Effects of fusion with an unstructured and hyper-acidic polypeptide on the nature of recombinant proteins expressed in bacteria

構造をとらない超酸性ポリペプチドの融合が組換え発現蛋白質の性状に与える影響

Takeshi SANGAWA

Graduate school of Science
Osaka University

June 2013

Contents

Abbreviations

Aβ	Amyloid β
AP	Alkaline phosphatase
ApoER2	Apolipoprotein E receptor-2
APP	Amyloid precursor protein
BCA	Bicinchoninic acid
BSA	Bovine serum albumin
CAPPD	Central APP domain
CBB	Coomassie brilliant blue
CHO	Chinese hamster ovary
CIAP	Calf intestine alkaline phosphatase
CuBD	Copper binding domain
Dkk1	Dickkopf-1
DTT	Dithiothreitol
ELISA	Enzyme-linked immunosolvent assay
ER	Endoplasmic reticulum
FAAR	*N*-(3-[2-furyl] acryloyl)-Ala-Arg
FAPP	*N*-(3-[2-furyl] acryloyl)-Phe-Phe
FATT	FLAG-acidic region-TARGET-Tag
FN9-10	Fibronectin 9-10th type-III repeat
FN10	Fibronectin 10th type-III repeat
Gdn-HCl	Guanidium hydrochloride
GFLD	Growth-factor like domain
GFP	Green fluorescent protein
GSH	reduced glutathione
GSSG	oxidized glutathione
GST	Glutathione-*S*-transferase
hCPB	human carboxypeptidase B
IgG	Immunoglobulin G
IPTG	isopropyl β-D-1-thiogalactopyranoside

LA1	LDLR class A module-1
LB	Lysogeny-Broth
MBP	Maltose binding protein
MeCPA	*Metarhizium* carboxypeptidase A
PAGE	Polyacrylamide electrophoresis
PAR	Protease-activated receptor
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PMSF	Phenylmethylsulfonyl fluoride
*p*NPB	*p*-nitrophenyl phosphonate
PVDF	Polyvinylidene difluoride
scFv	Single-chain variable fragment
SDS	Sodium dodecyl sulfate
TBS	Tris-buffered saline

Standard three-letter of one-letter abbreviations were used for common amino acids.

I. Introduction

Escherichia coli is the most universally used expression system

The DNA recombinant technology combined with heterologous expression systems is one of the most powerful and essential tools in the field of protein science. It enables overexpression of proteins which have low physiological expression levels, production of eukaryotic proteins in bacterial cells, and various genetic manipulations aimed at enhancing expression level, solubility, structural stability or purification yield of a target protein (Baneyx F, 1999).

Unarguably, *Escherichia coli* is the most widely used expression host in both industrial application and basic research because of low cost medium, rapid growth, abundant expression level, and well-characterized genetic background. Started from the recombinant production of human insulin in 1982, now approximately 40% of biopharmaceutical proteins are produced using *E. coli* expression system (Terpe K *et. al.*, 2006). Statistical study of Protein Data Bank (PDB) shows 88.5% of the deposited structures are derived from protein samples produced in *E. coli* (**Fig.1**), indicating the critical requirement of the system in experiments that demand large quantity of proteins such as X-ray crystallography.

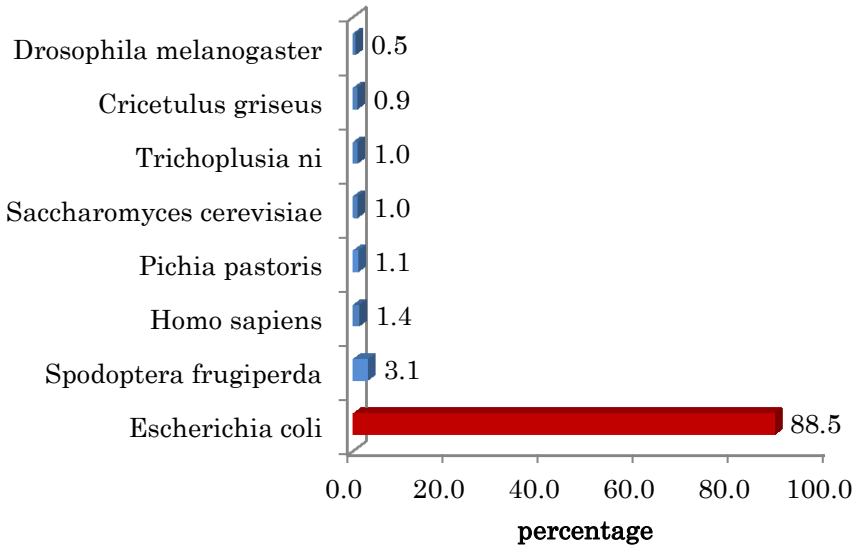

Figure 1. Classification of PDB-deposited data by expression organisms. The value beside bar represents usage of each organism in percentage in 82033 data calculated by RCSB PDB Statistics (http://www.rcsb.org/pdb/static.do?p=general_information/pdb_statistics/index.html). *Escherichia coli* strain was shown in red bar.

Three problems in the E. coli recombinant expression system

Despite the highly cost-effective nature of *E. coli* expression system, over 60% of pharmaceutical proteins are produced in other organisms (mostly mammalian cells and *Saccharomyces cerevisiae*). Apart from the obvious shortcoming such as the lack of post-translational modifications, there are three major hurdles in recombinant protein production in bacteria. First, expressed heterologous proteins often form insoluble aggregates within the bacteria cytosol. Second, one must come up with an efficient purification method for the target protein from the crude bacterial lysate, maintaining its native structure. When the protein is expressed as an insoluble form, one can still obtain native protein by "refolding", which leads to the third problem; one must empirically

determine the experimental condition to refold functional proteins from the denatured polypeptides *in vitro*. Therefore, 1) soluble expression, 2) purification, and 3) refolding are the three major bottlenecks in the *E. coli* expression system, sometimes forcing the researchers to abandon this great expression system.

Tagging strategy and protein purification

Tag fusion strategy is frequently used in protein expression projects. Especially, it is an excellent solution for purification. Most of the fusion tags have their specific interaction partners that enable one-step separation of the tagged protein from other contaminations (**Table 1**). The most commonly used tag is poly histidine tag (His-tag), using metal-chelating property of tandem histidine sequence (Porath J *et. al.*, 1975). His-tagged protein can be easily isolated from the crude lysate by flowing through a nickel complex-immobilized affinity column. FLAG (DYKDDDDK) (Hopp TP *et. al.* 1988), HA (YPYDVPDYA) (Field J *et. al.* 1988) and Myc (EQKLISEEDL) (Munro S *et. al.* 1984) are also common peptide tags. Proteins with these peptide tags can be isolated using immunoaffinity resins that had been coupled with their specific antibodies. Our lab had developed one such peptide tag called "TARGET-tag", utilizing an anti-peptide monoclonal antibody and its unique recognition sequence (Sangawa T *et. al.*, 2008, Nogi T *et. al.*, 2008, Tabata S 2010).

In addition to the short peptide, proteins which have larger molecular weight are also used as fusion tag. Glutathione-*S*-transferase (GST, 26 kDa) and maltose-binding protein (MBP, 40 kDa) are two most frequently used fusion partners (Benard V, *et. al.*, 2002, Ishii Y *et. al.* 1998). GST- and MBP- fusion proteins are specifically captured by glutathione- and amylose-immobilized resins, respectively.

As these “affinity resins” exploit their ability to interact specifically with the respective ligands, it is generally possible to isolate tagged proteins from crude starting material with high purity in a single step (Walters RR. 1985). One drawback is that these affinity resins, especially immunoaffinity resins, are generally expensive and lose capacity or specificity during the repeated use and regeneration, increasing the overall cost (Narayanan SR, 1994). Additionally, elution buffers often contain expensive components such as competitive peptide or compounds (e.g. epitope peptides). Alternatively, low or high pH elution buffers are employed to destroy the interaction between the resin and tag but their use often results in a serious damage of both the immobilized binders on the resin and the target protein.

Table 1. Commonly used affinity tags. Residue number, amino acid sequence, specific purification matrices and molecular size were described.

Tag name	Residues	Sequence	Purification matrix	Size (kDa)
Peptide tag				
Poly-His	Typically 6-10	HHHHHH	Ni- or Co- chelete	0.14 x 6-10
FLAG	8	DYKDDDDK	Anti-FLAG antibody	1.01
Myc	11	EQKLISEEDL	Anti-Myc antibody	1.20
HA	9	YPYDVPDYA	Anti-HA antibody	1.10
Strep-tag II	8	WSHPQFEK	Modified streptavidin	1.06
Proteinacious tag				
GST	211	Protein	Glutathione	26.00
MBP	396	Protein	Cross-linked amylose	40.00

In contrast to the proprietary affinity systems described above, traditional chromatographies such as ion exchange, hydrophobic interaction, and size exclusion are group-specific separation methods which have been used in the field even before the introduction of DNA recombinant technology (Hamilton PB *et. al.*, 1960, Lindqvist B *et. al.*, 1955, Porath J *et. al*., 1959). In these methods, proteins are fractionated according to their physicochemical properties such as electrostatic charges, hydrophobicity or molecular size. These traditional resins are advantageous in several regards. First, they are much less expensive compared to most of the modern affinity chromatographic resins. Second, binding capacity of traditional adsorption chromatographies is very high (e.g. binding capacity of anion exchange resin Q-sepharose FF is 120 mg /mL, in contrast to that of Ni-NTA and anti-FLAG resins of only 5-10 mg/mL, and 1-2 mg/mL, respectively). Third, these resins either do not require regeneration or are tolerable to harsh regeneration condition, making the repeated use possible. These traditional methods however generally suffer from the low purification efficiency due to the lack of strict specificity. As a result, combination of multiple chromatographic steps is usually employed to achieve sufficient purification. Because any purification step is invariably associated with sample loss to some extent, there are trade-off relationships among the cost, the purity and the yield. Despite the development of various sophisticated tagging technologies, there still remains a demand for better protein purification methods.

Insoluble expression

Insoluble expression is recognized as one of the major problem in bacterial expression systems. Proteins are synthesized on a ribosome and fold into their native structure in the appropriate cellular environments. According to the Anfinsen's dogma, this folding

process is considered to occur spontaneously because native structure is located at the bottom of the funnel-like energy landscape (Anfinsen CB, 1972). Although this dogma is generally accepted for events that happen in the infinitely diluted condition (except for intrinsically disordered proteins), it may not be directly applicable to a protein folding process in living cells that occurs in a crowded cytosol filled with other proteins and chemical compounds (McGuffee SR et. al., 2010). It is therefore natural to expect that physiological protein folding process is greatly affected by conditions such as concentration of proteins, redox reagents and molecular chaperones (Baneyx F, 2004). This could be one the reason why heterologously expressed proteins often have difficulty in the correct folding (Ellis RJ *et. al.*, 1999).

Misfolded proteins often form insoluble and inactive aggregation granule called inclusion body (Kane JF et. al., 1988). Formation of inclusion body is particularly common when eukaryotic proteins are expressed in prokaryotic organism such as *E. coli*. It is estimated that over 30% of human genome products forms inclusion body when expressed in *E. coli* (Braun P *et. al.*, 2002). Although there are many aspects that differ between eukaryotic and prokaryotic cells, one of the most important differences is their redox environment. This difference is crucial when target protein has disulfide bonds in the native structure. Such proteins, which require oxidative environments for proper folding, are transferred into the lumenal compartment of the endoplasmic reticulum (ER) in eukaryotic cells. In prokaryotes, synthesized polypeptides are emitted into the cytoplasm, which has reducing environment (Kadokura H *et. al.*, 2003). This is the main reason why disulfide-containing proteins are difficult to express in bacterial recombinant systems.

Other difference is the speed of polypeptide synthesis. Translation speed of prokaryotic

ribosome is faster than eukaryotes (prokaryotes: 10-20 amino acids/min, eukaryotes: 3-8 amino acids/min, respectively) (Siller E *et. al.*, 2010). Folding of a nascent polypeptide occurs post-translationally in prokaryotes and co-translationally in eukaryotes. Because the average length of polypeptides in eukaryotes is longer than prokaryotes (Rost B, 2002), and eukaryotic proteins tend to contain multiple domains in a single polypeptide, it is hypothesized that eukaryotic cells have evolved to have slow peptide synthesis with gradual folding scheme in order to reduce the number of possible (mis)folding states. In fact, it is reported that delaying translation speed by appropriate concentration of ribosome inhibitor (streptomycin) to *E. coli* improved the ratio of correctly-folded recombinant eukaryotic proteins (Siller E *et. al.*, 2010). In addition to the speed of production, the total quantity of produced proteins is also an important factor because protein aggregation is greatly dependent on the protein concentration (Kane JF *et. al.*, 1988). Therefore, there is a dilemma that high expression level which is clearly desirable from the viewpoint of protein production has to be avoided when expressing disulfide-containing, aggregation-prone proteins in *E. coli*.

There are several established ways to get around the problem of inclusion body formation. The simplest one is to lower the expression levels by controlling the incubation time and temperature after the induction. This approach is effective when the aggregation simply occurs in concentration-dependent manner. However, such approach is not effective when the protein has aggregation tendency as its intrinsic character. In the case of disulfide-containing proteins, attachment of bacterial secretion signal sequence to the protein can sometimes solve the problem by secreting the target protein into to periplasmic space, an oxidative folding space in gram-negative bacteria (Bowden GA *et. al.*, 1990). However, secretion efficiency of the signal-attached proteins depends

greatly on each case and the production yield is generally low, partly because of the limited available space of periplasm. Second approach for disulfide formation is an engineered *E. coli* strain that has oxidative condition in cytoplasm. Origami is a strain which has mutations in genes of two reducing enzymes thioredoxin B (*TrxB*) and glutathione reductase (*gor*) resulting in an oxidative cytoplasmic environment. Certain enzymes with critical disulfide bonds are successfully expressed and purified in a native form with this strain (Bessette PH *et. al.* 1999, Prinz WA *et. al.* 1997).

In certain cases aggregation-prone proteins can be "tamed" by co-expression with chaperones such as GroEL / ES, DnaK, DnaJ, and GrpE, because they can support or enhance correct folding (de Marco A *et. al.*, 2007). Folding enhancers can also be introduced as part of the tagging strategy. For example, fusion of highly-soluble protein tags such as GST, MBP and alkaline phosphatase (AP) are effective solubility enhancers (Smith DB *et. al.*, 1988, di Guana C *et. al.*, 1988, Manoil C *et. al.*, 1990, Sun P *et. al.* 2011). MBP is also reported to have a chaperone-like activity when fused to N-terminus (to fold prior to the target protein) of aggregation-prone protein (Bach H et. al., 2001).

Another route: protein refolding from inclusion bodies

Although the protein that formed inclusion body is insoluble and nonfunctional, it is known that inclusion body contains large quantity of relatively pure protein. Thus, instead of trying to avoid inclusion body formation, one can deliberately express the target protein as inclusion body and refold it to its native state. The concept of refolding is simple and reasonable (**Fig.2**). First, aggregated protein is solubilized by adding high concentration of denaturing agent to achieve complete unfolding. As the concentration of the denaturant decreases, target protein will gain partial native structure. As described

above, protein folding theory suggests that a polypeptide will fold spontaneously in the absence of denaturant. So the most effective way of facilitating the refolding would be to dilute the protein solution infinitely. In a real situation, however, simple dilution does not always work because the folding kinetics and appropriate conditions differ markedly between individual proteins, making the refolding experiment generally challenging.

Both folding and misfolding are governed by the same principle; the reactions proceed to minimize free energy, by burying hydrophobic surface, forming hydrogen bonds and disulfide bonds to gain enthalpic advantage (Dobson CM, 2004). Anfinsen's dogma states that native structure is the lowest free energy point in the folding funnel, but does not guarantee that there are no other local minima (Clark PL, 2004). Aggregation also occurs by the same reason as the folding reaction. Therefore, careful condition optimization is required to facilitate proper folding while avoiding misfolding and aggregation.

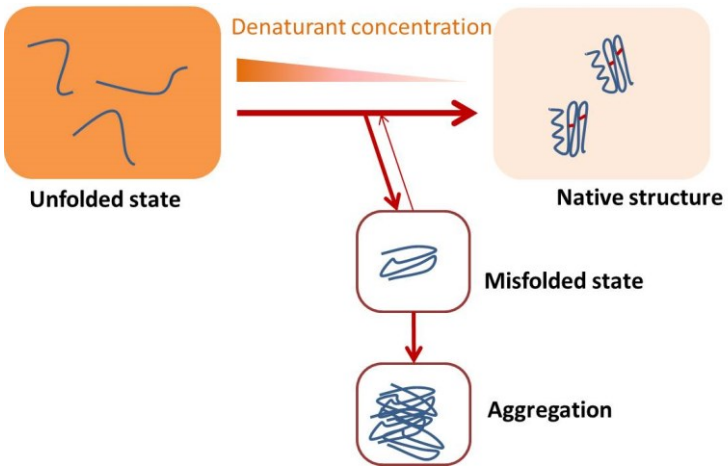

Figure 2. The schematic view of protein refolding. The blue line represents polypeptide. As concentration of denaturant (orange) reduced, unfolded polypeptides are folded into native structure spontaneously. Misfolding and aggregation compete with the proper folding.

