THE PROTEIN SOCIETY

METHODS AND APPLICATIONS

A multipurpose fusion tag derived from an unstructured and hyperacidic region of the amyloid precursor protein

Takeshi Sangawa,[1] Sanae Tabata,[1] Kei Suzuki,[1] Yasushi Saheki,[2] Keiji Tanaka,[2] and Junichi Takagi[1]*

[1]Laboratory of Protein Synthesis and Expression, Institute for Protein Research, Osaka University, Suita, Osaka 565-0871, Japan

[2]Laboratory of Protein Metabolism, Tokyo Metropolitan Institute of Medical Science, Setagaya-ku, Tokyo 156-8506, Japan

Received 11 January 2013; Accepted 14 March 2013
DOI: 10.1002/pro.2254
Published online 23 March 2013 proteinscience.org

Abstract: Expression and purification of aggregation-prone and disulfide-containing proteins in *Escherichia coli* remains as a major hurdle for structural and functional analyses of high-value target proteins. Here, we present a novel gene-fusion strategy that greatly simplifies purification and refolding procedure at very low cost using a unique hyperacidic module derived from the human amyloid precursor protein. Fusion with this polypeptide (dubbed FATT for Flag-Acidic-Target Tag) results in near-complete soluble expression of variety of extracellular proteins, which can be directly refolded in the crude bacterial lysate and purified in one-step by anion exchange chromatography. Application of this system enabled preparation of functionally active extracellular enzymes and antibody fragments without the need for condition optimization.

Keywords: protein expression; fusion tag; inclusion body; refolding; amyloid precursor protein

Introduction

Large-scale recombinant protein production and purification are critical processes in biotechnology and biomedical research, and *Escherichia coli*-based expression coupled with the use of various fusion tag systems has been the primary choice of method for these processes.[1] Although numerous fusion tag technologies that enable efficient expression, easy purification, or functional (re)folding of a target protein from the bacterial lysate have been developed and commercialized, no single technology is universally applicable to all expression projects.[2]

Widely used fusion tags include glutathione-*S*-transferase (GST), maltose-binding protein (MBP), polyhistidine tag, and epitope tags such as Flag peptide. Proteins fused with these tags can be purified on affinity matrices immobilized with cognate ligand or binding agent for the fusion tag, resulting in a simple purification. These affinity resins, however, generally tend to lose their activity over repeated regeneration cycles, making them expensive compared to more traditional matrices such as

Grant sponsors: Platform for Drug Discovery, Informatics, and Structural Life Science (PDIS) and the "Target Proteins Research Program (TPRP)" grant from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT).

*Correspondence to: Junichi Takagi, Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan. E-mail: takagi@protein.osaka-u.ac.jp

ion-exchange resins, and thus increasing the overall cost of protein production.[3] Another important issue that needs to be considered when expressing proteins heterologously in *E. coli* is the formation of insoluble aggregates within the cells. Several fusion partners including GST and MBP are reported to facilitate soluble expression of aggregation-prone proteins,[2] but their effects have yet to be tested empirically. More importantly, it is often observed that such "solubilized" proteins precipitate upon the removal of the fusion partners, indicating the critical importance of obtaining native structure before the tag removal. This "refolding" process is also crucial in cases where proteins are deliberately expressed as inclusion bodies in bacteria for the large-scale production.[4] Furthermore, the bacterial production of extracellular proteins with multiple disulfide bonds is particularly challenging because of the absolute requirement for an optimized oxidative refolding condition. As with efforts to aid general refolding, the use of special expression systems involving a modified cytosolic redox potential has been reported to aid disulfide bond formation in the expressed protein, but these solutions are effective in a limited number of cases.

We aimed at developing a versatile fusion tag that facilitates soluble expression, easy purification, and target-independent refolding of bacterially expressed proteins in a cost-efficient way. To this end, we used a unique hyperacidic module derived from the human amyloid precursor protein (APP) extracellular region.[5] Unlike its notoriously amyloidogenic sister fragment produced from the same precursor (i.e., amyloid β peptide or Aβ), this polypeptide exhibits very strong antiaggregation tendency when fused to various aggregation-prone proteins. This polypeptide segment of ~100 residues is predicted to behave as an intrinsically disordered protein with a theoretical isoelectric point (pI) of 3.2 and possesses an unusually large hydrodynamic radius in solution. Moreover, purification of the fusion proteins can be accomplished solely by anion exchange chromatography, without requiring any special or proprietary affinity resins. Most importantly, "direct refolding" in the bacterial lysate without prior purification of the target protein combined with one-step concentration from dilute solution using anion-exchange resin enabled us to simplify the protocol for obtaining various disulfide-containing extracellular proteins in their functional state.

