Doctoral Thesis
Doctor of Engineering

Biophysical analysis of protein solubility and aggregation using short amino acid peptide tags

Mohammed Monsur Alam Khan

Submitted to

Tokyo University of Agriculture and Technology

TABLE OF CONTENTS

Chapter -1

General Introduction

Chapter 2

Analysis of amino acid contribution to protein solubility using short peptide tags fused to a simplified BPTI variant

Chapter-3

Analysis of protein aggregation kinetics using short amino acid peptide tags

LIST OF FIGURES

Chapter -1

General Introduction

Chapter 2

Analysis of amino acid contribution to protein solubility using short peptide tags fused to a simplified BPTI variant

Chapter-3

Analysis of protein aggregation kinetics using short amino acid peptide tags

Page

LIST OF TABLES

Chapter -1

General Introduction

Chapter 2

Analysis of amino acid contribution to protein solubility using short peptide tags fused to a simplified BPTI variant

Chapter 3

Analysis of protein aggregation kinetics using short amino acid peptide tags

LIST OF ABBREVIATIONS

BPTI	: Bovine Pancreatic Trypsin Inhibitor
Simplified BPTI	: A single-disulfied-bonded bovine pancreatic trypsin inhibitor variant whose sequence was simplified by multiple alanine replacement
BPTI-19	: Simplified BPTI containing 19 alanines out of 58 residues
DSC	: Differential Scanning Calorimetry
HPLC	: High Performance Liquid Chromatography
SD	: Standard deviation

Biophysical analysis of protein solubility and aggregation using short amino acid peptide tags

Summary

Protein aggregation is a concern in biophysical studies. It has been reported that amyloidegenic protein aggregation is associated with several neurodegenerative diseases, whereas, disordered protein aggregation is a concern in biotechnology and pharmaceutical industries. Minimizing aggregation consequences is a challenge for the scientists. The pioneer one approach is optimization of protocols for expression, purification and solubilization. Surface supercharging mutation by replacing surface exposed hydrophobic amino acid with a hydrophilic one is another approach to increase protein solubility or to avoid aggregation tendency. Addition of highly soluble protein, domain or short peptide as fusion tag with protein is another way to minimize aggregation propensity of aggregation prone proteins. However, according to literature, still it is very difficult to select the effective measure for rationalizing protein aggregation tendency. For an effective measure to take, it is necessary to understand protein aggregation clearly, especially in connection with amino acids contributions to protein aggregation propensity. Protein aggregation problem and how individual amino acid contributes to protein aggregation tendency remain fully unsolved. Understanding protein solubility, and consequently protein aggregation, is an important issue both from an academic and biotechnological application viewpoint. In chapter I, I reviewed the general aspect of protein aggregation, aggregation consequences, minimizing approach, and identified the guide lines experimental methods.

In chapter 2, I report the effect of ten representative amino acids on the solublization property of protein. This effect was determined by measuring the solubility of a simplified bovine pancreatic trypsin inhibitor (BPTI) variant, to which a short artificial tag containing the amino acids of interest were added at its C-terminus. We first characterize the impact of tags on host protein and confirm that the short peptide tags remained independent of tags on host protein. We also confirm that the short peptide tags have no effect on structure, thermal stability, and biochemical function of the host protein. As anticipated, positively charged poly-Lys significantly increases the solubility of the model protein over ten-fold, at both pH 4.7 and pH 7.7, whereas, very hydrophobic poly-Ile markedly reduced the solubility of simplified BPTI. Poly-Asp and poly-Glu barely affected BPTI solubility at pH 4.7, but induced an 8-10 fold increase at pH 7.7, attributable to the ionization of their side chains. Although Pro is the most soluble amino acid, poly-Pro did not affect the protein solubility. The effect of the other tags on BPTI solubility ranged from negative to an eight fold increase. These observations suggest that this approach is valuable for measuring the solubility propensities of amino acids, which could eventually allow the calculation of a polypeptides relative solubility form its amino acid sequence.

In chapter 3, I reported the aggregation kinetics of tagged variants by determining the solubility as a function of equilibration time (20 min to 48 hrs) and with a total protein concentration raging form 10 µg/ml to 25 mg/ml. We observed, as anticipated, that proteins precipitated promptly when the total protein co9ncentration exceeded some

critical value. However, when total protein concentration was higher, the apparent solubility reached a concentration higher than the above critical value and decreased upon increasing equilibration time. These observations were confirmed using a wild type lysozyme. Like transient solubility the aggregation initiation concentration and the rate of protein aggregation was dependent on the type of amino acid composing the tags. These observations clearly demonstrated that total protein concentration and equilibration time need to be consider while defining protein solubility and short poly amino acid tags can be an effective tool to gain insight into the complex phenomenon of protein aggregation.

In chapter 4 (conclusion), the result achieved in the present study was summarized, and the future aspects of the research was described. our findings on solubility values of variants fused with tags of different types of amino acids could provide a "solubility propensity scale", which could eventually allow the calculation of a polypeptide's relative solubility from its amino acid sequence. In addition, findings related to total protein concentration and equilibration time dependent aggregation kinetics will help to take effective measures for minimizing aggregation consequences. Moreover, to the best our knowledge, this is the first study to focus on the contribution of different types of amino acids on protein aggregation propensity.

Chapter -1

General Introduction

1.1. Protein solubility

Protein solubility is traditionally defined as the ability of a protein to dissolve in aqueous solution [1]. Solubility is measure by the equilibrium concentration of protein in supernatant of a saturated or supersaturated solution [1]. This ability of protein is determined by it amino acid sequences under a given set of experimental conditions [2]. Three major forces such as electrostatic repulsion, hydrophobic protein-protein interaction, and hydration or solvation repulsion determine how much protein will retain in solvent as soluble [3].

1.2. Protein aggregation

Protein aggregation is a general term that denotes gathering of monomers of protein molecule resulting from several types of interactions. Protein aggregation is attracting much attention in biophysical studies [4-5]. Amyloidegenic protein aggregation is associated with several neurodegenerative diseases [6-7]. On the other hand, disordered protein aggregation is a concern in biotechnology and pharmaceutical industries, especially in protein production, shipment and the storage processes [8].

1.3. Types of protein aggregation

Protein aggregation can be classified in numerous ways, soluble/insoluble, covalent/non-covalent, reversible/irreversible, and native/denatured [9]. The aggregation which can not be seen by unaided eyes and can not be removed by filtering with a filter of pore size 0.22μm is soluble. Aggregation in which more than two monomers are associated by chemical bonds are termed as covalent aggregations. Usually oxidation of tyrosines to bityrosine, and thiols to disulfide are responsible covalent aggregations. Aggregation from which in favorable environmental condition

protein molecules can be dissociated to monomeric status is termed as reversible aggregation. This type of association forms due to the weak non covalent protein interactions. Aggregation of chemically or physically conformational altered monomers is termed as denatured aggregation [9].

1.4. Stages of protein aggregation formation

Aggregation can occur at every stage of protein processing. Accumulation of high amount of newly produced proteins may cause aggregation during cell culture [9] (Figure 1.1). The reasons may be either interaction of high amount of unfolded nascent peptide or inefficient recognition by the chaperones [10]. During purification, acidic condition on HPLC column mechanical stress resulted from filtration, may leads to aggregation formation. Adverse environment during formulation and filling may also cause protein aggregation [11] (Figure 1.2).

1.5. Adverse effects of low protein solubility

Low protein solubility or protein aggregation is a concern for biopharmaceutical industries. Due to low solubility, proteins can be aggregated at every step of manufacturing processes, production, purification, formulation, and filling [9]. High yields of total genomic proteins showed to have reverse relationship with quantity of soluble fractions, which may restrict the availability of many therapeutic potential proteins [12]. Low solubility or aggregation tendency of therapeutic proteins can be followed by the degradation pathways which ultimately compromise with the therapeutic activity of pharmaceutics [13]. Inducing immune response is a common incidence for most of the biopharmaceuticals but sometimes consequence can be

severe or even lethal. The reason behind the immunogenicity is protein aggregation tendency [14-15]. As recombinant proteins are being used widely for developing drugs the consequence may cause a loss of efficacy of the desire protein or even worse. IFN is an example of the losing of efficacy [16]. Many therapeutically important proteins can not be used simply because of aggregation. For example, wild type human calcitonin (hCT) has successfully being used after decreasing its self association tendency by introducing N-terminal acetylating and C-terminal amidating [17]. Stress-induced aggregation profile of three therapeutically important protein, human growth hormone (hGH), granulocyte colony-stimulating factor (G-CSF) and human leptin were converted to more soluble form by fusing with acidic tail of synuclein (ATS) for the purpose of better use [18]. High throughput structural genomic studies are commonly being hampered by low protein solubility, for why, it is required to improve solubility of desire proteins for study [19]. Biotechnological application of human enteropeptidase light chain (hEPL) was facilitated by improving its solubility and yield through surface supercharging [20]. Due to low solubility, human leptin cannot readily be crystallized. Determination of crystal structure of the obese protein leptin-E100 was possible by increasing it solubility with a single amino acid substitution of Glu for Trp at position 100 [21]. Preparation of highly soluble NMR-suitable protein sample is still a challenge. NMR study has been facilitated by increasing solubility of desire protein sample using highly soluble peptide tags as fusion partners [22] (Table 1.1).

1.6. Adverse effects of protein aggregation

Aggregation of proteins having distinct amino acid sequences but common phenomena of formation of highly ordered cross-amylodegenic structure are reported to be linked with human pathogenesis [23]. For example, Huntington`s disease, Alzheimer`s

Disease, Parkinson`s Disease, Prion`s Disease, amyotrophic lateral sclerosis (ALS), Prion`s Disease and many other lethal diseases are being caused by the deposition of aggregated proteins [6, 24]. The severity and early onset of symptoms of Huntington`s disease are closely associated with the aggregation of poly(Gln)-containing polypeptides [25]. Deposition of a 42-residue peptide called as amyloid β (Aβ) 42 leads to the neurodegenerative symptoms of Alzheimer`s Disease [26], ά-synuclein toxicity of dopamine neurons by the formation of aggregated ά-synuclein leads to the on-set of Parkinson`s Diseases [27]. Christopher AR., et. al. in 2004 compiles the causative proteins, pathology and affected tissues for protein aggregation [6] (Table 1.2).

1.7. Factors responsible for protein aggregation

Protein aggregation behaviors such as onset of aggregation, rate of aggregation, and the morphology of the aggregated state (i.e., amorphous precipitates or fibrils) depend on several factors. These factors can be divided into two groups; extrinsic and intrinsic. Extrinsic factors include temperature, pH, salt type, salt concentration, cosolutes, preservatives, and surfactants [28-30]. Heat can induce protein aggregation by initiating disulfide bridge formation between the free thiol groups [31]. At iso-electric point (pI) of protein, the net charge becomes equal to zero. The water-protein interaction reach to minimum level on the other hand protein-protein interaction increase to maximum level which ultimately leads to protein aggregation [32] (Figure1.3). Detail analysis on the effect of pH and salt concentration on solubility of lipoprotein from hen egg yolk is being conducted by Sousa RDS *et. al.,* [33] (Table 1.3). However precipitation does not solely depends on the pI value, rather being depends on the characteristic of the system. Wang W. *et. al.* in 1999 summarize the

factors responsible for protein aggregation [10]. Comprehensive aggregation analysis of the entire ensemble of the *E. coli* proteins revealed that aggregation tendency of protein is depending on peptide primary sequence [34]. For example protein containing higher amount of negatively charged residues (Asp and Glu) are more soluble, whereas, aromatic amino acid (Phe, Tyr and Trp) containing proteins are aggregation prone. Same study also showed that, proteins having high molecular weight are more aggregation prone in compare with low molecular proteins [34] (Figure 1.4).

1.8. Strategies to minimize protein aggregation

In varieties of applications ranging from pharmaceutical applications to general biochemical studies high protein concentration in solution is required. Scientists were trying to minimize the consequences of protein aggregation in various ways which include optimization of solution conditions, surface supercharging, fusion of soluble tags [35]. For an effective application one can run more than one measures simultaneously. It has been reported that each process has its own advantages and disadvantages.

1.8.1. Optimizing solution conditions

Among theses the pioneer one is optimizing the solution conditions. Optimization of protocols for expression, purification and solubilization is common strategy to improve the yield the soluble protein. Addition of equimolar amounts of Arg and Glu is one of the examples to improve the solubility of proteins [19, 35-37]. Running experiment at pH away from the pI of the sample protein is traditional way to avoid aggregation.

1.8.2. Surface supercharging

Surface supercharging is another approach to increase protein solubility. In this approach surface exposed hydrophobic amino acids are being replaced with a hydrophilic one to increase total charge of the target protein [38-39] (Table-1.4). Application of this approach has some limitations, especially for the proteins of unknown structure. There is chance of mutation to be occurred in buried region which may result in decreased stability rather than increasing solubility [36]. Trial and error basis searching for the proper amino acid to be replaced for every protein is another limitation of this approach.

1.8.3. Fusion tags to minimize aggregation tendency.

Fusion tags have become important tools in the fields of structural and functional proteomics. The affinity tag system has been proven to extent positive impact in yield, solubility and folding of the fusion partner [40]. The common features for the tags are, minimal effect on tertiary structure, less effect on host protein`s activity, easy to remove form the partner, leave the partner in native state after removal, commonly usable to a number of proteins [41]. Protein, domains, or peptides of different size can be used as fusing tags. The advantage of Small peptides such as poly-Arg, poly-His, c-myc, FLAG is that, these tags may not need to remove form the partner. The Arg-tag usually consist of five or six arginine molecules. Poly-His composed of six histidine is being widely used. On the other hand large peptide or protein may interact with the structure of the host and must need to remove before use. 40-kDa protein maltose-binding protein (MBP) and 26kDa glutathionine S-transferase (GST) are commonly used proteins tags. However, every tag has it`s own advantage and disadvantage [40] (Table-1.5).

1.9. Aggregation Mechanisms

Commonly assumed mechanisms for disordered protein aggregation in a physico-chemically isolated system include the association of natively folded proteins, the irreversible aggregation of partially unfolded proteins or chemically altered proteins [42]. Sedimentation assay of insulin and proinsulin revealed the self association mechanism of protein aggregation [43]. This type of self-association results from the non-covalent force and are reversible. Conformational change of protein native monomer to a non-native monomer leads to irreversible association of protein aggregation mechanism. Conformational change can occur due to heat or any other physical stress such as fridge dry. For pharmaceuticals and laboratory nonnative aggregation is particularly problematic one. This type of aggregation involves at least two steps, start with conformational change of native form to non native form and then assembly of structurally altered protein molecules into aggregates [11]. The mechanism for irreversible association of chemically altered proteins is similar with the previous type but in this case conformational change happen due to presence of chemicals as cosolute. Usually chemically degradation such as deamination, oxidation of methionine, protiolytic breakdown are mainly responsible for alternation of protein structure. However, these aggregation mechanisms are not mutually exclusive, rather for a particular protein sample more than one mechanisms may be active simultaneously [44] (Figure 1.5).

1.10. Studies on protein aggregation

Various studies have been conducted to minimize the consequences of protein aggregation, but biophysical mechanism of protein remains to be uncovered. Toward understanding of protein aggregation kinetics, protein itself is the most obscured

intrinsic factor along with other extrinsic factors [9-11]. Furthermore, difficulties for isolating the aggregation intermediates hinders study [45]. Nevertheless, techniques studying protein aggregation kinetics mostly rely on the detection of aggregates either as a function of the course of time or as a function of environmental factors (denaturant agent, pH, temperature etc.) [46-47]. Since the mechanism of protein aggregation includes the stage of intermediate seed formation (nucleation) and then the growth of aggregation nucleus, initial protein concentration may play an important role on aggregation kinetics, but no systematic study was made in these regards. To gain more insight into the protein aggregation kinetics and to determine aggregation mechanism in this context, it is necessary to investigate the course of time the aggregates form over a wide range of initial protein concentrations. Thus far, basic studies on aggregation mechanisms include (i) characterizing the effect on structure, (ii) determining the aggregation prone peptide sequence, (iii) specifying environmental effects, (iv) analyzing reversibility of aggregation, and (v) kinetics of aggregation [46]. Among these, kinetic study, can dictate viable processing, formulation, and the storage condition for biotechnological products [48]. Using a kinetic approach when analyzing aggregation mechanisms demands more attention.

1.11. Short peptide as a tool

The solubility of amino acids range widely, from a few score micromolar to several hundred millimolar [49], which presents a technical hurdle, among many others, for the construction of a solubility propensity scale, because such a wide range of values is difficult to measure accurately with a single protocol. Peptide sequences consisting of a single amino acid type (poly-amino-acid peptides) can be useful in partly

overcoming this problem, because they tend to amplify the adhesive, aggregation, polymerization, and solubility properties of the amino acids.

Recently, we used the biochemical properties of charged amino acids to improve protein solubility by fusing short poly-Lys or poly-Arg tags to a bovine pancreatic trypsin inhibitor (BPTI) variant, which increased its solubility by over six-fold without altering its NMR spectrum or activity [49] (Figure 1.6). We therefore reasoned that poly-amino-acid peptide tags, which can significantly modify a protein's solubility, could be used to determine the contribution of individual amino acids to protein solubility.

1.12. "Host-guest" approach for solubility study

The "host-guest" approach, has successfully being applied to peptides and proteins, as an experimental method for measuring the tendencies of amino acids to adopt certain states or conformations. This approach was first used by Scheraga's group in the 1960**s** [50], and later by Baldwin's group [51], to determine the helix propensities of amino acids. For determining the structural/biochemical context dependencies of helix [52], or β-sheet formation within a protein [53], this "host-guest" approach was also used. However, to analyze the contributions of amino acids to protein solubility a host-guest-like approach has rarely been used, and when it has been applied, the results were confounded by effects arising from the surrounding structural/biochemical environments of the host protein [54].

Here, we report a host-guest-like approach in which we used short poly-amino-acid tags as guest to investigate the effects on host protein solubility. We have chosen ten amino acid types representing the full range of biophysical properties (acidic, basic, polar, and hydrophobic) to use as tags. We fused short poly-amino-acid tags (guests), consisting five residues of one of the selected amino acids, to the C-terminus of our host protein, a simplified BPTI variant [55-56]. Wild type BPTI has mutated to simplified one, in which most of the surface residues were alanines, which we expected would minimize the interactions between the poly-amino-acid tag and the host protein. In our study later we ensured that the peptide tag did not affect the structural, functional, or thermodynamic properties of the host protein and the contributions of the amino acids to the protein's solubility were context independent. We observed that the solubilization effects of tags to the host protein consistent with hydrophobicity of tagged amino acids, those anticipated from hydrophobicity and hydrophilicity scales. These observations suggest that using short peptide tags as a part of "host-guest" model can provide an effective tool to develop a genuine solubility propensity scale to predict protein aggregation tendency or solubility.

1.13. Amino acid contribution to protein aggregation

Aggregation tendency or solubility of a protein is usually estimated in terms of a hydrophobicity-derived scale. However, as hydrophobicity refers to the transfer of amino acid from an aqueous to a nonpolar solution this scale can not provide the actual propensity of protein solubility because in a real protein solution molecules remain in same solution. [57-58] Studies have revealed that, hydrophobic interaction is the critical driving force behind aggregation formation [10-11, 59-60] and protein containing more hydrophobic amino acids is more aggregation prone [34, 39].

Surface supercharging through hydrophobic to hydrophilic mutation has become a popular approach to avoid aggregation tendency or to increase solubility of target proteins [20, 36, 54, 61]. Rationalizing protein solubility by the amino acid mutations on protein surface have been limited by various intrinsic and extrinsic factors that influence protein solubility [9-11]. Moreover, net charge of protein is an determinant factor of protein aggregation tendency [32, 62]. The charge of protein at a particular pH also depends on amino acids composition. Through protein aggregation tendency is encoded in its amino acid sequence, physio-chemical properties of protein aggregation at the amino acids residue level is not fully studied yet [34, 63-65]. In this perspective, a genuine solubility propensity scale for amino acids might allow the prediction of protein solubility or aggregation tendency from the primary sequence of the protein, more accurately. One of the goals of this study is to understand the contribution of individual amino acids to the aggregation propensity of protein.

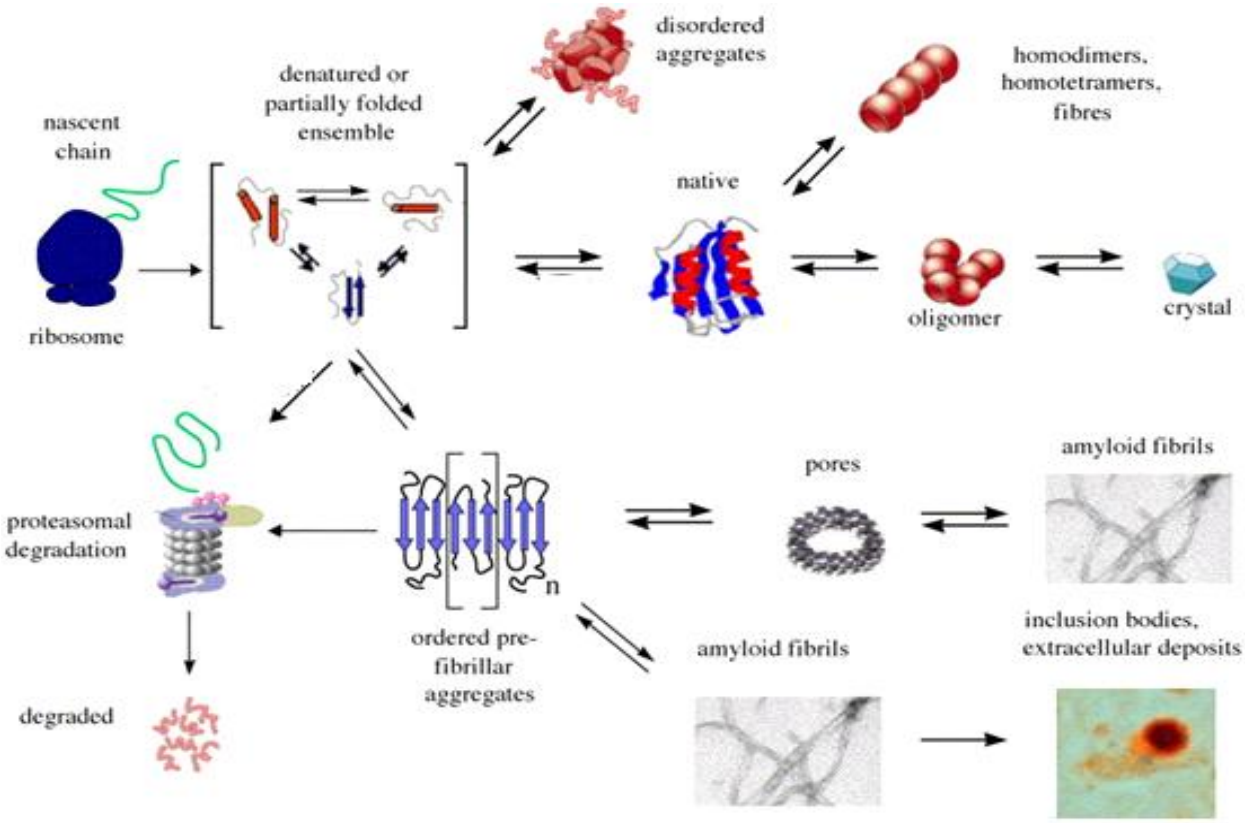


Figure 1.1. Protein aggregation within cell during cell culture. This figure is adopted from the literature cite by Stefani M and Dobson CM *(Stefani M and Dobson CM 2003 J Mol Med, 81;p. 678-699,* [7]).

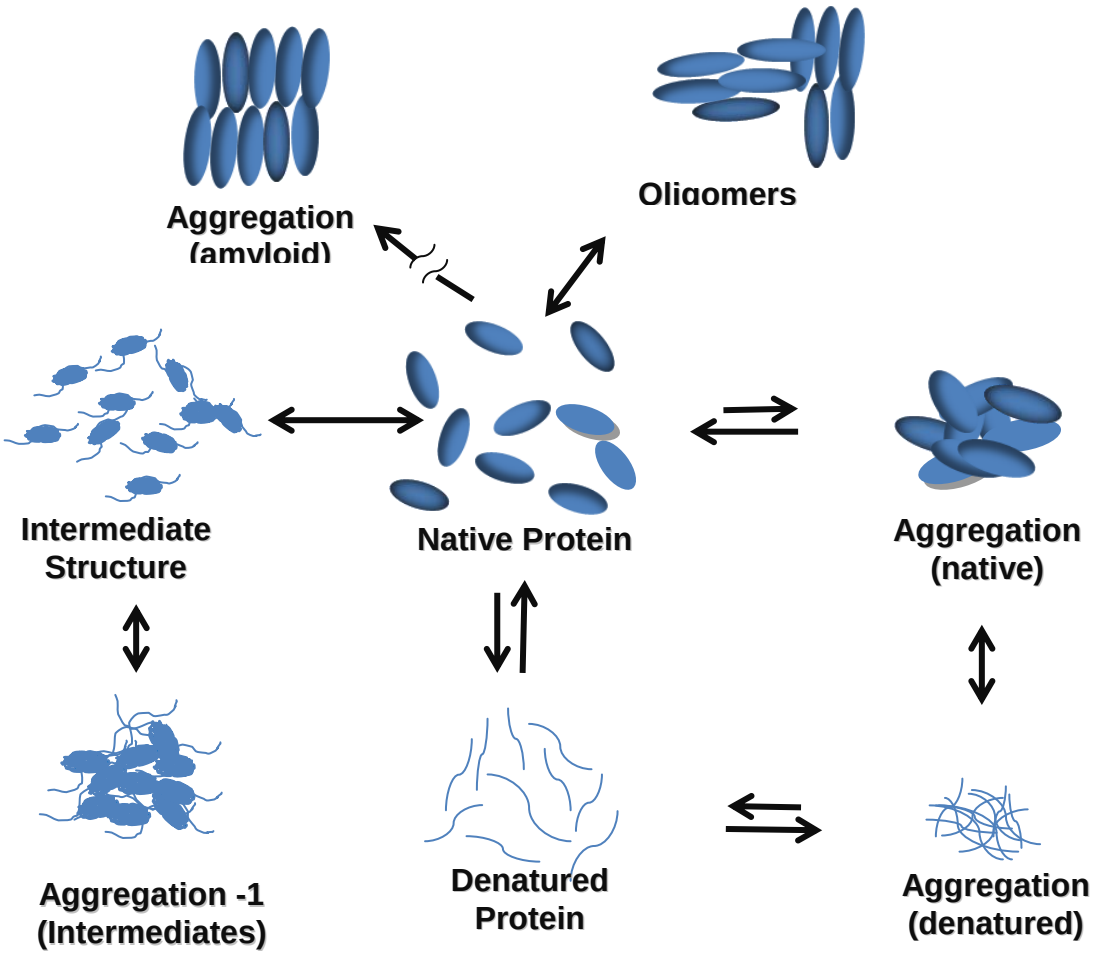


Figure 1.2. *In vitro* protein aggregation during formulation.