There are various parameters that need to be determined for particular refolding experiment (**Table 2**). First, denaturation conditions including type and concentration of denaturant and reducing agent, denaturation time and pH (Singh SM et. al., 2005), must be explored. Next, one has to choose from three major methods (dilution, dialysis, and on-column renaturation) for the purpose of lowering denaturant concentration (Vallejo LF *et. al.*, 2004, Basu A *et. al.*, 2011). These can be used in combination. Furthermore, process for lowering the denaturant concentration includes two alternative ways, in a step-wise or a continuous manner. In both cases, rate of the reduction of denaturant concentration is also an important factor. Chemical compounds called "refolding additives" are often added in the refolding buffer. L-arginine, polyethylene glycol (PEG), glycerol or low concentration detergents are used as an additive for aggregation suppression (Ishibashi M *et. al.*, 2005, Cleland JL *et. al.* 1992). If the protein has disulfides, redox agents such as glutathione, cysteine, cystine, dithiothreitol (DTT), or 2-mercaptoethanol are required (Mamathambika BS *et. al.*, 2008). Reducing and oxidizing agents are often mixed in various ratios to induce correct disulfide formation (called "redox shuffling") (Patil G *et. al.* 2008). The optimum duration of redox reaction in the refolding process may also vary significantly. Finally, the initial concentration of denatured protein is one of the most critical factors in many cases (Eiberle MK *et. al.*, 2010). In most cases, refolding reaction is conducted at 4°C at neutral to slightly basic pH condition. For proteins containing disulfide bonds, buffers with pH higher than 8 is favored because of the pK_A of thiol group. Physical condition can also affect the refolding efficiency. Attempts to facilitate refolding under high pressure (200 MPa) have seen some successes (Okai M *et. al.*, 2012). Although some papers state that the refolding process is conducted under continuous stirring, my personal experience and

words from colleagues do not support the idea of stirring being helpful in refolding reaction. This point needs to be explored more.

Table 2. Various refolding conditions. As illustrated in Fig.2, refolding procedure consists of three steps; denaturation of misfolded / aggregated proteins, reduction of denaturant concentration (this step is called "refolding" in a narrow sense), and purification. Typical conditions used in REFOLD database were described in the table.

Step	Variables	Typical examples
Denaturation	Denaturant	Gdn-HCl, urea, detergent
	Reducing agent	DTT, β-ME, GSH, Cysteine
	Time and temperature	4°C-37°C, no time - overnight
Refolding	Basic method	Dilution, Dialysis, On-column
	Buffer exchange method	Stepwise, Gradient
	Time and temperature	4°C-37°C, 30min - several days
	Redox agent	DTT, GSH/GSSG, Cysteine/Cystine
	Additives	L-Arg, PEG, Glycerol, Salt
	Final protein conc.	< 90 μg/mL
	Buffer pH	7-9.5
Purification	Pre- or Post- refolding	Ni-NTA in denaturation condition

The "successful" refolding methods reported in papers vary widely and are all case-dependent. Correlation between effective refolding protocol and properties of proteins e.g. amino acid sequence, structural classes or molecular weight is poorly understood. In fact, even for the single-chain antibodies, which share the same overall structure with minor variations in the CDR loop conformations, there is no generalized method for efficient refolding. Therefore, refolding condition must be determined by iterative trials and errors, which could take months. For this reason, refolding has not become researchers' favorite method of choice except when efficient refolding protocol has been already established.

In 2005, a research group of Monash University in Australia established the REFOLD database, an archive of protein information and successful refolding protocols (Chow MK *et. al.*, 2006). Although it has not been updated since January 2012, 1165 data were stored in the database. Statistical study of the REFOLD database revealed that *E. coli* was used as an expression host in 1088 trials (**Fig.3A**) (Khodarahmi R *et. al.*, 2008). Statistical analysis from this database revealed some tendency in current refolding studies. For example, dilution and dialysis are two most commonly used refolding methods (approximately 50% and 30% of successful trials, respectively) (**Fig.3B, right panel**). Dialysis method tends to be employed when the target protein have no disulfide bond in their native structure (**Fig.3B, left panel**). The mixture of reduced and oxidized glutathione (GSH / GSSG) is the most frequently used redox agent for proper disulfide formation during renaturation (**Fig.3C**). Although careful inspection of this database may provide general useful information, it only gives an "average method" but not necessarily provide experimental directions to one's own project. Therefore, there still exists a great demand for the establishment of target-independent refolding method.

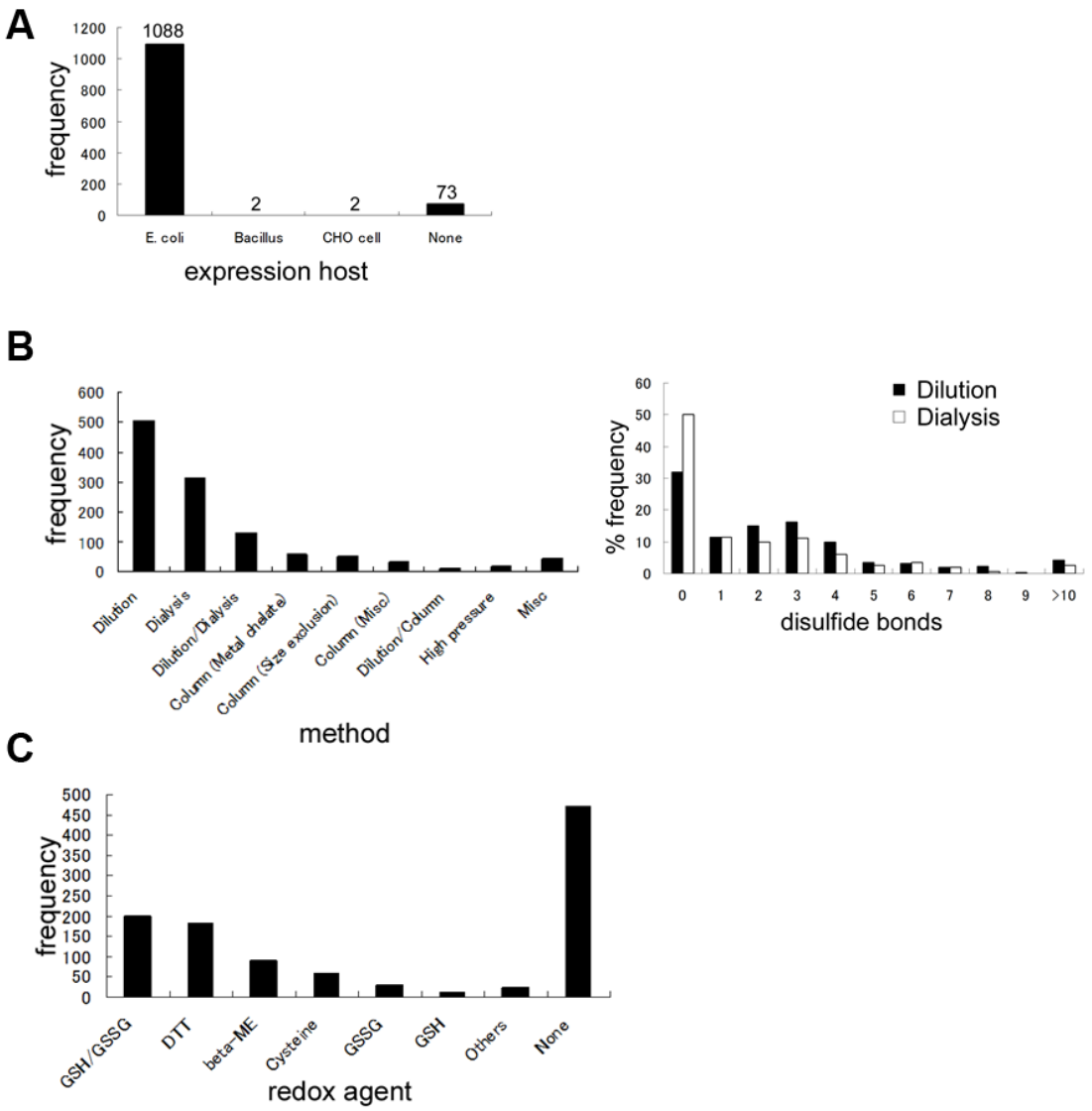

Figure 3. Statistical studies of REFOLD database. The vertical axis represents frequency of expression hosts in 1165 data stored in the database. **(A)** Distribution of expression organisms. "None" represents cases which use purchased or preliminary purified proteins. **(B)** Usage of refolding methods. Right panel shows frequency of respective methods. Left panel shows relationship between usage of two methods (dilution and dialysis) and number of disulfide bonds in the native target protein. Black bar represents dilution and white bar represents dialysis. **(C)** Usage of redox reagents in renaturation procedure.

Purpose of this study

Although *E. coli* expression system is a powerful method for protein production, there are three unsolved problems; (1) trade-off relationship between cost, purity and yield in purification procedure, (2) insoluble expression, (3) lack of generalized method for functional refolding. I first focused on the problem of purification. To achieve low cost, quick, and efficient purification without using the proprietary affinity tag systems, I sought for methods that utilize traditional purification methods such as ion exchange or size exclusion chromatography. These group-specific separation methods do not usually allow isolation of the target protein from contaminating proteins with similar pI or molecular size. With a prediction that a fusion partner with an extreme electrostatic charge would make it possible to isolate the target protein by an anion exchange chromatography in a single step, I chose to test the effect of adding an extracellular hyper-acidic region derived from the Amyloid Precursor Protein (APP). Moreover, I found that this fusion tag system also greatly helped to overcome the two remaining problems; insoluble expression and difficulties in protein refolding. Thus I applied the system to various proteins to characterize the property of the tag, and to understand the basic principle of the unique behavior of the tag polypeptide.

II. Results

Amyloid precursor protein and hyper acidic region

Amyloid Precursor Protein (APP) contains 5 extracellular regions (**Fig.4A**) (Leong SL *et. al.,* 2011). Starting from the N-terminus, there are Growth-Factor Like Domain (GFLD), Copper-Binding Domain (CuBD), hyper-acidic region, Central APP Domain (CAPPD), and Amyloid β region, which is half-buried in a cellular membrane.

The hyper-acidic region spans residue 190 to 295, of which 49% are negatively charged amino acids (Glu and Asp). Majority of this region is predicted to be natively unstructured by multiple disorder prediction tools including PONDR (**Fig.4B**).

Adding FLAG epitope (DYKDDDDV), TARGET-tag (YPGQVGYPGQVGYPGQV) for detection and factor Xa cleavage site, I established a novel hyper-acidic fusion tag "FATT", which is an acronym of FLAG-Acidic region-TARGET-Tag (**Fig.5**).

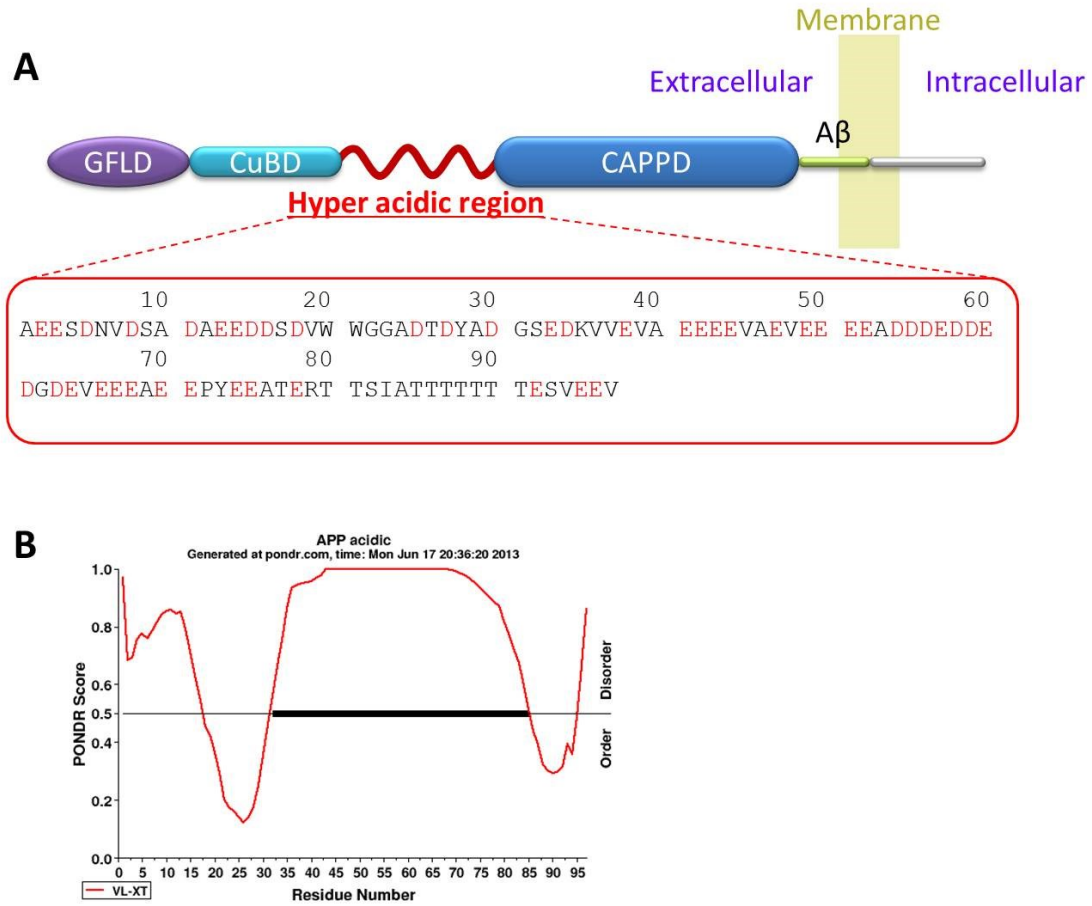

Figure 4. Characteristics of APP acidic region. **(A)** Illustration of APP acidic region. From N-terminus, GFLD, CuBD, hyper-acidic region, CAPPD and Aβ. Acidic residues in the sequence of the hyper-acidic region are shown in red letters. **(B)** Disorder prediction of APP hyper-acidic region using PONDR (http://www.pondr.com/index). The vertical axis represents PONDR score, which corresponds to disorder tendency and horizontal axis represents residue number corresponding to the number displayed in (A). The residue with score of >0.5 was concluded as a "disordered residue". 75.26% of the sequence was predicted to be disordered.

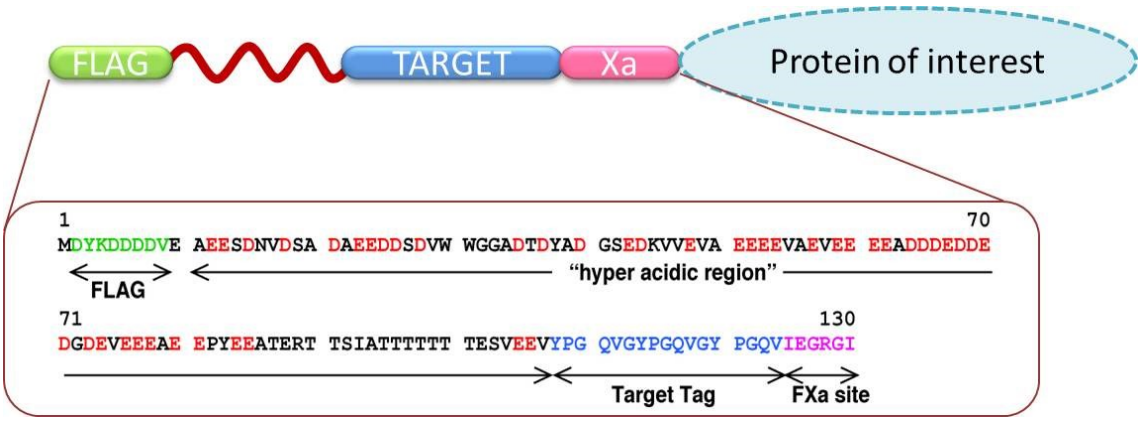

Figure 5. Schematic illustration and whole amino acid sequence of FATT. The sequence was colored as below; FLAG-tag: light green, hyper acidic region; red, TARGET-tag; blue, factor Xa recognition sequence: pink. (This figure was reproduced from Figure 1A in Sangawa *et. al*. 2013)

Strategy for affinity chromatography-like purification using anion exchange chromatography

Because of the highly acidic primary structure, theoretical isoelectric point of FATT is very low (3.2). First I hypothesized that fusion of FATT to a target protein would enable a single-step purification using anion exchange chromatography. Calculation of theoretical pI of various proteins before and after the fusion of FATT confirmed the strong acidifying effect by FATT-fusion. Remarkably, all proteins tested are estimated to have pI values lower than 4.5 (**Table 3**). As 99% of genome products in *E. coli* have pI value higher than 4.5 (Niwa T *et. al.*, 2005), in theory we should be able to purify FATT-fusion proteins using anion exchange chromatography in a single step.

Table 3. Theoretical pI of FATT-fusion proteins. The values were calculated using ExPASy Compute pI / MW tool (http://web.expasy.org/compute_pi/) by resolution of "average".

	Without FATT		With FATT	
Protein	M.W. (Da)	Theoretical pI	M.W. (Da)	Theoretical pI
Dkk1_C	10378	9.15	24719	4.20
GFPuv	28375	6.14	43219	4.35
MeCPA	44755	6.66	58999	4.35
Human CPB	47368	6.16	61709	4.44
Cutinase	26168	8.35	40510	4.33
P20.1 scFv	28591	6.50	42933	4.18
A5201F scFv	29158	6.15	43500	4.13
SG/19 scFv	29723	775	44065	4.27

Construction and expression of FATT-tagged proteins

To evaluate effects of FATT-fusion in protein expression level in *E. coli*, two model proteins GFP_{UV} (UV-excited green fluorescent protein, 246 residues, pI 5.81) and Dkk1_C (C-terminal fragment of human Dkk1, 91 residues, pI 9.15) (Haniu M *et. al.*, 2011) were expressed with or without N-terminal FATT (**Fig.6A**). FATT-fusion did not change the expression level of Dkk1 _C in the bacterial cytosol. Although FATT-tagged version of GFP_{UV} showed approximately 50% reduction in the expression level, such reduction was not observed in a majority of the cases (see **Fig.11** in the later section). Due to inclusion of epitope motifs at both N- and C-terminal of the acidic region, FATT-fusion proteins could be specifically detected by Western blotting using anti-FLAG polyclonal antibody and P20.1 IgG (anti-TARGET tag monoclonal antibody) (**Fig.6B**).