Results and Discussion

Design of the fusion tag

APP is a Type I membrane protein of ~700 residues, and its extracellular domain consists of several modules, including an N-terminal growth factor-like domain, a copper-binding domain, a hyperacidic region, and an α-helical central APP domain [Fig. 1(A)].[5] The last portion of the ectodomain and the first half of the transmembrane region constitute ~40 residue amyloidogenic Aβ peptide, which forms senile plaques in the brain tissue of patients with Alzheimer's disease. The acidic region spans residues 190–286 (residue numbering is based on the sequence of the neuronal isoform APP695) and shows an unusually high content of acidic residues (48%), resulting in the predicted pI value of 3.2. The low complexity nature of the sequence predicts that this region lacks any permanent secondary structure.[6] Furthermore, our own biophysical experiments suggest that an APP ectodomain fragment containing the acidic region does not behave as a compact globular protein in solution (data not shown). We reasoned that the strong negative charge and the unique hydrodynamic characteristics of this polypeptide would radically change the physicochemical property of a protein when appended as a fusion partner and may facilitate easy purification using conventional chromatographic separation methods.

To this end, we constructed a bacterial expression vector containing the 97-residue APP acidic region sequence preceded by a FLAG epitope tag and followed by a Target tag,[7] Factor Xa cleavage site, and a multiple cloning site [Fig. 1(B)]. The vector was used to express several target proteins that were fused C-terminal to the artificial fusion cassette (designated as FATT for Flag-Acidic-Target Tag). As shown in Figure 1(C), FATT-fusion did not change the expression level of a C-terminal fragment of human Dkk1 (Dkk1_C, 91 residues, pI = 9.15) in the bacterial cytosol. Although FATT-tagged version of another model protein UV-excited green fluorescent protein (GFP_{UV}; 246 residues, pI = 5.81) showed ~50% reduction in the expression level, such reduction was not observed in a majority of the cases (e.g., see Fig. 3 in the following section), indicating that FATT moiety does not have strong expression-suppressing effect. Because FATT is flanked by two different epitope tags, fusion proteins can be readily identified by the Western blotting using antibodies against these tags [Fig. 1(D)].

Effect of FATT on the chromatographic behavior of expressed protein

One obvious application of the FATT-fusion strategy is to achieve simple and universal purification using anion exchange resin. When the cytosol fractions from bacteria expressing FATT-Dkk1_C were passed through HiTrapQ anion exchange column, the FATT-fusion protein bound strongly to the column and eluted at ~600 m*M* NaCl [Fig. 1(E)]. The peak fraction for the FATT-fusion protein (fraction No.43) was well separated from that for the majority of the endogenous proteins, and contained only a few