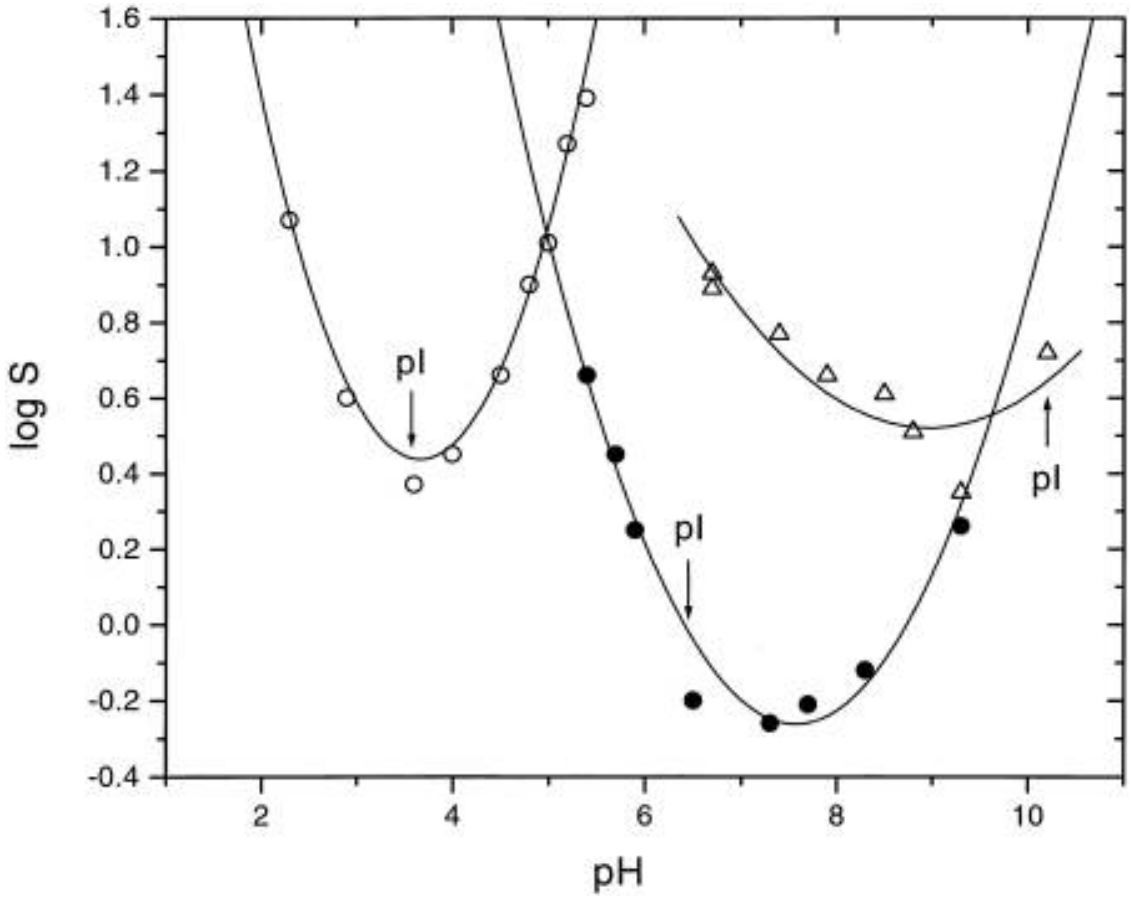


Figure 1.3. Effect of pH on protein solubility. Solubility of RNase Sa (○) and two if its mutants the 3K (•) and 5K (Δ) as a function of pH. This figure is adopted from the literature cite by Shaw, K.L., et al., *(Shaw, K.L., et al., 2001, Protein Sci, 10(6), p.1206-15* [32]*).*

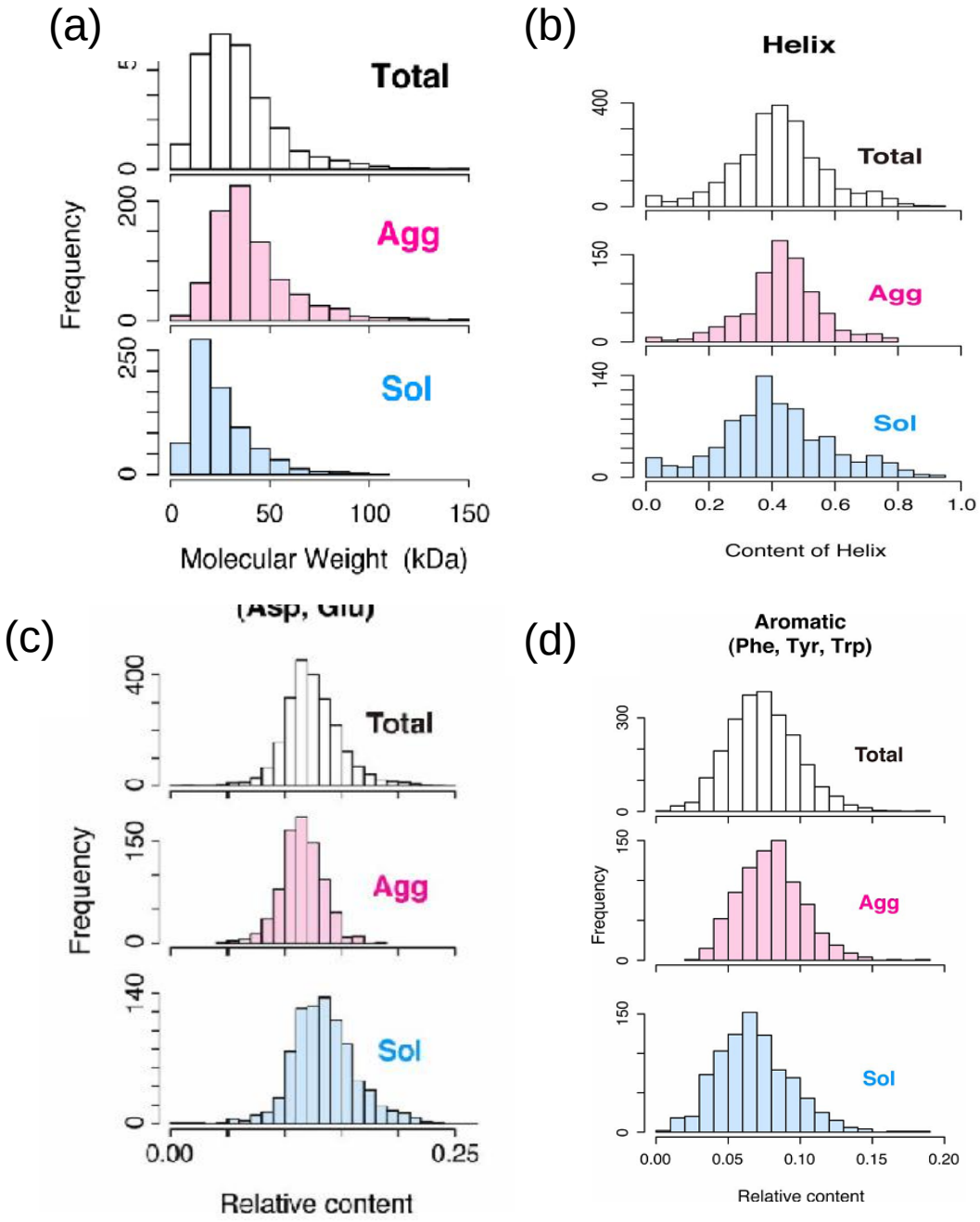


Figure 1.4. Correlation between solubility and physio-chemical properties. (a) Histograms of molecular mass in the total, aggregated and soluble fraction of proteins, (b) histogram of the relative contents of helix, (c) histogram of the relative contents of negatively charged residues, and (d) histogram of the relative contents of aromatic amino acids. This figure is adopted from the literature cite by Niwa, T., et al., *(Niwa, T., et al., 2009 PNAS,106(11);p.4201-4206* [34]).

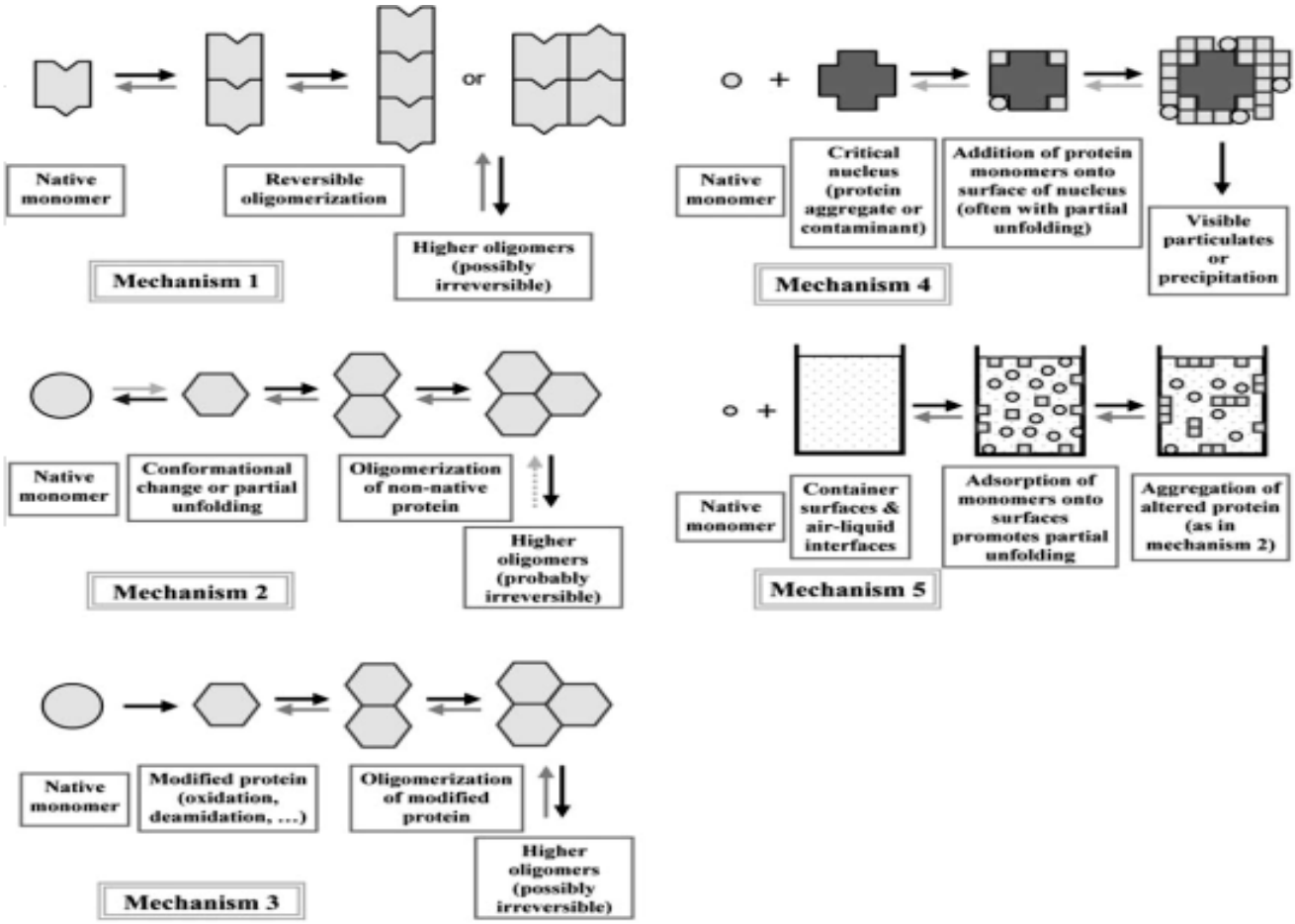


Figure 1.5. Schematic illustrations of five common aggregation mechanisms. This figure is adopted from the literature cite by Phil, JS and T. Arakawa *(Phil, JS and T. Arakawa 2009 Curr. Pharm Biotechnol 10(4);p.348-51)* [44].

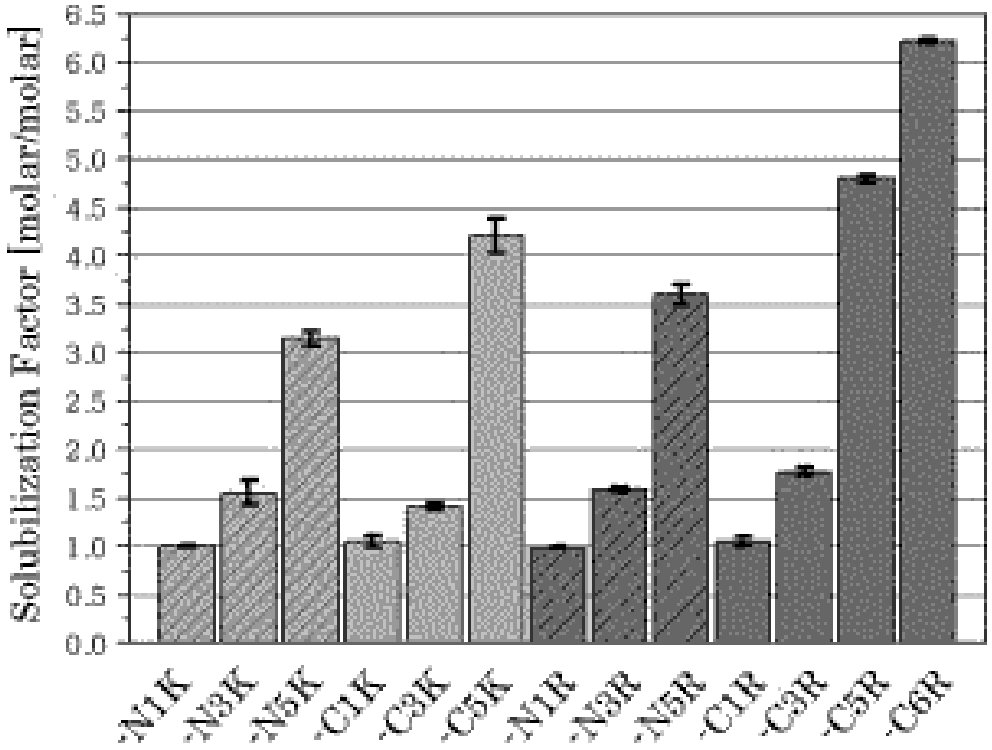


Figure 1.6. Effect of short peptide tags (poly-Lys and poly-Arg) on protein solubility. Tags of different lengths (1,3, 5and 6 amino acids) were added at C and N terminus end of simplified BPTI (BPTI-22). This figure is adopted from the literature cite by Kato A. et al., *(Kato A. et al., 2006 Biopolymer 85;p.13-18* [49] *).*

Problem	Examples	References
Out of study	33-50% of all expressed proteins are insoluble 25-57 % of remaining forms aggregation	[19]
Less production	GFP gene expression	[12]
Less recovery	Lysozyme	[66]
Experimental problem	human leptin	[21]
Developing drug	human IgG	[67]
Storage, shipment & processing	various therapeutic proteins	[10]
Introducing immunogenicity	Intravenous immune globulin (IVIG)	[15]

Table 1.1. Protein aggregation problems for biophysical study.

Disease	Etiology	Regions most affected	Characteristic pathology	Disease proteins deposited
Huntington's disease	Huntingtin (dominant)	Striatum, other basal ganglia, cortex, other regions	Intranuclear inclusions and cytoplasmic aggregates	Huntingtin with polyglutamine expansion
Other polyglutamine diseases (DRPLA, SCA1–3, etc., SBMA)	Atrophin-1, ataxin-1–3, etc.; androgen receptor (AR)	Basal ganglia, brain stem cerebellum, and spinal cord	Intranuclear inclusions	Atrophin-1, ataxins or AR
Alzheimer's disease (AD)	Sporadic (ApoE risk factor)	Cortex, hippocampus, basal forebrain, brain stem	Neuritic plaques and neurofibrillary tangles	A β peptide (from APP) and hyperphosphorylated tau
	Amyloid precursor protein (APP) (dominant)	Same as sporadic	Same as sporadic	Same as sporadic
	Presenilin 1, 2 (dominant)	Same as sporadic	Same as sporadic	Same as sporadic
Fronto-temporal dementia with Parkinsonism	Tau mutations (dominant)	Frontal and temporal cortex, hippocampus	Pick bodies	Hyperphosphorylated tau protein
Parkinson's disease (PD)	Sporadic	Substantia nigra, cortex, locus ceruleus, raphe, etc.	Lewy bodies and Lewy neurites	α-Synuclein
	α-Synuclein (dominant)	Similar to sporadic, but more widespread	Similar to sporadic	α-Synuclein
	Parkin (also DJ-1, PINK1) recessive (some dominant)	Substantia nigra	Lewy bodies absent (or much less frequent)	α-Synuclein (when present)
Amyotrophic lateral sclerosis (ALS)	Sporadic	Spinal motor neurons and motor cortex	Bonina bodies and axonal spheroids	Unknown (neurofilaments)
	Superoxide dismutase-1 (dominant)	Same as sporadic	Same	Unknown
Prion diseases (kuru, CJD, GSS disease, fatal familial insomnia)	Sporadic, genetic and infectious	Cortex, thalamus, brain stem, cerebellum, other areas	Spongiform degeneration, amyloid, other aggregates	Prion protein

ApoE, apolipoprotein E; APP, amyloid precursor protein; CJD, Creutzfeldt–Jakob disease; DRPLA, dentato-rubral and pallido-Luysian atrophy; GSS, Gerstmann–Straussler–Scheinker; SBMA, spinal and bulbar muscular atrophy; SCA, spino-cerebellar ataxia.

Table 1.2. Neurodegenerative diseases: proteins and pathology, *Roos CA. al. 2004 Nature Medicine 10, S10–S17* [6].

	pH				
	3.0	**4.03**	**6.5**	**8.7**	**10.0**
NaCl (*mol/l*)					
0.05	55.75±1.30	44.18±0.93	19.85±2.53	32.06±0.65	80.89±0.37
0.10	42.10±2.28	33.70±3.58	28.88±2.07	34.53±4.60	81.39±1.98
0.20	25.11±1.57	32.81±0.38	58.67±6.18	33.12±4.92	85.39±4.74
0.30	19.82±0.77	30.14±1.67	71.29±1.82	34.41±1.08	82.04±5.49
0.50	12.45±0.43	23.41±6.91	72.77±6.41	45.02±0.28	83.21±0.20
Na_2SO_4 (*mol/l*)					
0.05	14.04±1.34	17.10±0.98	51.95±1.30	36.58±0.93	51.58±0.24
0.10	13.13±0.94	14.94±1.71	76.73±0.85	42.24±1.88	52.60±1.22
0.20	8.63±0.46	17.75±1.13	90.04±0.12	55.94±2.05	52.59±0.17
0.30	12.39±0.26	16.99±0.04	95.07±1.93	56.82±0.56	54.10±1.40
0.50	1.28±0.97	16.71±1.80	98.78±1.41	49.10±1.96	45.40±4.13
$(NH_4)_2SO_4$ (*mol/l*)					
0.05	19.06±1.10	15.82±0.82	28.06±1.46	6.20±1.37	33.46±1.12
0.10	17.84±0.72	13.25±0.71	36.90±2.89	6.00±0.96	35.89±0.87
0.20	17.84±2.94	13.34±0.41	50.01±1.45	6.14±0.15	48.47±6.45
0.30	14.48±0.09	13.70±0.23	64.21±0.99	8.15±0.82	61.33±0.67
0.50	13.87±0.79	14.23±1.46	62.52±0.79	5.60±0.87	66.47±0.89

Table 1.3. Solubility of eggyolk (g/100 g) as a function of pH and salt concentrations.

*Sousa, R.D.S. et. al. 2007 Lwt-Food Sci. and Tech. 4**0(7)**, p1253-1258* [33].

Protein	Mutations	Effect
Type S1 dihydrofolate reductase	N48E; N130D	+
Potassium channel KcsA	W26E-V93E-L24D-L81R-L116R	+
Xylose isomerase	R202M-Y218D-V275A; K407E	+
Human translation initiation factor eIF2α	A27Q-L46H-V71K	+
HIV type 1 integrase	F185K	+
Potassium channel KchAfu104	I22E-Y25K-L28S-Y29A-L33S-V36D-I29K-I30K-L33E	+
Human apolipoprotein E C-terminal domain	F257A-W264R-V269A-L279Q-V287E	+
Moloney murine leukemia virus reverse transcriptase	L435K	+
SIV integrase	F185H	+
Human apolipoprotein D	W99H-I118S-L120S	+
Cholera toxin A1 subunit	F132S	+
HhaI methyltransferase	V213S	+
CD58	F1S-V9K-V21Q-V58K-T85S-L93G	+
Catalytic domain of beta4gal-T1	A155E-N160K-M163T-A168T-T242N-N255D-A259T	+
Human leptin	W100E	+
Human leptin	W100Q-W138Q	+
Xylose isomerase	S388T	−
HhaI methyltransferase	M51K	o

Table 1.4. Surface mutations to rationalize protein solubility, *Trevino S.R., et. al. 2008 J Pharm Sci, 97(10): p. 4155-66* [36].

Tag	Advantages	Disadvantages
GST	Efficient translation initiation; Inexpensive affinity resin; Mild elution conditions	High metabolic burden; Homodimeric protein; Does not enhance solubility
MBP	Efficient translation initiation Inexpensive affinity resin; Enhances solubility; Mild elution conditions	High metabolic burden
NusA	Efficient translation initiation; Enhances solubility	High metabolic burden
Thioredoxin	Efficient translation initiation; Enhances solubility	Not an affinity tag
Ubiquitin	Efficient translation initiation; Might enhanc solubility	Not an affinity tag
FLAG	Low metabolic burden; High specificity	Expensive affinity resin; Harsh elution conditions
BAP	Low metabolic burden; Mild elution conditions	Expensive affinity resin; Variable efficiency of enzymatic biotinylation; Does not enhance solubility
His6	Low metabolic burden; Inexpensive affinity resin; Mild elution conditions; Tag works under both native and denaturing conditions	Specificity of IMAC is not as high as other affinity methods; Does not enhanc solubility
STREP	Low metabolic burden; High specificity; Mild elution conditions	Expensive affinity resin; Does not enhance solubility
SET	Enhances solubility	Not an affinity tag
CBP	Low metabolic burden; High specificity; Mild elution conditions	Expensive affinity resin; Does not enhance solubility
S-tag	Low metabolic burden; High specificity	Expensive affinity resin; Harsh elution conditions (or on-column cleavage); Does not enhance solubility

Table 1.5. Commonly used fusion tags and their advantages and disadvantages, D*avid S.W.; 2005 TRENDS in Biotech, 23(6)* [40].

Chapter – 2

Analysis of amino acid contributions to protein solubility using short peptide tags fused to a simplified BPTI variant

2.1.1. Introduction

Protein solubility is an important, but often overlooked, property. It is associated with several human diseases [6], and *in-vitro* protein solubility is becoming an important issue in several areas of biotechnology, including the production of protein pharmaceuticals [17, 68]. Solubility is usually estimated in terms of a hydrophobicity-derived scale. However, hydrophobicity is not an actual measure of solubility propensity, because it refers to the transfer of amino acids from an aqueous to a nonpolar solution [57-58]. Biophysical attempts to rationalize the effects of amino acid mutations on protein solubility have been limited by various intrinsic and extrinsic factors that influence protein solubility [69-71]. A genuine solubility propensity scale for amino acids might allow the prediction of protein solubility based on the amino acid sequence of the protein, more accurately than using the present hydrophobicity-derived scales.

The "host-guest" approach, as applied to peptides and proteins, provides an experimental method for measuring the tendencies of amino acids to adopt certain states or conformations. It was used by Scheraga's group in the 1960s [50] and later by Baldwin's group [51], to determine the helix propensities of amino acids. It was also used to probe the structural/biochemical context dependencies of helix [52], or β-sheet formation within a protein [45]. However, a host-guest-like approach has rarely been used to analyze the contributions of amino acids to protein solubility, and when it has been applied, the results were confounded by effects arising from the surrounding structural/biochemical environments of the host protein [53].

The solubility of amino acids range widely, from a few score micro molar to several hundred mili molar, which presents a technical hurdle, among many others, for the construction of a solubility propensity scale, because such a wide range of values is difficult to measure accurately with a single protocol. Peptide sequences consisting of a single amino acid type (poly-amino-acid peptides) can be useful in partly overcoming this problem, because they tend to amplify the adhesive, aggregation, polymerization, and solubility properties of the amino acids.

2.1.2. Motivation

Recently, we used the biochemical properties of charged amino acids to improve protein solubility by fusing short poly-Lys or poly-Arg tags to a bovine pancreatic trypsin inhibitor (BPTI) variant, which increased its solubility by over six-fold without altering its NMR spectrum or activity [49]. We therefore reasoned that poly-amino-acid peptide tags, which can significantly modify a protein's solubility, could be used to determine the contribution of individual amino acids to protein solubility. Here, we report a host-guest like approach in which we used poly-amino-acid tags to investigate the effects of amino acid mutations/additions on protein solubility. We chose 10 amino acid types representing the full range of biophysical properties (acidic, basic, polar, and hydrophobic). To this end, we fused short poly-amino-acid tags (guests), consisting of one of the selected amino acids, to the C-terminus of our host protein, a simplified BPTI variant [55-56], in which most of the surface residues were alanines, which we expected would minimize the interactions between the poly-amino-acid tag and the host protein. We ensured that the measured contributions of the amino acids to the protein's solubility were context independent by confirming that the peptide tag did not affect the structural, functional, or thermodynamic properties of the host protein.

