Strategies for Alleviating *in vitro* Low Protein Solubility in Applied Biotechnology and Microbiology

Aiya Chantarasiri*
Faculty of Science, Energy and Environment, King Mongkut's University of Technology North Bangkok, Rayong Campus, Rayong, Thailand

Vithaya Meevootisom, Suthep Wiyakrutta
Department of Microbiology, Faculty of Science, Mahidol University, Bangkok, Thailand

* Corresponding author. E-mail: aiyac@kmutnb.ac.th
Received: 19 August 2014; Accepted: 17 December 2014; Published online: 4 February 2015

Abstract
From research to application, high protein solubility is usually a desired property yet sometimes difficult to achieve. The in vitro low solubility of the fully folded proteins is relevant to applied microbiological studies, biochemical studies, biopharmaceutical studies, high-resolution structural studies, and applications demanding high protein concentration. This insufficient protein solubility depends largely on the surface property of the protein molecule. To alleviate this problem, approaches emphasized on the improvement of water-binding ability or prevention of protein aggregation were employed including the use of chemical additives, fusion with solubility enhancement tags, and molecular engineering of the surface amino acid residues. With the availability of the three-dimensional structure of the target proteins, the effect of different surface amino acid residues on protein solubility could be systematically investigated. With the applications of advanced bioinformatics tools and guided by protein three-dimensional structure, solubility-improving mutagenesis can be designed and executed with a high chance of success. Integrating rational molecular engineering with other available approaches will be the effective strategy for alleviating in vitro low solubility of important proteins in the future.

Keywords: *Protein solubility, Protein aggregation, Molecular engineering, Solubility-improving mutagenesis*

1 Introduction

Protein engineering has been developed for more than half a century and still plays increasingly significance roles in several fields, particularly applied microbiology, biochemistry, biotechnology, and biopharmaceutical applications of proteins and enzymes. High kinetic activity, high protein solubility, and long-term stability are properties generally required in research and applications. Investigations on improving of kinetic activity and stability have been the topics of active research, but study for increasing protein solubility receives less attention. Cumulative documents reveal that up to 80% of the identified non-membrane proteins are low protein solubility [1-3]. This problem often obstructs the preparation of concentrated protein for high-resolution structural determination, quantitative binding assays, and characterization of the novel proteins. Likewise, more than 90% of all pharmaceutical proteins are unsuitable for pre-clinical study because of their low solubility [4]. It is evident that low protein solubility is one of the main

Please cite this article as: A. Chantarasiri, V. Meevootisom, and S. Wiyakrutta, "Strategies for Alleviating *in vitro* Low Protein Solubility in Applied Biotechnology and Microbiology," *KMUTNB Int J Appl Sci Technol*, Vol. 8, No. 2, pp. 137-143, Apr.-June 2015, http://dx.doi.org/10.14416/j.ijast.2014.12.003

bottlenecks in protein technology, which can happen during production, purification, preparation, shipping, storage, and downstream application steps. Various techniques for increasing protein solubility have been developed. Solubility of purified protein could be improved by dissolving in optimized buffer formulated with selected salts and additives [5]. Genetic fusion with short peptide (solubility enhancement tag) is an easy and commercially available method for improvement of protein solubility. Recently, rational mutagenesis of surface amino acid residues to enhance protein solubility has been developed [6,7]. Beside those experimental techniques, bioinformatics tools for molecular modeling and predicting protein solubility from the amino acid sequence or three-dimensional structure [8,9] contribute significantly to the success in solubility-enhancing protein engineering. In this review, the definition of low protein solubility and the strategies for alleviating *in vitro* low protein solubility, especially the novel approaches are discussed.