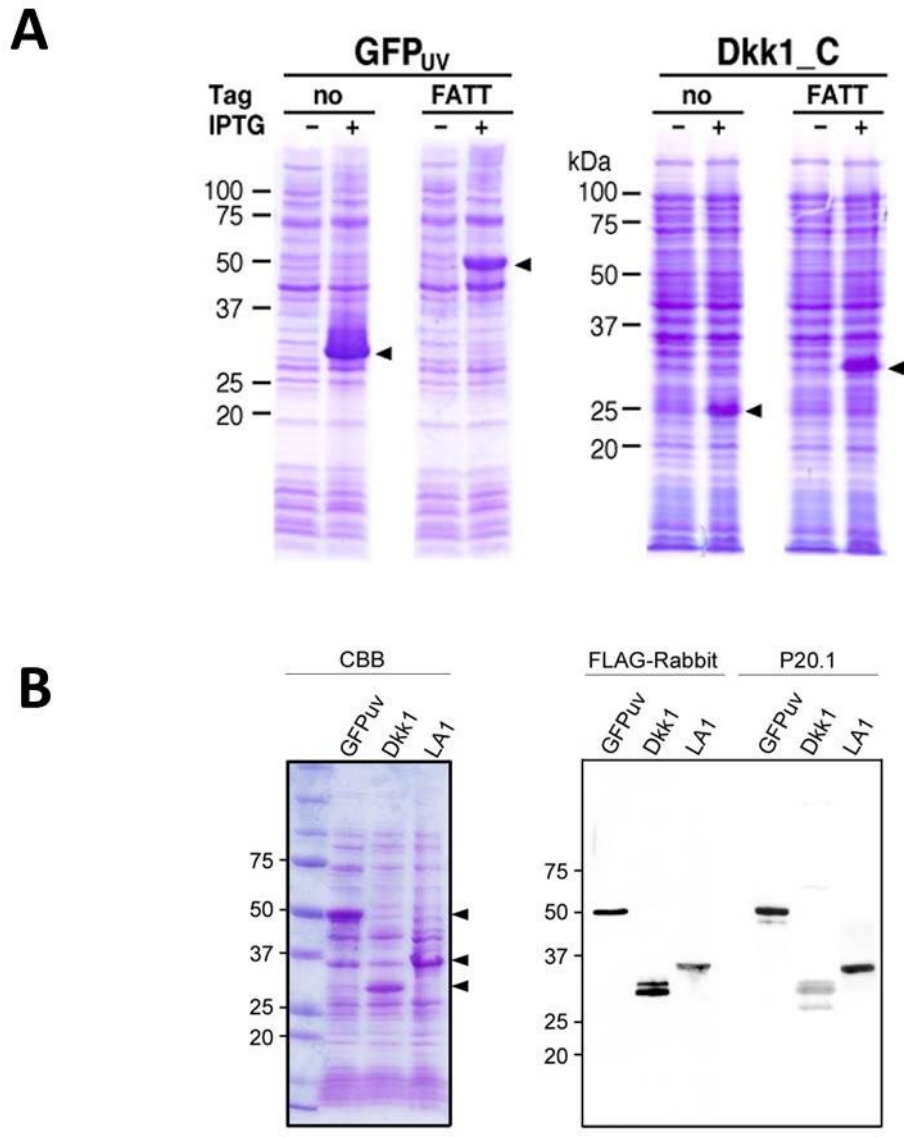

Figure 6. Basic properties of FATT. **(A)** SDS-PAGE analysis of total bacterial lysates expressing either untagged or FATT-tagged Dkk1_C (left panel) and GFP_{UV} (right panel). Sample volume was adjusted so that each lane contains roughly equal amount of *E. coli* endogenous proteins before and after induction. FATT-fusion proteins are indicated with a black arrow. **(B)** Western blot analysis of FATT-fusion proteins. Left panel represents CBB-staining gel. FATT-fusion proteins are indicated with a black arrow. Right panel represents Immunostained PVDF membrane. Anti-FLAG rabbit polyclonal antibody or P20.1 was used as a primary antibody, respectively.

(This figure was reproduced from Figure 1C and D in Sangawa *et. al*. 2013)

One-step purification of FATT-fusion proteins using anion exchange chromatography

I evaluated the anion-exchange chromatographic behavior of FATT-tagged protein using FPLC-pile format. The soluble fraction of cells expressing FATT-Dkk1_C was directly loaded onto HiTrap Q HP connected to an AKTA FPLC system (**Fig.7**). NaCl concentration in the running buffer was gradiently charged from 0 mM to 1 M. FATT-Dkk1_C was found in fraction 43 (NaCl concentration: approximately 550 mM) (**Fig.7**). This corresponded to the last eluting protein fraction in the chromatogram, because of the last two peaks after fraction 43 constituted primarily with nucleic acids as shown by the lack of CBB staining and the typical UV absorption profile (data not shown).

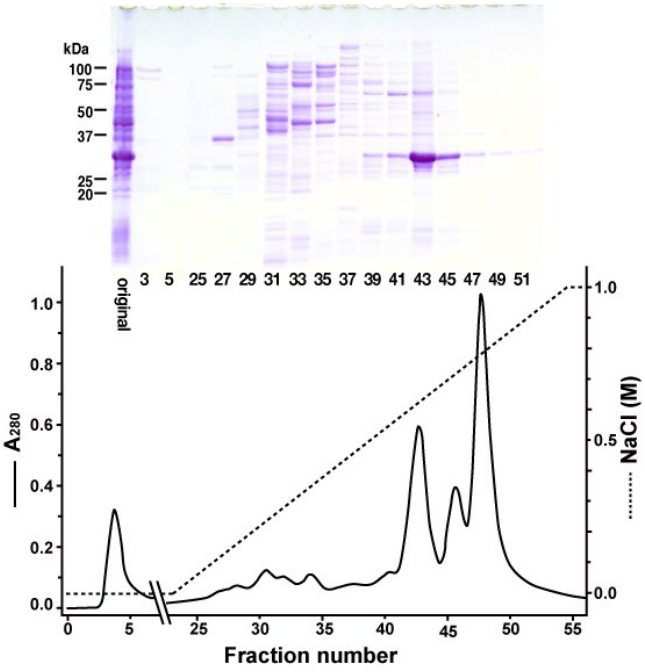

Figure 7. Anion exchange chromatography of cytosolic proteins from *E. coli* expressing FATT-Dkk1_C. SDS-PAGE analysis (upper panel) shows that FATT-Dkk1_C peak at fraction 43, well separated from the majority of other proteins. Lower panel represents chromatogram of HiTrap Q on AKTA FPLC. The solid line indicates A_{280} and the dotted line indicates NaCl concentration. (This figure was reproduced from Figure 1E in Sangawa *et. al*. 2013)

Next I examined the possibility of one-step purification in a gravity flow column chromatography. A soluble fraction of FATT-GFP_{UV}-expressing cell lysate was passed through a small column of Q-Sepharose FF. As the sample was dissolved in a buffer containing 150 mM NaCl, most proteins did not bind to the column and were recovered in the flow through fraction, except for the ~50 kDa FATT-GFP_{UV} (**Fig.8A**, compare lanes 1 and 2). No proteins were present in the wash fraction with a wash buffer containing 400 mM NaCl (lane 3), and FATT-GFP_{UV} was finally eluted with the elution buffer containing 1 M NaCl (lane 4). Purity of the target protein in the elution fraction was over 95% (estimated from band intensity of CBB-staining gel) (**Fig.8A**).

As FATT is a ~100 residue polypeptide and was unusually high electronegative character, it is important that it can be removed from the target protein after the purification. Upon incubation with factor Xa at 4°C for 16h, purified FATT-GFP_{UV} could be completely digested and split into ~30 kDa GFP_{UV} and ~17 kDa FATT portion (**Fig.8B**, compare between "-" and "+"). No significant difference in fluorescence property was observed before and after tag digestion (**Fig.8A**, lower photographs), suggesting that the presence of FATT segment does not have adverse effect on the structure and function of GFP_{UV}. Furthermore, GFP_{UV} and the liberated FATT moiety could be easily separated by simply passing through the second Q-Sepharose anion exchange column. Importantly, cleaved GFP_{UV} recovered in the flow through fraction was highly pure (Fig.8B, lane "IEX"), because everything else including FATT, uncleaved FATT-fusion protein (if any), and other contaminants all have strong tendency to bind to Q-Sepharose and were captured. The one-step purification, successful cleavage of the FATT portion, and the subsequent easy separation of the tag-free protein were applicable to many other proteins (described in the later section).

As the purification steps described above only utilize conventional anion exchange resin such as Q-Sepharose and does not require any other special equipment or reagents, these results confirm my prediction that FATT-fusion strategy is a powerful way to achieve cost-effective recombinant protein purification.

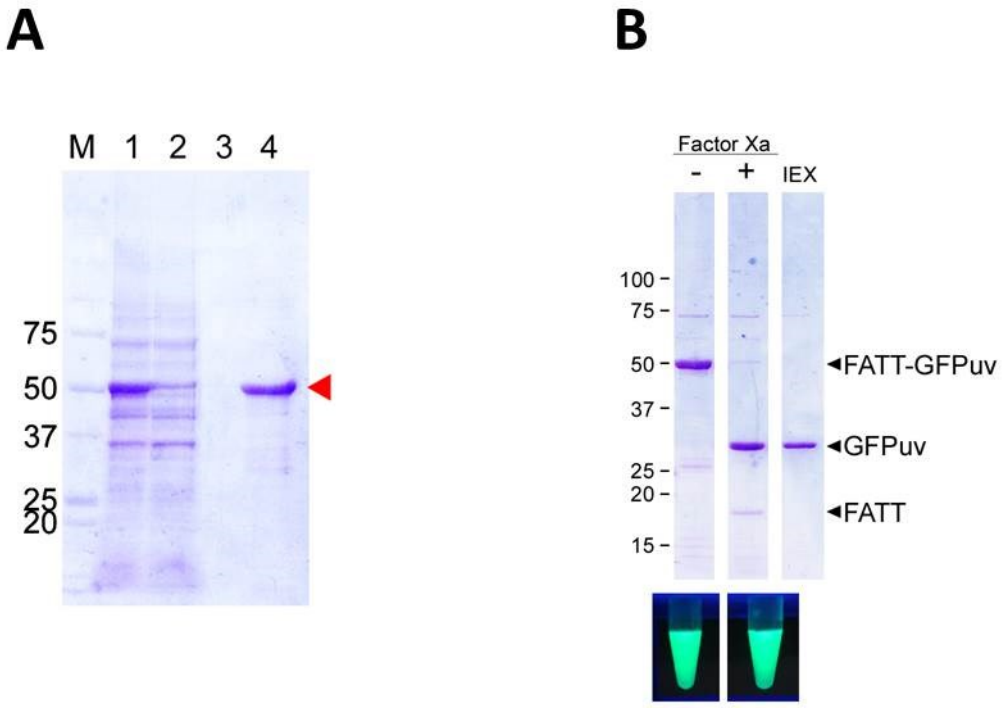

Figure 8. Purification of FATT-tagged GFP_{UV} using Q-sepharose. **(A)** Purification from bacterial lysate. M: molecular marker, lane 1: soluble fraction, lane2: flow through, lane 3: wash fraction, 4: elution fraction. Red arrow represents FATT-GFP_{UV}. **(B)** SDS-PAGE analysis before and after factor Xa treatment. (-): before cleavage, (+): after cleavage. Pictures below represent the fluorescence of GFP_{UV} under ultraviolet ray. IEX: flow through fraction of Q-sepharose purification after cleavage. The band of FATT was disappeared in IEX lane.

(This figure was reproduced from Figure 2A and B in Sangawa *et. al*. 2013)

Physicochemical property of FATT polypeptide

Analytical size exclusion chromatography of FATT-GFP_{UV} resulted in a monodisprese peak at the elution position corresponding to a 158 kDa globular protein, despite the fact that theoretical relative molecular mass was only 44 kDa (**Fig.9**, upper panel). After factor Xa digestion, two peaks were observed. Chromatogram monitored at 395 nm (maximum absorbance of GFP_{UV}) indicates that the later peak (at the position of 40 kDa) is the active GFP_{UV}. This position is considered to be reasonable because the theoretical molecular mass of GFP_{UV} is 30 kDa. On the other hand, another peak, expected as a digested FATT, appeared at 78 kDa position despite its true molecular mass of only 14 kDa (**Fig.9**, lower panel).

The unusually high apparent molecular size of the FATT (>5 times larger than the actual size) described above does not seem to be the result of oligomerization, because homooligomerization of FATT should be disfavored because of the strong electrostatic repulsion between negatively charged molecules. FATT sequence is expected to be unstructured, and it is known that unstructured proteins after denaturation have apparent molecular size compared to the theoretical value in size exclusion chromatography (Uversky VN, 1993, Uversky VN *et. al.*, 2010). It is therefore likely that FATT polypeptide exists as a monomer with large Stokes radius in solution. To confirm this hypothesis, cleaved FATT and GFP_{UV} were analyzed using size exclusion chromatography under native and denaturing (containing 6 M Gdn-HCl in protein solution and running buffer) conditions. As shown in **Fig.10**, the peak position of GFP_{UV} was drastically changed upon denaturation to 1.52 mL (corresponding to ~176 kDa), indicating the conversion from the native compact structure to the unfolded state. In contrast, the peal position of FATT remained unchanged (**Fig.10**). These results

strongly support the idea that FATT is in fact a natively disordered protein.

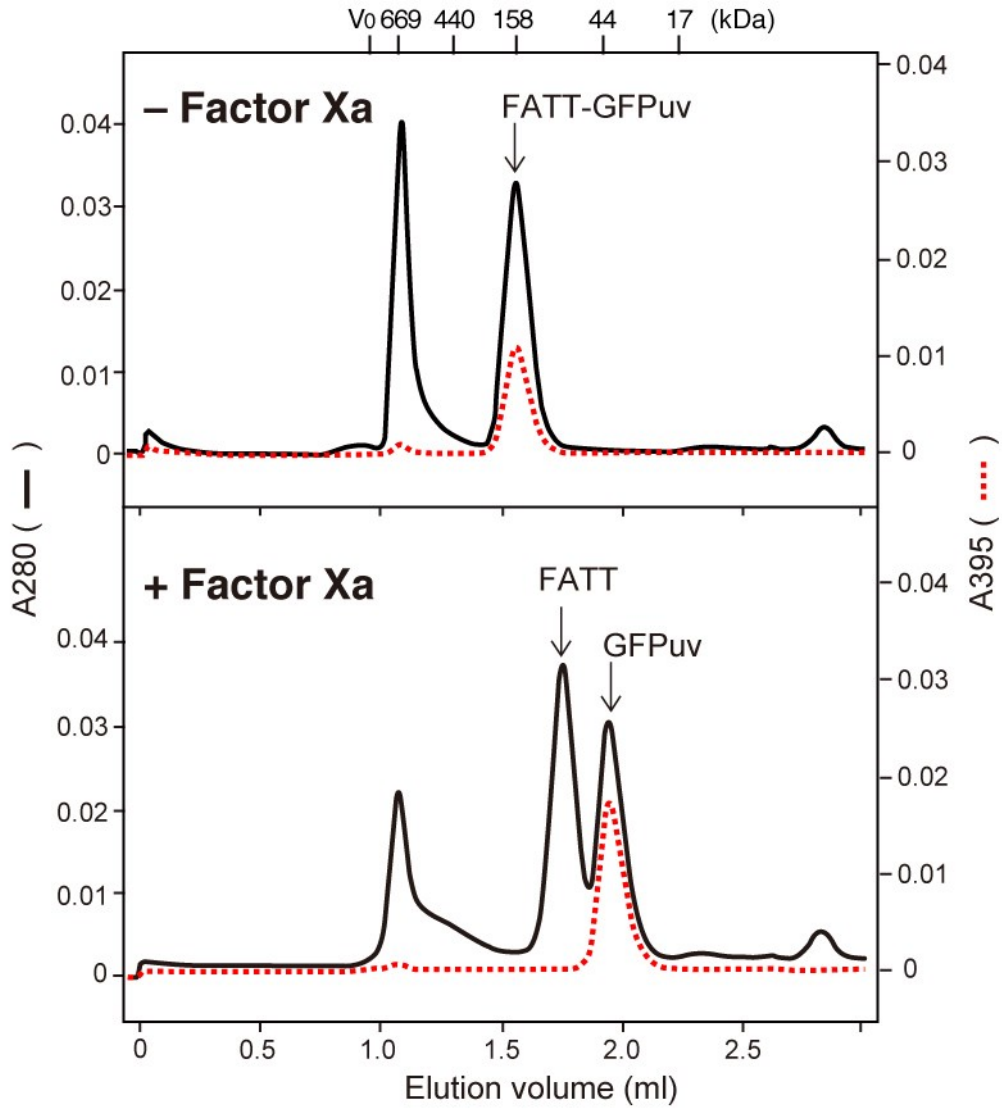

Figure 9. Solution behavior of FATT. Intact (upper panel) or factor Xa-treated (lower panel) FATT-GFP_{UV} samples were subjected to size exclusion chromatography on Superdex200. Elution positions for molecular weight standards including thyroglobulin (669 kDa), ferritin (440 kDa), aldolase (158 kDa), ovalbumin (44 kDa), and myoglobin (17 kDa) are indicated on the top.