2.1.3. Aim and objectives

The aim of this research (Paper 1) is to focus effect of tags of different types of amino acids on host proteins. The solubilization effects of amino acids were generally consistent with those anticipated from hydrophobicity and hydrophilicity scales. However, we also observed clear discrepancies, which suggest that the development of a genuine solubility propensity scale should provide new insight into the biophysics of protein solubility.

2.2. Methods and Materials

2.2.1. BPTI mutant design and expression

BPTI is a small globular protein of 58 residues with three disulfide bonds at positions (5-55), (14-38) and (30-51). In this study, we used a simplified BPTI in which most of the surface residues were alanines [14-15], as a host protein. This is because alanine is regarded as a biochemically and structurally neutral residue [72-73] and because alanine on the protein's surface is thus anticipated to minimize interactions between the host protein and the tag peptide. Seven of the ten positive residues and two of the four negative residues in the wild type are left in BPTI-19.

BPTI-19 was tagged with different poly-amino-acid peptides composed of a single amino acid type. We designed 10 poly-amino-acid-tagged variants, which we named according to the number and type of amino acid included [49]. For example, C5R denotes a BPTI-19 variant with a five-Arg-residue tag added to its C-terminus. The poly-amino-acid peptide tags could be added to either terminals, but the solubilization effect was largely independent on the site of addition [49].

The 10 amino acids types were chosen to include the full range of amino acids biochemical properties and because of their occurrence in natural and synthetic poly-amino-acid sequences. Polar (N, Q, S), hydrophobic (I), positively charged (R, K), and negatively charged (D, E) amino acids, together with Pro (P), were chosen as representative amino acids (Table 2.3.1). We also included His because His tags are commonly being used in molecular biology, as well as Gln because of its relationship to Huntington's disease [25]. We also added two Gly residues to all of the variants as a spacer between the poly-amino-acid tag and BPTI to ensure the flexibility of the poly-amino-acid tag and to reduce any putative interaction between the tag and the molecular surface of BPTI. A BPTI variant with two Gly residues added to its C-terminus was used as the reference molecule.

All BPTI mutants described herein are constructed in a pMMHa vector, start with His-tag-TrpΔLE leader followed by a single methionine, a BPTI mutant sequence and a termination codon [56] (Figure 2.2.1). BPTI mutants were designed based on an alanine scanning experiments [73]. All mutants identities were confirmed by DNA sequencing (using the Big-dye sequencing protocol).

BPTI variants were expressed using the pMMHa expression vector in an *Escherichia coli* cell line as inclusion bodies. In short, 1 µl of plasmid DNA was mixed with 50µl of JM109(DE3)pLysS cell, kept on ice for 30 minutes and transformed using heat shock at 42ºC for 45 seconds, followed by being kept on ice for 2 minutes. Transformed cells were spread on a LB plate supplemented with Ampicillin (Amp, 50mg/ml) and Chlorampenicol (Cp, 35mg/ml) and incubated for overnight at 37ºC. A single colony was then transferred into 50 ml of LB medium (supplemented with Amp

and Cp) and cultured at 37ºC and 220 rpm for 7~8 hours. Finally, 50ml of the pre-culture was added into 2L of LB medium and cultured at 37ºC, at 120 rpm for 15~16 hours (Figure 2.2.2).

Cells were collected by centrifugation at 6000 rpm for 15 minutes at 4ºC and immediately frozen at -85ºC. Frozen cells were suspended in 1X Lysis Buffer (50mM Tris-HCl pH8.7, 150mM NaCl) and sonicated (ASTRASON^{TM} XL2020, Misonix) at output7, 1min for three times keeping on ice. Pellets were collected by centrifugation at 7000 rpm for 20min at 4ºC, and re-sonicated in 1X Lysis Wash Buffer (50mM Tris-HCl pH8.7, 1%v/v NP-40, 0.1%v/v Deoxycholic acid, 1mM EDTA) for twice and pellets were collected by centrifugation at 7000 rpm for 20min at 4ºC (Figure 2.2.2).

2.2.2. BPTI mutant purification

After cell lysis by sonication, the cysteines were air oxidized overnight in guanidine hydrochloride (6MGdn-HCl, 50mM Tris-HCl pH8.7; 50ml Gdn-HCl for 2L culture) at room temperature. Refolded mutants were then extensively dialyzed against water (with three to four times water exchanges) at 4ºC using Spectra/Por dialysis membrane (with MWCO of 14000Da). Refolded mutants tagged with a His-tag TrpΔLE leader were then collected by centrifugation at 7000 rpm for 20min at 4ºC.

Then the fusion partner, a His-tagged TrpΔLE leader, was cleaved by cyanogen bromide treatment of methionine present in-between the His-tagged-TrpΔLE-leader and BPTI mutant sequences. The CNBr-treatment was carried out at 10mg/ml concentration of CNBr in 70% formic acid for 4~5 hours at room temperature. Provided, for 2L culture, the CNBr-treatment was conducted in two 50ml falcon tubes

each containing protein collected from 1L culture in 20ml of 70% formic acid. Then, the CNBr and some of the formic acid were removed by running Speed vac. for 2 hours and the protein samples were dialyzed overnight against water at 4ºC (four times water exchanges) using Spectra/Por dialysis membrane (with MWCO of 3500Da). The cleaved partner was removed as precipitate upon dialysis in 10mM Phosphate Buffer, pH6.0 at 4ºC. The BPTI mutant and His-tagged TrpΔLE leader were separated by centrifuging at 7000 rpm for 20 minutes at 4ºC and confirmed by SDS-PAGE (Figure 2.2.3).

Finally, mutants were purified by reverse phase HPLC. In short, supernatant obtained after the CNBr treatment followed by His-tagged TrpΔLE leader separation, was supplemented with 10% acetic acid followed by injection into reverse phase HPLC column (YMC-Pack PROTEIN-RP S-5μm x 250 mm, YMC) at 25% acetonitrile flow in a total flow of 8ml/min. After completion of the acetic acid peaks, a time-gradient program (25%~60% acetonitrile, at 1%/min) was run (Solution A: Water with 0.1% TFA: solution B: acetonitrile with 0.05% TFA). Mutant-specific HLPC illusions were collected and their identities were confirmed by MALDI mass spectroscopy (Figure 2.2.3). Purified mutants were then lyophilized, and preserved at -85ºC until used.

2.2.3. Trypsin inhibition activity measurement

The trypsin inhibition activity of the BPTI variants was assayed by monitoring the hydrolysis of N-α-benzoyl-D-L,-arginine p-nitroanilide (BAPA) in 0.2M phosphate buffer (pH 7.0) at room temperature [74]. In short, BPTI variants and trypsin were mixed at equimolar concentrations (280nM) and incubated for five minutes. BAPA

(1mg/ml in water with little DMSO added) was added to the sample, and the activities of the BPTI variants were measured by monitoring the changes in absorbance at 405 nm for five minutes. The same procedure was followed for determination of the background spectra in the absence of trypsin and BPTI, while the hydrolysis of BAPA by trypsin was measured in the absence of BPTI variants (Figure 2.2.4).

2.2.4. Protein solubility measurement

Protein solubility was determined as the maximum protein concentration in the supernatant of super-saturated protein solutions, as described previously [49]. In short, solubility measurements were performed at 25ºC in 50mM buffers at pH4.7 and pH8.7 in the presence of 1.3M and 1.5M ammonium sulfate, except when indicated. The samples with different initial concentrations were mixed thoroughly and allowed to equilibrate for 20 minutes at 25ºC. After equilibration, the samples were centrifuged at 20000xg for 20 minutes at 25ºC and protein concentrations of the supernatants were determined by UV analysis using a NANO DROP UV-spectrophotometer. In order to ensure reliability and reproducibility of the final solubility data, all protein solubility values were measured at least three times, in different working days, and with different sets of reagents (Figure 2.2.5). Experiments were carried out at pH 4.7 and pH 8.7. For pH 4.7 we used 50mM using acetate buffer, whereas, for pH 8.7 we used 50mM Tris-HCl buffer. In this Chapter-2 (first paper) we report the buffer pH as our experimental pH which is 8.7. Whereas in Chapter-3 (second paper) we used same buffer but reported the pH value observed after experiment. We observed a change in case of higher pH, pH 8.7 reduced to pH 7.7 after experiment. We did not find any change in pH value for lower pH (pH4.7) after running experiments.

2.2.5. Thermal stability measurement

Thermal stabilities of all the variants were monitored using the Circular Dichroism (CD) signal at 220nm. Samples for CD measurements were prepared by dissolving lyophilized proteins at 5-10 μM concentrations in 20mM sodium acetate (pH4.7) and 20mM Tris buffer (pH8.7) in the presence of 0.0, 0.5, 0.75, 1.0 and 1.5M ammonium sulfate. Measurements were carried out using a 1 cm cuvette with a Jasco J-820 spectropolarimeter, as previously described [49, 56]. Thermal denaturation experiments were conducted at a scan rate of 1ºC/min in the temperature range of 5-75ºC. The reversibility of the thermal denaturation was confirmed by cooling the sample to 5ºC followed by a re-heating to the highest temperature.

2.2.6. Crystallization and structures determination

Crystals of BPTI variants were grown at 20°C using the hanging drop vapor diffusion technique. The X-ray diffraction data were recorded from single crystals using a synchrotron beam line at the Photon Factory (PF, Tsukuba, Japan). The coordinates and structure factors of BPTI-19 and its His-tagged variant (BPTI-19-C5H) have been deposited in the Protein Data Bank under the PDB entry code 3AUB and 3AUE, respectively. This crystallization part of research was facilitated by co-researchers

2.3. Results and Discussion

2.3.1. Effects of poly-amino-acid tags on protein solubility

We measured protein solubility as the maximum protein concentration in the supernatant after centrifugation at 20,000 g for 20 min in 50 mM acetate buffer (pH 4.7) or 50 mM Tris-HCl buffer (pH 8.7) in the presence of 1.3 or 1.5 M ammonium sulfate. Poly-Lys increased protein solubility twelve to thirteen times (Figure 2.3.1,

Table 2.3.1), similar to our previous observation [49]. The solubility increase caused by poly-Arg was more modest (six-fold). In contrast, the negatively charged poly-Asp and poly-Glu tags barely affected protein solubility at pH 4.7 (Figure 2.3.1, Table 2.3.1). The polar (N, Q, S) and poly-His tags increased protein solubility by three to eight-fold. As expected, the addition of five Ile residues significantly reduced the solubility of BPTI. However, the addition of the very soluble Pro altered the solubility of BPTI only slightly (Figure 2.3.1, Table 2.3.1).

The solubilization effects of most poly-amino-acid tags were pH independent (Fig. 2.3.2), and those of the poly-Ile, -Pro, -Gln, -Arg, and -Ser tags were essentially the same at pH 4.7 and pH 8.7. However, the solubilization effects of poly-Asp and poly-Glu increased dramatically, by four to ten-fold, at pHs higher than 7.5 (Fig. 2.3.3). The pH dependence of the effects of the Asp and Glu tags on protein solubility is attributed to the tendency of proteins to aggregate at pH close to their isoelectric points (calculated pI of BPTI, 6.62), whereas at high pH, the ionized Asp and Glu side chains provide a net negative charge, which contributes to an increase in the overall protein solubility. Conversely, the poly-His-tag dramatically reduced the protein's solubility when the pH was increased from pH 4.7 to pH 8.7 (Figure 2.3.2, Figure 2.3.3). This is because the typical residue p*K*a of His is approximately 7.0, at which both the acidic and basic forms coexist; a change in the pH will affect the tautomeric equilibrium, reducing the solubility of the protein. The poly-Asn tag increased the solubility of the protein when the pH was increased, but no such effect was observed for poly-Gln, whose solubilization effect was pH independent.

Our observations were generally consistent with the solubility changes anticipated from their hydrophobicities [57-58], and the solubilities of the individual amino acids 102. For example, peptide tags composed of charged or hydrophilic residues increased the solubility of the protein, as anticipated (Figure 2.3.4). However, there were several striking discrepancies that cannot be overlooked when predicting the solubility of a polypeptide from its amino acid sequence. First, in quantitative terms, the solubilization effects of the tags composed of charged or polar residues did not correlate well with either their hydropathy [57], and/or hydrophilicity [75] scales or the solubilities of the individual amino acids [76] (Figure 2.3.5), as demonstrated by the near-zero correlation coefficients between these scales (nor did the other scales correlate with one another). Furthermore, hydropathy (as well as hydrophobicity) is usually a pH-independent scale, but the data for Asp, Glu, His, and to a lesser extent Asn clearly indicate or confirm that pH is an important factor influencing protein solubility. The Pro tag barely affected protein solubility, which is consistent with its low hydrophobicity, but is in sharp contrast to its solubility (Figure 2.3.5), which is reported to be as high as 1600 g/L 102. These observations demonstrate that the actual contribution made by an amino acid to protein solubility can differ from the effects anticipated based on either hydrophobicity or the solubility of individual amino acids.

2.3.2. Ammonium salts and reliability and reproducibility of the measurements

Although our protocol allows us to measure, in principle, protein solubility in a pure aqueous solution, the addition of ammonium sulfate was essential to circumvent technical difficulties associated with determining protein solubility at very high protein concentrations. For some tagged BPTI variants, measuring their solubility in the absence of salt required extremely high protein concentrations, which not only

necessitated a large amount of purified protein, but more importantly, caused poor phase separation or even gel formation. The addition of ammonium sulfate helped to achieve a well-defined separation of the aqueous and solid phases, with high reproducibility. Ammonium sulfate also alleviated a peculiar phenomenon whereby the final protein solubility depended on the initial amount of dissolved protein [36, 54]. Our measurements in the absence of salt or at low salt concentrations confirmed this observation. However, the solubility limit became almost independent from the initial amount of protein when ammonium sulfate was added to the solution, and the solubility limit was reached consistently. We speculate that the solubility limit dependence on the initial protein concentration at low salt concentrations might be related to the extremely slow kinetics of aggregation or to residual water and/or buffer molecules contained in the lyophilized protein powder. In either case, these effects were minimized by the addition of ammonium sulfate, ensuring highly consistent and reproducible measurements. All the solubility measurements reported in this study were repeated 3-5 times using different initial protein concentrations, and the final solubility limits of all the variants were within 10% of one another (Figure 2.3.1-Figure 2.3.5).

2.3.3. Protein stability and independence of the tag peptide from the host protein

Protein solubility dramatically decreases when the protein unfolds, and it is therefore important to ensure that the host protein remains fully folded during the solubility measurements. Under all conditions (0.0-1.5 M ammonium sulfate at pH 4.7-8.7), all the variants exhibited a two-state reversible thermal denaturation curve when probed with circular dichroism (Figure 2.3.6).

The stability of all variants showed a small increase with increasing ammonium sulfate concentrations, probably attributable to a salting-in/salting-out effect. The melting temperatures (T_m) of the poly-amino-acid-tagged variants were identical to that of the reference C2G variant, both in the presence and absence of ammonium sulfate, within an experimental error of ± 0.5-1°C (Figure 2.3.7 and Figure 2.3.8). Moreover, in the presence of 1.5M ammonium sulfate, the thermal folding/unfolding of the tagged variants was almost perfectly reversible (Figure 2.3.6). These observations suggest that the residues of the poly-amino-acid tag did not interact with the surface of the BPTI variant, and that all the variants were fully folded at 25°C, the temperature at which solubility was measured.

2.3.4. Structural and functional independence

The crystal structures of BPTI tagged with the poly-amino-acid tags provided a reliable and direct assessment of the minimal interactions that occur between the poly-amino-acid tags and the molecular surface of the host BPTI protein (Figure 2.3.9). In all the poly-amino-acid-tagged variants, both the backbone and side-chain structures of the host protein were almost perfectly retained, with an average backbone deviation of < 0.3 Å (BPTI-19, C5S, C5N, C5H, C5P, C3E, C3R, and C3K; manuscript in preparation) (Figure 2.3.10).

The trypsin inhibition activities of all of the tagged variants were completely identical to that of the untagged BPTI (Figure 2.2.4 and Table 2.3.1). Because the tag is located close to the BPTI-trypsin binding site, the unaltered trypsin inhibitory activity confirms that the poly-amino-acid tags did not interact with the molecular surface of the host protein.

2.3.5. Insight into polypeptide solubility

A major difficulty in predicting protein solubility is that the effects of several factors are intertwined. Moreover, technical hurdles originating from high protein concentrations must be overcome. Our strategy for circumventing these factors was to take advantage of the near-neutral biochemical characteristics of our alanine-simplified BPTI, and the tendency of the poly-amino-acid tags to amplify the effects of the particular amino acid on protein solubility. Another advantage of this strategy is that the poly-amino-acid tags are fully independent of the host BPTI protein, and we can therefore expect to measure context-independent solubility (i.e., independent of the surrounding protein environment).

Although high concentrations of ammonium sulfate are well known to reduce protein solubility, we consider that the relative amino acid solubilities were not strongly affected by its presence. Although our present results were obtained in the presence of ammonium sulfate, they nevertheless provide some useful insights into how amino acid composition influences the relative solubility of a polypeptide. In particular, the behavior of the Asp and Glu tags provide an explanation of a protein's tendency to aggregate at pHs close to its isoelectric point (pI), which is not explained by a pH-independent hydrophobicity scale. According to our measurements, the solubility of acidic proteins, which contain many Glu, Asp, and Asn residues, should increase at high pH, but the solubility of basic proteins would be little affected in the pH range 4.7-7.7. Accurate predictions of solubility will require the determination of the relative solubilities at various temperatures and under different conditions, including low salt concentrations. Furthermore, the contribution of cross terms, which originate from mixing two types of amino acids in a tag, must be assessed. Such corrections could, in principle, be readily accommodated in the solubility prediction.

2.4. Conclusions

Although solubility is an important property of proteins, few, if any, systematic studies have been conducted into how each of the 20 natural amino acids modulates protein solubility. Several factors may have hampered the construction of a proper "solubility propensity scale": the strongly intertwined nature of the intrinsic and extrinsic factors that influence protein solubility, the wide range of solubility values of amino acids, and technical difficulties associated with reliably measuring high protein concentrations. However, this study indicates that poly-amino-acid peptide tags fused to a simplified host protein can provide a valuable method for measuring the effects of amino acids on protein solubility, in a context-independent manner. Such measurements could yield a genuine solubility propensity scale, which could eventually allow the calculation of the relative solubility of a polypeptide from its amino acid sequence.

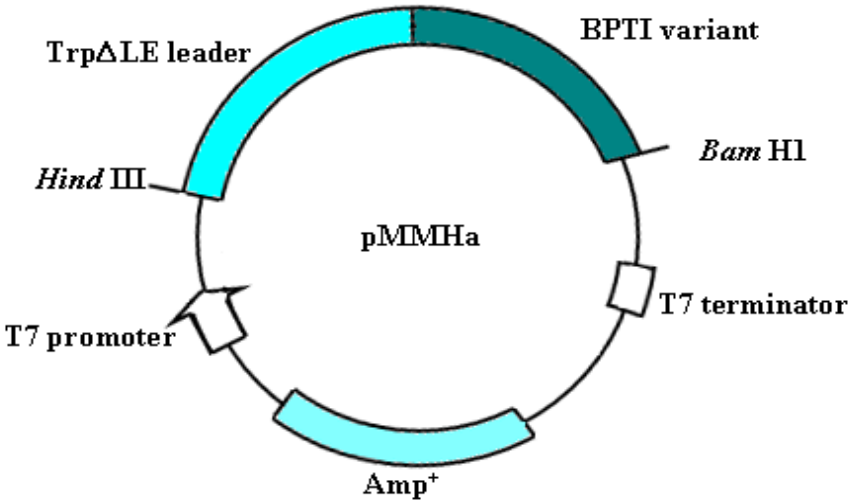


Figure 2.2.1. The plasmid for protein expression.

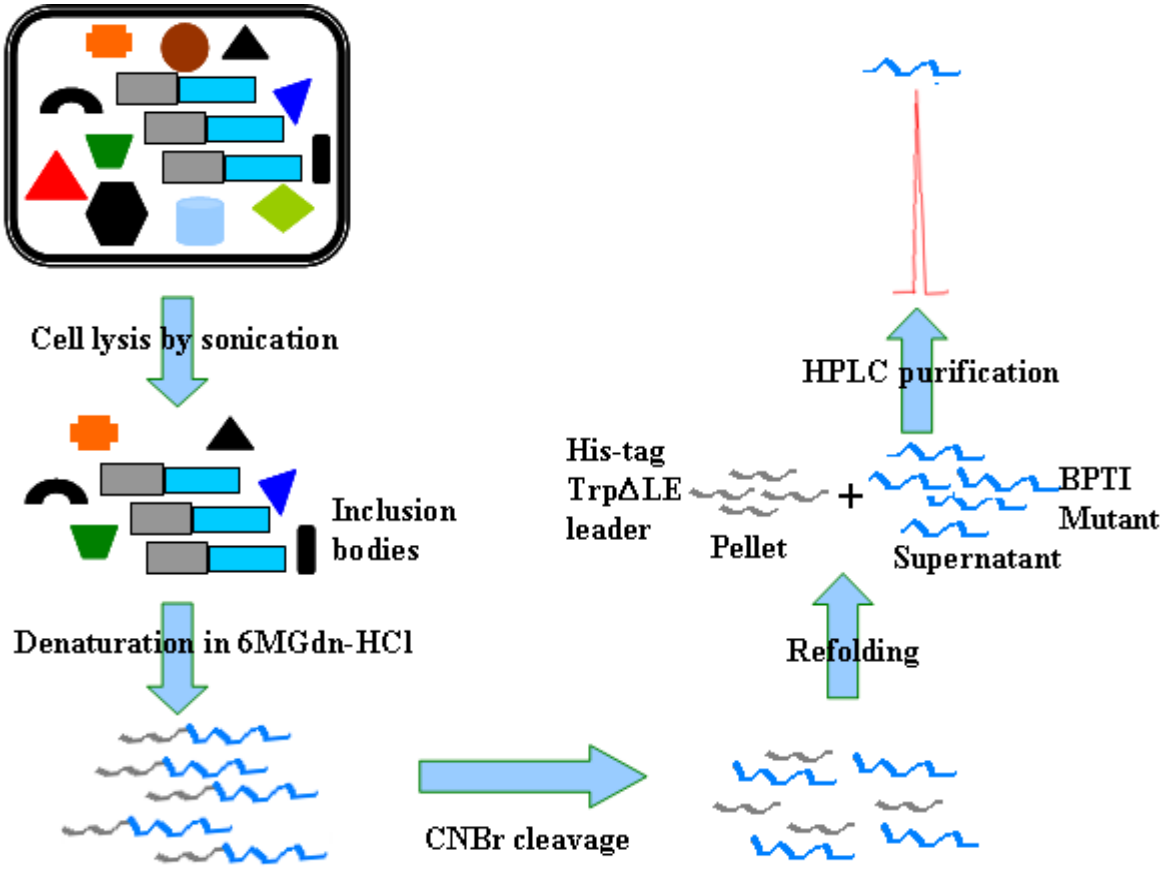
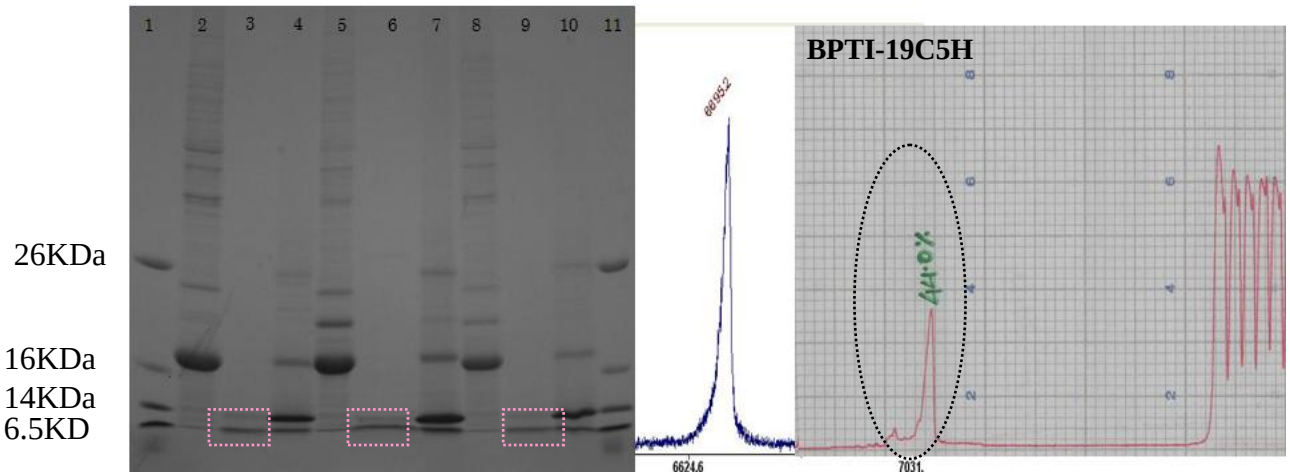


Figure 2.2.2. Schematic representation of purification protocol for BPTI variants.



SDS-PAGE Analysis

Lanes (1,11):marker;(2,5,8):after cell lysis by sonication; (3,6,9): supernatant; and (4,7,10): pellet obtained after CNBr-cleavage followed by dialysis.

HPLC Condition

Column: YMC-Pack/Protein-RP(250x20mmID);Solution A: milli-Q(0.1%TFA); Solution B: CH_3CN (0.05% TFA); Flow rate: 8ml/min; Sample vol: 4X3ml; Time program: Solution B from 15 to 60% at the rate of 1%/min .

Figure 2.2.3. Purification of BPTI variants**.** SDS-PAGE analysis (left) and HPLC chromatogram (right) are shown while confirmation of the HPLC purified sample by molecular mass determination is shown in blue rectangle (middle). The measured molecular weight of the variant C5H (6783.5Da) is very close to the mass calculated from the amino acid sequence (668.9Da).