2 Definition and Background of Protein Solubility

Thermodynamically, solubility is the relation between the nature of the solute and the solvent defined as an amount of a substance (a solute) that can be totally dissolved in a given amount of a solvent. Accordingly, protein solubility is defined as an amount of protein that can be completely dissolved in the water or buffer under the given environmental conditions and solution components. In biochemical and pharmaceutical study of proteins, the equilibrium solubility which is the concentration of a solute in its saturated solution exists in a state of equilibrium with pure solid solute is usually considered [10]. The concentration of the solute in the solution is constant as long as the environment factors and solution components are maintained. In practice, protein solubility is determined from the concentration of soluble protein in equilibrium with the solid phase under given conditions of pH, temperature, buffer concentration, and additives [11]. If the concentration of a protein exceeds its equilibrium solubility limit, the solution becomes supersaturated and protein in the solution moves to the insoluble phase either as an amorphous precipitation (disordered aggregation) or as a microcrystalline form [12] until a new solution equilibrium is established.

Protein solubility is a complex phenomenon affected by multiple intertwined factors and remains to be fully understood [13]. Factors influencing protein solubility can be classified into two groups, the extrinsic and intrinsic factors. Extrinsic or environment factors include conditions of pH, temperature, ionic strength, ionic composition, and solution composition. The pH of the solution is generally considered as the important extrinsic factor that influences protein solubility [14]. The intrinsic factor corresponds with the pH effect of the solution is the isoelectric point (p*I*) of the protein. The p*I* is the pH at which the protein has a net charge of zero or the balance of positive and negative charges. Protein solubility reaches a minimum when the pH of the environment is equal to the p*I*. Moving away from the p*I* value, either to the higher or lower pH, usually increases the protein solubility. Thus, knowing the p*I* of a protein is crucial for manipulating its solubility. Intrinsic properties of a protein are governed by the amino acid sequence that determines the folding process, the conformational structure, the solvent accessible surface area, and the net charge of the protein, all of which can affect protein solubility.

3 Classification of Low Protein Solubility

There are several types of low protein solubility reported in the literature, but the classification has been ambiguous and confusing [11]. Up to the present, the most useful low protein solubility classification is that proposed by Trevino [11] which discriminates low protein solubility into four types based on pharmaceutical applications. These are (i) low *in vitro* solubility, (ii) low *in vivo* solubility, (iii) amyloid formation, and (iv) low protein solubility due to conformational changes. In the present review, however, the classification of low protein solubility is re-organized into only two types which are (i) the *in vivo* low protein solubility caused by the failure to reach the fully folded state during expression in the host cell and (ii) the *in vitro* low solubility of the fully folded proteins after isolation from the host cell. The advantage of this new classification is to prevent the confusion in terms of the definition and the determination of strategies for improving protein solubility relevant to biotechnology and microbiology.

3.1 *In vivo low protein solubility*

Cytoplasmic and periplasmic inclusion body (IB) formation is a common problem of high-level expression of heterologous proteins in *Escherichia coli*. Generally, the over-expressed level of the heterologous proteins, the hydrophobic characteristic of the proteins, insufficient number of molecular chaperones, and the reducing condition in the cytoplasm of *E. coli* are the major causes of imperfect folding or misfolding of the proteins. When the proteins unfold or incorrectly fold, hydrophobic amino acid residues (Ala, Val, Ile, Leu, Met, Phe, Trp, and Cys) which are normally buried in the folded state become exposed [15]. The contacts between nonspecific regions intra- or inter-molecularly lead to amorphous (disordered) aggregations through non-covalent hydrophobic or ionic interactions or the combination of both [16] forming inclusion bodies. The inclusion bodies are dynamic structures formed by an unbalanced equilibrium between aggregated (solid phase) and soluble proteins in the expressing host cell [17]. In the review by Trevino [11], this low protein solubility type was classified in the group of low *in vivo* solubility.

Evidence suggested an association between inclusion body and the conformational instability of protein folding process. Using statistical analysis of the amino acid compositions of 81 proteins, it was reported that proteins in their partially unfolded conformation especially at early intermediate stage could generate insoluble inclusion body [18]. Based on mutational analysis, it was proposed that the success of the folding process of a protein depends on it having sufficient conformational stability [19].

