R. G. (1999). New fusion protein systems designed to give soluble expression in *Escherichia coli*. *Biotechnology and Bioengineering*, 65(4), 382–388.
34. Dixon, M., & Webb, E. C. (1961). Enzyme fractionation by salting-out: a theoretical note. *Advances in Protein Chemistry*, 16, 197–219.
35. Dobson, C. M. (1999). Protein misfolding, evolution and disease. In *Trends in Biochemical Sciences* (Vol. 24, Issue 9, pp. 329–332). [https://doi.org/10.1016/s0968-0004\(99\)01445-0](https://doi.org/10.1016/s0968-0004(99)01445-0)

36. Doi, N., Kakukawa, K., Oishi, Y., & Yanagawa, H. (2005). High solubility of random-sequence proteins consisting of five kinds of primitive amino acids. *Protein Engineering, Design & Selection: PEDS*, 18(6), 279–284.
37. Du, G., Fang, Q., & den Toonder, J. M. J. (2016). Microfluidics for cell-based high throughput screening platforms - A review. *Analytica Chimica Acta*, 903, 36–50.
38. Endo, A., & Kurusu, Y. (2007). Identification of in vivo substrates of the chaperonin GroEL from *Bacillus subtilis*. *Bioscience, Biotechnology, and Biochemistry*, 71(4), 1073–1077.
39. Enogieru, A. B., Omoruyi, S. I., Hiss, D. C., & Ekpo, O. E. (2019). GRP78/BIP/HSPA5 as a Therapeutic Target in Models of Parkinson's Disease: A Mini Review. *Advances in Pharmacological Sciences*, 2019, 2706783.
40. Eun, Y.-J., Utada, A. S., Copeland, M. F., Takeuchi, S., & Weibel, D. B. (2011). Encapsulating bacteria in agarose microparticles using microfluidics for high-throughput cell analysis and isolation. *ACS Chemical Biology*, 6(3), 260–266.
41. Evans, P., Wyatt, K., Wistow, G. J., Bateman, O. A., Wallace, B. A., & Slingsby, C. (2004). The P23T cataract mutation causes loss of solubility of folded gammaD-crystallin. *Journal of Molecular Biology*, 343(2), 435–444.
42. Foster, P. R., Dunnill, P., & Lilly, M. D. (1976). The kinetics of protein salting-out: Precipitation of yeast enzymes by ammonium sulfate. In *Biotechnology and Bioengineering* (Vol. 18, Issue 10, pp. 1496–1496). <https://doi.org/10.1002/bit.260181019>
43. Fowler, S. B., Poon, S., Muff, R., Chiti, F., Dobson, C. M., & Zurdo, J. (2005). Rational design of aggregation-resistant bioactive peptides: reengineering human calcitonin. *Proceedings of the National Academy of Sciences of the United States of America*, 102(29), 10105–10110.
44. Fox, S. W., & Foster, J. F. (1957). *Introduction to Protein Chemistry*.
45. Galvin, J. E., Uryu, K., Lee, V. M., & Trojanowski, J. Q. (1999). Axon pathology in Parkinson's disease and Lewy body dementia hippocampus contains alpha-, beta-, and gamma-synuclein. *Proceedings of the National Academy of Sciences of the United States of America*, 96(23), 13450–13455.
46. Goh, C.-S., Lan, N., Douglas, S. M., Wu, B., Echols, N., Smith, A., Milburn, D., Montelione, G. T., Zhao, H., & Gerstein, M. (2004). Mining the structural genomics pipeline: identification of protein properties that affect high-throughput experimental analysis. *Journal of Molecular Biology*, 336(1), 115–130.