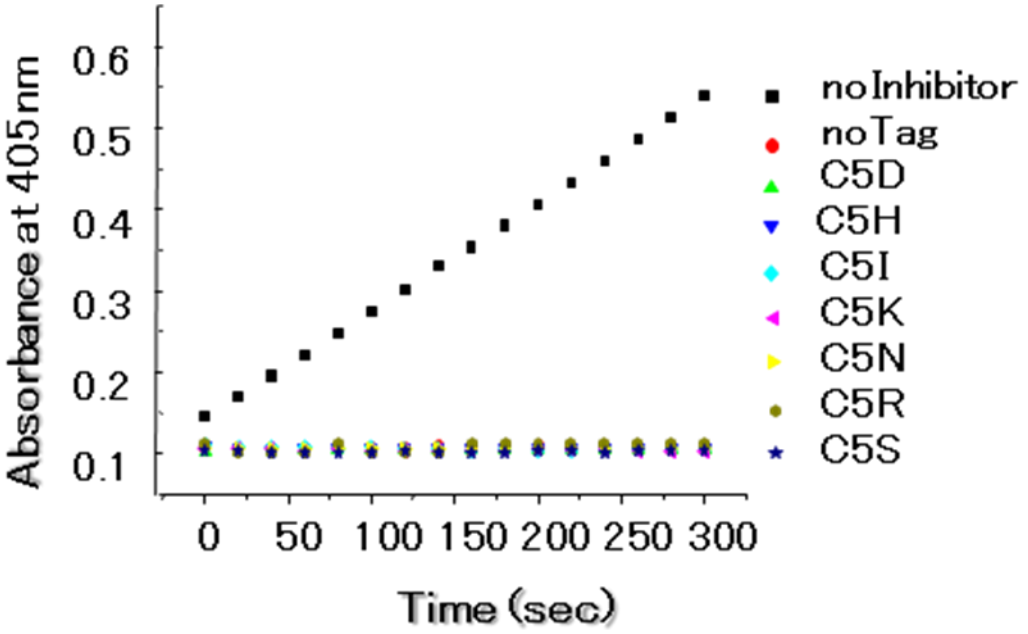


Figure 2.2.4. Trypsin inhibition activity of BPTI variants. Black tetra angular represent the activity of trypsin on N^{α}-benzoyl-D,L-arginine *p*-nitroanilide (BAPA) in absence of BPTI, while color symbols represent the activity of trypsin on BAPA in presence of BPTI-19 mutants at equimolar concentration (280nM) of Trypsin and BPTI variants.

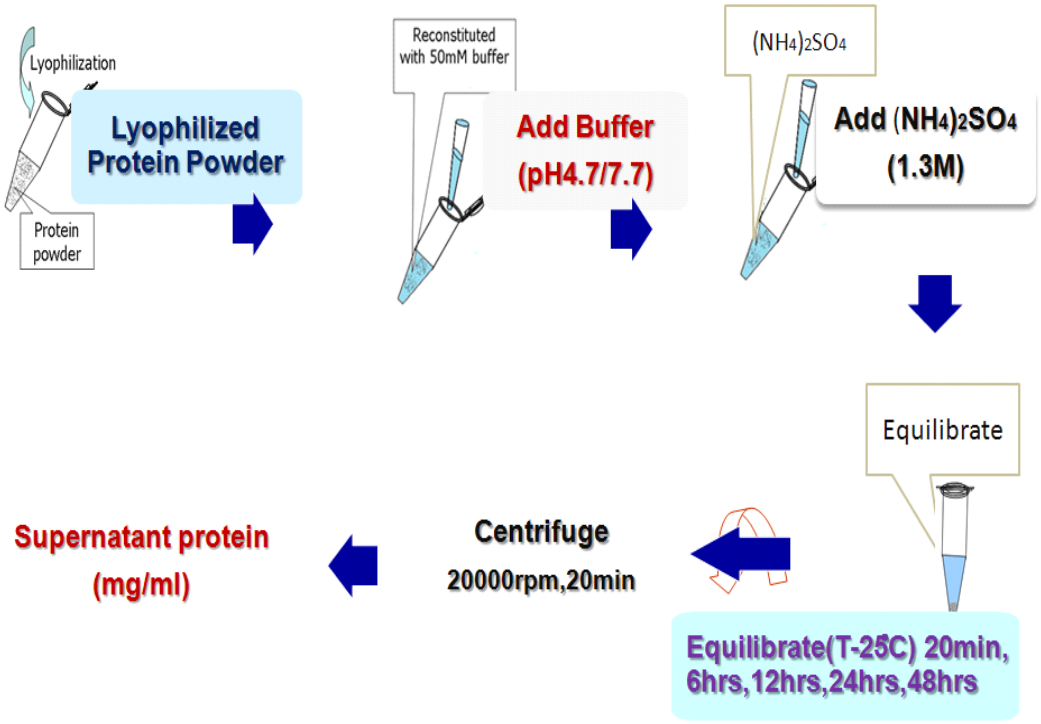


Figure 2.2.5. Solubility measurement protocol. Dissolving lyophilized protein powder in buffer then adding ammonium sulfate. Mixture were equilibrate at 25^{o}C then centrifuge. Finally protein concentrationin supernatant were measured by NANO DROP UV-spectrometer.

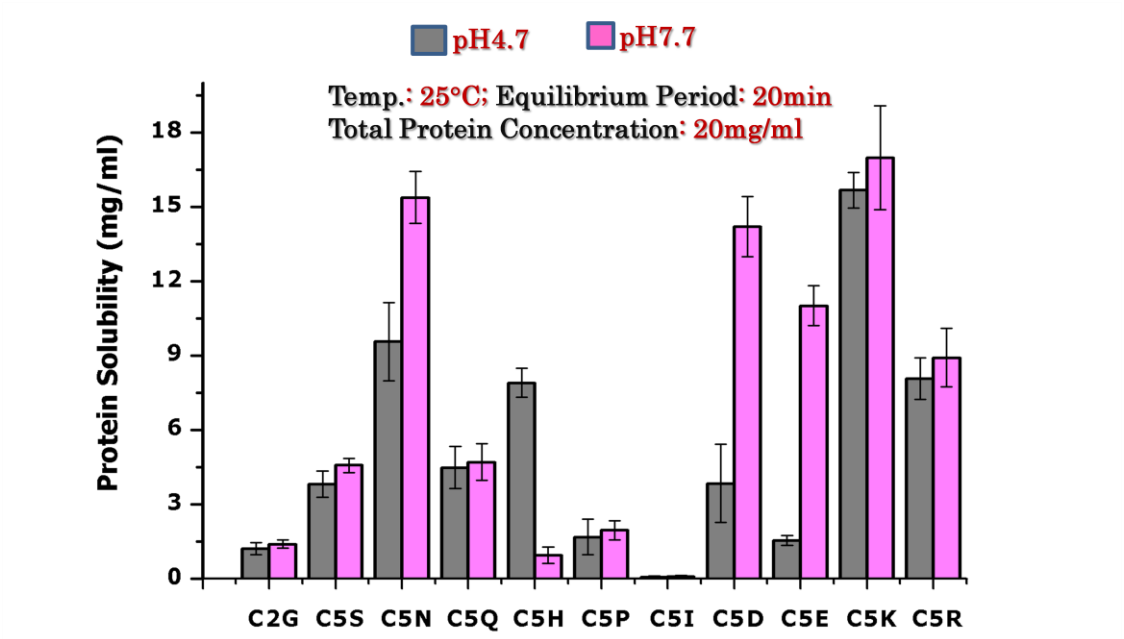


Figure 2.3.1. Effect of poly amin acid tags on protein solubility. Protein solubility in 50mM acetate, pH4.7 and in 50mM TrisHCl, pH8.7 in the presence of 1.3M ammonium sulfate at 25ºC.

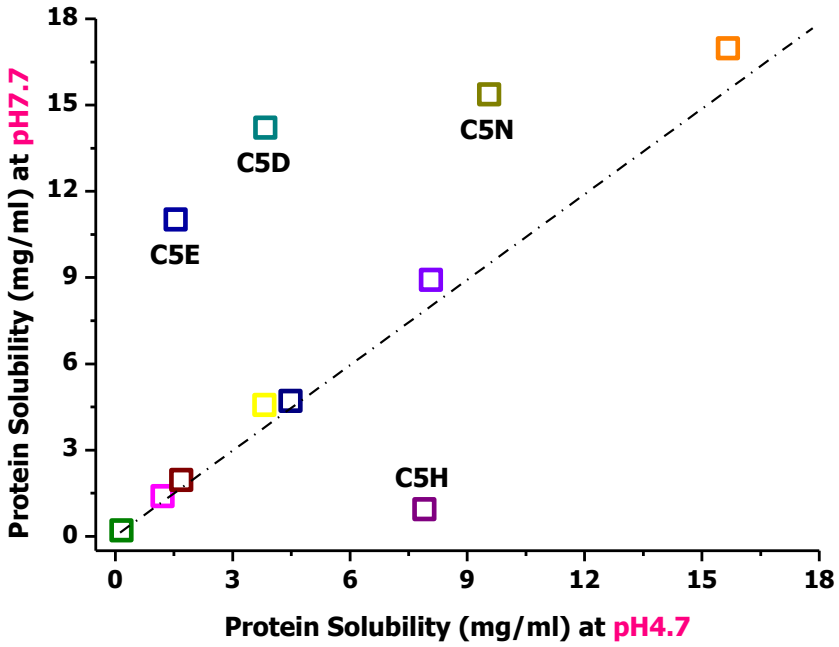


Figure 2.3.2. Correlation plot of protein solubility as function of pH (pH4.7 vs pH8.7).

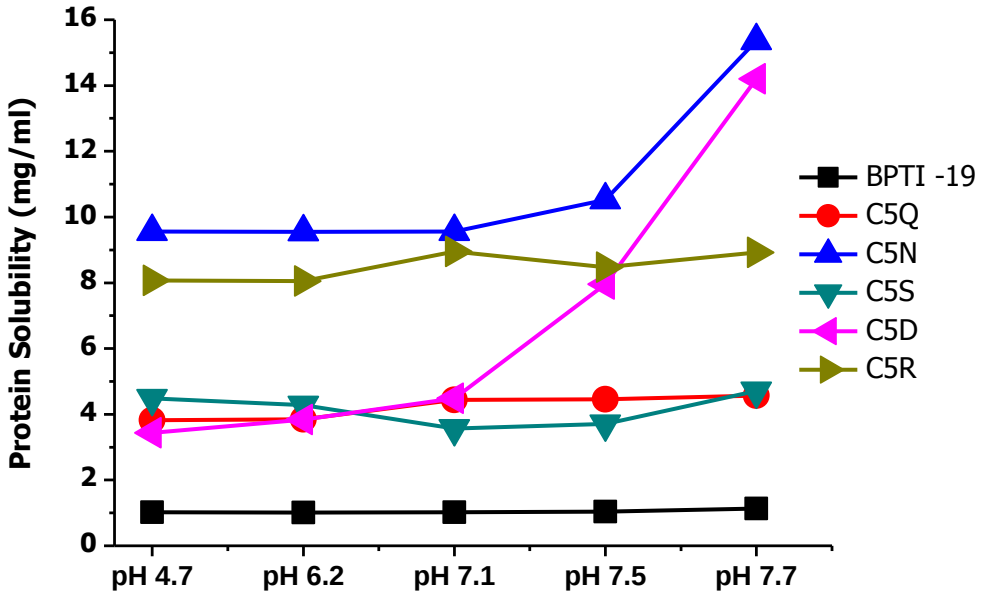


Figure 2.3.3. pH dependency of protein solubility. Measurements were conducted in 50mM acetate (pH4.7 and 5.6), phosphate (pH7.0) and TrisHCl (pH7.5, 8.0 and 8.7) buffers (■:BPTI-19; ●:C5Q, ▲:C5N, ▼:C5S, ◀:C5D ▶:C5R). All measurements were started with the same total protein concentration. Values reported here were measured independently from those reported in Table 2.3.1.

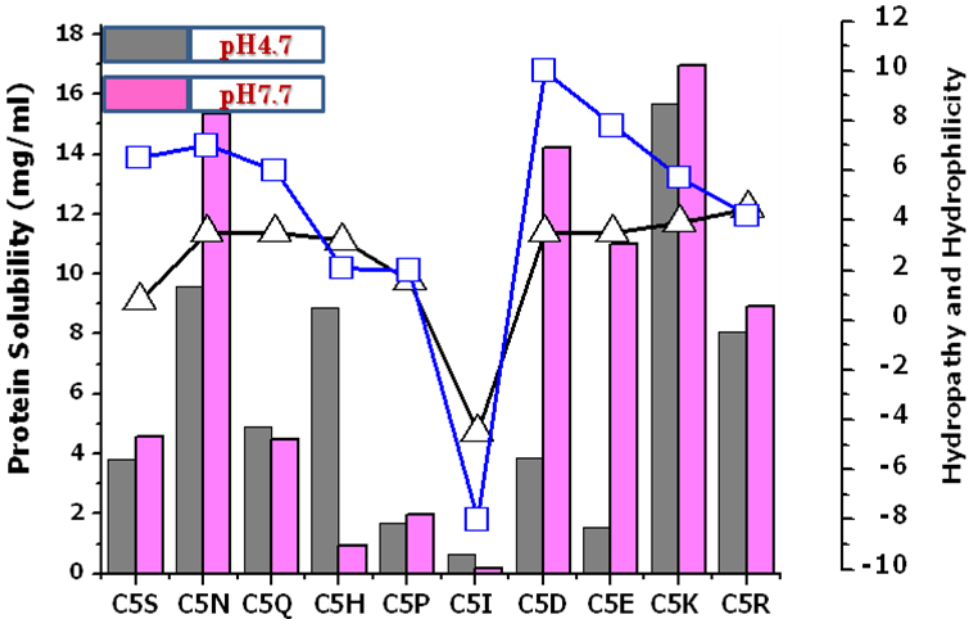


Figure 2.3.4. Comparative view of protein solubility with amino acid hydropathy (Δ), and hydrophilicity (□), *(Hydrophilicity; Parker JM.; et.al.; 1986 [69] and Hydrophobicity; Kyte J.; et.al.; 1982* [57].

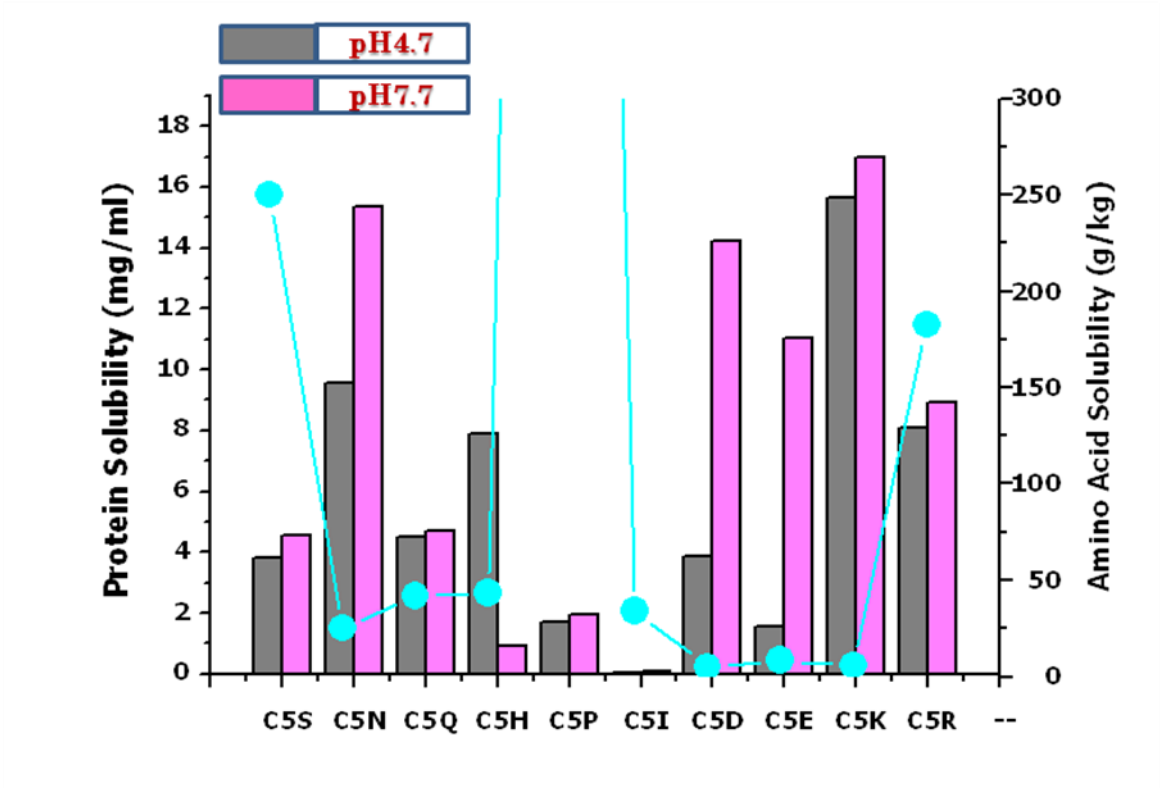


Figure 2.3.5. Comparative view of protein solubility with amino acid solubility. (*Amino acid`s solubility* (●), *(CRC handbook 90th edition, 2009)* [76].

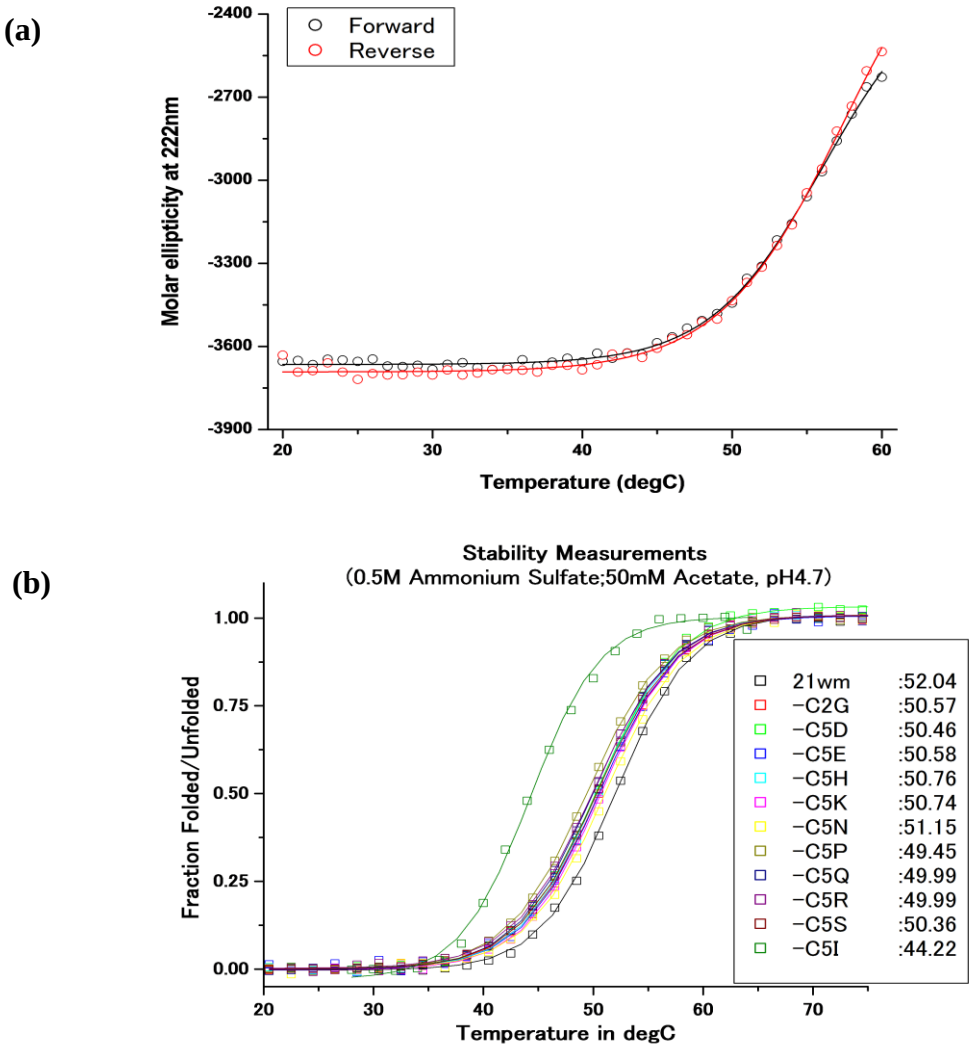


Figure 2.3.6. Thermal stability of the untagged and poly-amino acid tagged BPTI variants in the presence of ammonium sulfate. (a) Thermal stability of the poly-Asp tagged variant in the presence of 1.5M ammonium sulfate at pH8.7. CD spectra at 222nm were first recorded from 20-60ºC (forward) and then 60-20ºC (reverse) at a scan rate of 1ºC/min. (b) Thermal stability in the presence of 0.5M ammonium sulfate in 50mM acetate buffer of pH4.7. The melting temperatures are shown in parentheses. In both panels, symbols represent raw data while continuous lines stand for fitted data.

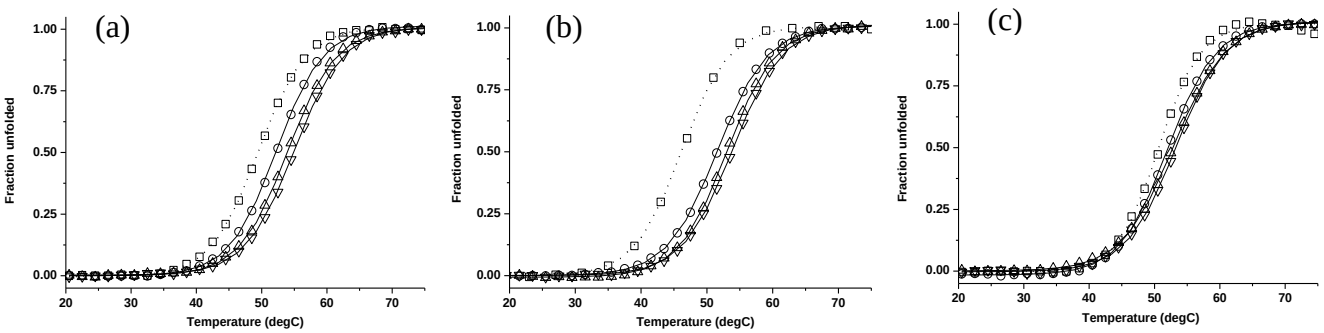


Figure 2.3.7. Thermal stability of poly-amino-acid tagged variants in the presence of ammonium sulfate in 50mM TrisHCl (pH8.7) using circular dichroism. (a) Untagged BPTI, (b) poly-Asp and (c) poly-Lys variants are shown. All measurements were conducted at a protein concentration of 5μM with 0.00M (□), 0.50M (○), 0.75M (△), and 1.00M (▽) ammonium sulfate. Symbols represent the raw data while continuous lines represent the fitted data. Dotted lines represent thermal denaturation curves in the absence of ammonium sulfate. Similar patterns of protein stability were observed at pH 4.7 in 50mM acetate buffer.

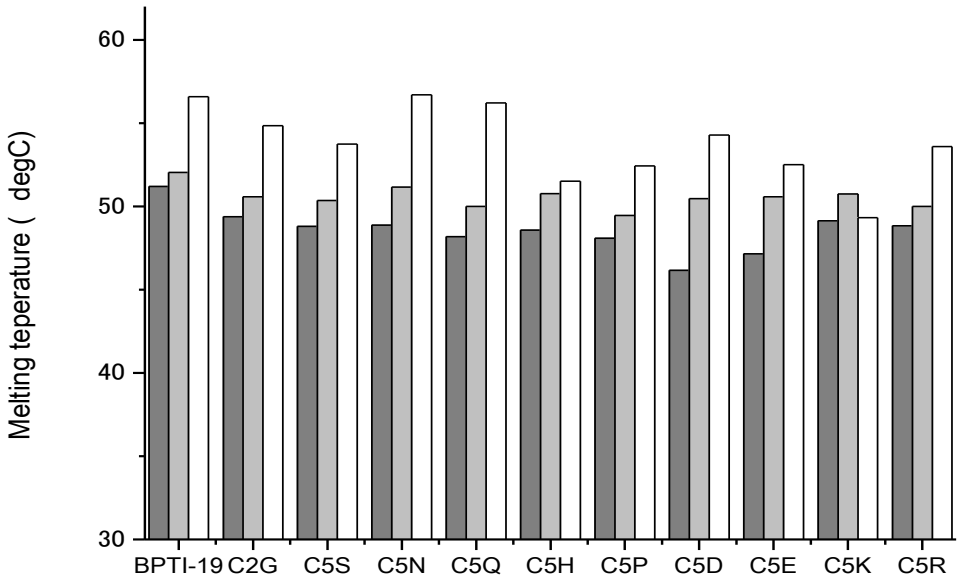


Figure 2.3.8. Melting temperatures BPTI variants. Calculated from the previously mentioned denaturation curves (Ammonium sulfate concentration: ■:0.0M; ■:0.5M and □:1M).

```
(a)            1         10         20         30         40         50        58
Wild type   : RPDFCLEPPY TGPGKARIIR YFYNAKAGLC QTFVYGGCRA KRNNFKSAED CMRTCGGA
BPTI-19-C5X: RPAFCLEPPY AGPGKARIIR YFYNAAAGAA QAFVYGGVRA KRNNFASAAD ALAACAAA-GGXXXXX
```

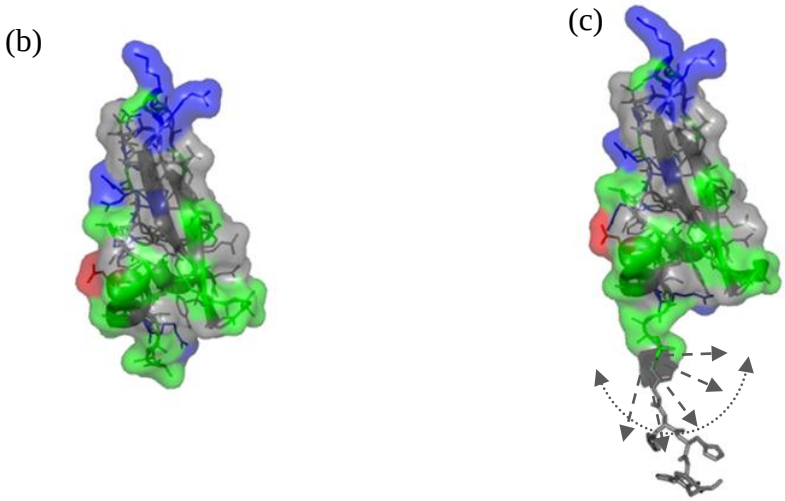


Figure 2.3.9. Sequence and structure of BPTI variants. (a) Sequence of the wild type BPTI and the poly-amino-acid tagged BPTI-19 variants. Alanine are in green, positively and negatively charged residues are in blue and red respectively. The poly-amino-acid tag residues are underlined. Two Gly were added as spacer residues, followed by five amino acids (X) of the same type. Ribbon and surface representation of (b) BPTI-19 (PDB ID-3AUB) and (c) BPTI-19-C5H (PDB ID-3AUE). The poly-His tag was partially visible, but fully extended pointing outward the BPTI-19's molecular surface.