(This figure was reproduced from Figure 2C in Sangawa *et. al*. 2013)

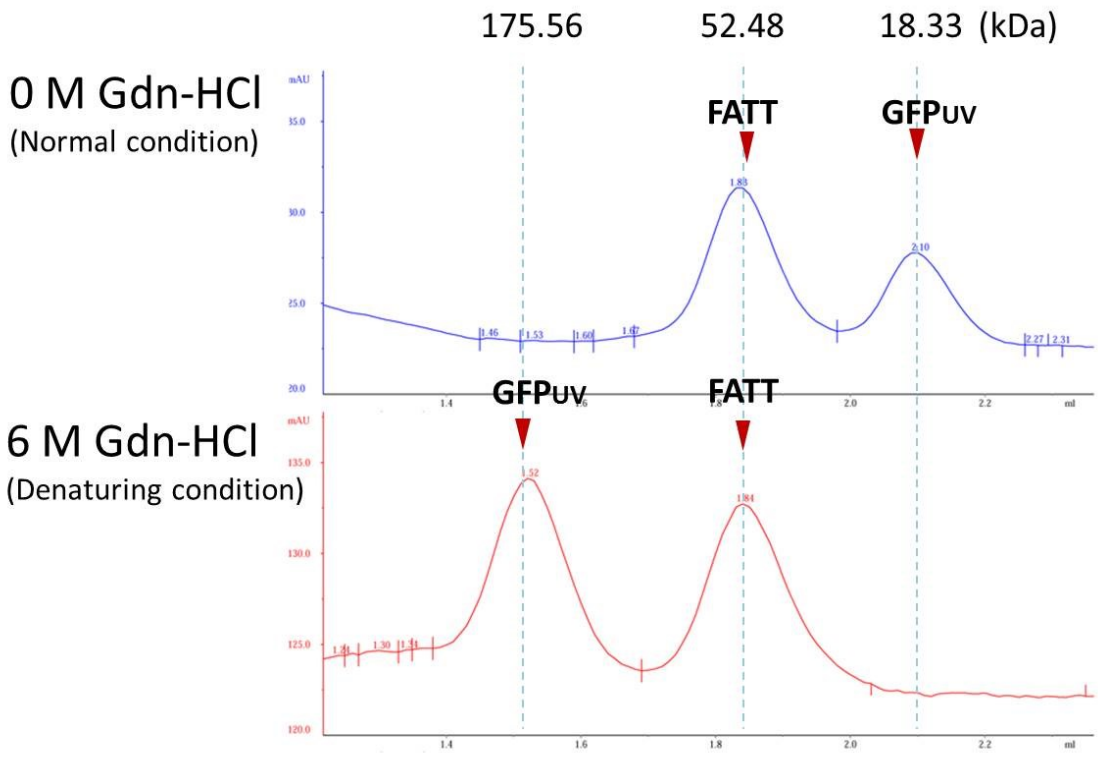

Figure 10. Analytical size exclusion chromatography of FATT and GFP_{UV} under absence (upper panel) and presence (lower panel) of 6 M Gdn-HCl. The vertical axis represents 280 nm adsorption (mAU), and the horizontal axis represents elution volume (mL). Molecular weights estimated from calibration curve were indicated above the chromatogram.

Soluble expression of FATT-fusion proteins

Extracellular proteins often contain disulfide bonds that complicate bacterial recombinant expression. I tried to enhance the solubility of those proteins using FATT-fusion strategy. Solubility of MeCPA (carboxypeptidase A derived from *Metarhizium anisopliae*), cutinase, and several single-chain Fv (scFv) antibody fragments with or without FATT-fusion were examined (**Fig.11**). Those proteins are aggregation-prone proteins that expressed in completely insoluble form in *E. coli* recombinant system without tag (**Fig.11, compare S and P under "No Tag"**). However, fusing FATT resulted in near-perfect soluble expression in all cases. Expression levels were not compromised by FATT-fusion.

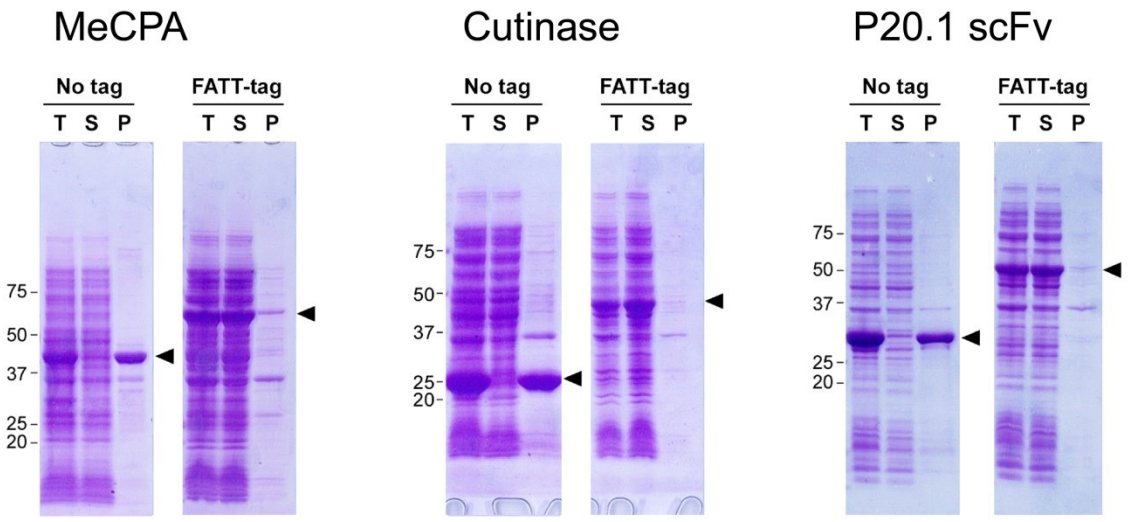

Figure 11. Switch from insoluble to soluble expression of three extracellular proteins. *E. coli* cells expressing indicated proteins with or without FATT tag were lysed by sonication to obtain total cell lysate (T), followed by centrifugation to separate soluble supernatant (S) and insoluble pellet (P) fractions. Proteins are separated on SDS-PAGE gels. Arrowheads point to the expressed proteins. Note that FATT-tagged protein bands are virtually absent from the insoluble pellet fraction.

(This figure was reproduced from Figure 3A in Sangawa *et. al*. 2013)

Functional refolding of extracellular proteins with disulfide bonds in a single protocol

Although several tags have been reported to enhance protein solubility, none of them were used as a "refolding enhancement" tag. For example, on-column refolding of polyhistidine-tagged proteins using metal chelete column was proven successful in numerous reports, but it is hard to expect that His-tag directly affects protein folding. It is also natural to assume that peptide tags such as FLAG-tag or His-tag are too short to change overall property of a protein. Similarly, large proteinacious tags such as GST or MBP, whose solubility is highly dependent on their higher-order structure, may not function as refolding enhancer because the denaturation step would destroy their structure to begin with, and attachment of such proteinacious tags would only increase the number of possible misfolded states, leading to the decreased overall refolding efficiency. In fact, there are only few examples of successful refolding with proteinacious tags (**Fig.12**). In this regard, FATT is sufficiently large but expected to have natively unfolded nature. Thus it is expected that the solubility-enhancement effect of FATT is retained during refolding conditions. So I tried to apply FATT for protein refolding. Three extracellular enzymes (MeCPA, human CPB, and cutinase) and three scFv fragments (P20.1, A5201A, and SG/19) were used as examples. Although these proteins were shown to be successfully expressed in the soluble fraction of the bacterial cytosol in the previous section, all of them have 2 disulfide bonds in their native structure that could not be formed in the reducing condition, and thus the expressed proteins were inactive. Therefore the refolding protocol coupled with oxidization of thiol groups was required.

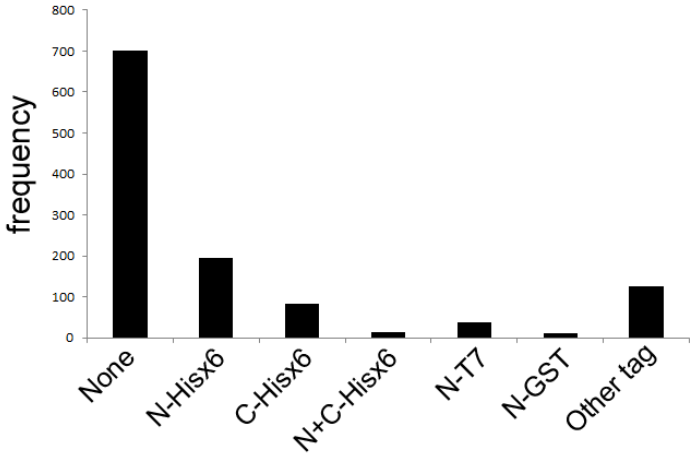

Figure 12. Usage of fusion tags in REFOLD database. The vertical axis represents frequency of each tag in 1165 trials in the database. N-terminal GST, which occupies <3% is almost only proteinacious tag used in refolding procedure.

To establish the generally-applicable protocol, I chose guanidium hydrochloride (Gdn-HCl) as a denaturant. Gdn-HCl is more favorable than another popular denaturant urea due to high denaturing ability (Rashid F *et. al.*, 2005). Additionally, urea often causes carbamoylation of proteins via production of cyanate which is unfavorable for protein folding (Qoronfleh MW *et. al.*, 2007). Dilution method was employed for reduction of denaturant concentration to keep protein concentration low during refolding procedure. Oxidized glutathione was used for oxidizing agent. L-arginine, the most frequently used refolding additive was chosen. L-arginine is a guanidium-group containing compound which is known as a weak protein denaturant. It is postulated that addition of L-arginine at certain concentration destabilizes the semi-stable misfolded state, but relatively inert for stable native structure, increasing the population of correctly-folded proteins (Arakawa T *et. al.*, 2007).

Generally, increasing volume of refolding mixture is troublesome because it

necessitates sample concentration step at later stage, which could be difficult depending on the nature of target protein. However, concentration of diluted FATT-fusion proteins can be easily accomplished by capturing onto anion exchange resin Q-Sepharose FF even from a very large volume of solution, which also serves as an extra purification step. Thus, I established the protocol described in **Fig.13** as a generally applicable refolding method for FATT-fusion proteins.

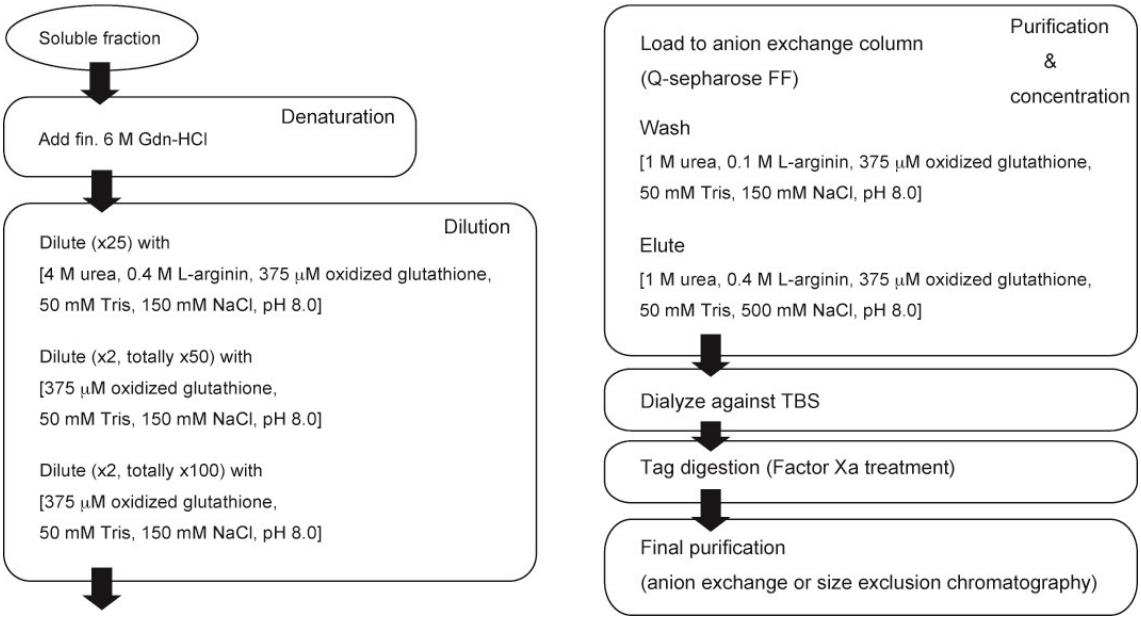

Figure 13. Scheme for the refolding procedure established in this research and subsequent isolation protocol. The total volume of the sample at each step is indicated in parenthesis, assuming the starting volume of bacterial lysate as 10 mL. For detail, see Experimental procedure.

(This figure was reproduced from Figure 4A in Sangawa *et. al*. 2013)

After the anion exchange purification, purified and ~ 50-fold concentrated FATT-fusion protein was obtained in the elution fraction containing 1 M NaCl. The fraction was then dialyzed against TBS to completely remove denaturants and refolding additives. As in the case of GFP_{UV}, FATT portion was completely cleaved from refolded proteins with factor Xa treatment for 16h at 4°C. No or only a few precipitation was observed after tag digestion. Tag and remaining FATT-fusion protein were easily removed by second anion exchange chromatography. **Fig.14** shows the SDS-PAGE analysis of the samples from each purification step.

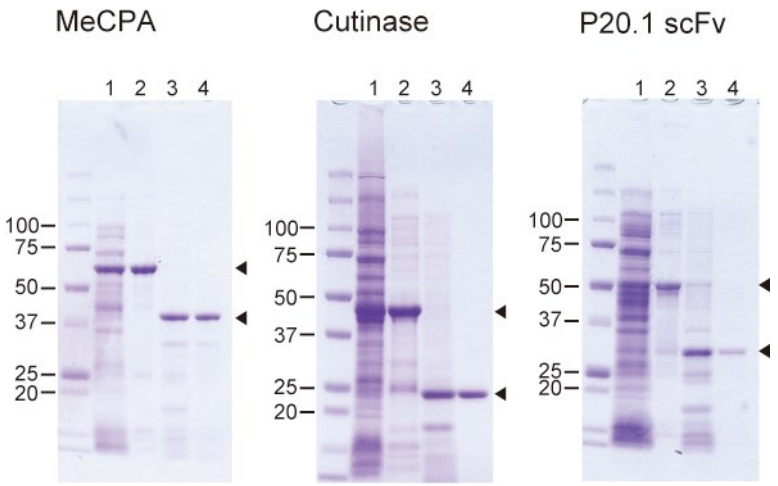

Figure 14. SDS-PAGE analysis of refolding process. Lane 1; soluble fraction, lane 2; elution fraction of first anion exchange purification, lane 3; after factor Xa treatment, lane 4; after second anion exchange chromatography. The target protein with and without FATT are indicated by black arrows.

(This figure was reproduced from Figure 4B in Sangawa *et. al*. 2013)

Monodisprese peaks were observed at appropriate elution volume in all tested cases. The peak fractions were collected and used for the following activity assays. Final yields of individual protein were shown in **Table 4**. At least several milligrams of well-behaving and homogeneous protein was obtained from 1 L culture.

Table 4. Final yield of extracellular proteins per 1 L cell culture. Each protein was refolded as illustrated in **Fig.13**, and then additionally purified by size exclusion chromatography to collect structurally homogeneous fraction. Then yields were measured using bicinchoninic acid protein assay (BCA method).

Protein	Yield (mg/L culture)
MeCPA	11.3
hCPB	4.2
Cutinase	11.3
P20.1	7.6
A5201F	5.1
SG/19	3.9

Activity of extracellular enzymes

MeCPA is an ortholog of the zinc metalloprotease carboxypeptidase A derived from Metarhizium anisopliae identified in 1994 (St. Leger RJ *et. al.*, 1994). It is known to exhibit relatively low specificity toward C-terminal amino acids of substrate proteins as compared to other carboxypeptidases (Austin BP *et. al.*, 2011). MeCPA was reported to cleave all amino acids except for basic residue (Arg, Lys) and a residue after proline (Pro-Xxx) making it valuable bioengineering tool as a trimming enzyme. However, recombinant production of this extracellular enzyme in bacteria has proven difficult. Even with an *E. coli* strain genetically manipulated to accommodate disulfide bond formation in the cytosol, the yield of the enzyme was low (0.5 mg/L culture) (Joshi L *et. al.*, 1999, Austin BP *et. al.*, 2012). Using the FATT-fusion strategy, I successfully obtained >10 mg of soluble MeCPA precursor (ProMeCPA) from 1 L of bacterial culture. ProMeCPA has 82-residue propeptide which inhibits enzyme activity. Although thermolysin was reported as an activating protease which cleaves the propeptide, it is a zinc-dependent metalloprotease as MeCPA that makes it difficult to inactivate selectively. Thus I used the serine protease chymotrypsin. With 1/83 molar ratio of chymotrypsin, propeptide was completely removed from proMeCPA after 1h incubation at room temperature.

In the case of human proCPB, the purified precursor was activated using trypsin as previously reported (Folk JE *et. al.*, 1961). Carboxypeptidase activity was determined by using synthetic peptide substrates conjugated with the N-(3-[2-furyl] acryloyl) (FA)-moiety at N-terminus, by measuring the decrease in the absorbance at 340 nm. As shown in **Fig.15A and B**, activated MeCPA and hCPB both exhibited concentration-dependent activity toward their respective substrate (FA-Phe-Phe for

MeCPA and FA-Ala-Arg for hCPB).

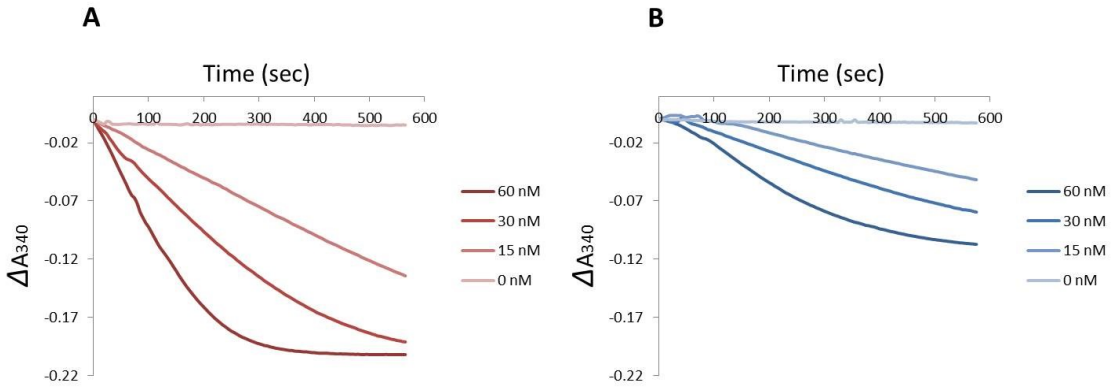

Figure 15. FAPP and FAAR assay for carboxypeptidases. **(A)** FAPP assay for MeCPA. **(B)** FAAR assay for hCPB. Enzyme was serially diluted into final concentration of 60-15 nM. Reduction of 340 nm absorbance by digestion of substrate dipeptide was recorded over time.