47. Guilloteau, J.-P., Rie`s-Kautt, M. M., & Ducruix, A. F. (1992). Variation of lysozyme solubility as a function of temperature in the presence of organic and inorganic salts. In *Journal of Crystal Growth* (Vol. 122, Issues 1-4, pp. 223–230). [https://doi.org/10.1016/0022-0248\(92\)90249-i](https://doi.org/10.1016/0022-0248(92)90249-i)
48. Han, X., Ning, W., Ma, X., Wang, X., & Zhou, K. (n.d.). *Improve Protein Solubility and Activity based on Machine Learning Models*. <https://doi.org/10.1101/817890>
49. Headen, D. M., García, J. R., & García, A. J. (2018). Parallel droplet microfluidics for high throughput cell encapsulation and synthetic microgel generation. In *Microsystems & Nanoengineering* (Vol. 4, Issue 1). <https://doi.org/10.1038/micronano.2017.76>
50. Hind, C. R., Tennent, G. A., Evans, D. J., & Pepys, M. B. (1983). Demonstration of amyloid A (AA) protein and amyloid P component (AP) in deposits of systemic amyloidosis associated with renal adenocarcinoma. *The Journal of Pathology*, 139(2), 159–166.
51. Howard, S. B., Twigg, P. J., Baird, J. K., & Meehan, E. J. (1988). The solubility of hen egg-white lysozyme. In *Journal of Crystal Growth* (Vol. 90, Issues 1-3, pp. 94–104). [https://doi.org/10.1016/0022-0248\(88\)90303-x](https://doi.org/10.1016/0022-0248(88)90303-x)
52. Huang, B., Eberstadt, M., Olejniczak, E. T., Meadows, R. P., & Fesik, S. W. (1996). NMR structure and mutagenesis of the Fas (APO-1/CD95) death domain. *Nature*, 384(6610), 638–641.
53. Huebner, A., Srisa-Art, M., Holt, D., Abell, C., Hollfelder, F., deMello, A. J., & Edel, J. B.

- (2007). Quantitative detection of protein expression in single cells using droplet microfluidics. *Chemical Communications*, 12, 1218–1220.
54. Jakiela, S., Kaminski, T. S., Cybulski, O., Weibel, D. B., & Garstecki, P. (2013). Bacterial growth and adaptation in microdroplet chemostats. *Angewandte Chemie*, 52(34), 8908–8911.
55. Kakalis, L. T., & Regenstein, J. M. (1986). Effect of pH and Salts on the Solubility of Egg White Protein. In *Journal of Food Science* (Vol. 51, Issue 6, pp. 1445–1447). <https://doi.org/10.1111/j.1365-2621.1986.tb13830.x>
56. Kalra, A., Tugcu, N., Cramer, S. M., & Garde, S. (2001). Salting-In and Salting-Out of Hydrophobic Solutes in Aqueous Salt Solutions. In *The Journal of Physical Chemistry B* (Vol. 105, Issue 27, pp. 6380–6386). <https://doi.org/10.1021/jp010568+>
57. Katsuragi, T., Tanaka, S., Nagahiro, S., & Tani, Y. (2000). Gel microdroplet technique leaving microorganisms alive for sorting by flow cytometry. In *Journal of Microbiological Methods* (Vol. 42, Issue 1, pp. 81–86). [https://doi.org/10.1016/s0167-7012\(00\)00179-2](https://doi.org/10.1016/s0167-7012(00)00179-2)
58. Kaushik, A. M., Hsieh, K., & Wang, T.-H. (2018). Droplet microfluidics for high-sensitivity and high-throughput detection and screening of disease biomarkers. *Wiley Interdisciplinary Reviews. Nanomedicine and Nanobiotechnology*, 10(6), e1522.
59. Kawaguchi, Y., Okamoto, T., Taniwaki, M., Aizawa, M., Inoue, M., Katayama, S., Kawakami, H., Nakamura, S., Nishimura, M., Akiguchi, I., Kimura, J., Narumiya, S., & Kakizuka, A. (1994). CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. In *Nature Genetics* (Vol. 8, Issue 3, pp. 221–228). <https://doi.org/10.1038/ng1194-221>
60. Khurana, S., Rawi, R., Kunji, K., Chuang, G.-Y., Bensmail, H., & Mall, R. (2018). DeepSol: a deep learning framework for sequence-based protein solubility prediction. *Bioinformatics*, 34(15), 2605–2613.