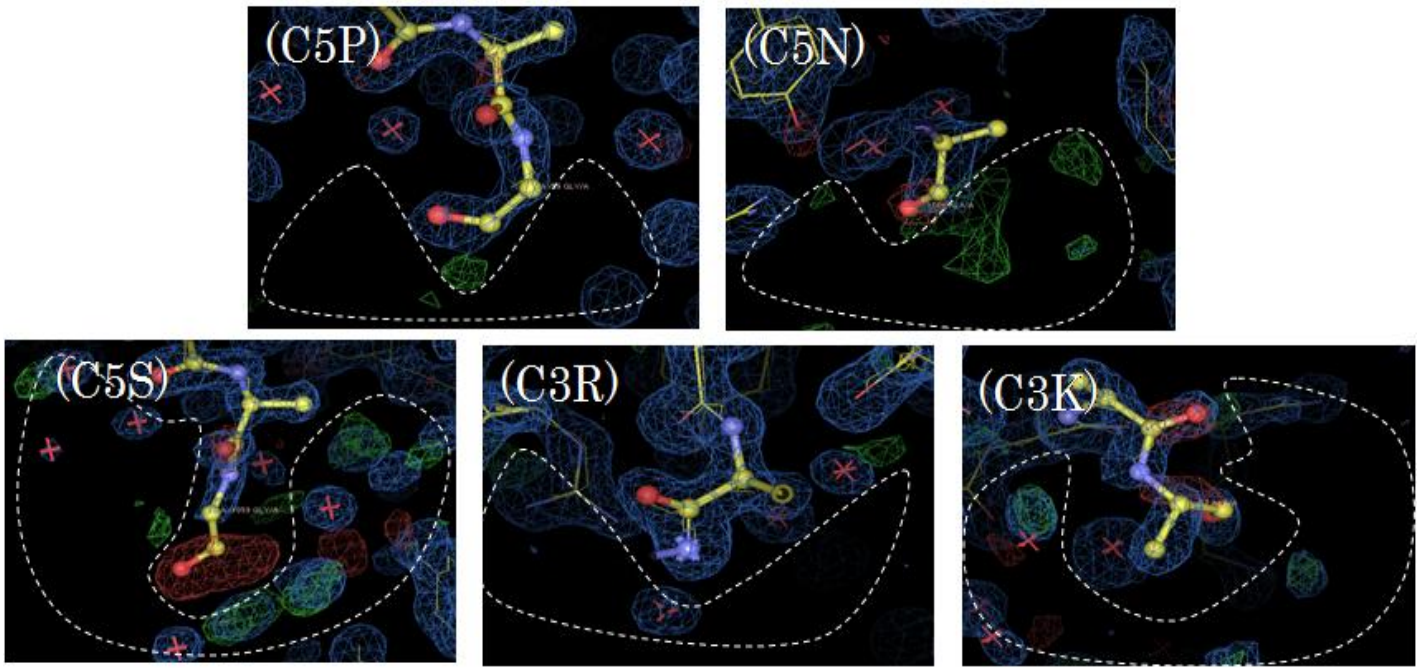


Figure 2.3.10. Structure of BPTI variants (C5P, C5N, C5S, C3R and C3K). The electron densities of the other poly-amino-acid tag were mostly invisible, indicating their highly flexible conformation.

Mutants	Protein Solubility[a] (mg/ml)		Stability[b] *T*m (°C)	pI[c]	Activity[d] (%)	Mutants	Protein Solubility[a] (mg/ml)		Stability[b] *T*m (°C)	pI[c]	Activity[d] (%)
	pH4.7	pH8.7					pH4.7	pH8.7			
BPTI-19	1.14±0.05	1.43±0.15	54.04	9.7	100	C5P	1.69±0.71	1.95±0.38	52.44	9.7	100
C2G	1.21±0.25	1.39±0.16	54.85	9.7	100	C5I	0.65±0.15	0.20±0.05	44.22	9.7	100
C5S	3.81±0.52	4.57±0.29	53.74	9.7	100	C5D	3.84±0.52	14.20±1.21	54.29	6.6	100
C5N	9.56±1.57	15.37±1.04[e]	52.44	9.7	100	C5E	1.54±0.21	11.02±0.80	52.51	6.6	100
C5Q	4.48±0.84	4.71±0.74	54.22	9.7	100	C5K	15.67±0.71[e]	16.98±2.10[e]	50.74	10.2	100
C5H	7.90±0.57	0.94±0.33	51.51	9.7	100	C5R	8.07±0.84	8.92±1.17	53.59	11.2	100

Table 2.3.1. Solubility of the poly-amino-acid tagged BPTI variants.

[a]Solubility was measured in the presence of 1.3M ammonium sulfate in 50mM acetate (pH4.7) and 50mM Tris-HCl (pH8.7), and with an initial protein concentration of 20mg/ml. All measurements were replicated more than three tim es.

[b]Thermal stabilities were determined in 20mM Acetate Buffer of pH4.7 by circular dichroism (CD) in the presence of 1.0M ammonium sulfate. *T*m stands for melting temperature.

[c]pI (iso-electric Point) was calculated using PROTEIN CALCULATOR v3.3 (http://www.scripps.edu/~cdputnam/protcalc.html).

[d]Trypsin inhibitory activity was measured by monitoring the hydrolysis of N-benzoyl-D,L-arginine-p-nitroanilide (BAPA) at equimolar concentrations of BPTIs and trypsin (280 nM). BPTI variants and trypsin were incubated in buffer for 5 minutes before addition of BAPA, and the hydrolysis of BAPA was monitored by measuring absorbance changes over 5 minutes at 405nm.

[e]Equilibrium might not have been reached, and this value might increase when measured with initial protein concentrations higher than 20mg/ml.

Chapter – 3

Analysis of protein aggregation kinetics using short amino acid peptide tags

3.1. Introduction

3.1.1. Introduction

Protein aggregation is attracting much attention in physico-chemical studies [4]. Amyloidogenic protein aggregation, which is associated with several neurodegenerative diseases, is the focus of intensive research [6-8]. On the other hand, non-amyloidogenic or amorphous protein aggregation is much less characterized, but it is a major concern in biotechnology and pharmaceutical industries, especially in the production and storage of therapeutic proteins [8]. This is because the aggregation tendency (or low solubility) of therapeutic proteins, not only reduces their production and therapeutic efficacy, but may also increase the risk of immunogenetic reactions [15].

Though several aspects of amorphous and amyloidogenic protein aggregation might be related, few studies have specifically addressed the mechanisms of amorphous aggregation, which are barely understood. Amorphous protein aggregation can be either reversible or virtually irreversible if aggregates are formed by partially or fully unfolded or chemically altered proteins [44]. For example, insulin forms reversible aggregates [43], whereas, the monoclonal antibody (IgG2) aggregates irreversibly through chemical modification [68]. In addition, a protein might aggregate simultaneously through different mechanisms, as observed for an interleukin-1 receptor antagonist [77].

Except for general trends, the physico-chemical aspects of amorphous protein aggregation are not well understood [78-79]. Generally, hydrophobic proteins are aggregation prone [34], whereas, proteins containing many charged residues on their surfaces are highly soluble [20, 61]. Some studies also reported that large proteins are likely to be more

aggregation prone than small ones [34], but this observation might simply reflect their tendency to partially unfold. Additionally, aggregation is generally thought to occur through a nucleation extension model similar to the helical polymerization model used to describe F-actin polymerization rather than a simple linear extension model [80]. However, the difficulty to isolate nuclei of amorphous aggregates and distinguish them un-ambiguously from small aggregates [45] has hampered the biophysical analysis of amorphous aggregation.

Mutational analyses are expected to help rationalizing the impact of amino acid's physico-chemical properties on protein aggregation tendency [81-82]. In one such example, the solubility of Ribonuclease Sa was examined by systematically mutating Thr76, which is located on its molecular surface, to all of the 20 natural amino acids [54]. In another example, Ankyrin repeat protein's solubility was analyzed by substituting solvent exposed hydrophobic Leu with positively charged Arg, eventually yielding highly soluble variants [39]. However, insights from mutational analysis remain relatively limited, as the effect of the local physico-chemical environment on solubility are difficult to disentangle from genuine amino acid solubility properties [83].

Aggregation kinetics is a determining factor for processing proteins in the biotechnological and pharmaceutical industries [9], but it is often overlooked in biochemical and biophysical studies, partly because aggregation can be very slow to occur. Although some studies have stressed its importance in rationalizing protein solubility and aggregation [84], few mutational analyses of aggregation kinetics have been reported [17]. In another example, the addition of an acidic tail of synuclein (ATS) fused to hGH, G-CSF and leptin

derivatives significantly reduced the aggregation rates of the host protein [18]. However, a systematic approach for rationalizing protein aggregation kinetics remains to be developed.

Recently, we used a simplified BPTI variant [56] as a model protein and reported short poly-amino acid peptide tags attached to its termini that can manipulate protein solubility in an essentially context-independent manner [49], and we used them to measure the contribution of individual amino acids on protein solubility for ten different types of amino acids [85]. In this work, we expanded this strategy to analyze how the amino acid types contribute to the kinetics of protein aggregation. The major finding of the present study is that, for each amino acid, at least two solubility parameters (which we coined the *AIC* and the *LS*) and presumably one or more aggregation rates (that we did not determine here) would be necessary for rationalizing protein solubility from its amino acid component as a function of time.

3.1.2. Motivation

Highly soluble fusion tags are widely being used to reduce aggregation tendency of target proteins [41], though larger fusion partners have inherent tendency to perturb the structure and function of the host proteins [40-41] (Table-1.5). In first part of our study, we reported that short amino acid peptide tags could influence protein solubility without affecting native structure, stability and functional properties [49, 85]. Contextually-independent existence of short poly-amino-acid peptide tags could be an effective tool to analyze the contribution of individual amino acids on aggregation propensity.

3.1.3. Aims and objectives

The aim of the second part of this research is to understand the aggregation kinetics and to gain insight into how individual amino acids contribute to protein aggregation tendency. In this work we describe the influence of equilibrium time and initial protein concentration on the kinetics of aggregation of simple monomeric globular protein. We observed that the protein solubility and aggregation are subject to both the initial protein concentration and equilibrium time. We report that the contribution of peptide tags to the aggregation propensity rate of model protein is amino acid type dependent. We believe that these results will provide insight into the understanding of protein aggregation kinetics and to the best of our knowledge, this is the first study to focus on the contribution of different types of amino acids on protein aggregation propensity.

3.2. Methods and materials

3.2.1. Protein aggregation measurement

Protein aggregation was determined as a function of the reduction of protein in supernatant with increase of incubation time. Like solubility measurement, protein samples with different initial protein concentration were incubated with 1.3M ammonium sulfate in 50mM buffers at pH 4.7 and pH 7.7 at 25ºC [85]. For lysozyme ammonium sulfate concentration was 1.8M in same buffers. Protein samples were mixed thoroughly and allowed to equilibrate for different equilibrium times (20 min, 6hrs, 12hrs, 24hrs, 48hrs (for BPTI mutants) and further over a duration of 7days (for lysozyme)). After equilibration, the samples were centrifuged at 20000xg for 20 minutes at 25ºC and the protein concentration of the supernatants was determined by UV analysis using Thermo Scientific NANO DROP spectrophotometer 2000 (Figure 2.2.5). Reliability and reproducibility of the final solubility

data were ensured by taking average of at least three different experiments in different working days, and with different sets of reagents. As it has been mentioned earlier we reported pH of this chapter-3 (second paper) is pH recorded after experiments. For lower pH there was no change in pH value of buffer (4.7), whereas, at higher pH, experiments caused reduction of buffer pH from 8.7 to 7.7.

3.2.2. Activity of incubated mutants

The trypsin inhibition activity of the incubated BPTI variants was assayed by following similar protocol for fresh sample previously mentioned. After experiment, incubated sample were diluted with phosphate buffer and then mixed with trypsin at equimolar concentrations (280nM) and incubated for five minutes. BAPA (1 mg/ml in water with little DMSO added) was added to the sample, and the activities of the BPTI variants were measured by monitoring the changes in absorbance at 405 nm for five minutes. The same procedure was followed for determination of the background spectra in the absence of trypsin and BPTI, while the hydrolysis of BAPA by trypsin was measured in the absence of BPTI variants (Figure 3.2.1).

3.2.3. Determining structural integrity of incubated mutants

Structural integrity of incubated BPTI mutants were assessed by comparing analytical HPLC elution profile of incubated samples with the elution profile of the fresh sample. In short, after completion of solubility experiments, sample was diluted to desire concentration with demonized water and then supplemented with 10% acetic acid followed by injection into analytic reverse phase HPLC column at 25% acetonitrile flow in a total flow of 1ml/min. After completion of the acetic acid peaks, a time-gradient program (25%~60% acetonitrile, at 1%/min) was run (Solution A: Water with 0.1% TFA: solution

B: acetonitrile with 0.05% TFA). Similarly, analytic HLPC was run for fresh sample with same gradient program. Finally elution peak of incubated sample was compared with the elution peak of fresh sample (Figure 3.2.2).

3.3. Results, Discussion and Conclusions

3.3.1. Solubility of BPTI variants

At first, we measured the solubility of BPTI variants as the amount of proteins in the supernatant of a supersaturated protein solution, as previously described [85]. We started with a total protein concentration of 20.0 mg/ml and an equilibration time of 20 minutes. Depending on the types of tagged amino acids the solubility values ranged from 0.2 mg/ml to 19.9 mg/ml (Figure 3.3.1). The solubilization effect of Lys was very high, and the saturation level was not reached even at a total protein concentration of 20.0 mg/ml (Figure 3.3.2). These amino acid type-dependent solubilization effects were in line with our previously published data [85].

Besides amino acid type, the solubility of BPTI variants depended on the equilibration period. For example, the amount of protein in the supernatant of the His tagged variant, C5H, reduced sharply from 8.4 mg/ml to 2.1 mg/ml with the extension of equilibration time from 20 minutes to 48 hours (Figure 3.3.3), which reflected the slow aggregation kinetics. In our previous study, we also noticed some time dependence in the solubility of the BPTI variants [85], but did not document this preliminary observation as our measurements were performed with a 20 minute equilibration time. In this study, we exhaustively investigated the kinetics of aggregations over a wide range of total protein concentrations and varying periods of equilibration time, and the results are explained in the following sections.

3.3.2. Aggregation kinetics dependence on total protein concentration

In addition to the slow kinetics, we observed that the kinetics depended on the total protein concentration. The protein concentration in the supernatant or "apparent solubility" reduced faster for the solutions with a high total protein concentration (Figure 3.3.4 and Figure 3.3.5). This concentration dependent aggregation kinetic explains a peculiar observation that for several variants, the apparent solubility, measured as the protein concentration in the supernatant after 20 minutes equilibration, was lower for solutions with higher total protein concentration. C5D, at pH 7.7, provides a clear example and aggregated faster with a total protein concentration of 20.0 mg/ml than 17.5 mg/ml (Figure 3.3.5B(h)).

3.3.3. Identification of an aggregation initiation concentration

The time dependence analysis of protein concentration in the supenatant indicated the existence of a critical protein concentration above which aggregation formation started, which we termed Aggregation Initiation Concentration (*AIC*) and the *AIC* ranged from 0.5 mg/ml to 16.6 mg/ml depending on the sample`s pH and the types of amino acids attached as tags (Table 3.3.1, Figure 3.3.4 - Figure 3.3.6). For example, the supernatant's concentration of C5N reduced from 9.4 mg/ml to 4.5 mg/ml by extending the equilibration time from 20 minutes to 48 hours when the total protein concentration was 12.5 mg/ml, but remained unchanged when total protein concentration was 6.0 mg/ml (Figure 3.3.4B(d)). The existence of an *AIC* and the aggregation kinetics are thus in line with a nucleation elongation mechanism, which implies that aggregation kinetics depend on both total protein concentration and the nuclei's size [86].

3.3.4. Transient solubility, long-term solubility, *AIC*, and protein "solubility"

Protein solubility has traditionally been measured by centrifuging samples after allowing them to settle for some equilibration time, which was considered to be adequate for amorphous aggregation to equilibrate [54, 85]. Our present study indicated that in some cases, the kinetics of protein aggregation can be much slower than previously thought. Thus, protein concentration in the supernatant of a saturated sample, measured after 20 minutes does not persist for a long period, and hereby we termed this solubility "Transient Solubility (*TS*)". For samples with a total protein concentration higher than the *AIC*, protein's concentration in the supernatant measured after 48 hours was lower than or equal to the *AIC*, and we termed it the "Long-term Solubility" (*LS*) (Figure 3.3.4 – Figure 3.3.5). The noticeable exception was C5P, which had an *LS* larger than the *AIC* (Figure 3.3.4B(g) and Figure 3.3.5B(g)) possibly reflecting either an extremely slow aggregation kinetics or might somehow be related to the high solubility of Pro as an individual amino acid [77]. We believe that *LS* is close to the final equilibrium solubility and is an important parameter, although some variants might not have fully reached equilibrium even after 48 hours. Similarly, *AIC* is also an important parameter because the protein concentration in the supernatant remains constant over the long term if the total protein concentration is lower than the *AIC*.

With respect to the nucleation-extension or helical polymerization model [86], the *AIC* would correspond to the minimum concentration at which aggregation nucleus can form, whereas *LS* (or actually the final equilibrium solubility) is the minimum concentration at which extension to larger aggregate occurs. In some aspects, both the *LS* and *AIC* could be considered as the "solubility" of a protein. Finally, we observed a high correlation between *LS* and *AIC* ($R = 0.874$) and *TS* and *AIC* ($R = 0.801$) (Figure 3.3.7 and Figure 3.3.8).

3.3.5. Effect of amino acid type to aggregation kinetics

The aggregation patterns of all the tagged BPTI variants were similar, but *TS*, *AIC*, and *LS* and protein aggregation speeds were strongly amino acid type-dependent (Table-1 and Figure 3.3.6). For example, the aggregation kinetics was fast for C5Q at either pH (Figure 3.3.4B(e) and 3.3.5B(e)), but slow for C5N at higher pH (Figure 3.3.5B(d)). This is because, the addition of amino acid tags affects the physico-chemical properties (total charge, hydrophobicity, pI, and flanking length) of model protein [85, 87] and hence its solubility and aggregation kinetics (Table- 3.3.1). Like for solubilization [85], we observed a strong pH dependence of the aggregation kinetics for Asn, His and Asp (Figure 3.3.4 and Figure 3.3.5). Interestingly, the relatively high solubility of C5R variant strongly reduced when the equilibration period was extended (Figure 3.3.4B(k) and Figure 3.3.5B(j)). On the other hand, C5K, the most soluble variant, remained soluble, even after an equilibration period of over 48 hours (Figure 3.3.4A(j) and Figure 3.3.5A(i)) [85]. It appears that for most tagged BPTI variants, protein aggregation progresses rapidly during the first 6 hours and then becomes much slower. This may be explained by the decreasing amount of proteins in the supernatant as aggregation propagates, which will slow down aggregation.

One necessary condition for the observation of an *AIC* is that the aggregation kinetics are slow, otherwise only the equilibrium value (*LS*) is observed. At a molecular level, slow aggregation kinetics would reflect a reaction that occurs only after a large number of configurations are sampled, and it would thus be observed for large molecules rather than small ones that have inherently simple intermolecular interactions.

3.3.6. Aggregation kinetics of natural proteins

To assess how the above observations relate to the aggregation kinetics of natural proteins, we carried out similar experiments with two natural proteins: Lysozyme and Albumin. The overall aggregation patterns were conserved, and we determined the *TS*, *AIC*, and *LS* for Lysozyme in 1.7 M ammonium sulfate by monitoring its aggregation for up to 7 days (Figure 3.3.9 and Figure 3.3.10). At pH 4.7 for a solution with total protein concentration of 15.0 mg/ml the supernatant protein concentration was 9.5 mg/ml after equilibrating for 20 minutes, which reduced to 6.9 mg/ml after 48 hours and finally to 5.8 mg/ml after 7 days of equilibration, and the *AIC* was 6.5 mg/ml (Figure 3.3.9). The aggregation kinetic pattern of Albumin was similar to that of C5K, and it remained soluble over an equilibration period up to 48 hours for solutions with total protein concentrations up to 25.0 mg/ml in the presence of 2 M ammonium sulfate (Figure 3.3.11).

Furthermore our observations with BPTI provides a rational for the an apparently mysterious observation that Lysozyme's solubility was higher when the total concentration was higher [86], which we also noticed in our previous study [85]. For example, C5H precipitated, after a 20-minute equilibration, when the total protein concentration was 5.0 mg/ml and indicated an "apparent solubility" of 4.1 mg/ml. However, when the total protein concentration was raised to 15.0 mg/ml, the protein concentration in the supernatant reached 10.5 mg/ml, which was higher than the above C5H's solubility (Figure 3.3.1; at an even higher total protein concentration of 20.0 mg/ml, aggregation accelerated and the supernatant reached 8.4 mg/ml, as discussed above).

3.3.7. Reversibility of aggregation

We did measure aggregation status of the protein sample both in supernatant and pellet. First we separate supernatant having total protein concentration higher than *AIC* and dissolve with buffer and found that supernatant protein went to solution like fresh protein and no scattering reaction formation was observed. On the other hand, a large portion but not all of the aggregation from the pellet dissolve in buffer though take longer time than the aggregation form supernatant. To get back aggregated protein from pellet in a dissolve form may require removal of ammonium sulfate or to lower down the ammonium sulfate concentration by adding sufficient amount of buffer. These findings as well as native like activity of incubated sample indicate the formation of amorphous aggregation which are reversible in nature Figure 3.3.13).

3.4. Conclusion

Although protein aggregation has been traditionally hypothesized to follow a helical polymerization model[16], there are few direct evidence, but our observation of an *AIC* and an *LS* provides a rare experimental corroboration of the model. One important message for long term protein conservation/preservation in solution is that samples should be prepared with protein concentration under the *AIC*. *AIC* and *LS* can both be considered as the solubility of a protein and they can be distinguished only if the kinetics of aggregations are slow. Noteworthy, *AIC* correlated strongly with *LS*. Experiments with lysozyme indicated similar aggregation patterns, and strongly suggested that mechanisms similar to those deduced using our model protein can give insight into the aggregation kinetics of natural proteins.

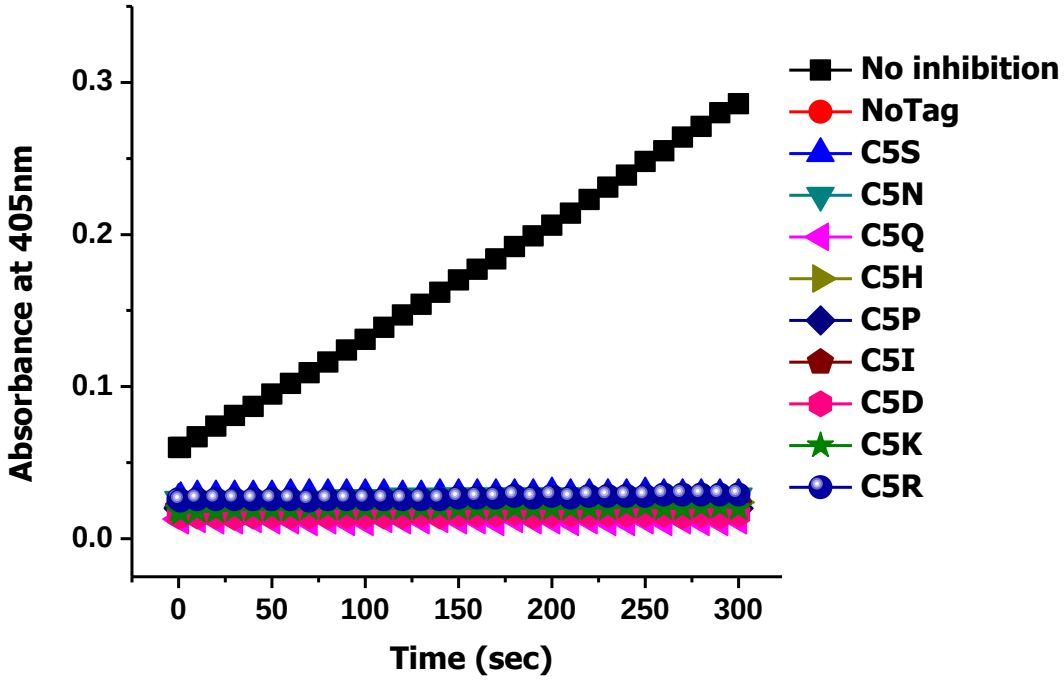


Figure 3.2.1. Trypsin inhibition activity of incubated BPTI variants. Black tetra angular represent the activity of trypsin on N^{α}-benzoyl-D-L-arginine *p*-nitroanilide (BAPA) in absence of BPTI, while color symbols represent the activity of trypsin on BAPA in presence of incubated BPTI-19 mutants at equimolar concentration (280nM) of Trypsin and BPTI variants.