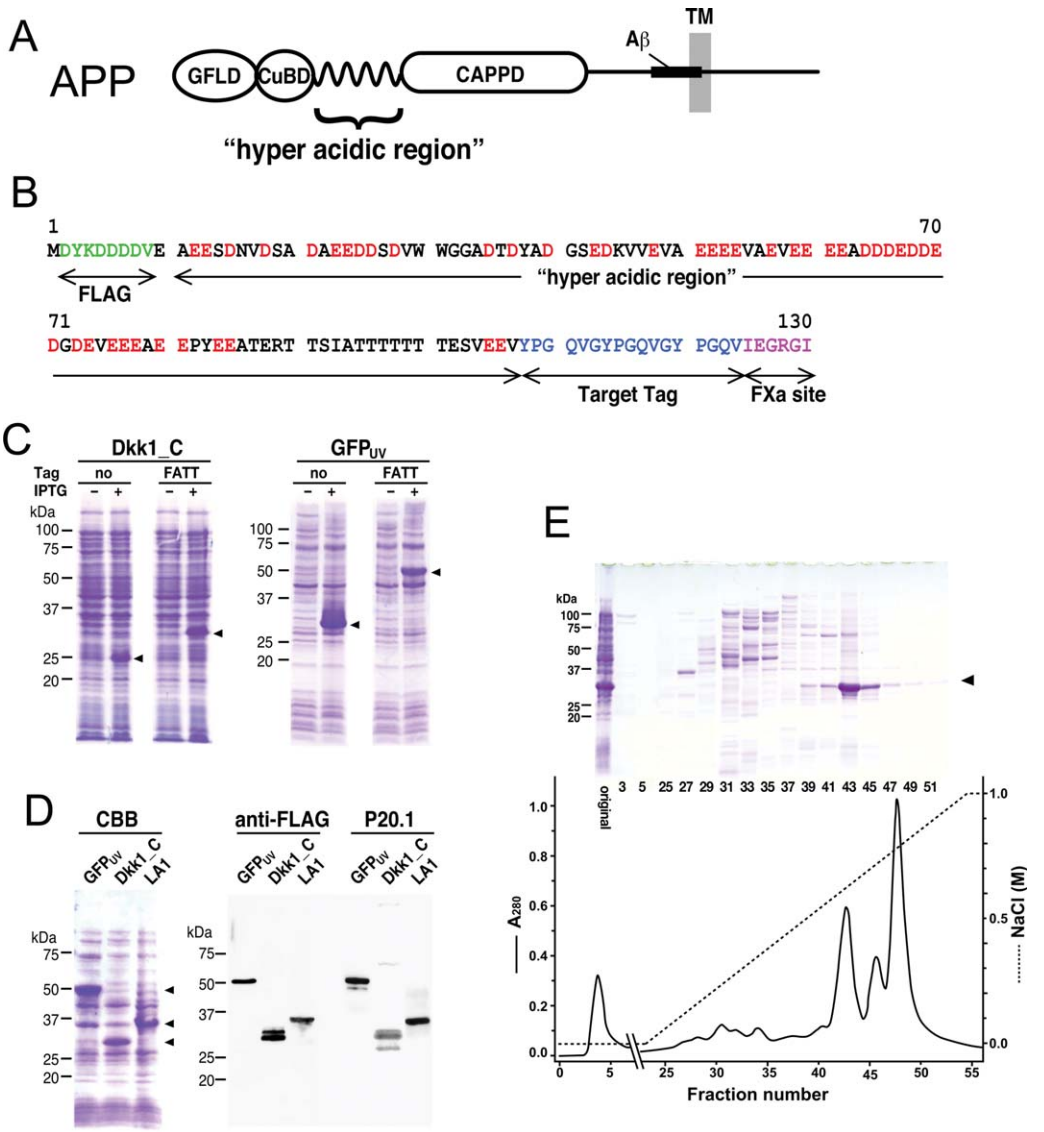

Figure 1. Basic properties of FATT-tag. (A) Schematic domain organization of amyloid precursor protein (APP). GFLD, growth factor-like domain; CuBD, copper-binding domain; CAPPD, central APP domain; Aβ, amyloid β; TM, transmembrane domain. The hyperacidic region located between CuBD and CAPPD is indicated by a squiggly line. (B) Amino acid sequence of the N-terminal FATT-tag fusion cassette used in this study. (C) SDS-PAGE analysis of total bacterial lysates expressing either untagged (no) or FATT-tagged (FATT) versions of Dkk1 C-terminal fragment (left panel) or 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 with 1 m*M* IPTG. The arrows point to the expressed proteins. (D) FATT-tagged proteins can be detected by Western blot with anti-FLAG polyclonal antibody and P20.1 (anti-TARGET tag). Tagged soluble proteins include GFP_{UV}, Dkk1_C, and LDLR class A module 1 (LA1) from apolipoprotein E receptor 2. (E) Anion exchange chromatography of cytosolic proteins from *E. coli* expressing FATT-Dkk1_C. SDS-PAGE analysis (upper panel) shows that FATT-Dkk1_C peaks at fraction 43 (arrow), well separated from the majority of other proteins.