The activity of MeCPA was also assessed by its ability to remove C-terminal hexahistidine tag from a model substrate protein. Activated MeCPA was added to the C-terminally hexahistidine-tagged substrate protein (P4-FN10-His_6) at 1:20 to 1:2000 enzyme: substrate ratio (**Fig.16A**). SDS-PAGE analysis showed that the band corresponding to the P4-FN10-His_6 was gradually shifted to lower position as enzyme concentration increases. This shift was not observed in the presence of 1, 10-phenanthroline, a chelating agent that inhibit zinc-dependent catalytic activity. Moreover, the substrate protein became undetectable by anti-polyHis antibody in a Western blot after the treatment with MeCPA, indicating the successful cleavage of the C-terminal His-tag (**Fig.16B**).

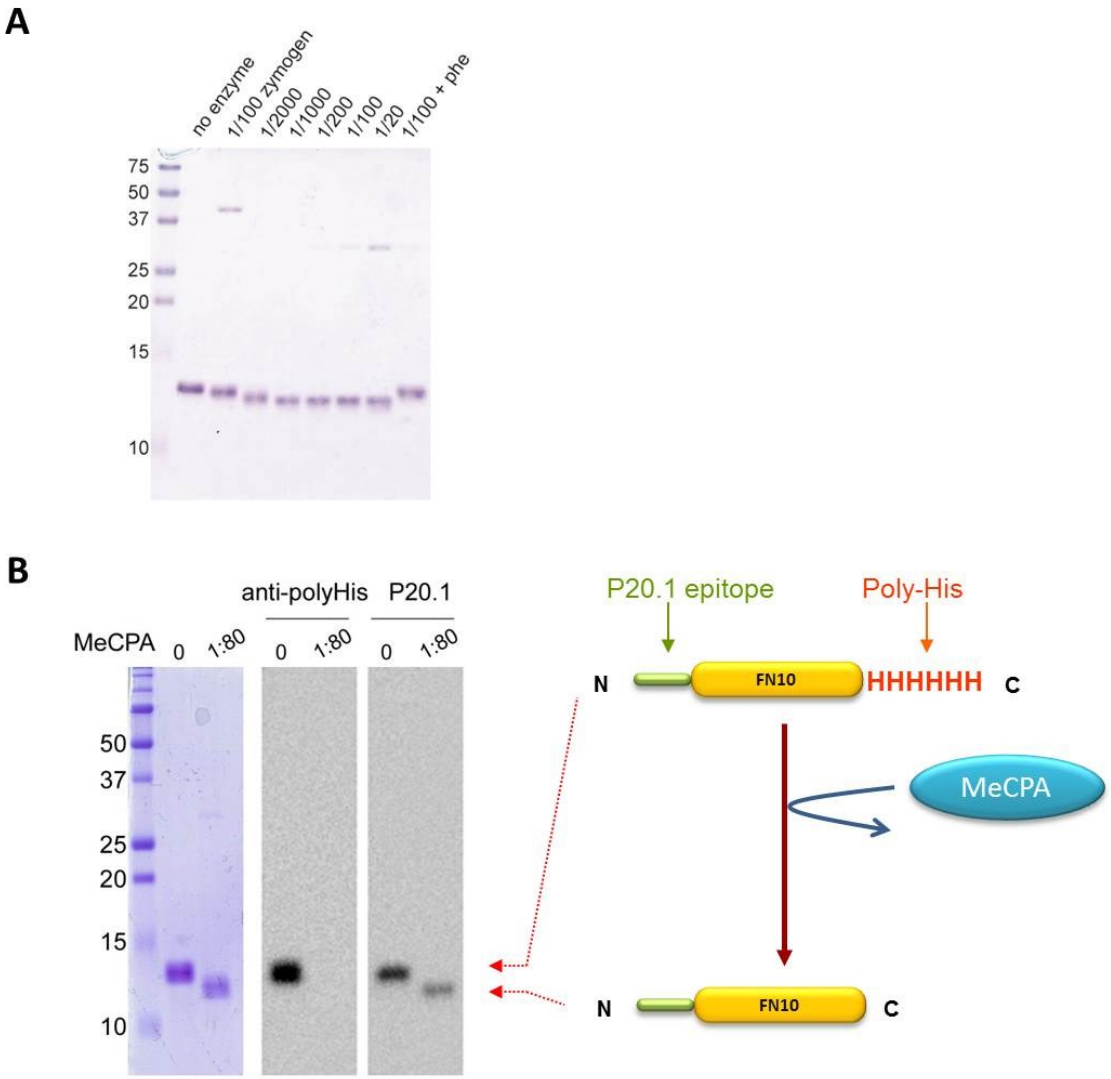

Figure 16. (A) The C-terminal peptide trimming activity of the activated MeCPA was assessed by a band-shift of the model substrate (P4-FN10-His_6) after incubating at varying enzyme/substrate ratio (wt/wt) in the absence or presence of 2.5 mM 1,10-phenanthroline (phe). **(B)** Digestion of His-tag was confirmed by detecting the substrate using anti-polyHis monoclonal antibody before and after MeCPA treatment. Anti-polyHis could not detect the substrate after treatment, that indicates lost of His-tag. Additionally, antibody against N-terminal P4-tag (P20.1) detected both bands before and after MeCPA treatment, suggests no cleavage was occurred to the N-terminus.

Cutinase, also known as cutin hydrolase is an enzyme derived from *Fusarium solani* that degrades cutin, the polyester compound found in higher plants. Cutinase is widely used in industrial application such as food production, cosmetics, and fine chemicals. The esterase activity was evaluated using the chromogenic substrate *p*-nitrophenyl butyrate. As shown in Fig.17, it was confirmed that the purified enzyme possessed high activity with a k_{cat}/K_m value of 1849 $mM^{-1}s^{-1}$, which was even higher than the value of 1159 $mM^{-1}s^{-1}$ obtained for the enzyme produced in bacterial periplasm in a native state (Liu Z *et. al.*, 2009).

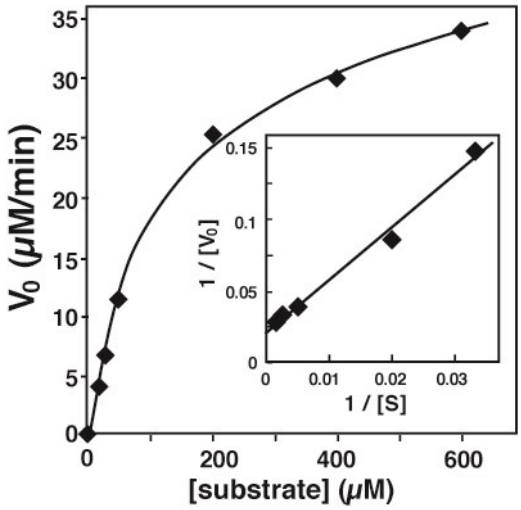

Figure 17. Enzymatic activity of refolded cutinase. Absorbance of the *p*-nitrophenol liberated from the substrate *p*NPB at 405 nm was recorded over time and the initial velocity was plotted against substrate concentration. K_M and V_{max} values were calculated from the double-reciprocal plot.

(This figure was reproduced from Figure 5B in Sangawa *et. al*. 2013)

Activity of single-chain antibody fragments

As an additional test to validate the FATT-fusion strategy in the production of functional extracellular proteins, we applied this method to several scFvs originated from different monoclonal antibodies including P20.1, A5201A, and SG/19. ScFv is an engineered antibody fragment where the variable regions of the light and heavy chain were linked with glycine-rich linker. In general, either periplasm in soluble expression or insoluble expression in inclusion bodies coupled with a denaturation/ refolding procedure are used for the recombinant production of scFv fragments. However, the optimum conditions for expression and purification differ considerably among different scFvs. Furthermore, it is often observed that a particular scFv will be resistant to conversion to a biologically active form even after applying sophisticated refolding strategies. When fused with the FATT, all three scFv fragments tested were successfully purified to homogeneity by using the protocol described in the previous section, and were found to actively recognize their respective antigens with good specificity. The P20.1 scFv showed specific binding toward the epitope peptide when analyzed by surface plasmon resonance, with a concentration dependency in almost identical manner to that obtained with the same scFv prepared from solubilized inclusion bodies by a conventional refolding method (**Fig.18**). The K_D value was 1.86 x 10^{-6} M, which was the same order as the scFv from inclusion body (1.49 x 10^{-6} M).

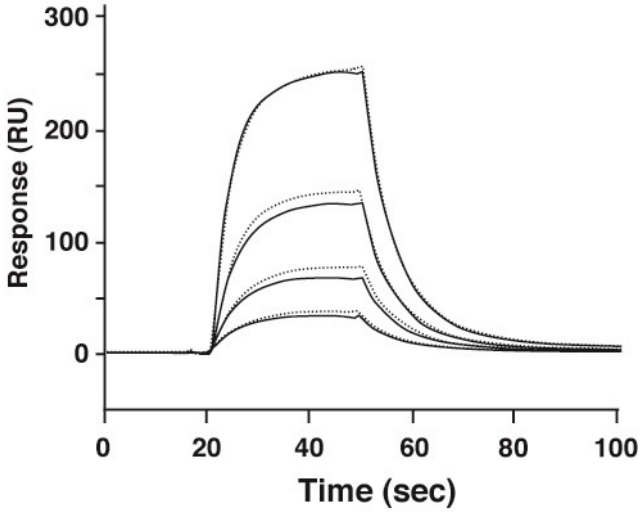

Figure 18. Binding kinetics of refolded P20.1 scFv toward antigen peptide assessed by surface plasmon resonance. Serially diluted scFv solutions (0.125, 0.25, 0.5, and 1 μg/ml) were flowed over sensor chip surface bearing "P4" antigen peptide (GYPGQV) for 30 sec. Dotted line represents the same scFv renatured from solubilized inclusion body.

(This figure was reproduced from Figure 5C in Sangawa *et. al*. 2013)

The A5201A and SG/19 scFvs prepared using the FATT fusion strategy could detect their native antigens in ELISA (**Fig.19 A and B**). Both scFvs were able to detect their specific antigen, nicastrin and integrin α6β1, respectively. The activity of SG/19 scFv was also demonstrated by indirect immunofluorescence microscopy (**Fig.19C**). As the refolding of the latter two scFvs from inclusion bodies was not possible due to a severe oligomerization and degradation tendency (data not shown), these results suggest that the FATT-fusion strategy may be applied to the preparation of scFv fragments in a case-independent manner.

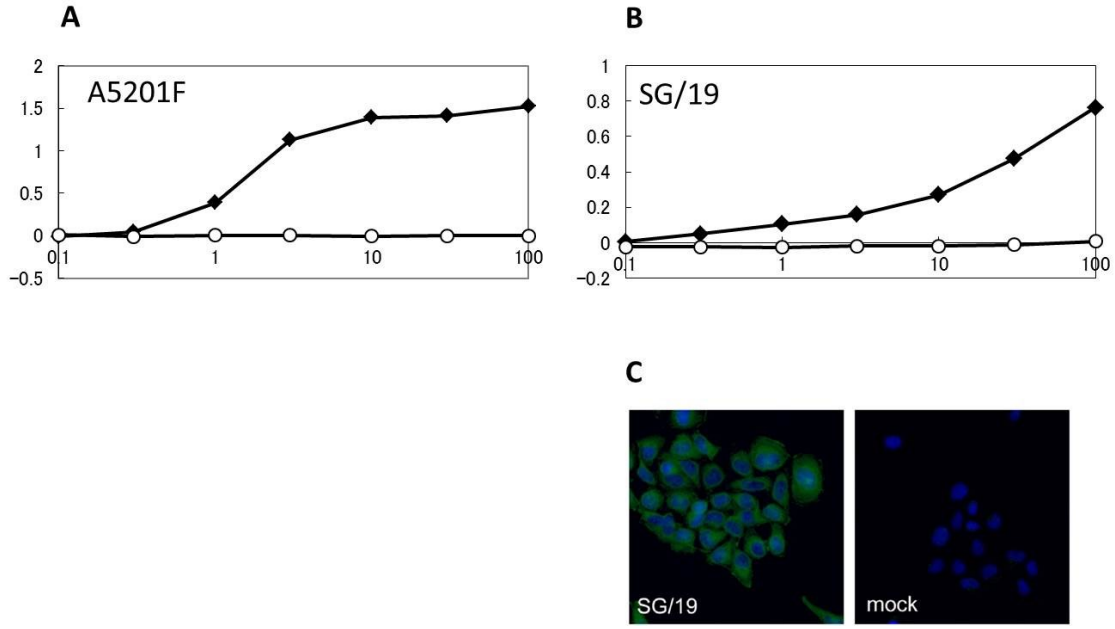

Figure 19. ELISA format binding assay for **(A)** anti-nicastrin A5201F and **(B)** anti-integrin β1 SG/19 scFv. C-terminally Myc-tagged scFvs were allowed to interact with antigen-coated wells and detected by anti-Myc antibody. A5201F and SG/19 are represented in (♦). P20.1 scFv (○) was used as a negative control in both experiments. **(C)** Indirect immunofluorescence staining of HeLa 3S cells. SG/19 scFv (right panel) or P20.1 scFv (control, left panel) were added to the cell to detect β1-integrins on cell surface. Bound scFvs were visualized by AlexaFluor 488-conjugated rabbit anti-Myc antibody (green). Nuclear staining by Hoechst 33342 is shown in blue.

(This figure was reproduced from Figure 5D and E in Sangawa *et. al*. 2013)

III. Discussion

FATT is intrinsically unstructured

In this study, I present strong evidence that FATT polypeptide behaves as an intrinsically disordered protein (IDP) in solution. FATT showed a large hydrodynamic radius in size exclusion chromatography, and the peak position did not change upon denaturation by the addition of 6 M Gdn-HCl, indicating that FATT is already unstructured under physiological buffer condition. I do not think that the high apparent molecular size of the FATT tag shown by the size exclusion chromatography is due to the oligomerization, because homooligomerization of FATT should be disfavored by its strong negatively charged character. Furthermore, FATT sequence is predicted to be disordered by multiple prediction programs. It is therefore likely that FATT polypeptide exists as a monomer with large Stokes radius in solution. Analytical ultracentrifugation experiment revealed that ectodomain fragment of APP containing the hyper acidic region (which corresponds to the major part of FATT) also had unusually large hydrodynamic radius, but behaved as a monomer (Nishikawa K, unpublished data). Therefore, I predict that the hyperacidic region of APP and FATT share the same unique character in that they are “intrinsically disordered” in a physiological buffer, and this may be critical for the ability of FATT to solubilize aggregation-prone proteins.

The mechanisms of the solubilization effect of intrinsically disordered proteins

It is previously reported that unstructured polypeptides inhibit protein aggregation in both fusion and non-fusion conditions. Santner *et. al*. reported that fusion of intrinsically disordered dehydrin family protein such as ERD10, ERD14 or Rab18 to various aggregation-prone proteins resulted in a strong enhancement of the solubility

when expressed in *E. coli* cytosol (Santner AA *et. al.*, 2012). Moreover, they reported that artificially designed polypeptides with a predicted unstructured nature also prevented protein aggregation (Santner AA *et. al.*, 2012). Addition of α-casein and α-synuclein, which are predicted to be disordered along their entire length, are reported to prevent heat- and reducing stress-induced protein aggregation (Bhattacharyya J *et. al.*, 1999, Rekas A *et. al.*, 2002). In these cases, it is not necessary to fuse the polypeptide to the target protein. Classical chaperones such as GroEL/ES and Hsp (Heat shock protein) family proteins are also known to contain unstructured region in many cases (Tompa P and Csermely P, 2004, Machida K *et. al.*, 2008). The predicted disordered regions typically span 25-50% of the entire chaperone sequence and are often required for the chaperone activity. p23, a co-chaperone of Hsp90, has C-terminal disordered region that is required for activity of Hsp90 but not to the Hsp90-p23 interaction (Weikl T *et. al.*, 1999). Hsp90 itself also contains disordered hinge region that is necessary for full chaperone function (Csermery P *et. al.*, 1998).

There are four possible mechanisms to explain how disordered polypeptides inhibit protein aggregation. The first one is that a large hydrophilic surface of disordered polypeptides compared to folded, globular proteins would increase the protein solubility. Generally, accessible surface areas of disordered proteins are 1.5-2 fold larger than ordered proteins because of the lack of well-packed protein core (Gunasekaran K *et. al.*, 2004). Moreover, Tompa (Tompa P, 2002) reported that average amino acid frequency of IDPs is highly biased compared to globular proteins. This analysis revealed that Glu, Lys and Pro are found much more frequently in IDPs than globular proteins, whereas contents of Asn, Cys, Leu, Phe, Trp and Tyr are significantly lower. Using the analytical result by Tompa et al, I calculated a “cumulative hydropathy index” for each group by

multiplying frequency of an amino acid occurrence by its hydropathy index (**Table 5**). By summing these values, I obtained the total hydrophobicity index of -94.8 for IDPs and -28.9 for globular proteins. This clearly indicates that sequences of IDPs are more hydrophilic than ordered proteins, and supports the idea that disordered polypeptides with large hydrophilic surface may enhance overall solubility when fused to the target proteins.

Table 5. Amino acid frequency and hydrophilicity of intrinsically unstructured and globular proteins. %frequency was referred from Tompa P, 2002. "Cumulative hydropathy index" of ordered and disordered proteins were calculated as follow; %frequency of each amino acid was multiplied by its hydropathy index. Then the products of all 20 amino acids were summed to compare the hydrophilicity of ordered and disordered proteins.