3.2 *In vitro low protein solubility*

Another type of solubility problem addressed in this review is low solubility of the fully folded protein (or low *in vitro* solubility). In this case, the proteins can be expressed, folded properly, and can be purified in their native form, but cannot be concentrated sufficiently for further applications such as biochemical studies and pharmaceutical use. Protein aggregation or precipitation is driven mainly by surface interaction of the folded state. It is different from the aggregation of partially or fully unfolded protein that is caused by the exposed hydrophobic cores. Proteins at the concentrations of several mg/mL are usually required for structural analyses such as X-ray crystallography and nuclear magnetic resonance technique. Therapeutic monoclonal antibodies need to be prepared at more than one hundred mg/mL in a small volume [6] for subcutaneous administration [20]. High protein concentration is difficult to achieve for most expressed proteins because they have limited solubility and propensity to aggregate or precipitate. Under any particular controlled extrinsic conditions, key intrinsic determinants for precipitation of the fully folded proteins include net charge, hydrophobicity, size, and conformation of the proteins.

The amino acid residues whose side chain are exposed on the surface of the folded protein molecule contribute to the net charge and hydrophobicity which are the two major intrinsic factors that determine solubility of the protein. The effect of net charge on protein solubility at the controlled pH was systematically analyzed using the wild-type and mutants of ribonuclease Sa (RNase Sa) [21]. The results confirmed that the net charge of a protein depends on the content of ionizable groups and p*Ka* values of the surface amino acid residues. High net charge and number of surface ionizable residues contributes favorably to solubility of the RNase Sa.

When dissolving solid protein in an aqueous buffer solution, solubility is also influenced by the form of the solid phase in equilibrium with the solution phase. In the amorphous solid form the protein molecules arrange randomly and loosely. In contrast, in the crystalline solid form the protein molecules arrange orderly and more tightly. Generally, solubility of protein in an amorphous form is higher than when it is in the crystalline form.

4 Strategies for Increasing *in vitro* Protein Solubility

The problem of *in vivo* low protein solubility associates mainly with protein folding process in the host cell during and after translation. Strategies for solving this problem emphasize how to improve yield of the correctly folded proteins. Several approaches have been employed and reviewed elsewhere, for instance, the development of different *E. coli* host strains, optimization of culture conditions, addition of chemical chaperones, and co-expression of molecular chaperones [22]. Molecular engineering of the target proteins can

be done to improve their solubility provided that the modifications do not affect their intended use. These include fusion with solubility enhancing peptide tag and protein modification by structure-guided directed mutagenesis.

For *in vitro* low protein solubility, the proteins can be successfully over-expressed and purified but cannot be concentrated sufficiently for their intended use such as for high-resolution structural studies and biopharmaceutical applications. To solve this problem, optimizations of extrinsic (environmental) factors and intrinsic factors (modification of the protein itself) have been developed.

4.1 *Use of chemical additives*

The *in vitro* solubility of fully folded proteins can be increased by supplementing some additives to the buffers during purification, concentration, and storage. The commonly used chemical additives for this purpose are detergents, sugars, and salts. Chloride salts at concentrations below 1 M have been used as the agent for improving protein solubility by the phenomenon known as salting-in effect.