contaminating protein bands. We confirmed that the last two peaks in the chromatogram (fractions 45–50) corresponded to oligonucleotide fragments originated from the bacterial genomic DNA and did not contain any proteins, indicating that FATT-Dkk1_C was the final protein eluting from the resin. We fused various proteins of differing size (14–50 kDa) and pI value (6–9) to FATT and confirmed that all tested fusion proteins eluted at 500–800 m*M* NaCl under the same condition as that in Figure 1(E). Due to the strong negative charge, fine salt gradient elution using prepacked column in an high-performance liquid chromatography-like format was unnecessary in most cases; near-complete purification was possible by a batch incubation of the bacterial lysate with anion exchange resin Q-Sepharose in the presence of 150 m*M* NaCl followed by elution with 1 *M* NaCl [Fig. 2(A)]. Typically, the FATT-fusion proteins thus prepared are 80–95% pure after just one chromatography step as judged by sodium

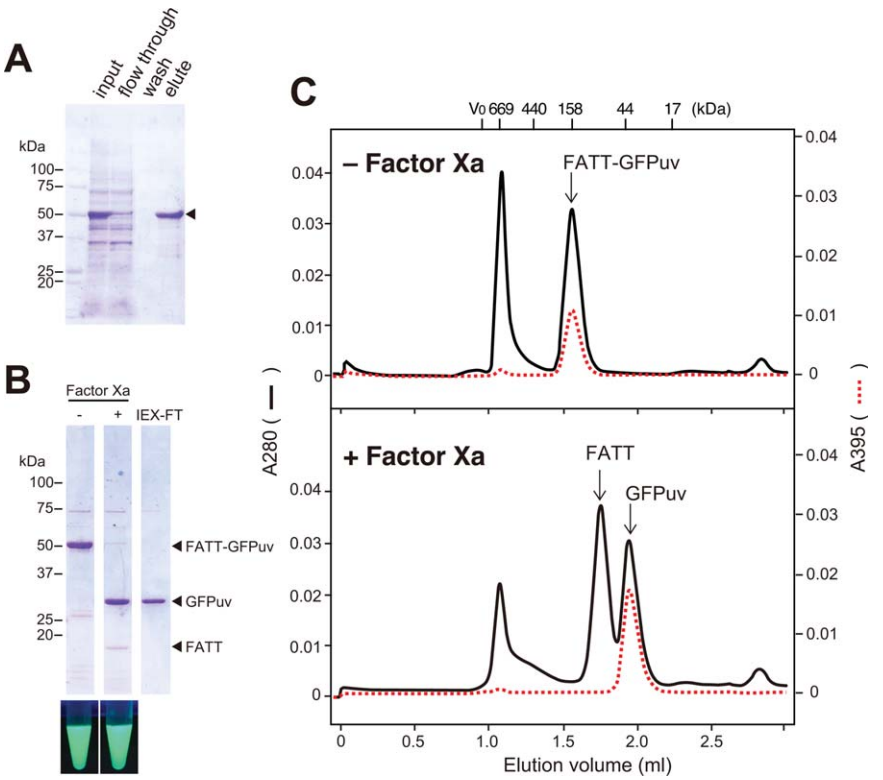

Figure 2. Purification and tag removal of FATT-GFP_{UV}. (A) One-step purification of FATT-GFP_{UV} from bacterial lysate using a small column of Q-Sepharose FF. Input: total lysate, flow through: unbound flow through, wash: 150 m*M* NaCl buffer wash, and elute: 500 m*M* NaCl buffer elution. (B) SDS-PAGE analysis of FATT-GFP_{UV} before (−) and after (+) factor Xa treatment. Note that the tag removal did not change the fluorescence property (bottom). Pure untagged GFP_{UV} can be recovered in the flow-through fraction from the second-round anion exchange step (IEX-FT). (C) Solution behavior of FATT tag. 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 ka), and myoglobin (17 kDa) are indicated on the top.

dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE; Fig. 4 and data not shown).