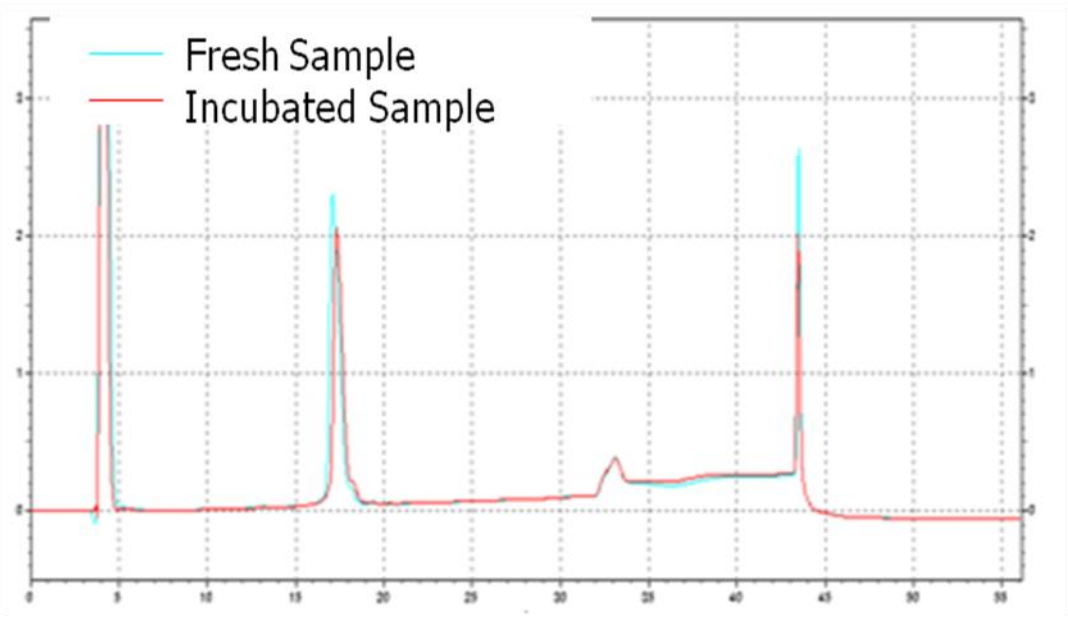


Figure 3.2.2. Structure integrity of incubated BPTI variants. Blue line represent the elution of fresh sample while red line for the incubated sample. Comparison is presented here with the elution of 50mg/ml of the mutant of C5N. Elution peak indicated that there is no degradation of BPTI mutants in presence of 1.3M ammonium sulfate.

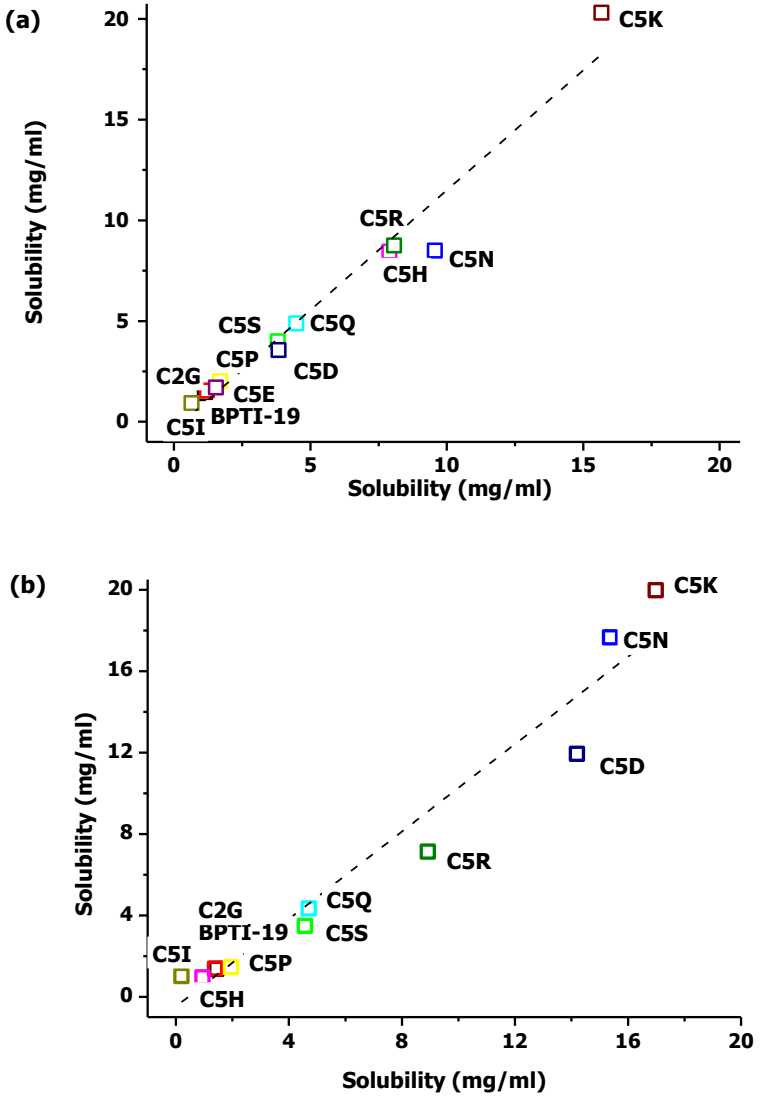


Figure 3.3.1. Comparison of solubility values with previously published values [85], X-axis represents the reference value, whereas, Y-axis represents values in the present study after equilibration period of 20 min, (a) at pH4.7 and (b) pH 7.7.

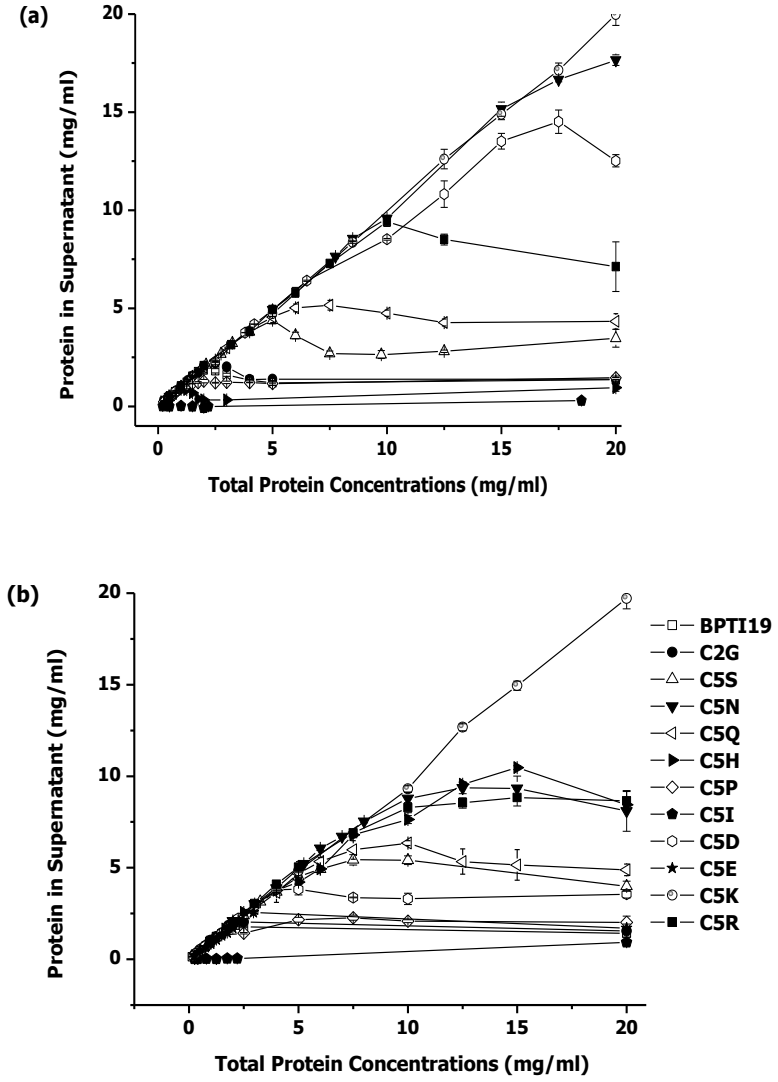


Figure 3.3.2. Solubility values with different total protein concentration after equilibration period of 20 minutes after equilibration period of 20 min, (a) at pH4.7 and (b) pH 7.7.

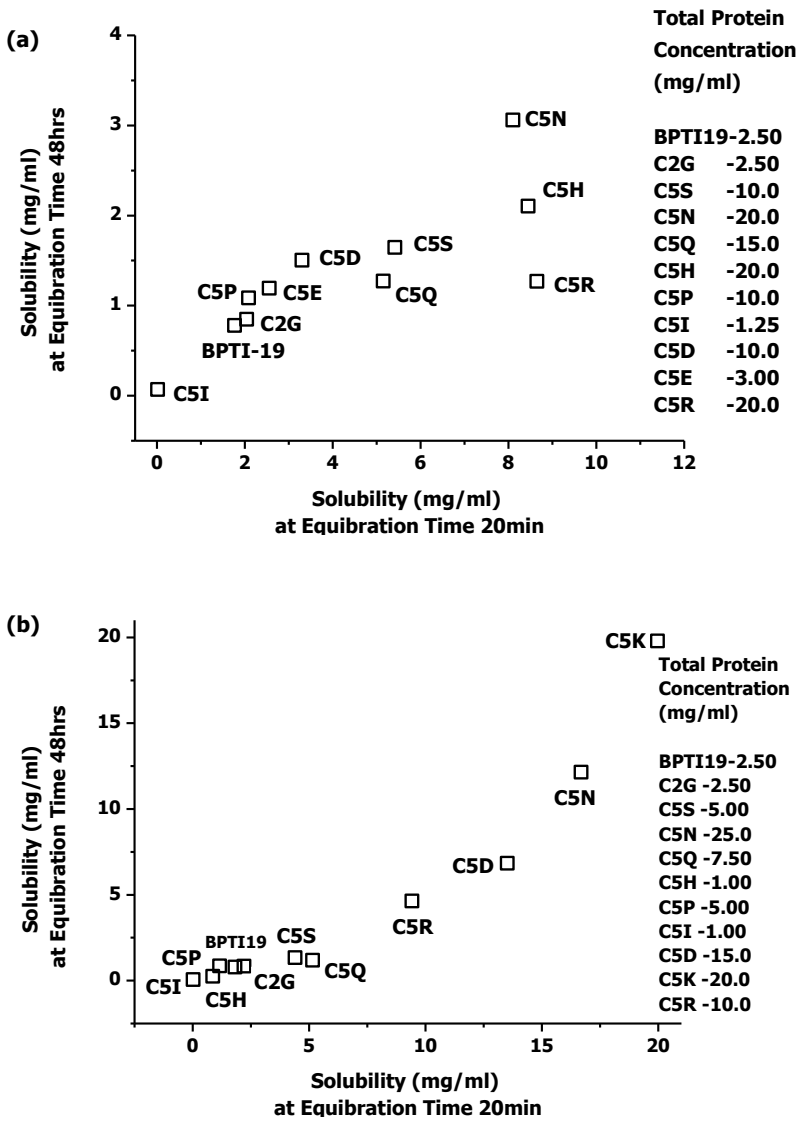


Figure 3.3.3. Solubility values for different equilibration time 20 min vs. 48 hrs, (a) at pH 4.7 (b) at pH 7.7 Total protein concentrations are indicated on the right side of the figure.

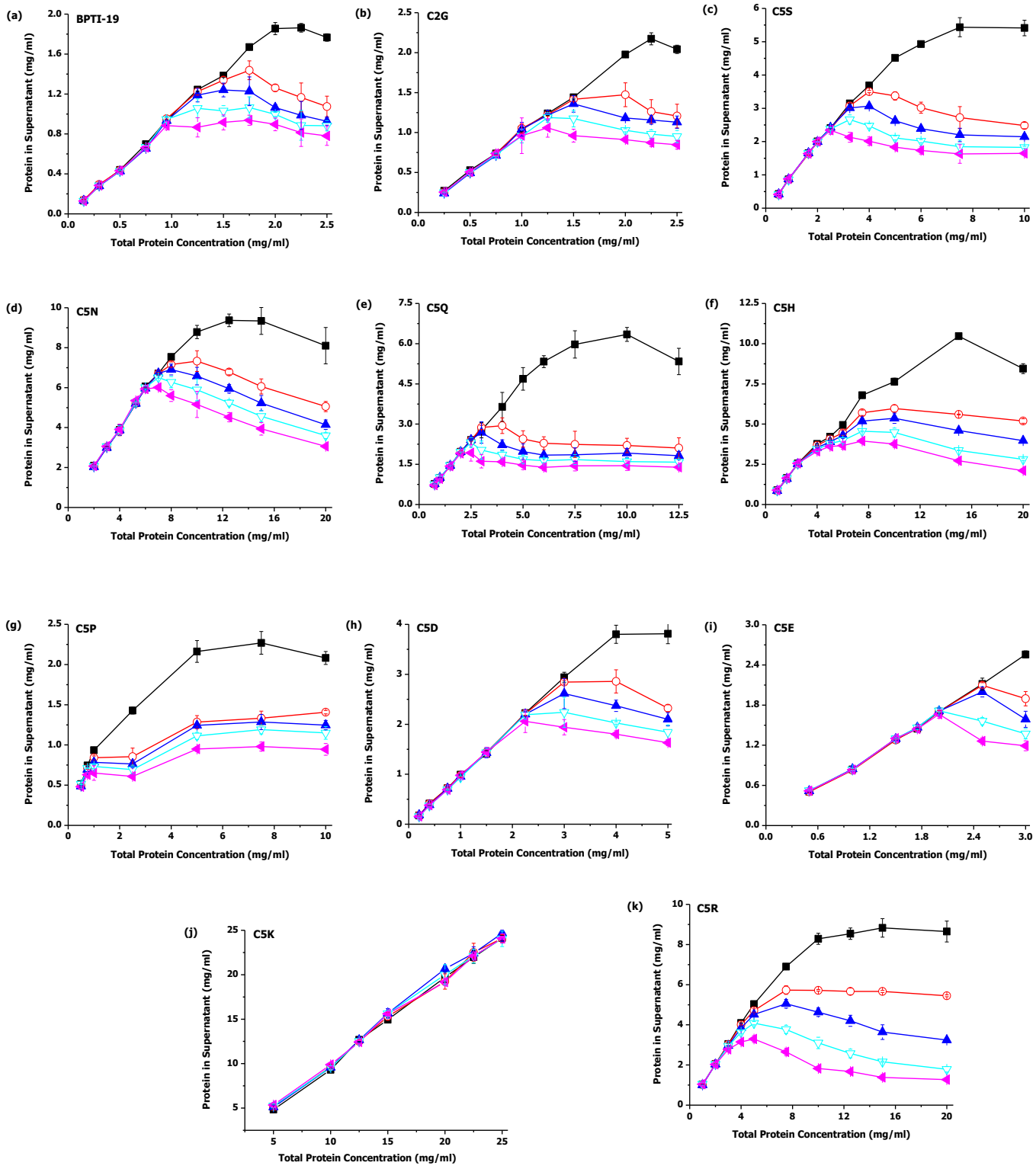


Figure 3.3.4A. Aggregation kinetics of poly-amino-acid tagged BPTI variants at pH 4.7. 3.3.3A. Presented as a function of equilibration time (a) BPTI-19, (b) C2G, (c) C5S, (d) C5N, (e) C5Q, (f) C5H, (g) C5P, (h) C5D, (i) C5E, (j) C5K, and (k) C5R and 20 min (■), 6 hrs (○), 12 hrs (▲), 24 hrs (▽), and 48 hrs (◀).

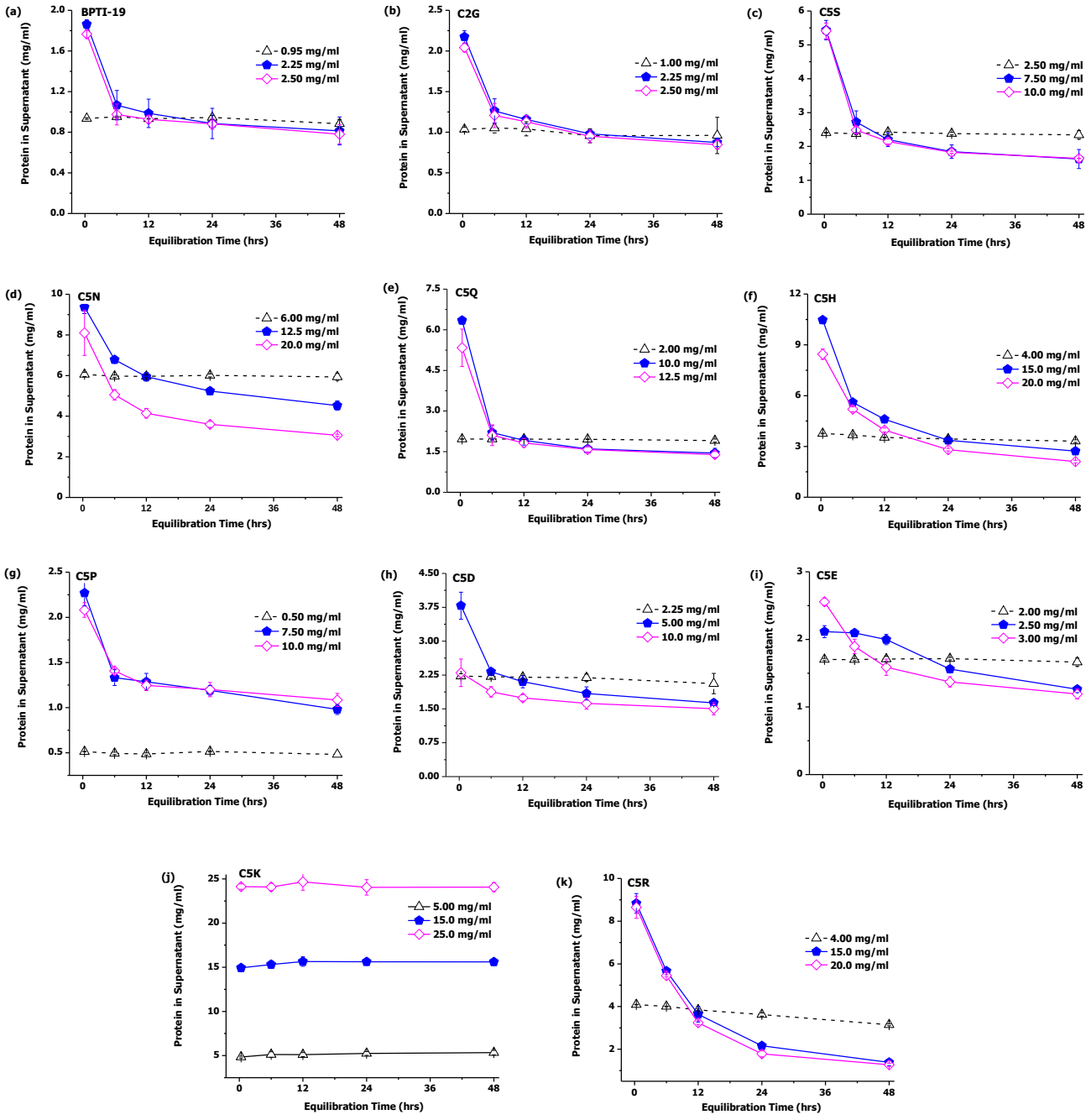


Figure 3.3.4B. Aggregation kinetics of poly-amino-acid tagged BPTI variants at pH 4.7. Presented as a function of total protein concentration. Values after 20 minutes equilibration are indicated as 0 (zero) hour. Values corresponding to the *AIC* are shown by dotted lines and total protein concentrations are indicated on the right side of the corresponding figure.

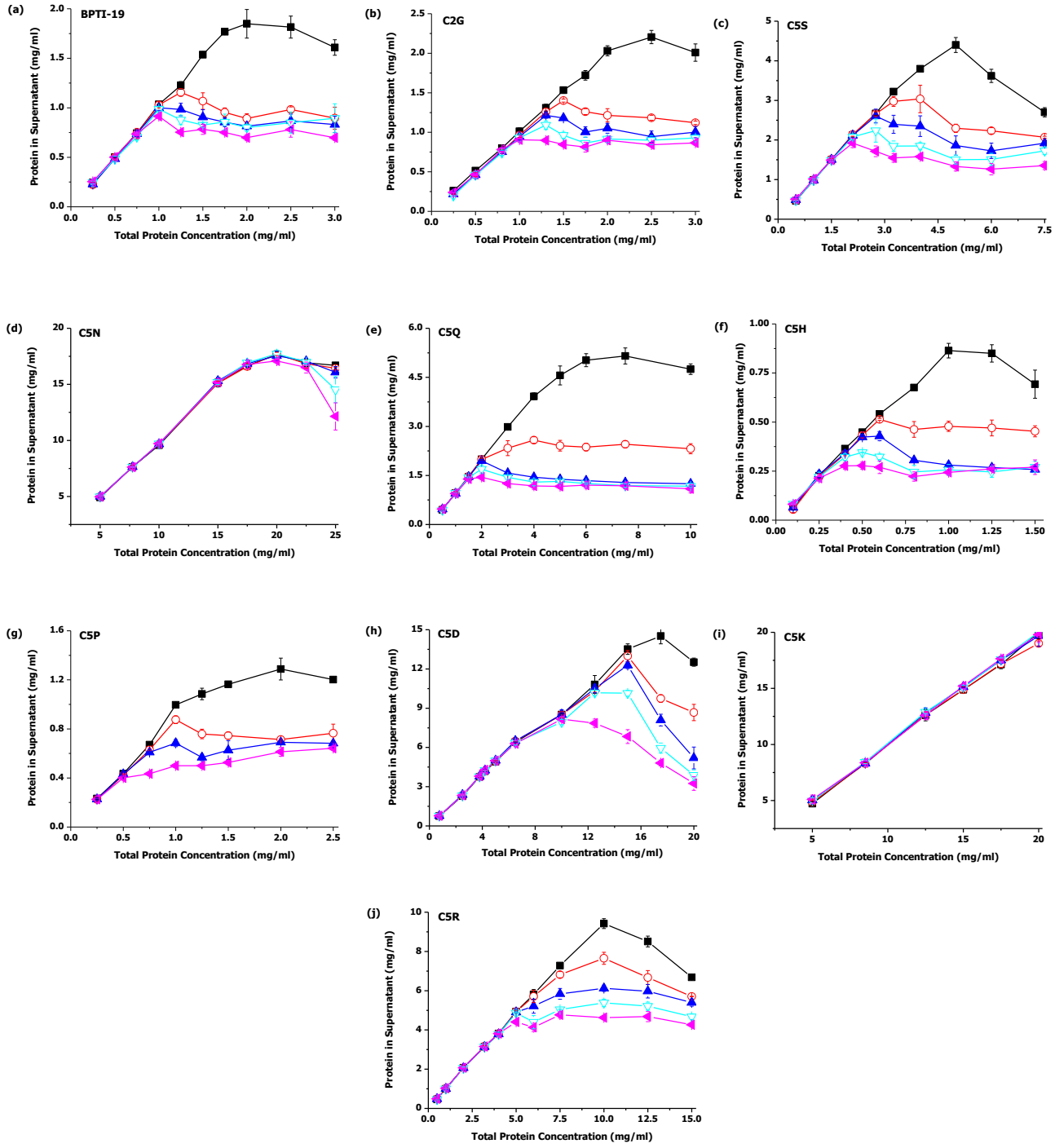


Figure 3.3.5A. Aggregation kinetics of poly-amino-acid tagged BPTI variants at pH 7.7. 3.3.3A. Presented as a function of equilibration time (a) BPTI-19, (b) C2G, (c) C5S, (d) C5N, (e) C5Q, (f) C5H, (g) C5P, (h) C5D, (i) C5E, (j) C5K, and (k) C5R and 20 min (■), 6 hrs (○), 12 hrs (▲), 24 hrs (▽), and 48 hrs (◀).

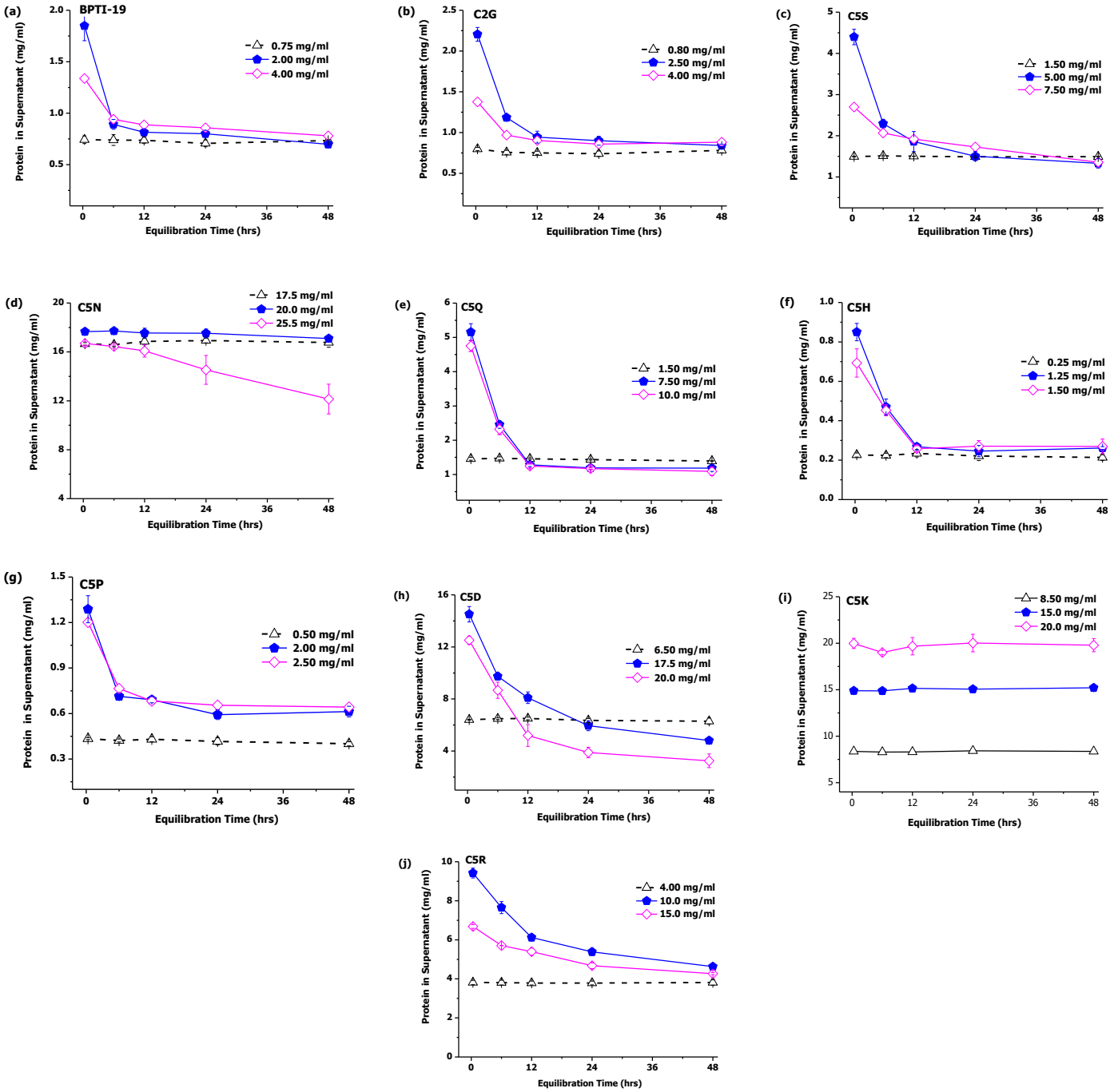
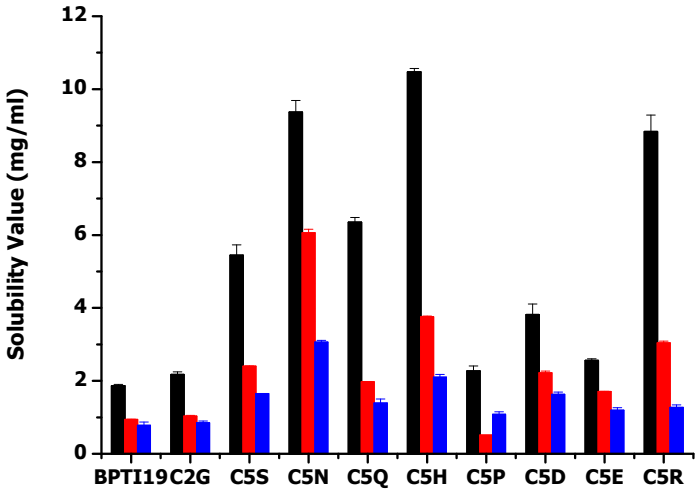


Figure 3.3.5B. Aggregation kinetics of poly-amino-acid tagged BPTI variants at pH 7.7. Presented as a function of total protein concentration. Values after 20 minutes equilibration are indicated as 0 (zero) hour. Values corresponding to the *AIC* are shown by dotted lines and total protein concentrations are indicated on the right side of the corresponding figure.