Apart from the commonly used additives such as sodium chloride, sucrose, and sorbitol, the addition of charged L-amino acids has been used for improving solubility of protein. Simultaneous addition of 50 mM L-arginine and L-glutamate to the buffer can significantly increase (up to 8.7 times) the solubility of several unrelated recombinant proteins with known solubility problems. With these amino acid additives, *in vitro* protein aggregation and precipitation were effectively prevented, long-term storage stability was dramatically increased, and the protein samples were protected from proteolytic degradation [23]. Arginine alone could suppress protein aggregation and protein-protein or protein-surface interaction during protein purification and *in vitro* protein re-folding process. L-Arg and L-Glu have been used as the excipients for helping the solubility of pharmaceutical proteins and peptides [24]. The solubility of fibroblast growth factor 20 (FGF-20) increased with the increasing concentration of arginine-sulfate in a low range but a salting out effect was observed at high arginine-sulfate concentration range [25]. The added L-Arg and L-Glu did not interfere with downstream processes such as SDS-PAGE analysis, solution-state studies by NMR, specific protein-protein interactions, and protein-RNA interactions. However, the presence of L-Arg and L-Glu in the solution does interfere with protein binding to the ion-exchange columns used for the protein purification.

4.2 *Fusion with solubility tags*

For *in vitro* solubility enhancement, usually the tag should remain with the target protein after expression, purification, and throughout the application. The first example is the non-cleavage solubility-enhancement tag (SET) derived from the highly stable and highly soluble 56-aa peptide of the B1 domain of protein G [26]. Fusion of this tag to the N-terminus of target proteins could significantly improve both solubility and stability of the heterodimeric complex between regulatory domains of human DNA fragmentation factor 40 (DFF40) and human DNA fragmentation factor 45 (DFF45) CIDE domains. The second example is the tag derived from the novel stabilizing peptide acidic tail of synuclein (ATS) [27]. By fusion to the C-terminus of human growth hormone (hGH), granulocyte colony-stimulating factor (G-CSF), and human leptin; the ATS was found to greatly increase the solubility, the storage stability and resistance to environmental stresses such as agitation and freeze/thaw of the fusion proteins. Alternatively, the highly positively-charged solubility enhancement peptide (SEP) tags containing short poly-Lys or poly-Arg were developed [13]. The SEP tags could increase the solubility of a bovine pancreatic trypsin inhibitor (BPTI) variant in a direct relation to the number of positively-charged residues of the tag. Recently, the same research group measured the effects of 10 short poly-amino-acid tags on the solubility of BPTI and determined the positively-charged or negatively-charged tags affected BPTI solubility under a close to neutral pH condition (pH 7.7) [28]. These tags did not alter the conformational structure, thermal stability or activity of fusion partner.

4.3 *Molecular modifications by directed mutagenesis*

Several studies have succeeded in improving the protein solubility of fully folded proteins by using site-directed mutagenesis as reviewed in Trevino et al [11]. This approach can be more useful in several target

proteins when compared to the fusion with solubility enhancement tag. Importantly, this approach does not significantly change the molecular weight of the target protein and does not affect the high-resolution structural studies. Principally, solubility enhancement mutagenesis is done by replacing the hydrophobic surface amino acid residues with charged, polar, hydrophilic [29] or less hydrophobic residues. The surface charged or polar amino acids enhance the protein-water interaction resulting in improvement of protein solubility, whereas the hydrophobic residues promote hydrophobic interaction between protein molecules that leads to protein aggregation. The naturally occurring F185K mutant type of HIV-1 integrase catalytic domain (INC) was found to have dramatically increased solubility. Structural analysis revealed that this mutation altered the salt bridge network that drove the conformational change of INC in such a way that some hydrophobic residues were buried while some other hydrophilic residues were exposed on the protein surface. With more hydrophilic and less hydrophobic surface residues, the protein-water was enhanced and this contributed to the solubility increase of INC [30]. Correspondingly, the structured-based engineering of the anti-IL-13 monoclonal antibody CNTO607 showed significantly more solubility when replacing an aggregation hot spot or surface hydrophobic patch containing three residues (F99-H100-W100a) in heavy-chain CDR3 to triple alanine residues [31].