		IDPs		Globular proteins	
Amino acid	Hydropathy index	% frequency	Cumulative hydropathy index	% frequency	Cumulative hydropathy index
Ala (A)	1.8	7.15	12.87	8.15	14.67
Arg (R)	-4.5	4.21	-18.945	4.61	-20.745
Asn (N)	-3.5	2.06	-7.21	4.66	-16.31
Asp (D)	-3.5	5.05	-17.675	5.78	-20.23
Cys (C)	2.5	0.61	1.525	1.64	4.1
Gln (Q)	-3.5	4.46	-15.61	3.69	-12.915
Glu (E)	-3.5	14.26	-49.91	5.98	-20.93
Gly (G)	-0.4	4.31	-1.724	7.99	-3.196
His (H)	-3.2	1.51	-4.832	2.33	-7.456
Ile (I)	4.5	3.67	16.515	5.43	24.435
Leu (L)	3.8	5.44	20.672	8.37	31.806
Lys (K)	-3.9	10.43	-40.677	6.05	-23.595
Met (M)	1.9	1.3	2.47	2.03	3.857
Phe (F)	2.8	1.66	4.648	3.95	11.06
Pro (P)	-1.6	12.07	-19.312	4.61	-7.376
Ser (S)	-0.8	6.91	-5.528	6.31	-5.048
Thr (T)	-0.7	5.14	-3.598	6.15	-4.305
Trp (W)	-0.9	0.32	-0.288	1.55	-1.395
Tyr (Y)	-1.3	1.42	-1.846	3.64	-4.732
Val (V)	4.2	8.02	33.684	7	29.4
total			-94.771		-28.905

The second explanation is that a large exclusion volume of unstructured polypeptides may lower the chance of molecular collisions, leading to the suppression of aggregation. The exclusion effect of disordered polypeptide is reported by Brown HF *et. al.* in 1997 (Brown HF *et. al.,* 1997). By using atomic force microscopy (AFM), they observed that neurofilaments with protruding unstructured polypeptides are surrounded by areas with no grains, indicating that the polypeptide "sidearms" excluded nearby molecules with their thermodynamic movement. This situation may be analogous to the effect of polyethylene glycol (PEG), which has extended structure and no electrostatic charge, and is often used as a stabilizer for proteins. Protein modification with PEG (PEGylation) increase solubility, stability, and circulating lifetime, decrease immunogenicity and sometimes affect activity of target proteins, probably through the exclusion of both specific and non-specific interaction partners (Kochendoerfer G *et. al.*, 2003).

The third explanation is the simple electric repulsion. Although the frequencies for acidic and basic amino acids in globular proteins are almost balanced (Asp + Glu: 11.76%, Arg + Lys: 10.66%, Arg + Lys + His: 12.99%, respectively), IDPs are more acidic (Asp + Glu: 19.31%, Arg + Lys: 14.64%, Arg + Lys + His: 16.15%, respectively) (**Table 5**). Thus IDPs may have tendency to repel each other.

The fourth explanation is a "chaperone" activity. The term "chaperone" contains three enzymatic activity; foldase, holdase and unfoldase. Foldase refers to an enzymatic actively that helps protein folding in an ATP dependent manner. Holdase is similar to foldase, but does not help protein folding actively. Holdase isolate unfolded protein from environment to prevent unfavorable protein-protein interaction during folding procedure. The last one, unfoldase refers to an activity that captures misfolded proteins

and convert them into the unfolded state to make them redo the proper folding process. The persuasive “entropy transfer” model was suggested by Tompa P and Csermely P in 2004 to explain chaperone activity of disordered proteins (Tompa P and Csermely P, 2004). The model predicts that the main function of disordered proteins as chaperones is “unfoldase”. The scenario is as follows. First, disordered protein changes conformation rapidly in solution, enjoying the entropic advantage. Second, once it binds to a misfolded protein, it partially folds around the interface. Third, the loss of entropy because of this partial folding is compensated by local unfolding of a misfolded protein. Although experimental certification of the model is difficult, a loss of flexibility upon binding of misfolded substrate has been confirmed for the flexible apical domain of GroEL and the C-terminal tail of α-crystallin by NMR spectrometry (Carver JA *et. al.*, 1992, Lindner RA *et. al.*, 1998).

In case of FATT-assisted solubilization and refolding

In case of FATT, the first scenario may apply because it contains only 21.3% of hydrophobic residues (Ala, Leu, Ile, Val, Met, Phe and Cys, which has plus value of hydropathy index). It is relatively low compared to the mean value of globular proteins (36.6%) and even lower than the average of IDPs (27.9%). Thus FATT has much larger hydrophilic surface than a globular protein or IDP with the same molecular size.

The second explanation, entropic exclusion, is also likely to play a role, considering the large hydrodynamic radius of FATT moiety. In fact, 14 kDa FATT shows Stokes radius corresponding to a 80 kDa globular protein in size exclusion chromatography.

The third explanation is clearly playing a role, because FATT tag has total charge of 49.0 at pH 7.0.

The last scenario (unfoldase activity with entropy transfer) may not operate in the case of FATT-mediated solubilization from the following reason. To achieve entropy mediated protein unfolding, FATT must primarily interact with a misfolded protein. This interaction must mainly be driven by hydrophobic interactions. However, FATT has poor hydrophobic amino acids contents. Additionally, acidic nature of FATT is unsuitable for interaction with partially-misfolded proteins because protein surfaces are generally tend to charge negatively.

Hypothetical view of FATT-assisted solubilization and refolding -the inverse "fly-casting" mechanism-

The "fly-casting" mechanism was proposed by Huang Y and Liu Z in 2009 to explain how IDPs can efficiently mediate protein-protein interaction even though they do not assume well-ordered structure (Huang Y and Liu Z, 2009). The hypothesis is that IDPs have greater on-rate for an interaction with their target molecules (on average ~2.9 fold higher than ordered proteins) because of a greater effective capture radius. This mechanism is reasonable when there are favorable interactions between disordered protein and its ligand. Conversely, when there is a repulsive force rather than favorable interaction between two molecules, disordered proteins are expected to repel other molecules more efficiently, because of the larger repulsive radius.

Here I suggest a hypothetical mechanism of FATT-assisted solubilization and refolding. FATT-fusion proteins repulse each other because of a large (-49.0) net negative charge. This electric repulsion can extend further in the distance compared to ordered fusion partners according to the "fly-casting" mechanism. During the refolding process, this repulsive volume isolates the folding intermediates from other molecules, preventing

the premature aggregate formation and facilitating the correct folding (Fig.20).

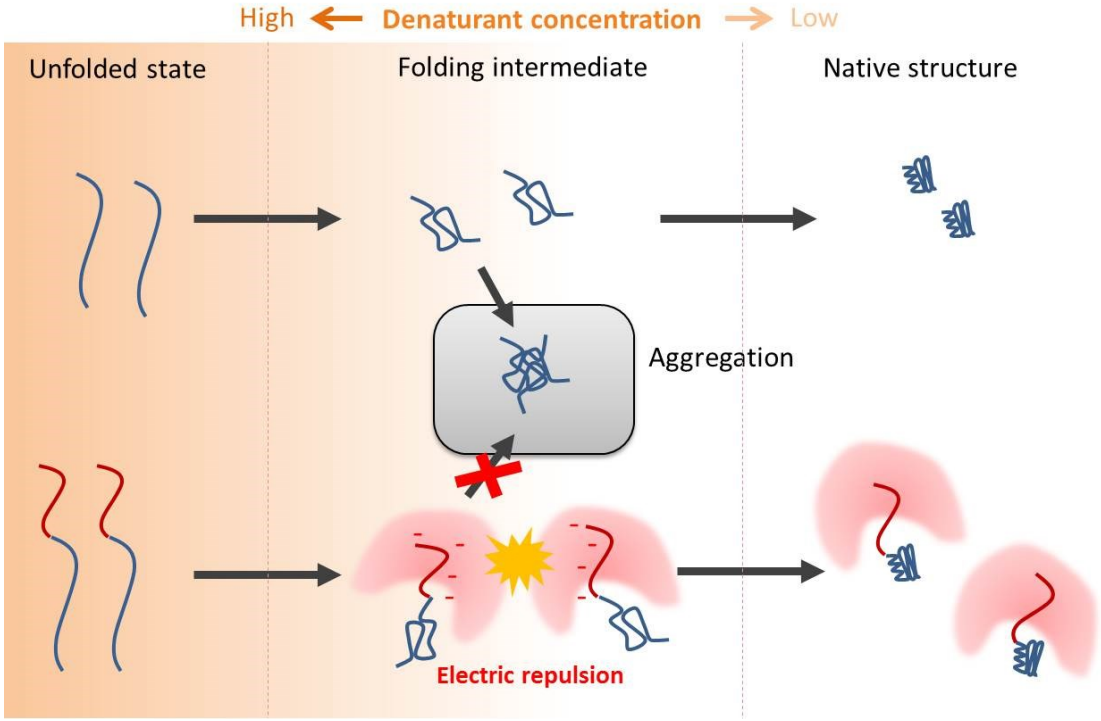

Figure 20. Hypothetical view of the effect of FATT on protein refolding. A protein with (lower) and without (upper) FATT are illustrated. Since FATT has large negative charge and dynamic radius, aggregation between FATT-fusion proteins are suppressed because of wide-range electric repulsion results in enhancement of proper folding.

IV. Conclusion

In this study, I showed that fusion of hyper-acidic FATT greatly changed electrostatic property of proteins and thus enables affinity-like purification using low cost anion-exchange chromatography. This aspect is also usable on the complete separation of digested tag and target protein moiety. Thus FATT-fusion strategy was established as a cost-effective and widely applicable purification method.

Furthermore, I found that FATT robustly improved solubility of various aggregation-prone proteins upon heterologous overexpression in *E. coli*. The solubility-enhancement effect was retained during denaturing / refolding procedure and it realized a generalized oxidative refolding protocol. As it was indicated that intrinsically disordered FATT has large thermodynamic radius in solution phase, I concluded that this solubility-enhancement effect is generated by “shielding” proteins from unfavorable interaction that lead them aggregate.

In summary, FATT was established as a fusion tag which overcomes the three major bottlenecks in *E. coli* heterologous expression systems; (1) trade-off relationship between cost, purity and yield in purification procedure, (2) insoluble expression, (3) lack of generalized method for functional refolding. Therefore FATT may be (and according to my hypothesis, other IDPs are also) used as a reasonable primary-choice fusion tag in bacterial expression.

V. Perspective

Modification of FATT and other unstructured tags

Since the beneficial property of FATT tag reported here is unlikely to be sequence-specific, it should be possible to re-design the tag by incorporating additional functional motifs or shuffling the amino acid sequence to obtain custom-made FATT tag with improved properties. The fusion topology can also be modified, because we confirmed that C-terminally FATT-tagged proteins can be expressed, purified, and refolded as efficient as the N-terminally tagged versions.

Here I illustrate some perspective in further modifications of FATT and other unstructured tags. First, extension of disordered and negatively-charged sequence possibly improves property of FATT with extended Stokes radius and repulsive force. Second, as it is expected that solubilization effect of FATT is not depend on chaperone (unfoldase) activity, fusion of chaperone-like unstructured polypeptides such as α-synuclein to FATT may results in further enhancement of solubility. It is also favorable for extending polypeptide length of unstructured region. Third, according to the “fly-casting” model, Fusion of affinity tags such as poly-histidine to the terminus of FATT may enhance the binding rate of the affinity tags to their specific partners with a large capturing radius.

Further usage of FATT

Another intriguing possibility is the potential use of FATT tag as a way to improve/modify the bioactivity of recombinant proteins. The biophysical property of the FATT tag suggests that its fusion may have similar effect as chemical modification by polyethylene glycol, which often improves in vivo stability and toxicity of the target

protein. Combined with the fact that the tag sequence is derived from abundant human protein and hence minimally immunogenic, this unique fusion strategy may find various useful applications in biotechnology field not limited to the one reported here.

VI. Materials and methods

Construction of FATT-tagged proteins

Residues 190-285 of APP695 gene (a gift from K. Yoshikawa) was amplified by polymerase chain reaction (PCR). Using extension PCR, genes encoding *NdeI* cleavage site and FLAG-tag were added to the 5' side of amplified APP fragment. TARGET-tag, factor Xa cleavage site and *BamHI* recognition site were added to the 3' side. This fragment (*NdeI*-FLAG-acidic region-TARGET-factor Xa-*BamHI*) was used as an insert after treatment of *NdeI* and *BamHI*.

pET11b (Novagen) was used as a vector. The vector was treated by *NdeI* and *BamHI*. The cleavage product was purified by cutting out from 0.8 % agarose gel electrophoresis and dephosphorylated by adding Calf Intestine Alkaline Phosphatase (CIAP, TOYOBO), followed by ligation using Ligation High ver.2 (TOYOBO).

The production was named "pFATT3" and used as a general vector for FATT-fusion constructs.

The following segments were used for the expression of each target protein; GFP_{UV} (Clontech), residues 1-238; human Dkk1_C (a gift from S. Sokol), residues 178-246; human ApoER2 LA1 (a gift from T. Yamamoto), residues 42-83; MeCPA (a gift from D. Waugh), residues 17-418; cutinase, residues 32-230. All scFv constructs contained V_H and V_L regions of respective monoclonal antibody intervened by a $(GGGGS)_3$ linker and followed by a hexahistidine tag and a Myc tag.

Those genes were amplified by PCR and used as an insert after digestion by appropriate restriction enzymes. Every inserts were ligated to the downstream of factor Xa recognition site of pFATT3. All constructs used in this study were shown in **Table 6** (at the end of this paper).

Expression of FATT-tagged proteins

Expression plasmids of FATT-fusion proteins (Target gene / pFATT3) were transformed into *E. coli* strain BL21 (DE3) and cultured overnight on a Lysogeny-Broth (LB) -agarose plate. A colony was picked and precultured overnight at 37˚C in liquid LB medium containing 50 μg/mL ampicillin. Then it was diluted into 1 L of fresh LB medium and was induced with 1 mM IPTG when OD_{600} was reached to 0.5, followed by additional 4 h incubation. Cultural mediums at the point before induction and harvest were sampled and centrifuged to collect bacterial pellet. Cells were then lysed with Bugbuster (Novagen). Expression levels of FATT-tagged and no-tagged proteins were compared by SDS-PAGE analysis. Sample volumes were adjusted in reference to OD_{600} so that concentration of contaminated proteins in each sample becomes approximately equal.

Western blot analysis

Three FATT-fusion proteins (GFPuv, Dkk1_C, and LA1) were expressed and cultural mediums at 4 h after induction were sampled and separated by SDS-PAGE using 5-20% gradient gel as described above. The gel was then blotted on a PVDF membrane. The membrane was blocked with Blocking-One (Nacalai tesque). Anti-FLAG polyclonal antibody (SIGMA) or P20.1 IgG diluted with Blocking-One was used as a primary antibody. The membrane was washed 3 times with TBS-1% BSA. Peroxidase-conjugated anti-rabbit or anti-mouse Ig was employed as a secondary antibody. Each antibody reaction was performed 1h at room temperature. Blots were visualized by ECL Plus reagent (GE-Healthcare) and recorded in ImageQuant LAS4000

mini system (GE healthcare).

Cell harvest and lysis

E. coli cells expressing FATT-fusion protein were harvested after culture as described below. Cultural medium was centrifuged at 5,000 g for 20 min. Cells were then resuspended in TBS (50 mM Tris, 150 mM NaCl) pH 8.0. 10 μM leupeptin, 1 μM pepstatin A and 0.25 mM phenylmethanesulfonyl fluoride (PMSF) were added to prevent unfavorable proteolysis. Cells were lysed using ultrasonic disruptor UD-201 (TOMY) and centrifuged at 20,000 g to remove cell debris. Benzonase (Novagen) was added to the supernatant and incubated for 30 min at room temperature to digest genomic DNA.

Anion exchange chromatography

Anion exchange purification in prepacked HiTrap Q HP (GE healthcare) column and open column Q-sepharose Fast Flow (GE healthcare) were individually tested.

Purification using prepacked column was performed as below. The soluble fraction of bacterial lysate was directly loaded onto HiTrap Q HP column at a flow rate of 1 mL/min. 50 mM tris, pH 8.0 was chosen as a primary buffer condition. Then NaCl concentration was gradiently increased from 0 mM to 1 M. The elution was fractionated every 1 mL.

Open column was used as below. Soluble fraction was loaded directly to the Q-sepharose FF column equilibrated with TBS pH 8.0, followed by 5 column volume of wash buffer (50 mM Tris, 400 mM NaCl, pH 8.0) and the same volume of elution buffer (50 mM Tris, 1 M NaCl, pH 8.0). The elution fraction was dialyzed against TBS pH 8.0

to reduce NaCl concentration. Then 2 U/mL of factor Xa (Novagen) and 2.5 mM $CaCl_2$ was added to cleave FATT. The reaction was performed at 4ºC. After 16 h incubation, the reaction was terminated by PMSF.

Analytical size exclusion chromatography

Size exclusion chromatography was performed using Superdex^{TM} 200 5/150GL column (GE healthcare) equilibrated with TBS pH 8.0 at a flow rate of 0.15 mL/min.

Solubility assessment of FATT-tagged proteins

The gene fragments coding aggregation-prone protein were inserted to pFATT3, transformed into BL21 (DE3), cultured and lysed as described above. The cell lysate was centrifuged for 10 min at 20,000 g to separate supernatant and precipitation. Supernatant was sampled as a soluble fraction. Precipitation was sampled as an insoluble fraction after re-solubilization by small aliquot of 8 M urea followed by dilution with TBS.