61. Kim, J. C., & Lund, D. B. (1998). MILK PROTEIN/STAINLESS STEEL INTERACTION RELEVANT TO THE INITIAL STAGE OF FOULING IN THERMAL PROCESSING. In *Journal of Food Process Engineering* (Vol. 21, Issue 5, pp. 369–386). <https://doi.org/10.1111/j.1745-4530.1998.tb00459.x>
62. Kim, W., & Hecht, M. H. (2006). Generic hydrophobic residues are sufficient to promote aggregation of the Alzheimer's Aβ₄₂ peptide. *Proceedings of the National Academy of Sciences of the United States of America*, 103(43), 15824–15829.
63. Knaust, R. K. C., & Nordlund, P. (2001). Screening for Soluble Expression of Recombinant Proteins in a 96-Well Format. In *Analytical Biochemistry* (Vol. 297, Issue 1, pp. 79–85). <https://doi.org/10.1006/abio.2001.5331>
64. Lai, J. K., Ambia, J., Wang, Y., & Barth, P. (2017). Enhancing Structure Prediction and Design of Soluble and Membrane Proteins with Explicit Solvent-Protein Interactions. *Structure*, 25(11), 1758–1770.e8.
65. Lan, F., Demaree, B., Ahmed, N., & Abate, A. R. (2017). Single-cell genome sequencing at ultra-high-throughput with microfluidic droplet barcoding. *Nature Biotechnology*, 35(7), 640–646.
66. Langendorff, V., Cuvelier, G., Launay, B., Michon, C., Parker, A., & De Kruijff, C. G. (1999). Casein micelle/ι-carrageenan interactions in milk: influence of temperature. In *Food Hydrocolloids* (Vol. 13, Issue 3, pp. 211–218). [https://doi.org/10.1016/s0268-005x\(98\)00087-3](https://doi.org/10.1016/s0268-005x(98)00087-3)
67. Leung, K., Zahn, H., Leaver, T., Konwar, K. M., Hanson, N. W., Pagé, A. P., Lo, C.-C., Chain, P. S., Hallam, S. J., & Hansen, C. L. (2012). A programmable droplet-based microfluidic device applied to multiparameter analysis of single microbes and microbial communities. *Proceedings of the National Academy of Sciences of the United States of America*, 109(20), 7665–7670.
68. Liu, E. Y., Jung, S., Weitz, D. A., Yi, H., & Choi, C.-H. (2018). High-throughput double emulsion-based microfluidic production of hydrogel microspheres with tunable chemical

- functionalities toward biomolecular conjugation. *Lab on a Chip*, 18(2), 323–334.
69. Liu, J.-L., & Li, C.-L. (2019). A generalized Debye-Hückel theory of electrolyte solutions. In *AIP Advances* (Vol. 9, Issue 1, p. 015214). <https://doi.org/10.1063/1.5081863>
 70. Long, W. F., & Labute, P. (2010). Calibrative approaches to protein solubility modeling of a mutant series using physicochemical descriptors. *Journal of Computer-Aided Molecular Design*, 24(11), 907–916.
 71. Luan, C.-H., & -H. Luan, C. (2004). High-Throughput Expression of *C. elegans* Proteins. In *Genome Research* (Vol. 14, Issue 10b, pp. 2102–2110). <https://doi.org/10.1101/gr.2520504>
 72. Luo, D., Pullela, S. R., Marquez, M., & Cheng, Z. (2007). Cell capsules with tunable transport and mechanical properties. *Biomicrofluidics*, 1(3), 34102.
 73. Macdonald, M. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. In *Cell* (Vol. 72, Issue 6, pp. 971–983). [https://doi.org/10.1016/0092-8674\(93\)90585-e](https://doi.org/10.1016/0092-8674(93)90585-e)
 74. Makrides, S. C. (1996). Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiological Reviews*, 60(3), 512–538.
 75. Martin, K., Henkel, T., Baier, V., Grodrian, A., Schön, T., Roth, M., Michael Köhler, J., & Metzger, J. (2003). Generation of larger numbers of separated microbial populations by cultivation in segmented-flow microdevices. *Lab on a Chip*, 3(3), 202–207.
 76. Mazutis, L., Gilbert, J., Ung, W. L., Weitz, D. A., Griffiths, A. D., & Heyman, J. A. (2013). Single-cell analysis and sorting using droplet-based microfluidics. *Nature Protocols*, 8(5), 870–891.