Next we explored how FATT affects the function and structure of soluble proteins using GFP_{UV} as a model protein. As shown in Figure 2(B), the FATT portion can be cleaved off by factor Xa, which splits the 50-kDa FATT-GFP_{UV} into 30-kDa GFP_{UV} and 18-kDa FATT. The fluorescence property of the GFP_{UV} was normal and unchanged by the treatment [Fig. 2(B), bottom images], indicating that the presence of highly acidic FATT sequence next to GFP_{UV} did not have an adverse effect on the structure/function of GFP_{UV}. The liberated GFP_{UV} can be easily purified to homogeneity by passing through a second anion exchange column in the presence of 150 m*M* NaCl [Fig. 2(B), lane "IEX-FT"], because everything else including the cleaved FATT will bind to the column. We found that the FATT-containing proteins exhibit very unusual solution behavior. Analytical gel filtration chromatography of FATT-GFP_{UV} shows that it exists as a monodisperse molecular species with an apparent molecular weight (MW) of 158 kDa, despite its actual size of 44 kDa [Fig. 2(C), upper panel]. This unusual property is endowed by the FATT portion, because the factor Xa-cleaved FATT (theoretical MW = 14 kDa) and GFP_{UV} (theoretical MW = 30 kDa) eluted at positions corresponding to 78 and 40 kDa, respectively [Fig. 2(C), lower panel]. Considering the potential electrostatic repulsion between FATT polypeptides and the tag's sharp and symmetric peak shape in the chromatograms, it is highly unlikely that FATT forms a stable and uniform oligomer. The large discrepancy between the actual and the observed molecular sizes for the FATT therefore suggests that it exists as a monomer with multiple rapidly interconverting conformations, in effect occupying a large hydrodynamic space.

Use of FATT as a strong solubility enhancer for aggregation-prone proteins

The strong electronegative property and the observed large exclusion volume of the FATT polypeptide led us to speculate that it may have an anti-aggregation effect when fused to aggregation-prone proteins. To test this possibility, we chose three extracellular proteins [a secreted carboxypeptidase A from *Metarhizium anisopliae* (MeCPA)], a bacterial esterase cutinase, and a single-chain Fv (scFv)