Oxidative refolding of FATT-tagged extracellular proteins

For the production of correctly folded proteins with disulfide bonds, various proteins tagged with FATT that had been expressed in the cytosol of bacteria were subjected to the "direct refolding" method coupled with one-step concentration by anion exchange resin. Guanidium hydrochloride (Gdn-HCl) was directly added to the soluble fraction so that the final concentration becomes 6 M. After 30 min incubation at room temperature and filtration (0.45 μm), it was diluted 25-fold with refolding buffer A, containing 4 M urea, 0.4 M L- arginine hydrochloride (L-Arg) and 375 μM oxidized glutathione

(GSSG). Then diluted solution was gently stirred for 4 h at 4°C, followed by 2-fold dilution with refolding buffer B (0.4 M L-Arg, 375 μM GSSG, 50 mM Tris pH 8.0) and incubated overnight at 4°C. Additional 2-fold dilution with refolding buffer B was performed to finally achieve 100-fold dilution.

Concentration and purification using anion exchange column

Renatured FATT-fusion protein was able to purify and concentrate in a single-step using batchwise capture by Q-sepharose FF (GE healthcare). The resin was added to the diluted sample solution and stirred gently for 1 h at 4°C. Then the solution was passed through empty column to collect resin. The column was washed 5 column volume wash buffer (50 mM Tris, 150 mM NaCl, 1 M urea, 0.1 M L-arginine-HCl, 375 μM GSSG, pH 8.0), followed by elution with the same volume of elution buffer (wash buffer containing 1 M NaCl). Buffer of elution fraction was exchanged by dialyzing against TBS pH 8.0.

Tag removal and final purification

FATT was digested by factor Xa treatment and removed by passing through anion exchange column as described above. To obtain structurally homologous protein, size-exclusion chromatography was performed. HiLoad™ 16/60 Superdex™200 prep grade (GE healthcare) equilibrated with TBS pH 8.0 was used for large-scale purification. The final yields of refolded proteins were quantified by Bicinchoninic acid (BCA) assay. Bovine serum albumin was used as a standard.

Assessment of Carboxypeptidase activities

Carboxypeptidase activities of *Metarhizium* carboxypeptidase A (MeCPA) and human Carboxypeptidase B (hCPB) were assessed using FAPP and FAAR assay, respectively (**Fig.21**) (Peterson LM *et. al.*, 1982). Because MeCPA was purified as an inactive proenzyme, propeptide of MeCPA was removed by chymotrypsin at 1/83 molar ratio. After 1 h incubation at room temperature, reaction was stopped by adding PMSF.

The enzyme solution was then diluted with 50 mM tris, 75 mM NaCl, pH 7.5 into various concentrations. The solution was mixed with 200 μM of low molecular weight substrate *N*-(3-[2-furyl] acryloyl)-Phe-Phe (FAPP, Bachem) in a quartz cuvette. Diminishment of 340 nm absorbance was measured over time. Digestion of C-terminal peptide tag was also tested. Human Fibronectin 10^{th} type-III repeat (residue 1447-1540) with C-terminal additional linker and hexahistidine-tag sequence (GSHHHHHH, FN10-His_6) was used as a substrate. MeCPA was activated as describe above, and was added to FN10-His_6 solution in various molar ratios. The reaction mixture was incubated 4 h at room temperature. A transition of polypeptide length was detected by SDS-PAGE using 15% acrylamide gel in reducing condition. Human CPB, which also has propeptide, was activated with trypsin in 1/100 molar ratio. After 30 min incubation at room temperature, PMSF was added to terminate the reaction. Enzymatic activity was measured as described above. *N*-(3-[2-furyl] acryloyl)-Ala-Arg (FAAR, Bachem) was used as a substrate instead of FAPP. Enzyme was used in 30 nM concentration.

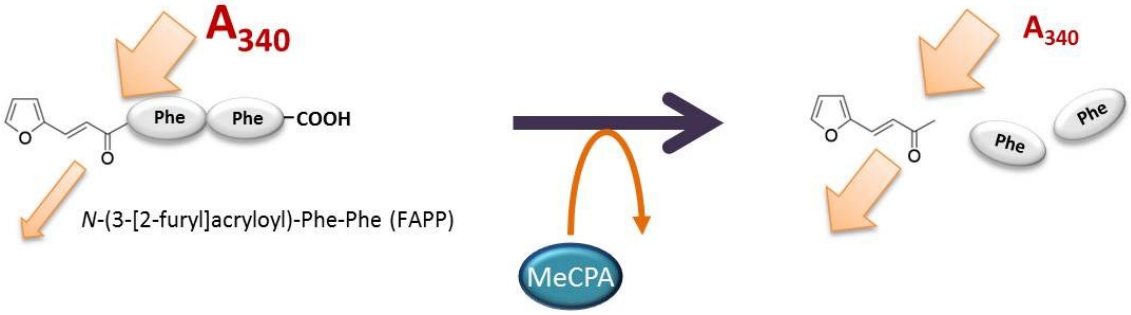

Figure 21. The schematic principle of FAPP assay was illustrated. FA-linked dipeptide FAPP (or FAAR) has absorbance at 340 nm. After digestion of amino acids by carboxypeptidase, A_{340} is dramatically diminished.

Assessment of esterase activity

The esterase activity of cutinase was measured as described below. Various concentrations of *p*-nitrophenyl butyrate (*p*NPB) were added to 2 nM cutinase diluted with 20 mM tris pH 8.0 in a quartz cuvette. 400 nm absorbance over time was measured. 15400 $M^{-1}cm^{-1}$ was used as a molar attenuation coefficient of *p*NPB at 400 nm (Hosie L *et. al.*, 1987).

Assessment of antibody activities

Antibody activity of P20.1 scFv was measured by Surface Plasmon Resonance method using BIACORE 2000 instrument. Fibronectin 9^{th}-10^{th} repeat with N-terminal epitope sequence (MPRGYPGQV) and C-terminal additional free cysteine (P4-FN9-10-Cys) was used as a ligand. P4-FN9-10-Cys was biotinylated using EZ-Link™ PEO-Maleimide Activated Biotin (PIERCE). Then it was immobilized approximately

2000 RU on a Sensor chip SA (GE healthcare). Various concentration of scFv was injected to the flow cell using TBST (50 mM tris, 150 mM NaCl, 0.01% Tween-20, pH 8.0) at 25°C and flow rate of 20 μl /min. The sensorgrams were fitted using Langmuir-binding model to determine the dissociation constant. Activities of anti-nicastrin antibody A5201F scFv and anti-integrin β1 antibody SG/19 scFv were measured using enzyme-linked immunosorbent assay (ELISA). Nicastrin ectodomain or integrin α6β1 was coated on a 96-well microtiter plate. The wells were blocked with 1% bovine serum albumin (BSA) in TBS for 1 h at room temperature. Then various concentrations of scFvs were added in the volume of 50 μl. P20.1 scFv was used as a negative control. After 1 h incubation at room temperature, the wells were washed three times with TBS. Then anti-Myc polyclonal antibody (MBL) was added in the volume of 50 μL and incubated 1 h at room temperature. After 3 times wash with TBS, antibodies were probed using anti-rabbit Ig-HRP conjugate (SIGMA) and visualized by ABTS Peroxidase Substrate (KPL). Activity of SG/19 scFv was evaluated by an indirect immunofluorescence microscopy of β1-expressing Hela3S cells. Briefly, HelaS3 cells were seeded onto 18-mm glass coverslips coated with poly-L lysine. The cells were fixed in phosphate-buffered saline (PBS) containing 4% formaldehyde for 15 min at room temperature. After blocking with 1% BSA in PBS, they were then incubated with 2 μg/ml SG/19 scFv or control P20.1 scFv at room temperature for 1h. After washing three times with PBS, cells were incubated with 1μg/ml Alexa 488-conjugated anti-Myc IgG (MBL) and Hoechst33342 (Invitorogen) for 2 hours at room temperature. The coverslips were mounted using Vectashield (Vector Laboratories) and the fluorescence images were obtained using a digital fluorescence microscope (model BZ-9000; Keyence, Osaka, Japan).

Table 6. FATT-fusion proteins used in this study. Markers represent as below. Red; FATT, light green; tags other than FATT, gray; propeptide. Magenta in FATT-P20.1 represents mutation in TARGET-tag to avoid self-recognition of renatured scFv.

FATT-GFP$_{UV}$

MDYKDDDDVE AEESDNVDSA DAEEDDSDVW WGGADTDYAD GSEDKVVEVA EEEEVAEVEE
EEADDDEDDE DGDEVEEEAE EPYEEATERT TSIATTTTTT TESVEEVYPG QVGYPGQVGY
PGQVIEGRGI PSAMAAAMSK GEELFTGVVP ILVELDGDVN GHKFSVSGEG EGDATYGKLT
LKFICTTGKL PVPWPTLVTT FSYGVQCFSR YPDHMKRHDF FKSAMPEGYV QERTISFKDD
GNYKTRAEVK FEGDTLVNRI ELKVLILRRW KHSRTQTRVQ LYSHNVYITA DKQKNGIKAN
FKIRHNIEDG SVQLADHYQQ NTPIGDGPVL LPDNHYLSTQ SALSKDPNEK RDHMVLLEFV
TAAGITHGMD ELYKTGAAAH HHHHHGAAEQ KLISEEDLNG AA

FATT-Dkk1_C

MDYKDDDDVE AEESDNVDSA DAEEDDSDVW WGGADTDYAD GSEDKVVEVA EEEEVAEVEE
EEADDDEDDE DGDEVEEEAE EPYEEATERT TSIATTTTTT TESVEEVYPG QVGYPGQVGY
PGQVIEGRGI PMEMYHTKGQ EGSVCLRSSD CASGLCCARH FWSKICKPVL KEGQVCTKHR
RKGSHGLEIF QRCYCGEGLS CRIQKDHHQA SNSSRLHTCQ RH

FATT-LA1

MDYKDDDDVE AEESDNVDSA DAEEDDSDVW WGGADTDYAD GSEDKVVEVA EEEEVAEVEE
EEADDDEDDE DGDEVEEEAE EPYEEATERT TSIATTTTTT TESVEEVYPG QVGYPGQVGY
PGQVENLYFQ GSAMAAAGPA KECEKDQFQC RNERCIPSVW RCDEDDDCLD HSDEDDCPK

FATT-MeCPA

MDYKDDDDVE AEESDNVDSA DAEEDDSDVW WGGADTDYAD GSEDKVVEVA EEEEVAEVEE
EEADDDEDDE DGDEVEEEAE EPYEEATERT TSIATTTTTT TESVEEVYPG QVGYPGQVGY
PGQVIEGRGI PMAAESPVSY DGYKVFRVPV VDDGTHIQSL IDHLNLNVWQ PPSKKGAFAD
IQVAPSQLAA FENAMKGRSF EIMHEDLGDS IAREGTIQAY AAGSANASWF TSYHPYNDHL
QWMKDIASQY PSNVKSVTSG TTGDGNTITG LHIFGSSGGG NKPAVVFHGT VHAREWIVAM
TLEYITNELL AKYATDSAVK AVVDKYDFYM FPIVNVDGFK YTQSSDRMWR KNRSRNQGSS
CLGTDPNRNW PYKWDGPGSS TNPCTETYRG ASAGNSPEVK SYIAFLDKIK KSQGVKLYID
WHSYSQLFMT PYGYSCSART PNNAALQALA KGASDAMRSV HGTTFAYGPV CNVIYQVAGG
SIDWVQDVLK ADNVFTIELR DKGRYGFVLP PDQIIPSGEE SFAGAMHLFQ QMSSSRV

FATT-hCPB

MDYKDDDDVE AEESDNVDSA DAEEDDSDVW WGGADTDYAD GSEDKVVEVA EEEEVAEVEE
EEADDDEDDE DGDEVEEEAE EPYEEATERT TSIATTTTTT TESVEEVYPG QVGYPGQVGY
PGQVIEGRGI PHHGGEHFEG EKVFRVNVED ENHINIIREL ASTTQIDFWK PDSVTQIKPH
STVDFRVKAE DTVTVENVLK QNELQYKVLI SNLRNVVEAQ FDSRVRATGH SYEKYNKWET
IEAWTQQVAT ENPALISRSV IGTTFEGRAI YLLKVGKAGQ NKPAIFMDCG FHAREWISPA
FCQWFVREAV RTYGREIQVT ELLDKLDFYV LPVLNIDGYI YTWTKSRFWR KTRSTHTGSS
CIGTDPNRNF DAGWCEIGAS RNPCDETYCG PAAESEKETK ALADFIRNKL SSIKAYLTIH
SYSQMMIYPY SYAYKLGENN AELNALAKAT VKELASLHGT KYTYGPGATT IYPAAGGSDD
WAYDQGIRYS FTFELRDTGR YGFLLPESQI RATCEETFLA IKYVASYVLE HLY

FATT-cutinase

MDYKDDDDVE AEESDNVDSA DAEEDDSDVW WGGADTDYAD GSEDKVVEVA EEEEVAEVEE
EEADDDEDDE DGDEVEEEAE EPYEEATERT TSIATTTTTT TESVEEVYPG QVGYPGQVGY
PGQVIEGRGI PMGHHHHHHH HHHSSGHIEG RHMGLGRTTR DDLINGNSAS CRDVIFIYAR
GSTETGNLGT LGPSIASNLE SAFGKDGVWI QGVGGAYRAT LGDNALPRGT SSAAIREMLG
LFQQANTKCP DATLIAGGYS QGAALAAASI EDLDSAIRDK IAGTVLFGYT KNLQNRGRIP
NYPADRTKVF CNTGDLVCTG SLIVAAPHLA YGPDARGPAP EFLIEKVRAV RGSAGHMYPG
QYPGQYPGQY PGQYPGQV

FATT-P20.1

MDYKDDDDVE AEESDNVDSA DAEEDDSDVW WGGADTDYAD GSEDKVVEVA EEEEVAEVEE
EEADDDEDDE DGDEVEEEAE EPYEEATERT TSIATTTTTT TESVEEVAPG QVGAPGQVGA
PGQVIEGRGI PMVQIQLVQS GPEVQKPGET VRISCKASGY TFTTAGMQWV QKMPGKSLKW
IGWINTRSGV PKYAEDFKGR FAFSLETSAS IAYLHINNLK NEDTATYFCA REGPGFVYWG
QGTLVTVSSG GGGSGGGGSG GGGSTQTVVT QESALTTSPG ETVTLTCRSS TGAVTTSNYA
NWVQEKPDHL FTGLIVGTNN RVPGVPPRFS GSLIEDKAAL TITGAQTEDE AIYFCALWYS
NHWVFGGGTK LTVLGAAAHH HHHHGAAEQK LISEEDLNGA A

FATT-01F

MDYKDDDDVE AEESDNVDSA DAEEDDSDVW WGGADTDYAD GSEDKVVEVA EEEEVAEVEE
EEADDDEDDE DGDEVEEEAE EPYEEATERT TSIATTTTTT TESVEEVYPG QVGYPGQVGY
PGQVIEGRGI PMVEVKLVES GGGLVQPGGS LKLSCATSGF TFSDYFLFWV RQTPEKSLEW
VAYIGYGGGS TYYPDTVKGR FTISRDNAKN TLYLQMSRLK SEDTAIYYCA RRDGYSFDYW
GQGTTLTVSS AGGGGSGGGG SGGGGSDIQM TQTTSSLSAS LGDRVTISCR ASQDISNYLN
WYQQKPDGSV KLLIYYTSRL HSGVPSKFSG SGSGTDFSLT ITNLEQEDIA TYFCQQAFSL
PWTFGGGTKL EIKAAARGGP EQKLISEEDL NSAVDHHHHH H

FATT-SG/19

MDYKDDDDVE AEESDNVDSA DAEEDDSDVW WGGADTDYAD GSEDKVVEVA EEEEVAEVEE
EEADDDEDDE DGDEVEEEAE EPYEEATERT TSIATTTTTT TESVEEVYPG QVGYPGQVGY
PGQVIEGRGI PMAAAMQVHL QQSGAELMKP GASVKISCKA TGYTFTSYWI EWVKQRPGHG
LEWLGEILPG SGYIHYNEKF KGKATFTTDT SSNTAYMQLS SLTSEDSAVY YCSRALALYA
MDYWGQGTSV TVSSGGGGSG GGGSGGGGST DIVMTQATPS IPVTPGESVS ISCRSNKSLL
HSNGNTYLYW FLQRPGQSPR LLIFRMSNLA SGVPDRFSGS GSGTAFTLRI SRVEAADVGI
YFCLQHLEYP FTFGAGTKLE LKRAAAHHHH HHGAAEQKLI SEEDLNGAA

References

Anfinsen C. B. (1973) Principles that govern the folding of protein chains. *Science 181* (4096): 223–230.

Arakawa, T., Ejima, D., Tsumoto, K., Obeyama, N., Tanaka, Y., Kita, Y., & Timasheff, S. N. (2007). Suppression of protein interactions by arginine: a proposed mechanism of the arginine effects. *Biophysical chemistry*, *127*(1), 1-8.

Austin, B. P., Tözsér, J., Bagossi, P., Tropea, J. E., & Waugh, D. S. (2011). The substrate specificity of *Metarhizium anisopliae* and *Bos Taurus* carboxypeptidases A: Insights into their use as tools for the removal of affinity tags. *Protein expression and purification*, *77*(1), 53-61.

Austin, B. P., & Waugh, D. S. (2012). Isolation of *Metarhizium anisopliae* carboxypeptidase A with native disulfide bonds from the cytosol of *Escherichia coli* BL21 (DE3). *Protein expression and purification*, *82*(1), 116-124.

Bach, H., Mazor, Y., Shaky, S., Shoham-Lev, A., Berdichevsky, Y., Gutnick, D. L., & Benhar, I. (2001). *Escherichia coli* maltose-binding protein as a molecular chaperone for recombinant intracellular cytoplasmic single-chain antibodies. *Journal of molecular biology*, *312*(1), 79-93.

Basu A, Li X, Leong S. S.J. (2011) Refolding of proteins from inclusion bodies: rational design and recipes. *Applied microbiology and biotechnology*, 92(2), 241-251.