 77. McDonald, J. C., Duffy, D. C., Anderson, J. R., Chiu, D. T., Wu, H., Schueller, O. J., & Whitesides, G. M. (2000). Fabrication of microfluidic systems in poly(dimethylsiloxane). *Electrophoresis*, 21(1), 27–40.
 78. McMeekin, T. L., Cohn, E. J., & Blanchard, M. H. (1937). Studies in the Physical Chemistry of Amino Acids, Peptides and Related Substances. X. The Solubility of Cystine in Solutions of Chlorides and Sulfates. In *Journal of the American Chemical Society* (Vol. 59, Issue 12, pp. 2717–2723). <https://doi.org/10.1021/ja01291a073>
 79. McPhaul, M. J. (n.d.). Alterations of Androgen Action Caused by Mutation of the Human Androgen Receptor. In *Androgens in Health and Disease* (pp. 103–122). <https://doi.org/10.1385/1-59259-388-7:103>
 80. Meyer, P.-F., McSweeney, M., Gonneaud, J., & Villeneuve, S. (2019). AD molecular: PET amyloid imaging across the Alzheimer's disease spectrum: From disease mechanisms to prevention. *Progress in Molecular Biology and Translational Science*, 165, 63–106.
 81. Mine, Y. (1995). Recent advances in the understanding of egg white protein functionality. In *Trends in Food Science & Technology* (Vol. 6, Issue 7, pp. 225–232). [https://doi.org/10.1016/s0924-2244\(00\)89083-4](https://doi.org/10.1016/s0924-2244(00)89083-4)
 82. Pace, C. N., Nick Pace, C., Fu, H., Fryar, K. L., Landua, J., Trevino, S. R., Schell, D., Thurlkill, R. L., Imura, S., Martin Scholtz, J., Gajiwala, K., Sevcik, J., Urbanikova, L., Myers, J. K., Takano, K., Hebert, E. J., Shirley, B. A., & Grimsley, G. R. (2014). Contribution of hydrogen bonds to protein stability. In *Protein Science* (Vol. 23, Issue 5, pp. 652–661). <https://doi.org/10.1002/pro.2449>
 83. Pace, C. N., Nick Pace, C., Martin Scholtz, J., & Grimsley, G. R. (2014). Forces stabilizing proteins. In *FEBS Letters* (Vol. 588, Issue 14, pp. 2177–2184). <https://doi.org/10.1016/j.febslet.2014.05.006>
 84. Pace, C. N., Treviño, S., Prabhakaran, E., & Scholtz, J. M. (2004). Protein structure, stability and solubility in water and other solvents. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 359(1448), 1225–1234; discussion 1234–1235.
 85. Pande, A., Annunziata, O., Asherie, N., Ogun, O., Benedek, G. B., & Pande, J. (2005).

- Decrease in Protein Solubility and Cataract Formation Caused by the Pro23 to Thr Mutation in Human γ D-Crystallin^{†,‡}. In *Biochemistry* (Vol. 44, Issue 7, pp. 2491–2500). <https://doi.org/10.1021/bi0479611>
86. Pédélecq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C., & Waldo, G. S. (2006). Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology*, 24(1), 79–88.
 87. Pédélecq, J.-D., Piltch, E., Liong, E. C., Berendzen, J., Kim, C.-Y., Rho, B.-S., Park, M. S., Terwilliger, T. C., & Waldo, G. S. (2002). Engineering soluble proteins for structural genomics. *Nature Biotechnology*, 20(9), 927–932.
 88. Pham, N., Radajewski, D., Round, A., Brennich, M., Pernot, P., Biscans, B., Bonneté, F., & Teychené, S. (2017). Coupling High Throughput Microfluidics and Small-Angle X-ray Scattering to Study Protein Crystallization from Solution. *Analytical Chemistry*, 89(4), 2282–2287.
 89. Prestrelski, S. J., Arakawa, T., & Carpenter, J. F. (1993). Separation of freezing- and drying-induced denaturation of lyophilized proteins using stress-specific stabilization. II. Structural studies using infrared spectroscopy. *Archives of Biochemistry and Biophysics*, 303(2), 465–473.