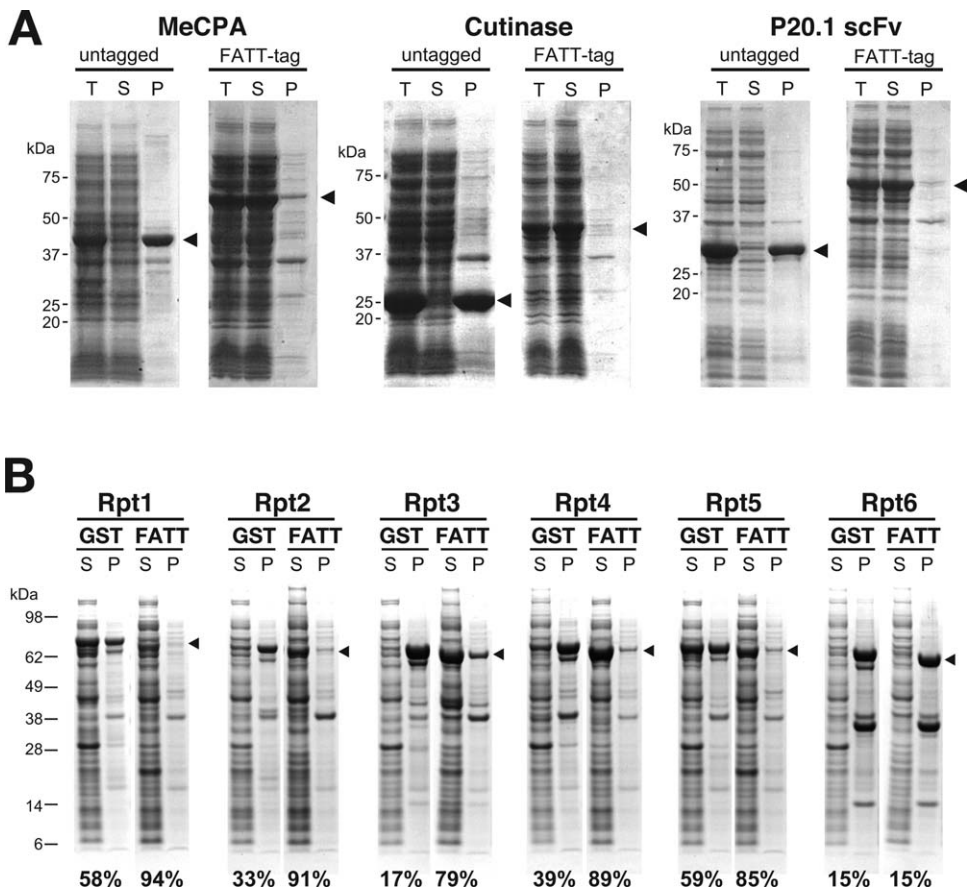

Figure 3. FATT-tag directs expressed proteins to the soluble fraction. (A) Switch from insoluble to soluble expression upon FATT fusion for three extracellular proteins. *E. coli* cells expressing the indicated proteins with or without FATT tag were lysed by sonication to obtain the total cell lysate (T), which was then centrifuged to separate soluble supernatant (S) and insoluble pellet (P) fractions. Same volume of samples were taken from each fraction and separated on SDS-PAGE gels followed by Coomassie blue staining. Arrows point to the expressed proteins. Note that bands for FATT-tagged proteins are virtually absent from the insoluble pellet fraction. (B) Comparison of the solubility-enhancing effect of FATT-tag and GST. Six test proteins (Rpt1–6) tagged with either FATT-tag or GST were expressed in *E. coli* and the soluble (S) and insoluble (P) fractions were analyzed by SDS-PAGE as in (A). Note that because of the similar size of GST and FATT-tag, the expressed fusion proteins migrated to roughly the same position (arrows) regardless of the tag used. Solubility indices (percentage of recovery in the soluble fraction) are shown at the bottom.

fragment from monoclonal antibody P20.1). When untagged versions of these proteins are expressed in bacteria, the induced proteins are solely recovered in the insoluble fractions of cell lysate [Fig. 3(A), panels "untagged"]. In sharp contrast to this, FATT-fused versions are well expressed in the soluble fraction and are nearly absent from the insoluble pellet [Fig. 3(A), panels "FATT-tag"]. To explore more about the general applicability of FATT-fusion as a way to facilitate soluble expression, six base subunits from the 19S proteasomal complex were individually expressed as either FATT- or GST-fusion proteins. Bacterial expression of these proteins (Rpt1–6) has been very difficult because the individual subunits have a strong tendency to aggregate,[8] and even the solubility-enhancing GST-tag showed very limited effects [Fig. 3(B), panels "GST"]. When tagged with FATT, however, near-complete soluble expression was achieved for five of the six target proteins [Fig. 3(B), panels "FATT"]. The subunit that was resistant to the strong solubilizing effect of FATT (Rpt6) had a very high pI value of 9.1, suggesting that the negative charge of FATT may have been cancelled.

Production of extracellular proteins with disulfide bonds using FATT-assisted expression and "direct refolding"

Although most FATT-fusion proteins apparently remained soluble after bacterial lysis, that does not warrant that the target proteins acquired their native conformation. In fact, the extracellular proteins solubilized by the FATT-fusion strategy shown in Figure 3(A) contain multiple intramolecular disulfide bridges in their native folds that cannot be formed in the bacterial cytosol, and thus these types of protein are expected (and also confirmed) to be nonfunctional. Therefore, we developed a simple way to "refold" the FATT-fused proteins based on the