(A)



(B)

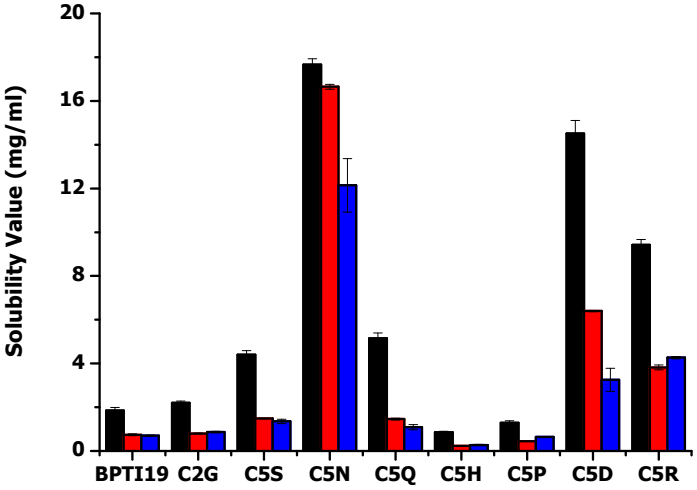


Figure 3.3.6. Different solubility values of poly-amino acid tagged variants, (A) at pH 4.7 and (B) at pH 7.7. *TS* *AIC* *LS*

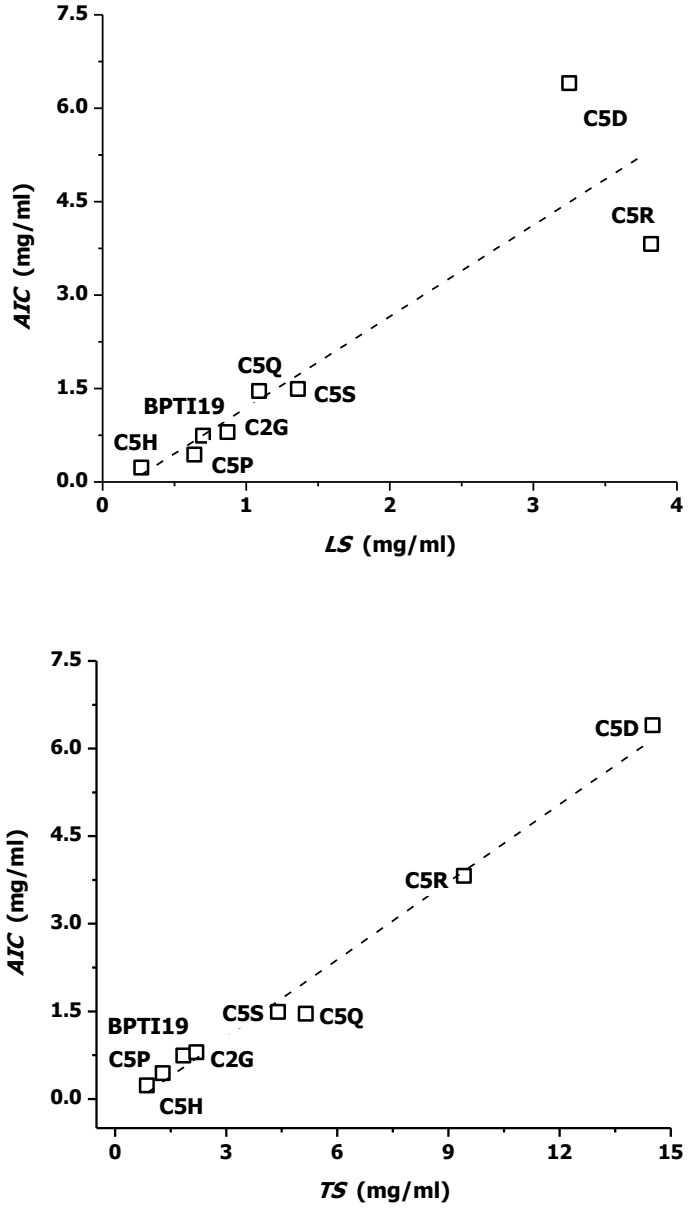
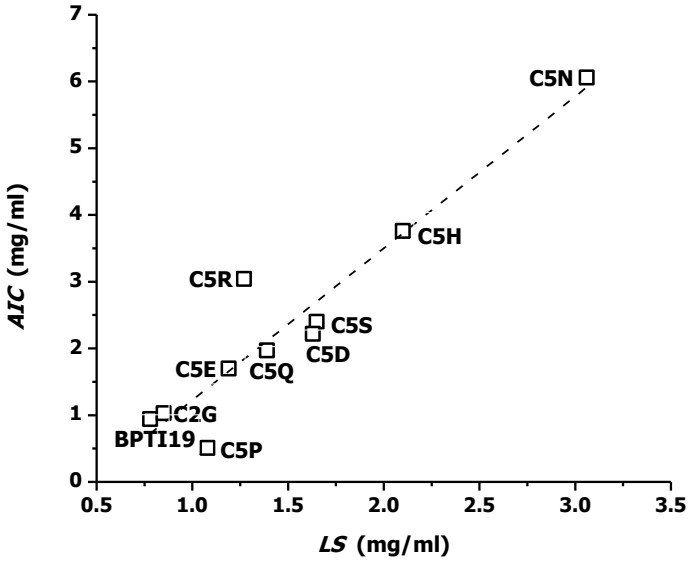


Figure 3.3.7. Correlation plot of solubility values at pH 7.7 (A) *LS* vs. *AIC* (R = 0.876), and (B) *TS* vs. *AIC* (R = 0.833).

(A)



(B)

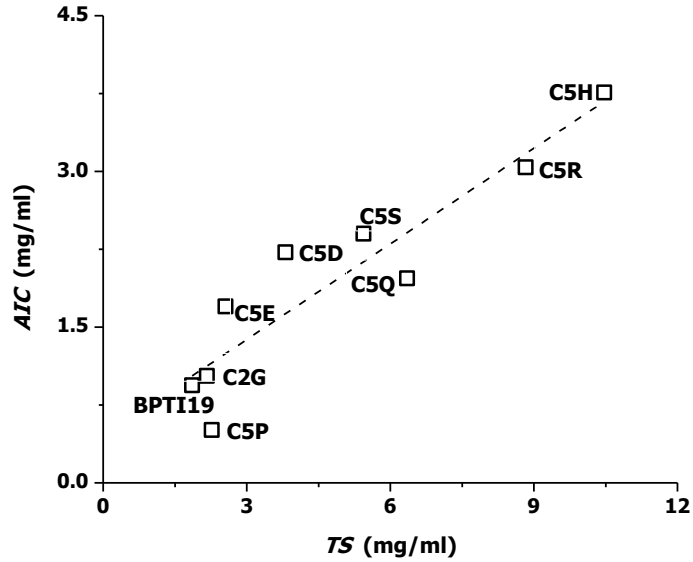
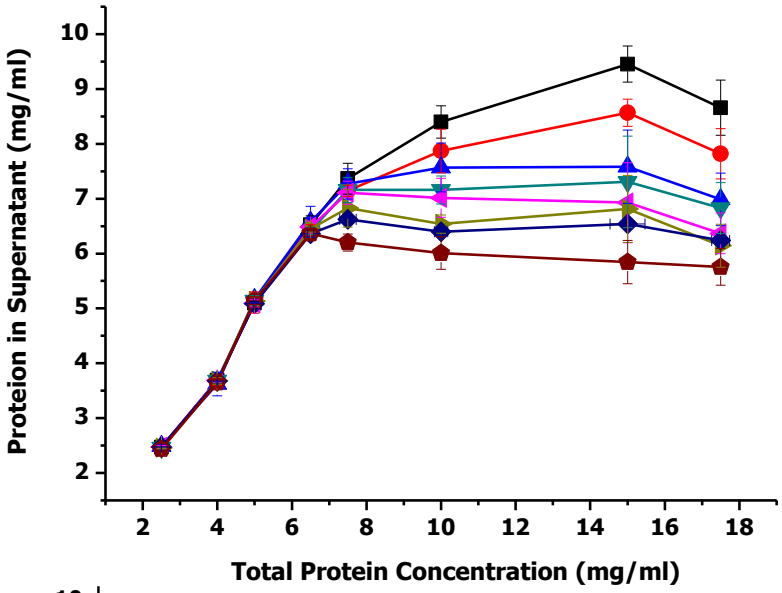


Figure 3.3.8. Correlation plot of solubility values at pH 4.7 (A) *LS* vs. *AIC* (*R* = 0.776), and (B) *TS* vs. *AIC* (*R* = 0.982).

(A)



(B)

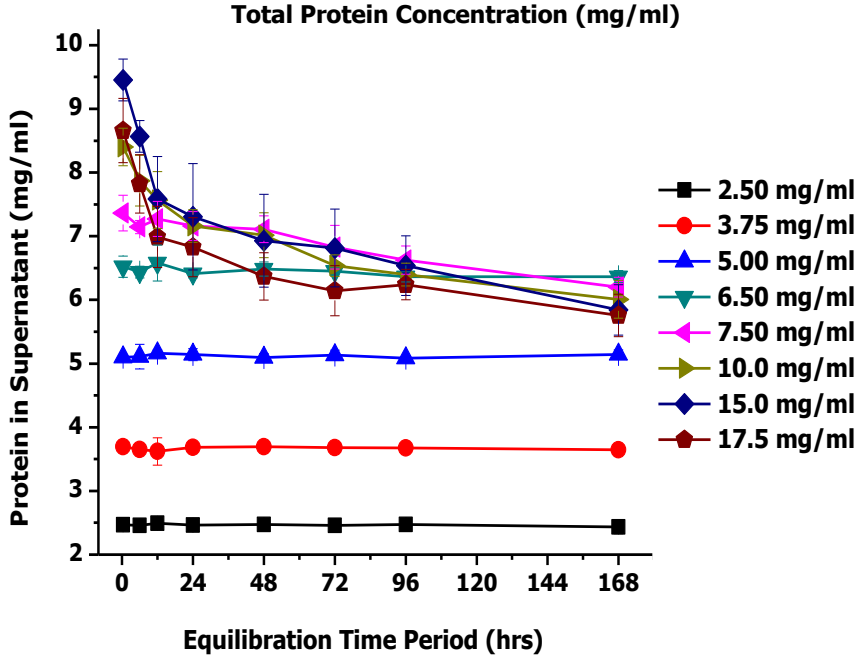
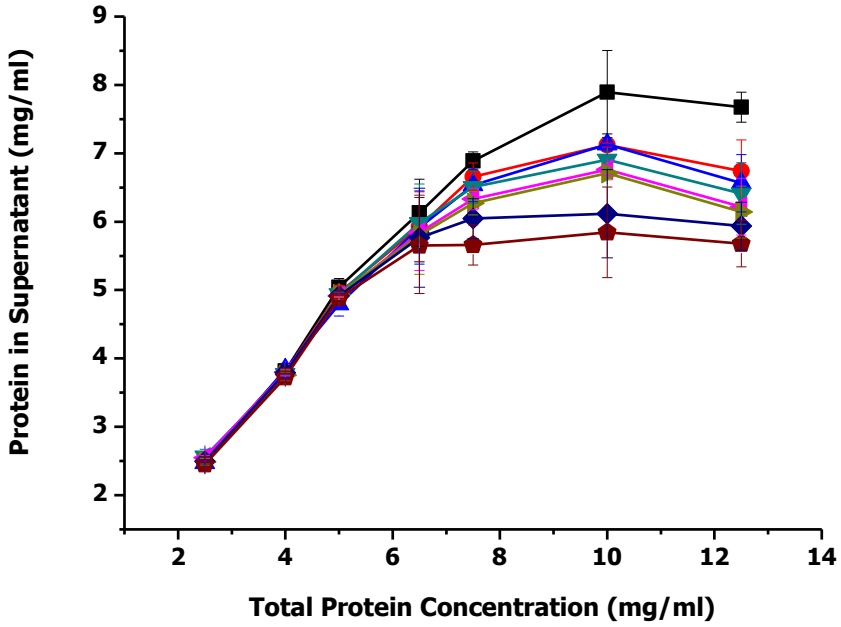


Figure 3.3.9. Aggregation kinetics of Lysozyme at pH 4.7 (A) presented as a function of equilibration time. Measurements were conducted in presence of 1.8M ammonium sulfate after different incubation periods, 20Min (■), 6Hrs (●), 12Hrs (▲), 24Hrs (▼), 48Hrs (◀), 72 Hrs (▶), 96Hrs (◆) and 7 Days (⬟) at 25^{o}C temperature. (B) presented as a function of total protein concentration. Concentration are given at the right side of the figure.

(A)



(B)

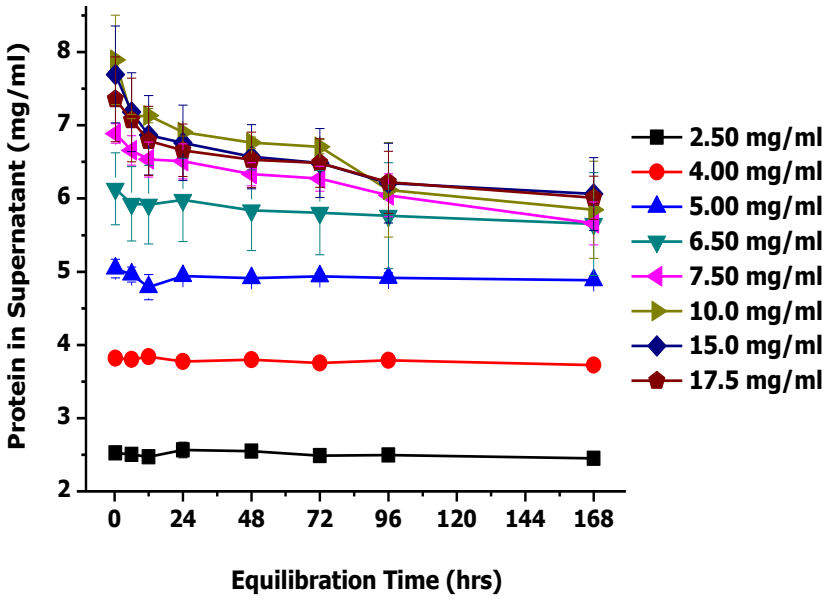


Figure 3.3.10. Aggregation kinetics of Lysozyme at pH 7.7 (A) presented as a function of equilibration time. Measurements were conducted in presence of 1.8M ammonium sulfate after different incubation periods, 20Min (■), 6Hrs (●), 12Hrs (▲), 24Hrs (▼), 48Hrs (◀), 72 Hrs (▶), 96Hrs (◆) and 7 Days (⬟) at 25^{o}C temperature. (B) presented as a function of total protein concentration. Concentration are given at the right side of the figure.

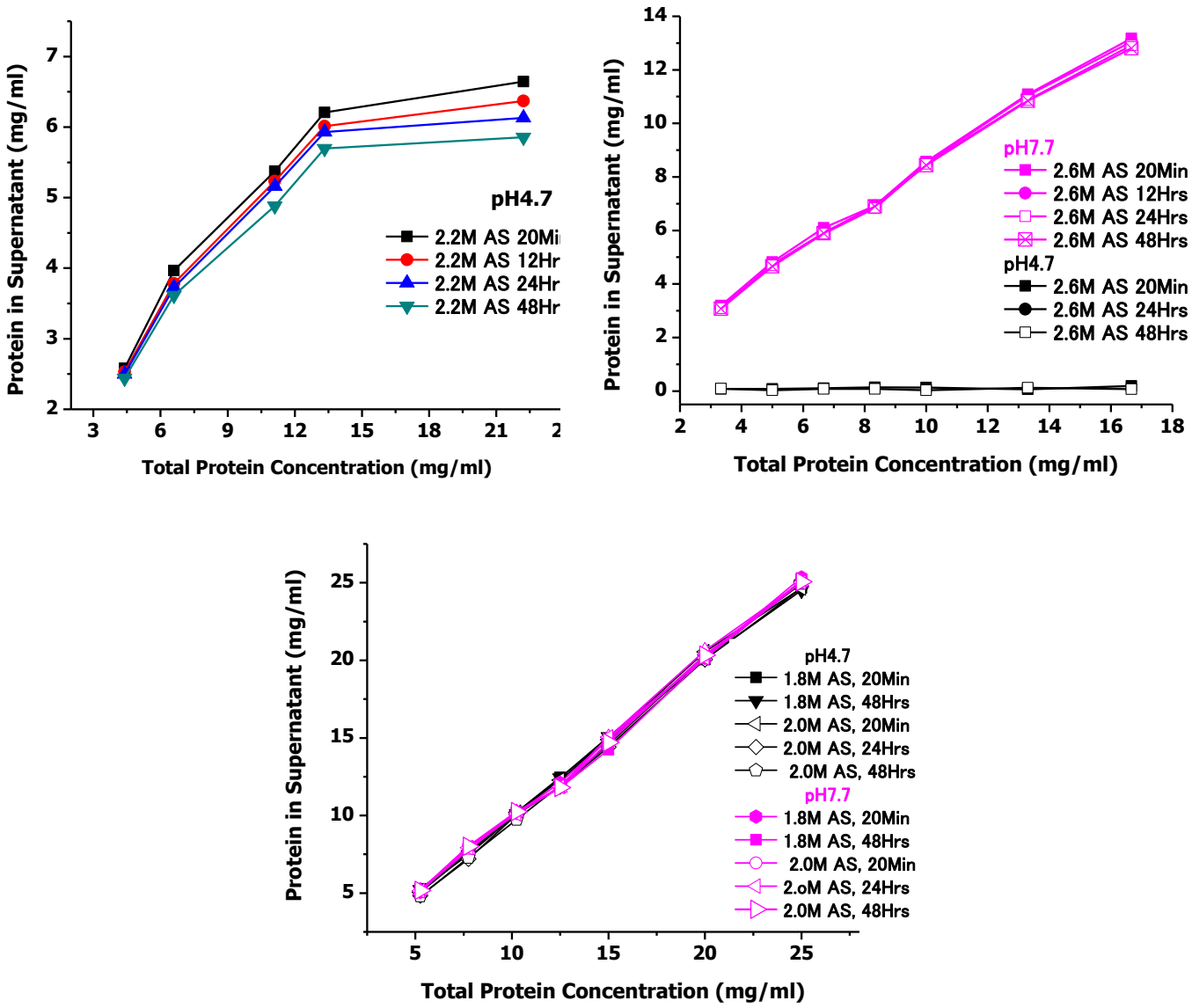
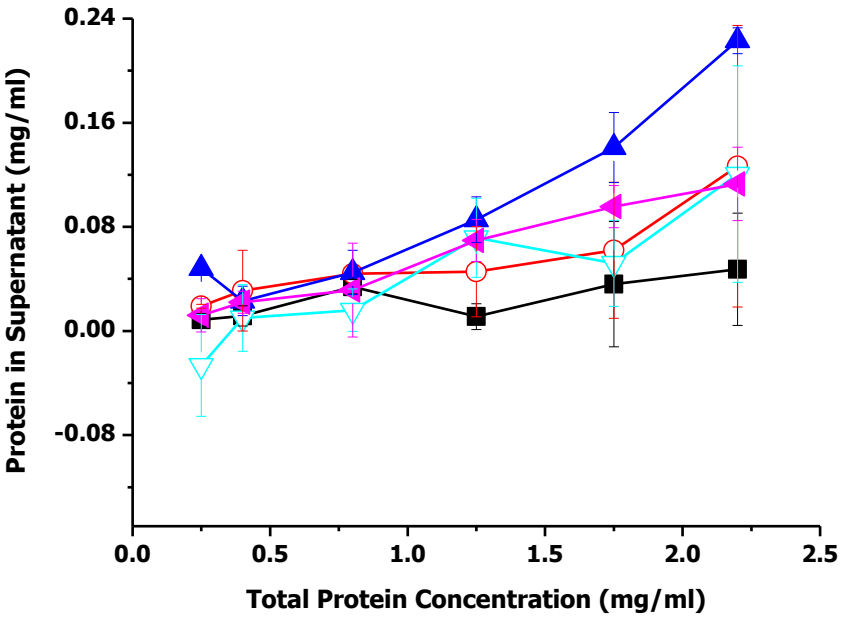


Figure 3.3.11. Aggregation kinetics of Albumin at different conditions. Ammonium Sulfate (AS) concentration, experimental pH and equilibration periods are mentioned at the right side of the figures.

(A)



(B)

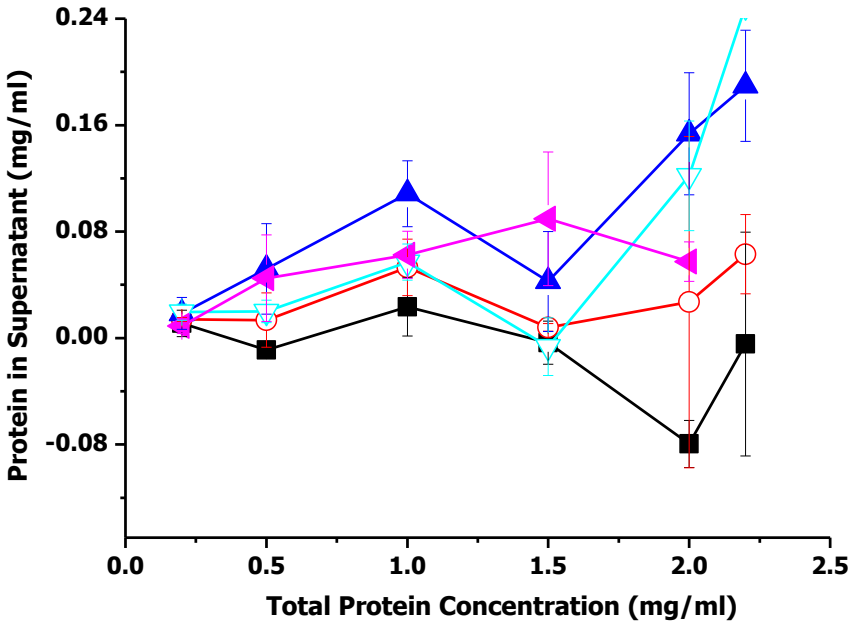
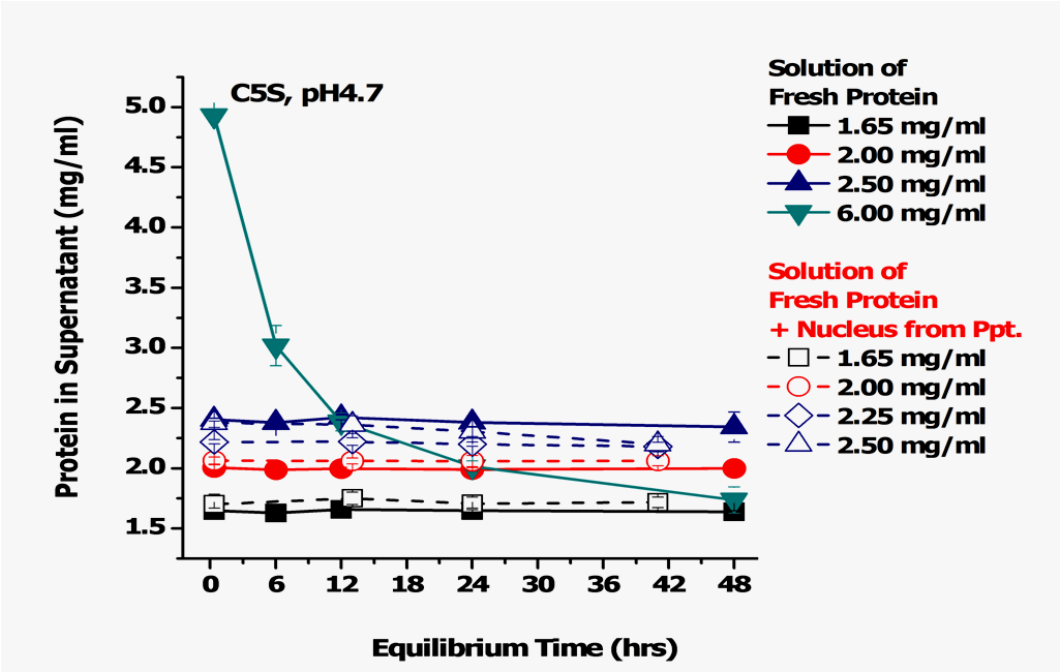


Figure 3.3.12. Aggregation kinetics of C5I (A) at pH 4.7 and (B) at pH 7.7. Supernatant protein amount was measured by deducting the amount of protein in the pellet from the corresponding initial total amount of protein.