 90. Qamar, S., Islam, M., & Tayyab, S. (1993). Probing the Determinants of Protein Solubility with Amino Acid Modification I. In *The Journal of Biochemistry* (Vol. 114, Issue 6, pp. 786–792). <https://doi.org/10.1093/oxfordjournals.jbchem.a124257>
 91. Rao, V. R., Cohen, G. B., & Oprian, D. D. (1994). Rhodopsin mutation G90D and a molecular mechanism for congenital night blindness. In *Nature* (Vol. 367, Issue 6464, pp. 639–642). <https://doi.org/10.1038/367639a0>
 92. Raveshi, M. R., Agnihotri, S. N., Sesen, M., Bhardwaj, R., & Neild, A. (2019). Selective droplet splitting using single layer microfluidic valves. In *Sensors and Actuators B: Chemical* (Vol. 292, pp. 233–240). <https://doi.org/10.1016/j.snb.2019.04.115>
 93. Regenstein, J. M., & Regenstein, C. E. (1984). Salting In and Salting Out. In *Food Protein Chemistry* (pp. 117–129). <https://doi.org/10.1016/b978-0-12-585820-5.50017-8>
 94. Ricci, M., & Brems, D. (2004). Common Structural Stability Properties of 4-Helical Bundle Cytokines: Possible Physiological and Pharmaceutical Consequences. In *Current Pharmaceutical Design* (Vol. 10, Issue 31, pp. 3901–3911). <https://doi.org/10.2174/1381612043382611>
 95. Riès-Kautt, M., & Ducruix, A. (1997). [3] Inferences drawn from physicochemical studies of crystallogenesis and precrystalline state. *Methods in Enzymology*, 276, 23–59.
 96. Sandstrom, W. M. (1930). Physico-chemical Studies on Proteins. IV. In *The Journal of Physical Chemistry* (Vol. 34, Issue 5, pp. 1071–1101). <https://doi.org/10.1021/j150311a016>
 97. Schein, C. H. (1990). Solubility as a Function of Protein Structure and Solvent Components. In *Nature Biotechnology* (Vol. 8, Issue 4, pp. 308–317). <https://doi.org/10.1038/nbt0490-308>
 98. Schein, C. H. (1993). Solubility and secretability. *Current Opinion in Biotechnology*, 4(4), 456–461.
 99. Shang, L., Cheng, Y., & Zhao, Y. (2017). Emerging Droplet Microfluidics. In *Chemical Reviews* (Vol. 117, Issue 12, pp. 7964–8040). <https://doi.org/10.1021/acs.chemrev.6b00848>
 100. Shire, S. J., Shahrokh, Z., & Liu, J. (2010). Challenges in the Development of High Protein Concentration Formulations. In *Current Trends in Monoclonal Antibody Development and Manufacturing* (pp. 131–147). https://doi.org/10.1007/978-0-387-76643-0_9
 101. Singh, H., & Ye, A. (2008). Interactions and functionality of milk proteins in food emulsions. In *Milk Proteins* (pp. 321–345). <https://doi.org/10.1016/b978-0-12-374039-7.00011-8>
 102. Smith, D. (2000). Functional properties of muscle proteins in processed poultry products. In *Poultry Meat Processing*. <https://doi.org/10.1201/9781420042177.ch11>
 103. Sormanni, P., & Vendruscolo, M. (2019). Protein Solubility Predictions Using the CamSol Method in the Study of Protein Homeostasis. In *Cold Spring Harbor Perspectives in Biology* (Vol. 11, Issue 12, p. a033845). <https://doi.org/10.1101/cshperspect.a033845>

104. Spada, A. R. L., La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E., & Fischbeck, K. H. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. In *Nature* (Vol. 352, Issue 6330, pp. 77–79). <https://doi.org/10.1038/352077a0>
105. Suea-Ngam, A., Howes, P. D., Srisa-Art, M., & deMello, A. J. (2019). Droplet microfluidics: from proof-of-concept to real-world utility? *Chemical Communications*, 55(67), 9895–9903.