Recently a systematic study to compare the contribution of all 20 amino acids to protein solubility using RNase Sa, which is a small enzyme with 96 amino acids residues and one disulfide bond, as a model. A completely solvent exposed position, Thr76, was chosen as the position for mutation by replacing it with the other 19 amino acids, and the solubility of each mutant was measured. The results showed that the acidic amino acids (Asp and, Glu) contributed significantly more favorably than any other hydrophilic amino acids including Arg, Lys, Asn, Gln, and Thr under both low and high net charge conditions [6,11,29]. This is supported by the solubility measurement study of seven proteins in ammonium sulfate and polyethyleneglycol 8000. It found that increased negative surface charge of the proteins had a positive effect on *in vitro* solubility [14] due to the strong water-binding properties of the acidic amino acids [32]. This finding provides valuable insight for the understanding of protein solubility and it was applied to a larger and unrelated protein D-phenylglycine aminotransferase (D-PhgAT), a homo-dimeric enzyme consisted of two identical 47-kDa subunits from *Pseudomonas stutzeri*. Double point mutations to change the highly solvent-exposed residue at position N439 and Q444 involved with protein crystal-packing contacts to the acidic amino acids resulted in a mutant enzyme having 5.9 times greater solubility [7], confirming the reliability of the approach.

The finding that acidic amino acids contribute most favorably to protein solubility is in good agreement with the effect of the acidic solubility tag or negatively-charged tag mentioned above as well as the high prevalence of acidic amino acids on the surface of proteins from halophiles which would make the protein soluble in an environment of reduced water activity [33].

With current knowledge, rationale molecular engineering to improve protein solubility can be done with a higher chance of success than before. However, the ability to precisely identify the surface residue candidate for mutation will rely on the availability of the good three-dimensional structure of the candidate protein.

5 Future Prospects

The use of fusion tags will be more popular due to their effectiveness and versatility. Further development should be focused on the creation of multifunctional tags that can serve many purposes. In addition to enhancing solubility of the target protein, the tag can be designed to be able to enhance protein expression at the transcription-translation level, assist in protein folding, stabilization of the partner protein, simplify protein purification and tag removal. The new tags should be smaller in size so that they will not pose much burden on the resources of the host cells, but equally or more effective than the previous tags. Solubility and stability enhancing molecular engineering of proteins will gain more attention and utilization since the success of the rationale mutations is evident. This can be an effective approach provided that the mutations do not affect the final application of the proteins, especially in therapeutic use. The only bottleneck to this approach is the limited availability of the three-dimensional

structure of the target proteins for guiding the precise mutation scheme.

The final goal is not to develop a single best method to solve the protein insolubility problem. Rather, a combination of carefully selected approaches will be the only way to maximize the target protein production. This should be optimized on a case by case basis using all the available options.

6 Conclusions

Large amounts of proteins and enzymes are continually required for processes and applications in several industrial sectors, agriculture, consumer markets, pharmaceutical, and therapeutic use etc. In order to generate good supply of proteins and enzymes to meet these high demands with economically viable production cost, effective expression systems are of the prime importance. For proteins that can be expressed successfully, some may face the problem of *in vitro* low protein solubility, which can obstruct their further development that requires the protein in high concentrations. Enhancement of the *in vitro* protein solubility can be achieved by using chemical additives such as sugars or certain amino acids, fusion with a solubility tag that may remain with the protein until the final use and solubility enhancing mutagenesis. In alleviating the protein insolubility, the cause of the problem should be indentified and the methods for improving protein solubility should be carefully selected. Usually, the best result will be obtained from the proper combination of many methods. Thus, all the available old and new methods should be studied in details in order to formulate an effective operational approach.

References

[1] D. Christendat, A. Yee, A. Dharamsi, Y. Kluger, A. Savchenko, J.R. Cort, and et al., "Structural proteomics of an archaeon," *Nat Struct Biol*, vol. 7(10), pp. 903-909, 2000.

[2] A. Yee, X. Chang, A. Pineda-Lucena, B. Wu, A. Semesi, B. Le, and et al., "An NMR approach to structural proteomics," in *Proc. Natl Acad Sci U S A*, 2002, vol. 99(4), pp.1825-1830.