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

Baneyx, F., & Mujacic, M. (2004). Recombinant protein folding and misfolding in *Escherichia coli*. *Nature biotechnology*, *22*(11), 1399-1408.

Benard, V., & Bokoch, G. M. (2002). Assay of Cdc42, Rac, and Rho GTPase activation by affinity methods. *Methods in enzymology*, *345*, 349-359.

Bessette, P. H., Åslund, F., Beckwith, J., & Georgiou, G. (1999). Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. *Proceedings of the National Academy of Sciences*, *96*(24), 13703-13708.

Bhattacharyya, J., & Das, K. P. (1999). Molecular chaperone-like properties of an unfolded protein, αs-casein. *Journal of Biological Chemistry*, *274*(22), 15505-15509.

Bowden, G. A., & Georgiou, G. (1990). Folding and aggregation of beta-lactamase in the periplasmic space of *Escherichia coli*. *Journal of Biological Chemistry*, *265*(28), 16760-16766.

Braun, P., Hu, Y., Shen, B., Halleck, A., Koundinya, M., Harlow, E., & LaBaer, J. (2002). Proteome-scale purification of human proteins from bacteria. *Proceedings of the National Academy of Sciences*, *99*(5), 2654-2659.

Brown, H. G., & Hoh, J. H. (1997). Entropic exclusion by neurofilament sidearms: a mechanism for maintaining interfilament spacing. *Biochemistry*, *36*(49), 15035-15040.

Buchner E, (1897) Alcoholic Fermentation Without Yeast Cells. *Berichte der Deutschen Chemischen Gesellschaft **30**; 117–124.*

Carver, J. A., Guerreiro, N., Nicholls, K. A., & Truscott, R. J. (1995). On the interaction of α-crystallin with unfolded proteins. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, *1252*(2), 251-260.

Chow, M. K., Amin, A. A., Fulton, K. F., Fernando, T., Kamau, L., Batty, C., ... & Buckle, A. M. (2006). The REFOLD database: a tool for the optimization of protein expression and refolding. *Nucleic acids research*, *34*(suppl 1), D207-D212.

Clark, P. L. (2004). Protein folding in the cell: reshaping the folding funnel. *Trends in biochemical sciences*, *29*(10), 527-534.

Cleland, J. L., Builder, S. E., Swartz, J. R., Winkler, M., Chang, J. Y., & Wang, D. I. (1992). Polyethylene glycol enhanced protein refolding. *Nature Biotechnology*, *10*(9), 1013-1019.

Dobson C. M. (2004) Principles of protein folding, misfolding and aggregation. In Seminars in cell & developmental biology (Vol. 15, No. 1, pp. 3-16). Academic Press.

de Marco, A., Deuerling, E., Mogk, A., Tomoyasu, T., & Bukau, B. (2007). Chaperone-based procedure to increase yields of soluble recombinant proteins produced in E. coli. *BMC biotechnology*, *7*(1), 32.

di Guana, C., Lib, P., Riggsa, P. D., & Inouyeb, H. (1988). Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. *Gene*, *67*(1), 21-30.

Eiberle, M. K., & Jungbauer, A. (2010). Technical refolding of proteins: Do we have freedom to operate?. *Biotechnology journal*, *5*(6), 547-559.

Ellis, R. J., & Hartl, F. U. (1999). Principles of protein folding in the cellular environment. *Current opinion in structural biology*, *9*(1), 102-110.

Field, J., Nikawa, J., Broek, D., MacDonald, B., Rodgers, L., Wilson, I. A., ... & Wigler, M. (1988). Purification of a RAS-responsive adenylyl cyclase complex from Saccharomyces cerevisiae by use of an epitope addition method. *Molecular and cellular biology*, *8*(5), 2159-2165.

Folk J. E., Gladner J. A. (1961) Influence of cobalt and cadmium on the peptidase and esterase activities of carboxypeptidase B. *Biochimica et biophysica acta*, *48*(1), 139-147.

Gunasekaran, K., Tsai, C. J., & Nussinov, R. (2004). Analysis of ordered and disordered protein complexes reveals structural features discriminating between stable and unstable monomers. *Journal of molecular biology*, *341*(5), 1327-1341.

Hamilton P. B., Bogue D. C., Anderson R. A. (1960) Ion exchange chromatography of amino acids. Analysis of diffusion (mass transfer) mechanisms. *Analytical Chemistry*, *32*(13), 1782-1792.

Haniu, M., Horan, T., Spahr, C., Hui, J., Fan, W., Chen, C., ... & Lu, H. S. (2011). Human Dickkopf‐1 (huDKK1) protein: Characterization of glycosylation and determination of disulfide linkages in the two cysteine‐rich domains. *Protein Science*, *20*(11), 1802-1813.

Hopp, T. P., Prickett, K. S., Price, V. L., Libby, R. T., March, C. J., Cerretti, D. P., ... & Conlon, P. J. (1988). A short polypeptide marker sequence useful for recombinant protein identification and purification. *Biotechnology*, *6*(1205), 5.

Hosie, L., Sutton, L. D., & Quinn, D. M. (1987). *p*-Nitrophenyl and cholesteryl-*N*-alkyl carbamates as inhibitors of cholesterol esterase. *Journal of Biological Chemistry*, *262*(1), 260-264.

Huang, Y., & Liu, Z. (2009). Kinetic advantage of intrinsically disordered proteins in coupled folding–binding process: a critical assessment of the “fly-casting” mechanism. *Journal of molecular biology*, *393*(5), 1143-1159.

Ishibashi, M., Tsumoto, K., Tokunaga, M., Ejima, D., Kita, Y., & Arakawa, T. (2005). Is arginine a protein-denaturant? *Protein expression and purification*, *42*(1), 1-6.

Ishii, Y., Sonezaki, S., Iwasaki, Y., Tauchi, E., Shingu, Y., Okita, K., ... & Kondo, A. (1998). Single-Step Purification and Characterization of MBP (Maltose Binding Protein)-DnaJ Fusion Protein and Its Utilization for Structure-Function Analysis. *Journal of biochemistry*, *124*(4), 842-847.

Joshi, L., & Leger, R. J. S. (1999). Cloning, expression, and substrate specificity of MeCPA, a zinc carboxypeptidase that is secreted into infected tissues by the fungal entomopathogen *Metarhizium anisopliae*. *Journal of Biological Chemistry*, *274*(14), 9803-9811.

Kadokura, H., Katzen, F., & Beckwith, J. (2003). Protein disulfide bond formation in prokaryotes. *Annual review of biochemistry*, *72*(1), 111-135.

Kane, J. F., & Hartley, D. L. (1988). Formation of recombinant protein inclusion bodies in *Escherichia coli*. *Trends in biotechnology*, *6*(5), 95-101.

Khodarahmi, R., Beyrami, M., & Soori, H. (2008). Appraisal of casein's inhibitory effects on aggregation accompanying carbonic anhydrase refolding and heat-induced ovalbumin fibrillogenesis. *Archives of biochemistry and biophysics*, *477*(1), 67-76.

Kochendoerfer, G. (2003). Chemical and biological properties of polymer-modified proteins. *Expert opinion on biological therapy*, *3*(8), 1253-1261.

Leong S. L., Barnham K. J., Multhaup G., Cappai R. (2011) Amyloid precursor protein. *Encyclopedia of Inorganic and Bioinorganic Chemistry.*

Lindner, R. A., Kapur, A., Mariani, M., Titmuss, S. J., & Carver, J. A. (1998). Structural alterations of α‐crystallin during its chaperone action. *European journal of biochemistry*, *258*(1), 170-183.

Lindqvist B., Storgårds T. (1955) Molecular-sieving properties of starch. *Nature*. *175*: 511-512.

Liu, Z., Gosser, Y., Baker, P. J., Ravee, Y., Lu, Z., Alemu, G., ... & Montclare, J. K. (2009). Structural and functional studies of *Aspergillus oryzae* cutinase: enhanced thermostability and hydrolytic activity of synthetic ester and polyester degradation. *Journal of the American Chemical Society*, *131*(43), 15711-15716.

Machida, K., Kono-Okada, A., Hongo, K., Mizobata, T., & Kawata, Y. (2008). Hydrophilic residues 526KNDAAD531 in the flexible C-terminal region of the chaperonin GroEL are critical for substrate protein folding within the central cavity. *Journal of Biological Chemistry*, *283*(11), 6886-6896.

Mamathambika, B. S., & Bardwell, J. C. (2008). Disulfide-linked protein folding pathways. *Annual review of cell and developmental biology*, *24*, 211-235.

Manoil, C. O. L. I. N., Mekalanos, J. J., & Beckwith, J. (1990). Alkaline phosphatase fusions: sensors of subcellular location. *Journal of bacteriology*, *172*(2), 515-518.

McGuffee, S. R., & Elcock, A. H. (2010). Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS computational biology*, *6*(3), e1000694.

Munro, S., & Pelham, H. R. (1984). Use of peptide tagging to detect proteins expressed from cloned genes: deletion mapping functional domains of *Drosophila* hsp 70. *The EMBO journal*, *3*(13), 3087.

Narayanan, S. R. (1994). Preparative affinity chromatography of proteins. *Journal of Chromatography A*, *658*(2), 237-258.

Niwa, T., Ying, B. W., Saito, K., Jin, W., Takada, S., Ueda, T., & Taguchi, H. (2009). Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of Escherichia coli proteins. *Proceedings of the National Academy of Sciences*, *106*(11), 4201-4206.

Nogi, T., Sangawa, T., Tabata, S., Nagae, M., Tamura‐Kawakami, K., Beppu, A., ... & Takagi, J. (2008). Novel affinity tag system using structurally defined antibody‐tag interaction: Application to single‐step protein purification. *Protein Science*, *17*(12), 2120-2126.

Ohlson, M., Sörensson, J., Lindström, K., Blom, A. M., Fries, E., & Haraldsson, B. (2001). Effects of filtration rate on the glomerular barrier and clearance of four differently shaped molecules. *American Journal of Physiology-Renal Physiology*, *281*(1), F103-F113.

Okai, M., Ohtsuka, J., Asano, A., Guo, L., Miyakawa, T., Miyazono, K. I., ... & Tanokura, M. (2012). High pressure refolding, purification, and crystallization of flavin reductase from *Sulfolobus tokodaii* strain 7. *Protein expression and purification*, *84*(2), 214-218.

Rekas, A., Ahn, K. J., Kim, J., & Carver, J. A. (2012). The chaperone activity of α‐synuclein: Utilizing deletion mutants to map its interaction with target proteins. *Proteins: Structure, Function, and Bioinformatics*, *80*(5), 1316-1325.

Patil, G., Rudolph, R., & Lange, C. (2008). *In vitro*-refolding of a single-chain Fv fragment in the presence of heteroaromatic thiols. *Journal of biotechnology*, *134*(3), 218-221.

Peterson, L. M., Holmquist, B., & Bethune, J. L. (1982). A unique activity assay for carboxypeptidase A in human serum. *Analytical biochemistry*, *125*(2), 420-426.

Pinna, M. C., Salis, A., Monduzzi, M., & Ninham, B. W. (2005). Hofmeister Series: The Hydrolytic Activity of Aspergillus n iger Lipase Depends on Specific Anion Effects. *The Journal of Physical Chemistry B*, *109*(12), 5406-5408.

Porath, J., Carlsson, J. A. N., Olsson, I., & Belfrage, G. (1975). Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature*, *258*, 598-599.

Porath J, Flodin P (1959). Gel filtration: a method for desalting and group separation. *Nature*, *183*(4676), 1657-1659.

Prinz, W. A., Åslund, F., Holmgren, A., & Beckwith, J. (1997). The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm. *Journal of Biological Chemistry*, *272*(25), 15661-15667.

Qoronfleh, M. W., Hesterberg, L. K., & Seefeldt, M. B. (2007). Confronting high-throughput protein refolding using high pressure and solution screens. *Protein expression and purification*, *55*(2), 209-224.

Rashid, F., Sharma, S., & Bano, B. (2005). Comparison of guanidine hydrochloride (GdnHCl) and urea denaturation on inactivation and unfolding of human placental cystatin (HPC). *The protein journal*, *24*(5), 283-292.

Rekas, A., Ahn, K. J., Kim, J., & Carver, J. A. (2012). The chaperone activity of α‐synuclein: Utilizing deletion mutants to map its interaction with target proteins. *Proteins: Structure, Function, and Bioinformatics*, *80*(5), 1316-1325.

Rigby P, Shenk T. (1991) Paul Berg, on his 65th birthday. *Nucleic Acids Res*. *19*(25):7041.

Rost, B. (2002). Did evolution leap to create the protein universe?. *Current opinion in structural biology*, *12*(3), 409-416.

Sangawa, T., Nogi, T., & Takagi, J. (2008). A murine monoclonal antibody that binds N-terminal extracellular segment of human protease-activated receptor-4. *Hybridoma*, *27*(5), 331-335.

Sangawa, T., Tabata, S., Suzuki, K., Saheki, Y., Tanaka, K., & Takagi, J. (2013). A multipurpose fusion tag derived from an unstructured and hyperacidic region of the amyloid precursor protein. *Protein Science*. *22*(6):840-50.

Santner, A. A., Croy, C. H., Vasanwala, F. H., Uversky, V. N., Van, Y. Y. J., & Dunker, A. K. (2012). Sweeping away protein aggregation with entropic bristles: intrinsically disordered protein fusions enhance soluble expression. *Biochemistry*, *51*(37), 7250-7262.

Siller, E., DeZwaan, D. C., Anderson, J. F., Freeman, B. C., & Barral, J. M. (2010). Slowing bacterial translation speed enhances eukaryotic protein folding efficiency. *Journal of molecular biology*, *396*(5), 1310-1318.

Singh, S. M., & Panda, A. K. (2005). Solubilization and refolding of bacterial inclusion body proteins. *Journal of bioscience and bioengineering*, *99*(4), 303-310.

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

Stleger, R. J., Bidochka, M. J., & Roberts, D. W. (1994). Characterization of a Novel Carboxypeptidase Produced by the Entomopathogenic Fungus *Metarhizium anisopliae*. *Archives of biochemistry and biophysics*, *314*(2), 392-398.

Sun, P., Tropea, J. E., & Waugh, D. S. (2011). Enhancing the solubility of recombinant proteins in *Escherichia coli* by using hexahistidine-tagged maltose-binding protein as a fusion partner. In *Heterologous Gene Expression in E. coli* (pp. 259-274). Humana Press.

Tabata, S., Nampo, M., Mihara, E., Tamura-Kawakami, K., Fujii, I., & Takagi, J. (2010). A rapid screening method for cell lines producing singly-tagged recombinant proteins using the "TARGET tag" system. *Journal of proteomics*, *73*(9), 1777-1785.

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

Terpe, K. (2006). Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied microbiology and biotechnology*, *72*(2), 211-222.

Uversky, V. N. (1993). Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry*, *32*(48), 13288-13298.

Vallejo, L. F., & Rinas, U. (2004). Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. *Microbial Cell Factories*, *3*(1), 11.

Walters RR (1985). Affinity chromatography. *Analytical Chemistry, 57*(11), 1099A-1114A.

Weikl, T., Abelmann, K., & Buchner, J. (1999). An unstructured C-terminal region of the Hsp90 co-chaperone p23 is important for its chaperone function. *Journal of molecular biology*, *293*(3), 685-691.

Williams, D. C., Van Frank, R. M., Muth, W. L., & Burnett, J. P. (1982). Cytoplasmic inclusion bodies in Escherichia coli producing biosynthetic human insulin proteins. *Science*, *215*(4533), 687-689.

Yi, D. (2008). Cancer, viruses, and mass migration: Paul Berg's venture into eukaryotic biology and the advent of recombinant DNA research and technology, 1967–1980. *Journal of the History of Biology*, *41*(4), 589-636.

Zhang, T., Xu, X., Shen, L., Feng, Y., Yang, Z., Shen, Y., ... & Wang, X. (2009). Modeling of protein refolding from inclusion bodies. *Acta biochimica et biophysica Sinica*, *41*(12), 1044-1052.

Publication

Sangawa, T., Tabata, S., Suzuki, K., Saheki, Y., Tanaka, K., & Takagi, J. (2013). A multipurpose fusion tag derived from an unstructured and hyperacidic region of the amyloid precursor protein. *Protein Science*. *22*(6):840-50.

Acknowledgement

The present work was carried out under the direction of Prof. Junichi Takagi, Laboratory of Protein synthesis and expression, Institute for Protein Research, Osaka University. I would like to appreciate Prof. Takagi for enormous supports over many years and careful direction which were of inestimable value for me and my study.

I would also thank associate professor Kenji Iwasaki for giving valuable comments and teaching. I really thank Dr. Terukazu Nogi, associate professor in Department of supramolecular biology, graduate school of nanobioscience, Yokohama City University for his kind direction in the research field. Without his supports and invaluable direction, I could not have accomplished this work.

Special thanks also go to Dr. Kei Suzuki, Dr. Yasushi Saheki and Dr. Kenji Tanaka. They are indispensable contributors for this work. Ms. Sanae Tabata greatly contributed to the groundwork of the study.

I am grateful to assistant professor Yu Kitago, Ms. Emiko Mihara and all members of Takagi Lab for their insightful advices and continuous supports. I would also like to express my gratitude to former members of Takagi Lab, Dr. Masamichi Nagae, Dr. Hiroki Tanaka, Dr. Norihisa Yasui, Mr. Takuro Yasuda, Mr. Samuel Thompson and Ms. Maiko Nambo-Noda. Their comments and supports were innumerably valuable throughout the course of my study. I am also indebted to Ms. Keiko Tamura-Kawakami and Dr. Naoko Norioka for their excellent technical supports and advices. I would thank Ms. Mayumi Nakano and Mie Sakai

Lastly, my heartfelt appreciation goes to my parents and Ayumi Takahashi for their moral support and warm encouragements.

Takeshi Sangawa　August 13