106. Sugihara, J., & Baldwin, T. O. (1988). Effects of 3' end deletions from the *Vibrio harveyi* luxB gene on luciferase subunit folding and enzyme assembly: generation of temperature-sensitive polypeptide folding mutants. *Biochemistry*, 27(8), 2872–2880.
107. Tanford, C. (1961). *Physical Chemistry of Macromolecules*.
108. Thomas, P. J., Ko, Y. H., & Pedersen, P. L. (1992). Altered protein folding may be the molecular basis of most cases of cystic fibrosis. *FEBS Letters*, 312(1), 7–9.
109. Thomas, P. J., Qu, B. H., & Pedersen, P. L. (1995). Defective protein folding as a basis of human disease. *Trends in Biochemical Sciences*, 20(11), 456–459.
110. Touati, A., Creuzenet, C., Chobert, J. M., Dufour, E., & Haertl, T. (1992). Solubility and reactivity of caseins and β -lactoglobulin in protic solvents. In *Journal of Protein Chemistry* (Vol. 11, Issue 6, pp. 613–621). <https://doi.org/10.1007/bf01024961>
111. Trainor, K., Broom, A., & Meiering, E. M. (2017). Exploring the relationships between protein sequence, structure and solubility. *Current Opinion in Structural Biology*, 42, 136–146.
112. Trevino, S. R., Martin Scholtz, J., & Nick Pace, C. (2007). Amino Acid Contribution to Protein Solubility: Asp, Glu, and Ser Contribute more Favorably than the other Hydrophilic Amino Acids in RNase Sa. In *Journal of Molecular Biology* (Vol. 366, Issue 2, pp. 449–460). <https://doi.org/10.1016/j.jmb.2006.10.026>
113. Trevino, S. R., Scholtz, J. M., & Pace, C. N. (2008). Measuring and increasing protein solubility. *Journal of Pharmaceutical Sciences*, 97(10), 4155–4166.
114. Varma, V. B., Ray, A., Wang, Z. M., Wang, Z. P., & Ramanujan, R. V. (2016). Droplet Merging on a Lab-on-a-Chip Platform by Uniform Magnetic Fields. *Scientific Reports*, 6, 37671.
115. Vojdani, F. (1996). Solubility. In *Methods of Testing Protein Functionality* (pp. 11–60). https://doi.org/10.1007/978-1-4613-1219-2_2
116. Weibel, D. B., Diluzio, W. R., & Whitesides, G. M. (2007). Microfabrication meets microbiology. *Nature Reviews. Microbiology*, 5(3), 209–218.
117. Wigley, W. C., Stidham, R. D., Smith, N. M., Hunt, J. F., & Thomas, P. J. (2001). Protein solubility and folding monitored in vivo by structural complementation of a genetic marker protein. *Nature Biotechnology*, 19(2), 131–136.
118. Wilkinson, D. L., & Harrison, R. G. (1991). Predicting the solubility of recombinant proteins in *Escherichia coli*. *Bio/technology*, 9(5), 443–448.
119. Wynn, R. M., Davie, J. R., Cox, R. P., & Chuang, D. T. (1992). Chaperonins groEL and groES promote assembly of heterotetramers (α 2 β 2) of mammalian mitochondrial branched-chain alpha-keto acid decarboxylase in *Escherichia coli*. *The Journal of Biological Chemistry*, 267(18), 12400–12403.
120. Xiao, T., & Song, X. (2017). A molecular Debye-Hückel theory and its applications to electrolyte solutions: The size asymmetric case. *The Journal of Chemical Physics*, 146(12), 124118.
121. You, M., Lim, S.-H., Kim, M.-J., Jeong, Y., Lee, M.-G., & Ha, S.-H. (2014). Improvement of the Fluorescence Intensity during a Flow Cytometric Analysis for Rice Protoplasts by Localization of a Green Fluorescent Protein into Chloroplasts. In *International Journal of Molecular Sciences* (Vol. 16, Issue 1, pp. 788–804). <https://doi.org/10.3390/ijms16010788>
122. Zhu, P., & Wang, L. (2016). Passive and active droplet generation with microfluidics: a review. *Lab on a Chip*, 17(1), 34–75.