[3] A. Yee, K. Pardee, D. Christendat, A. Savchenko, A.M. Edwards, and C.H. Arrowsmith, "Structural proteomics: toward high-throughput structural biology as a tool in functional genomics," *Acc Chem Res*, vol. 36(3), pp. 183-189, 2003.

[4] G.W. Caldwell, D.M. Ritchie, J.A. Masucci, W. Hageman, and Z. Yan, "The new pre-preclinical paradigm: compound optimization in early and late phase drug discovery," *Curr Top Med Chem*, vol. 1(5), pp. 353-366, 2001.

[5] L.K. Mosavi and Z.Y. Peng, "Structure-based substitutions for increased solubility of a designed protein," *Protein Eng*, vol. 16(10), pp. 739-745, 2003.

[6] S.R. Trevino, J.M. Scholtz, and C.N. Pace, "Amino acid contribution to protein solubility: Asp, Glu, and Ser contribute more favorably than the other hydrophilic amino acids in RNase Sa," *J Mol Biol*, vol. 366(2), pp. 449-460, 2007.

[7] A. Chantarasiri, V. Meevootisom, D. Isarangkul, and S. Wiyakrutta, "Effective improvement of D-phenylglycine aminotransferase solubility by protein crystal contact engineering," *J Mol Microbiol Biotechnol*, vol. 22, pp. 147-155, 2012.

[8] P. Smialowski, A.J. Martin-Galiano, A. Mikolajka, T. Girschick, T.A. Holak, and D. Frishman, "Protein solubility: sequence based prediction and experimental verification," *Bioinformatics*, vol. 23(19), pp. 2536-2542, 2007.

[9] H. Tjong and H.X. Zhou, "Prediction of protein solubility from calculation of transfer free energy," *Biophys J*, vol. 95(6), pp. 2601-2609, 2008.

[10] Y. Gong, D. Grant, and H.G. Brittain, "Principles of solubility," in *Solvent Systems and Their Selection in Pharmaceutics and Biopharmaceutics*, vol. 6, P. Augustijns and M.E. Brewster, Ed. New York: Springer Science, 2007, pp. 1-28.

[11] S.R. Trevino, J.M. Scholtz, and C.N. Pace, "Measuring and increasing protein solubility," *J Pharm Sci*, vol. 97(10), pp. 4155-4166, 2008.

[12] D. M. Bollag, M. D. Rozycki, and S. J. Edelstein, *Protein Methods*, New York: Wiley-Liss, 1996, pp. 1-432.

[13] A. Kato, K. Maki, T. Ebina, K. Kuwajima, K. Soda, and Y. Kuroda, "Mutational analysis of protein solubility enhancement using short peptide tags," *Biopolymers*, vol. 85(1), pp. 12-18, 2007.

[14] R.M. Kramer, V.R. Shende, N. Motl, C.N. Pace, and J.M. Scholtz, "Toward a molecular

understanding of protein solubility: increased negative surface charge correlates with increased solubility," *Biophys J*, vol. 102, pp. 1907-1915, 2012.

[15] C.N. Pace, S. Trevino, E. Prabhakaran, and J.M. Scholtz, "Protein structure, stability and solubility in water and other solvents," *Philos Trans R Soc Lond B Biol Sci*, vol. 359(1448), pp. 1225-1234, 2004.

[16] S.M. Singh and A.K. Panda, "Solubilization and refolding of bacterial inclusion body proteins," *J Biosci Bioeng*, vol. 99(4), pp. 303-310, 2005.

[17] A. Villaverde and M.M. Carrio, "Protein aggregation in recombinant bacteria: biological role of inclusion bodies," *Biotechnol Lett*, vol. 25(17), pp. 1385-1395, 2003.

[18] D.L. Wilkinson and R.G. Harrison, "Predicting the solubility of recombinant proteins in *Escherichia coli*," *Biotechnol*, vol. 9(5), pp. 443-448, 1991.