(A)



(B)

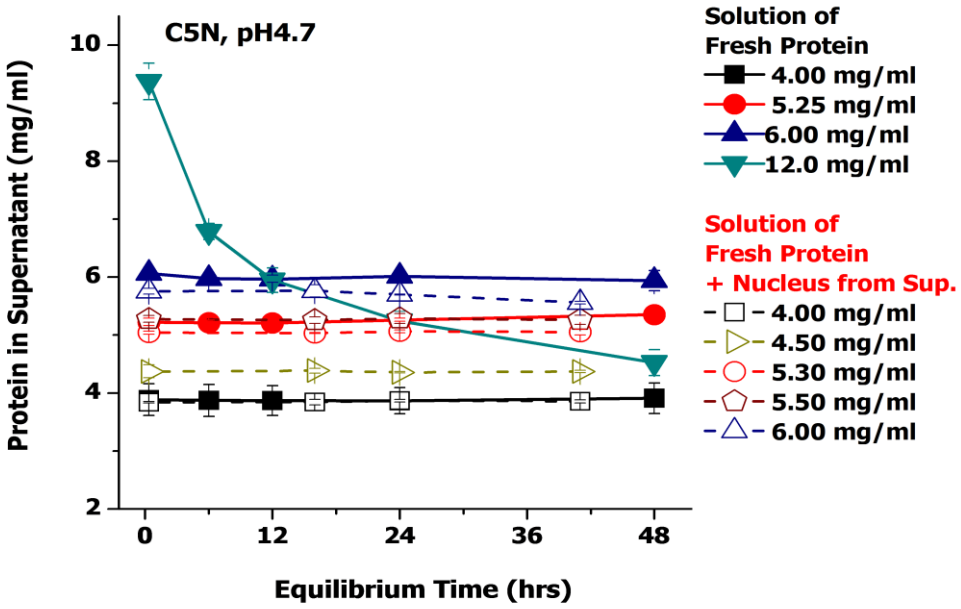


Figure 3.3.13. Reversibility of aggregated mutants at pH 4.7 (A) C5S, and (B) C5N.

Mutant ID	Hydro pathicity [a]	pH 4.7				pH 7.7			
		Total Charge[b]	***TS*** [c] (mg/ml)	***AIC*** [c] (mg/ml)	***LS*** [c] (mg/ml)	**Total Charge**[b]	***TS*** [c] (mg/ml)	***AIC*** [c] (mg/ml)	***LS*** [c] (mg/ml)
BPTI-19	0.167	5.7	1.86±0.04	0.94±0.01	0.78±0.09	4.4	1.85±0.14	0.74±0.04	0.70±0.03
C2G	0.148	5.7	2.17±0.08	1.03±0.02	0.85±0.05	4.4	2.20±0.08	0.80±0.02	0.87±0.03
C5S	0.075	5.7	5.44±0.29	2.40±0.01	1.65±0.00	4.4	4.40±0.19	1.49±0.01	1.36±0.10
C5N	- 0.132	5.7	9.37±0.32	6.06±0.10	3.06±0.05	4.4	17.66±0.2	16.65±0.11	12.14±1.22
C5Q	- 0.132	5.7	6.35±0.13	1.97±0.01	1.39±0.11	4.4	5.15±0.25	1.46±0.04	1.09±0.12
C5H	- 0.109	10.6	10.47±0.10	3.76±0.02	2.10±0.08	4.7	0.86±0.04	0.23±0.01	0.27±0.03
C5P	0.014	5.7	2.27±0.14	0.51±0.00	1.08±0.07	4.4	1.29±0.09	0.44±0.01	0.64±0.01
C5I [d]	0.483	5.7	0.0	0.0	0.0	4.4	0.0	0.0	0.0
C5D	-0.132	2.4	3.81±0.30	2.22±0.05	1.63±0.06	-0.6	14.51±0.59	6.40±0.03	3.25±0.53
C5E	-0.132	2.4	2.56±.05	1.70±0.01	1.19±0.07	-0.6	ND [e]	ND [e]	ND [e]
C5K [f]	- 0.163	10.7	>25.00	>25.00	>25.00	9.3	>25.00	>25.00	>25.00
C5R	- 0.209	10.7	8.83±0.46	3.04±0.05	1.27±0.07	9.4	9.42±0.25	3.82±0.11	3.82±0.05

Table 3.3.1 Physiochemical characteristic and solubility values of poly-amino acid tagged BPTI variants.

[a] Hydropathicity calculated using ProtParam tool (http://web.expasy.org/cgi-bin/protparam/protparam

[b] Total charge calculated using PROTEIN CALCULATOR v3.3 http://www.scripps.edu/cgi-bin/cdputnam/protcalc3

[c] Values measured in the presence of 1.3 M ammonium sulfate in 50 mM acetate (pH4.7) and 50mM Tris-HCl (pH7.7). All values are presented as average over more than three experiments. Error is the standard deviation.

[d] The concentration of C5I was determined by measuring the amount of protein that precipitated, because the concentration in the supernatant was too low for direct measurement and the values were very close to zero (Figure 3.3.12).

[e] ND: Not Determined; and [f] C5K remained soluble even after 48 hours at a total protein concentration of 25.0 mg/ml.

Chapter – 4

Conclusions

This dissertation has mainly focused on impact of short amino acid tag on host protein solubility and protein aggregation kinetics. We first characterize the effects of short peptide tags of ten different types of amino acids representing all different physiochemical properties of amino acids on protein. We observed that short peptides had no or very little effect on the stability, structure and function of proteins by DSC, X-ray crystallography and activity measurement. Poly-amino-acid tags remained very flexible, and structurally independent from the protein, enabling controlling protein solubility in a context-independent manner. The contribution of poly peptide tags on host protein solubility was amino acid type dependent. Tag with positively charged amino acid poly-Lys increased protein solubility largely. Poly-Agr also increased the solubility of host but modestly. In contrast, the negatively charged poly-Asp and poly-Glu tags barely affected protein solubility at pH 4.7 but increased at high pH which is attributable to the tendency of proteins to aggregate at pHs close to their isoelectric points. As expected, the addition of tag of hydrophobic amino acid poly- Ile significantly reduced the solubility of BPTI. These observations were generally consistent with the solubility changes anticipated from their hydrophobicity value. For example, peptide tags composed of charged or hydrophilic residues increased the solubility of the protein, as anticipated.

Few, if any, systematic studies have been conducted into how each of the 20 natural amino acids modulates protein solubility, though solubility is an important property of proteins and largely depends on amino acid compositions. Strongly intertwined nature of the intrinsic and extrinsic factors that influence protein solubility, the technical difficulties associated with reliably measuring high protein concentrations, the wide range of solubility values among amino acids may have hampered the construction of a proper “solubility propensity scale”.

Current study indicates that short peptide tags fused to a host protein can be effective tool for measuring the effect of amino acids on protein solubility, in a context-independent manner. Such measurements could help to develop a genuine solubility propensity scale on how individual amino acids contribute to protein solubility, which could eventually allow the prediction of relative solubility of a polypeptide from its primary sequence.

Studies of protein aggregation kinetics demonstrated that depending on the total protein concentrations and equilibration time a protein solution is equilibrated over a wide range of solubility value. Solution having considerably high amount of protein it is possible to attain maximum amount of dissolved protein in supernatant only for short period. With extent of equilibrium time proteins stared to aggregate. However, there is a critical value of protein concentration under which no aggregation formation occur even for long incubation. This critical protein concentration represent the minimum amount of protein required for initiating aggregation and hereby termed as "aggregation initiation concentration (*AIC*)" Once protein solution reach the *AIC* aggregation formation rate is faster for the solution with higher initial protein concentration. Protein of high solubility propensity shows a tendency to have high aggregation initiation concentration value.

We also observed that, *TS* and *AIC* were varying depending on types of tagged amino acids. The aggregation formation rate which we did not measure here seems to be quickest for C5Q at either pH, whereas, slowest for C5N at higher pH. Slow aggregation formation rate by the positively charged Arg tagged variant represent the logic behind the versatile use of

Arg to increase protein solubility either as solute or as fusion tag [88]. Like amino acid contribution to protein solubility, how individual amino acids contribute to protein aggregation propensity is yet lacking. The type-dependent role of poly peptide tags to aggregation total concentration and the rate of aggregation is a new focus to gain insight into the contribution of individual amino acids to protein aggregation propensity and thereby to develop a propensity scale on contribution to protein aggregation tendency. Such a propensity scale will help a lot in predicting aggregation tendency in the field of protein science.

Chapter –5

References

1. Tjong, H., Zhou, H.-Z., *Prediction of protein solubility from calculation of transfer free energy,* Biophysical Journal, 2008, **95**: p. 2601-2609.

2 Smialowski, P., et al., *Protein solubility: sequence based prediction and experimental verification,* Bioinformatics 2007, **23(19),** p. 2536-2542.

3. Oss, C.J van, Good, R.J., Chaudhury, M.K. *Solubility of protein* J of Pro. Chem. 1986, **5(6)**, p. 385-405.

4. Fink, A.L., *Protein aggregation: folding aggregates, inclusion bodies and amyloid.* Fold Des, 1998, **3**(1): p. R9-23.

5. Sabate, R., N.S. de Groot, and S. Ventura, *Protein folding and aggregation in bacteria.* Cell Mol Life Sci, 2010, **67**(16): p. 2695-2715.

6. Ross, C.A. and M.A. Poirier, *Protein aggregation and neurodegenerative disease.* Nat Med, 2004, **10 Suppl**: p. S10-17.

7. Stefani, M. and C.M. Dobson, *Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution.* Jour. Mol. Med (Berl.), 2003, **81**(11): p. 678-699.

8. Wang, W., *Instability, stabilization, and formulation of liquid protein pharmaceuticals.* Int. Jour. Pharm, 1999, **185(2)**: p. 129-188.

9. Cromwell, M.E., E. Hilario, and F. Jacobson, *Protein aggregation and bioprocessing.* Aaps Journal, 2006, **8(3)**: p. E572-579.

10. Wang, W., S. Nema, and D. Teagarden, *Protein aggregation--pathways and influencing factors.* Int J Pharm, 2010, **390(2)**: p. 89-99.

11. Chi, E.Y., et al., *Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation.* Pharm Res, 2003, **20(9)**: p. 1325-1336.

12. Martinez-Alonso, M., E. Garcia-Fruitos, and A. Villaverde, *Yield, solubility and conformational quality of soluble proteins are not simultaneously favored in recombinant Escherichia coli.* Biotechnol Bioeng, 2008, **101(6)**: p. 1353-1358.

13. Cleland, J.L., M.F. Powell, and S.J. Shire, *The development of stable protein formulations: a close look at protein aggregation, deamidation, and oxidation.* Crit. Rev. Ther. Drug Carrier Syst., 1993, **10(4)**: p. 307-377.

14. Kessler, M., D. Goldsmith, and H. Schellekens, *Immunogenicity of biopharmaceuticals.* Nephrol Dial Transplant, 2006, **21 Suppl 5**: p. v9-12.

15. Rosenberg, A.S., *Effects of protein aggregates: an immunologic perspective.* Aaps Journal, 2006, **8(3)**: p. E501-507.

16. Braun, A., et al., *Protein aggregates seem to play a key role among the parameters influencing the antigenicity of interferon alpha (IFN-alpha) in normal and transgenic mice.* Pharm Res, 1997, **14(10)**: p. 1472-1478.

17. Fowler, S.B., et al., *Rational design of aggregation-resistant bioactive peptides: reengineering human calcitonin.* Proc Natl Acad Sci U S A, 2005, **102(29)**: p. 10105- 10110.

18. Lee, E.N., et al., *Stabilizing peptide fusion for solving the stability and solubility problems of therapeutic proteins.* Pharm Res, 2005, **22(10**): p. 1735-1746.

19. Golovanov, A.P., et al., *A simple method for improving protein solubility and long-term stability.* J Am Chem Soc, 2004, **126(29)**: p. 8933-8939.

20. Simeonov, P., et al., *Surface supercharged human enteropeptidase light chain shows improved solubility and refolding yield.* Protein Eng Des Sel, 2011. **24**(3): p. Suppl, 2010, **190**: p. 82-87.

21. Zhang, F., et al., *Crystal structure of the obese protein leptin-E100.* Nature, 1997, **387(6629)**: p. 206-9.261-268.

22. Zhou, P. and G. Wagner, *Overcoming the solubility limit with solubility-enhancement tags: successful applications in biomolecular NMR studies.* J Biomol NMR, 2010, **46(1)**: p. 23-31.

23. Aguzzi, A. and T. O'Connor, *Protein aggregation diseases: pathogenicity and therapeutic perspectives.* Nat Rev Drug Discov, 2010, **9(3)**: p. 237-248.

24. Forman, M.S., Trojanowski, J.Q. and Lee, V.M-Y., *Neurodegenerative diseases: a decade of discoveries paves the way for therapeutic breakthroughs* Nature Medicine, 2004, **10(10)**: p. 1055-1063.

25. Chen, S., F.A. Ferrone, and R. Wetzel, *Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation.* Proc Natl Acad Sci U S A, 2002, **99(18)**: p. 11884-11889.

26. Kim, W. and M.H. Hecht, *Generic hydrophobic residues are sufficient to promote aggregation of the Alzheimer's Abeta42 peptide.* Proc Natl Acad Sci U S A, 2006, **103(43)**: p. 15824-15829.

27. Gundersen, V., *Protein aggregation in Parkinson's disease.* Acta Neurol Scand Suppl, 2010, **190**: p. 82-87.

28. Schein, C.H., *Solubility as a Function of Protein Structure and Solvent Components.* Nature Biotech. 1990, **8(4)**: p. 308-315.

29. Inyang, U.E. and A.O. Idhu, *Influence of pH and salt concentration on protein solubility, emulsifying and foaming properties of sesame protein concentrate,* Jour. of the Ame. Oil Che. Soc., 1996, **73(12)**: p. 1663-1667

30. Li, Y., e al., *Multi-variate approach to global protein aggregation behavior and kinetics: Effect of pH, NaCl, and Temperature for ά–Chymotrypsinogen A,* Jour. of Pharm. Sci. 2010, **99(2)**: p. 645-662

31. Monahan, F.J., J.B. German, and J.E. Kinsella, *Effect of Ph and Temperature on Protein Unfolding and Thiol-Disulfide Interchange Reactions during Heat-Induced Gelation of Whey Proteins,* Jour. of Agri. and Food Chem., 1995, **43(1)**: p. 46-52.

32. Shaw, K.L., et al., *The effect of net charge on the solubility, activity, and stability of ribonuclease Sa.* Protein Sci, 2001, **10(6)**: p. 1206-1215.

33. Sousa, R.D.S., et al., *Effect of pH and salt concentration on the solubility and density of egg yolk and plasma egg yolk.* Lwt-Food Science and Technology, 2007, **40(7)**: p. 1253-1258.

34. Niwa, T., et al., *Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of Escherichia coli proteins.* Proc Natl Acad Sci U S A, 2009, **106(11)**: p. 4201-4206.

35 Dale, G.E., C. Oefner, and A. D'Arcy, *The protein as a variable in protein crystallization.* Journal of Structural Biology, 2003, **142(1)**: p. 88-97.

36. Trevino, S.R., J.M. Scholtz, and C.N. Pace, *Measuring and increasing protein solubility.* J Pharm Sci, 2008, **97(10)**: p. 4155-4166.

37. Bondos, S.E., and A. Bicknell, *Detection and prevention of protein aggregation before, during, and after purification.* 2002, **316:** p.223-231.

38. Malissard, M. and E.G. Berger, *Improving solubility of catalytic domain of human beta-1,4-galactosyltransferase-1 through rationally designed amino acid replacements.* Eur J Biochem, 2001, **268**(**15)**: p. 4352-4258.

39. Mosavi, L.K. and Z. Peng, *Structure-based substitution for increased solubility of a designed protein,* 2003, **16(10)**: p. 739-745.

40. Waugh, D.S., *Making the most of affinity tags.* Trends Biotechnol, 2005, **23(6)**: p. 316-320.

41. Terpe, K., *Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems.* Appl Microbiol Biotechnol, 2003, **60(5)**: p. 523-533.

42. Frieden., C., *Protein aggregation process: In search of the mechanism,* Pro. Sci., 2007, **16,**: p. 2334-2344

43. Pekar, A.H. and B.H. Frank, *Conformation of Proinsulin - Comparison of Insulin and Proinsulin Self-Association at Neutral pH.* Biochemistry, 1972, **11(22)**: p. 4013-4016

44. Philo, J.S. and T. Arakawa, *Mechanisms of protein aggregation.* Curr Pharm Biotechnol, 2009, **10(4)**: p. 348-351.

45. Bauer, R., et al., *Characterization and isolation of intermediates in beta-lactoglobulin heat aggregation at high pH.* Biophysical Journal, 2000, **79(2)**: p. 1030-1038.

46. Morris, A.M., M.A. Watzky, and R.G. Finke, *Protein aggregation kinetics, mechanism, and curve-fitting: a review of the literature.* Biochim Biophys Acta, 2009, **1794(3)**: p. 375-397.

47. Fedurkina, N.V., et al., *Change in kinetic regime of protein aggregation with temperature increase. Thermal aggregation of rabbit muscle creatine kinase.* Biochemistry (Mosc), 2006, **71(3)**: p. 325-331.

48. Roberts, C.J., *Non-native protein aggregation kinetics.* Biotechnol Bioeng, 2007, **98(5)**: p. 927-938.

49. Kato, A., et al., *Mutational analysis of protein solubility enhancement using short peptide tags.* Biopolymers, 2007, **85(1)**: p. 12-18.

50. Scheraga, H.A., J.A. Vila, and D.R. Ripoll, *Helix-coil transitions re-visited.* Biophys Chem, 2002, **101(2)**: p. 255-265.

51. Chakrabartty, A., T. Kortemme, and R.L. Baldwin, *Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions.* Protein Sci, 1994, **3(5)**: p. 843-852.

52. Blaber, M., X.J. Zhang, and B.W. Matthews, *Structural basis of amino acid alpha helix propensity.* Science, 1993, **260(5114)**: p. 1637-1640.

53. Minor, D.L., Jr. and P.S. Kim, *Measurement of the beta-sheet-forming propensities of amino acids.* Nature, 1994, **367(6464)**: p. 660-663.

54. Trevino, S.R., 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, 2007, **366(2)**: p. 449-460.

55. Islam, M.M., et al., *Crystal structure of an extensively simplified variant of bovine pancreatic trypsin inhibitor in which over one-third of the residues are alanines.* Proc Natl Acad Sci U S A, 2008, **105(40)**: p. 15334-15339.

56. Kuroda, Y. and P.S. Kim, *Folding of bovine pancreatic trypsin inhibitor (BPTI) variants in which almost half the residues are alanine.* J Mol Biol, 2000, **298(3)**: p. 493-501.

57. Kyte, J. and R.F. Doolittle, *A simple method for displaying the hydropathic character of a protein.* J Mol Biol, 1982, **157(1)**: p. 105-132.

58. Nozaki, Y. and C. Tanford, *The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. Establishment of a hydrophobicity scale.* J Biol Chem, 1971, **246(7)**: p. 2211-2217.

59. Calamai, M., et al., *Relative influence of hydrophobicity and net charge in the aggregation of two homologous proteins.* Biochemistry, 2003, **42(51)**: p. 15078-15083.

60. De Young, L.R., K.A. Dill, and A.L. Fink, *Aggregation and denaturation of apomyoglobin in aqueous urea solutions.* Biochemistry, 1993, **32(15)**: p. 3877-86.

61. Lawrence, M.S., K.J. Phillips, and D.R. Liu, *Supercharging proteins can impart unusual resilience.* Journal of the American Chemical Society, 2007, **129(33)**: p. 10110-10112

62. Gudiksen, K.L., et al., *Increasing the net charge and decreasing the hydrophobicity of bovine carbonic anhydrase decreases the rate of denaturation with sodium dodecyl sulfate.* Biophys J, 2006, **91(1):** p. 298-310.

63. Chiti, F., et al., *Rationalization of the effects of mutations on peptide and protein aggregation rates.* Nature, 2003, **424(6950)**: p. 805-808.

64. De Groot, N.S. and S. Ventura, *Protein Aggregation Profile of the Bacterial Cytosol.* Plos One, 2010, **5(2)**: p. e1-17.

65. Monsellier, E., et al., *Aggregation propensity of the human proteome.* PLoS Comput Biol, 2008, **4(10)**: p. e1-9.

66. Goldberg, M.E., R. Rudolph, and R. Jaenicke, *A kinetic study of the competition between renaturation and aggregation during the refolding of denatured-reduced egg white lysozyme.* Biochemistry, 1991, **30(11)**: p. 2790-2797.

67. Luo, Q., et al., *Chemical modifications in therapeutic protein aggregates generated under different stress conditions.* J Biol Chem, 2011, **286(28)**: p. 25134-25144.

68. Ricci, M.S. and D.N. Brems, *Common structural stability properties **of 4**-helical bundle cytokines: possible physiological and pharmaceutical consequences.* Curr Pharm Des, 2004, **10(31)**: p. 3901-3911.

69. Arakawa, T., Y. Kita, and A.H. Koyama, *Solubility enhancement of gluten and organic compounds by arginine.* Int J Pharm, 2008, **355**(**12)**: p. 220-223.

70. Howe, P.W., *A straight-forward method of optimizing protein solubility for NMR.* J Biomol NMR, 2004, **30(3)**: p. 283-286.

71. Pace, C.N., G.R. Grimsley, and J.M. Scholtz, *Protein ionizable groups: pK values and their contribution to protein stability and solubility*. J Biol Chem, 2009, **284(20)**: p. 13285-13289.

72. Cunningham, B.C. and J.A. Wells, *High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis.* Science, 1989, **244(4908)**: p.1081-1085.

73. Yu, M.H., J.S. Weissman, and P.S. Kim, *Contribution of individual side-chains to the stability of BPTI examined by alanine-scanning mutagenesis.* J Mol Biol, 1995, **249(2)**: p. 388-397.

74. Kassell, B., *Bovine trypsin-kallikrein inhibitor (Basic Pancreatic Trypsin Inhibitor, Polyvalent Inhibitor from Bovine Organs,* Lorand L., editor. Methods in Enzymology, 1970, **19**: p. 844-852.

75. Parker, J.M., D. Guo, and R.S. Hodges, *New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites.* Biochemistry, 1986, **25(19)**: p. 5425-5432.

76. David R. L. *CRC Handbook of Chemistry and Physics*, Editor: 90th edition: section 7-1, 2009 .

77. Krishnan, S. A.A. Raibekas, *Multistep aggregation pathway of human interleukin-1 receptor antagonist: kinetic, structural, and morphological characterization*, Biophys. J. 2009, **96:** p.199-208.

78. Li, Y., C.J. Robert, *Protein aggregation pathways, kinetics and thermodynamics.* In Aggregation of Therapeutic Proteins (Wang, W & Roberts C.J. Eds), WILEY Hoboken, New Jersey, 2010, p. 63-97.

79. Wang, W., N. Li, S. Speaker, *External factors affecting protein aggregation.* In Aggregation of Therapeutic Proteins (Wang, W & Roberts C.J. Eds), WILEY Hoboken, New Jersey, 2010: p. 119-178.

80. Oosawa, F., M. Kasai, *A theory of linear and helical aggregations of macromolecules*, Jour. Mol. Biol., 1962, **4**: p.10-21

81. Kato, A., et al., *Thermodynamic properties of BPTI variants with highly simplified amino acid sequences,* Jour. Mol. Biol., 2007, **372:** p. 737-746.

82. Ricci, M.S. et al., *Mutational approach to improve physical stability of protein therapeutics susceptible to aggregation*. In misbehaving proteins (Murphy, R.M. & Tsai A.M. Eds), Springer, New York, 2006, p. 331-350.

83. Wang, G.Z. L.L. Chen, H.Y. Zhang, *Neighboring-site effects of amino acid mutation*, Biochem. Biophys. Res. Commun. 2007, **353:** p. 531-534.

84. Lyutova, E.M., A.S. Kasakov, B.Y. Gurvits, *Effects of arginine on kinetics of protein ggregation studied by dynamic laser light scattering and tubidimetry techniques*, Biotechnol. Prog., 2007, **23**: p.1411-1416.

85. Islam, M.M., M.A. Khan, Y. Kuroda, *Analysis of amino acid contributions to protein solubility using short peptide tags fused to a simplified BPTI variant*, Biochim. Biophys. Acta. 2012, **1824:** p.1144-1150.

86. Oosawa, F., S. Asakura, *Thermodynamics of the Polymerization of Protein.* Academic Press, London, 1975.

87. Zhang, Y.B. et al., *Protein aggregation during overexpression limited by peptide extensions with large net negative charge*, Protein Expr. Purif. 2004, **36:** p. 207-216.

88. Fuchs, S.M. and R.T. Raines, *Polyarginine as a multifunctional fusion tag.* Protein Sci, 2005, **14(6)**: p. 1538-1544.

Acknowledgments

I would like to express my heartiest gratitude to my supervisor Dr. Yutaka Kuroda for his continuous support and guidance throughout my studies by sharing his wide knowledge. His insightful advices during the challenging stages of my research have been invaluable and greatly appreciated.

I would like to thank Dr. Mohammad Monirul Islam of Chittagong University for his friendly advice throughout the course, specially, for introducing laboratory techniques during my initial stage of research. I want to thank Mrs. Patricia McGahan for English proof reading and her generous support and advices during my paper working. Further, I want to thank Keiko Takahashi for all the documentations.

Laboratory members to whom I am grateful are Dr. Tharangani Rathnayaka, Dr. Tetsuya Kamioka, Mr. Elahi Montasir, Mrs. Manjiri Ravindra Kulkarni, Mr. Shunta Kimura and Mr. Kei Kobayashi for their help while running experiments. Thanks to all members of my laboratory for being so supportive and good friends.

I greatly appreciate Bangladesh Government for deputing me to continue this research. My sincere thanks to the concern officials of the Bangladesh Government specially those who are working at relevant desks of Ministry of Public Administration.

Finally, I would like thank my parents, and other family members for all of their support, understanding and care. Special thanks to my wife, Mrs. Kohinoor Begum, my beloved sons, Ishfak Imam Khan (Kaif), and Shafkat Sharaf Khan (Soif). I am especially grateful to my wife, without her sacrifice it would have been impossible to continue this research.