[19] S.R. Trevino, S. Schaefer, J.M. Scholtz, and C.N. Pace, "Increasing protein conformational stability by optimizing beta-turn sequence," *J Mol Biol*, vol. 373(1), pp. 211-218, 2007.

[20] S.J. Shire, Z. Shahrokh, and J. Liu, "Challenges in the development of high protein concentration formulations," *J Pharm Sci*, vol. 93(6), pp. 1390-1402, 2004.

[21] K.L. Shaw, G.R. Grimsley, G.I. Yakovlev, A.A. Makarov, and C.N. Pace, "The effect of net charge on the solubility, activity, and stability of ribonuclease Sa," *Protein Sci*, vol. 10(6), pp. 1206-1215, 2001.

[22] M. Martinez-Alonso, N. Gonzalez-Montalban, E. Garcia-Fruitos, and A. Villaverde, "Learning about protein solubility from bacterial inclusion bodies," *Microb Cell Fact*, vol. 8(4), 2009 Doi:10.1186/1475-2859-8-4

[23] A.P. Golovanov, G.M. Hautbergue, S.A. Wilson, and L.Y. Lian, "A simple method for improving protein solubility and long-term stability," *J Am Chem Soc*, vol. 126(29), pp. 8933-8939, 2004.

[24] T. Arakawa, K. Tsumoto, Y. Kita, B. Chang, and D. Ejima, "Biotechnology applications of amino acids in protein purification and formulations," *Amino Acids*, vol. 33(4), pp. 587-605, 2007.

[25] H. Maity, C. Karkaria, and J. Davagnino, "Effects of pH and arginine on the solubility and stability of a therapeutic protein (Fibroblast Growth Factor 20): relationship between solubility and stability," *Curr Pharm Biotechnol*, vol. 10(6), pp. 609-625, 2009.

[26] P. Zhou, A.A. Lugovskoy, and G. Wagner, "A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins," *J Biomol NMR*, vol. 20(1), pp. 11-14, 2001.

[27] E.N. Lee, Y.M. Kim, H.J. Lee, S.W. Park, H.Y. Jung, J.M. Lee, and et al., "Stabilizing peptide fusion for solving the stability and solubility problems of therapeutic proteins," *Pharm Res*, vol. 22(10), pp. 1735-1746, 2005.

[28] M.M. Islam, M.A. Khan, and Y. Kuroda, "Analysis of amino acid contributions to protein solubility using short peptide tags fused to a simplified BPTI variant," *Biochim Biophys Acta*, vol. 1824, pp. 1144-1150, 2012.

[29] C.N. Pace, G.R. Grimsley, and J.M. Scholtz, "Protein ionizable groups: pK values and their contribution to protein stability and solubility," *J Biol Chem*, vol. 284, pp. 13285-13289, 2009.

[30] H.Q. He, J.P. Hu, B. Liu, W.Z. Chen, and C.X. Wang, "Activity, solubility comparison and molecular dynamics simulation analysis of wild type and F185K mutant type HIV-1 integrase catalytic domain," *Prog Biochem Biophys*, vol. 36(9), pp. 1146-1153, 2009.

[31] S.J. Wu, J. Luo, K. Neil, J. Kang, E.R. Lacy, G. Canziani, and et al., "Structure-based engineering of a monoclonal antibody for improve solubility," *Protein Eng Des Sel*, vol. 23(8), pp. 643-651, 2010.

[32] K.D. Collins, "Charge density-dependent strength of hydration and biological structure," *Biophys J*, vol. 72, pp. 65-76, 1997.

[33] X. Tadeo, B. Lopez-Mendez, T. Trigueros, A. Lain, D. Castano, and O. Millet, "Structural basis for the amino acid composition of proteins from halophilic archea," *PLoS Biology*, vol. 7(12), 2009, Doi:10.1371/journal.pbio.1000257