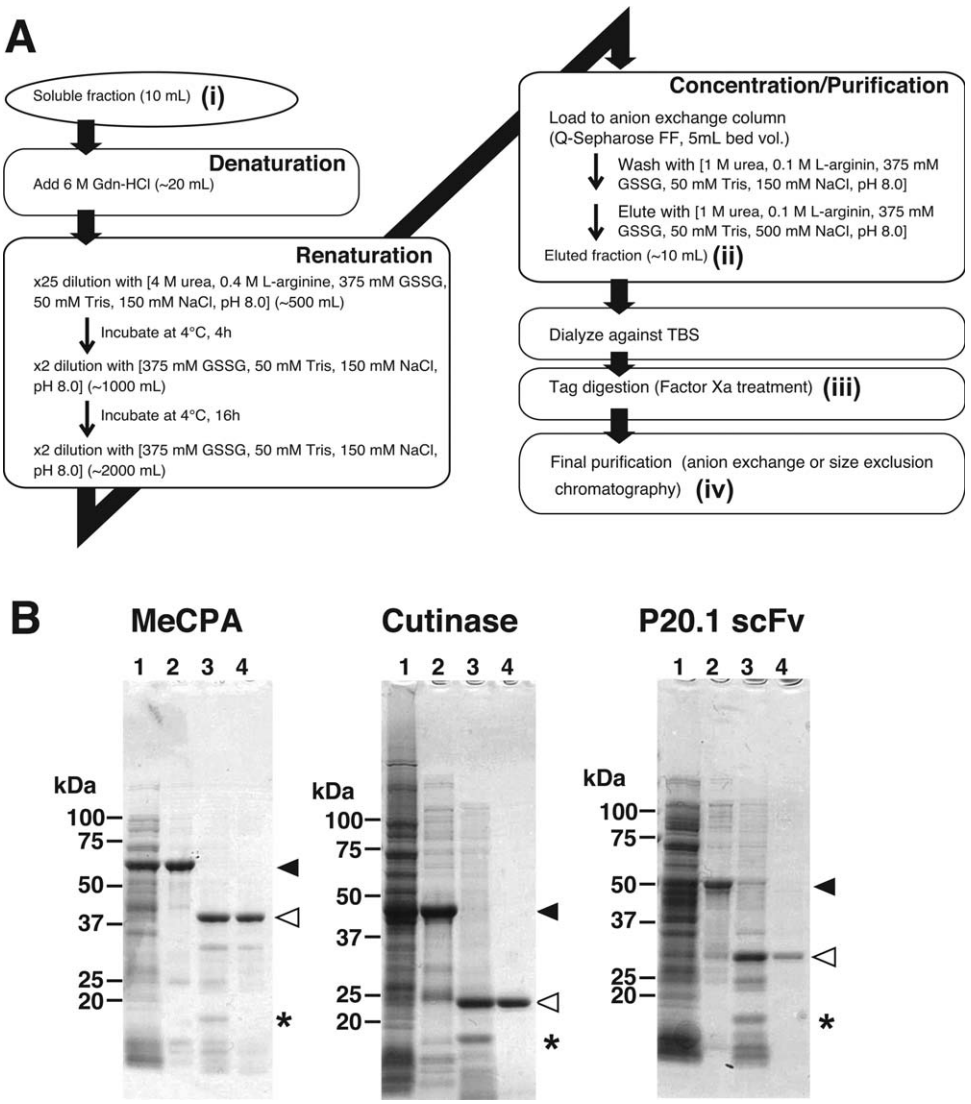

Figure 4. Simplified and universal refolding protocol for FATT-tagged extracellular proteins. (A) Scheme of the protocol for "direct refolding" and subsequent isolation. The total volume of the sample at each step is indicated in parenthesis, assuming a starting bacterial lysate volume of 10 mL. For details, see Methods section. (B) Refolding of three target extracellular proteins. Lane 1, cytosolic fraction of *E. coli* expressing each FATT-tagged protein before denaturation [Sample (i) in the scheme from (A)]; Lane 2, refolded sample after anion exchange purification [Sample (ii)]; Lane 3, factor Xa-digested sample [Sample (iii)]; Lane 4, sample after the second purification by anion exchange column [Sample (iv)]. FATT-tagged proteins, untagged target proteins, and the released FATT tag portion are indicated by filled triangles, open triangles, and asterisks, respectively.

denature/dilution procedure that is commonly used to refold proteins from inclusion bodies [Fig. 4(A)]. First, the total soluble fraction from bacteria expressing FATT-fused protein was treated with 6M guanidine hydrochloride to achieve complete denaturation, and then the proteins were allowed to renature by diluting 100-fold (final) in a Tris buffer containing oxidized glutathione. We discovered that an initial purification step was not necessary, making it possible to perform "direct refolding" from the crude lysate. This means that the presence of the endogenous bacterial proteins did not interfere with the refolding process, probably as a result of the shielding effect of FATT moiety. Although the dilution method is widely used and has been proven successful in numerous protein refolding projects,[9] one drawback of this method is the necessity to concentrate renatured protein from the diluted refolding mixture of very large volume (typically >1 L). While such concentration can be achieved by a tagging with His tag and passing through metal-chelate resin, endogenous proteins that binds to the resin still needs to be removed in the later steps. In contrast, the concentration process can be greatly simplified for FATT-fused proteins because they can be selectively adsorbed onto a small volume of anion exchange resin in the presence of physiological concentration of salt and eluted with high-salt buffer. When this method was applied to three extracellular proteins [Fig. 4(B)], it was shown that the

concentrating step also served as an efficient purification step, because the FATT-fused protein was highly enriched in the eluted fraction [Fig. 4(B), lane 2 in each panel]. As in the case of soluble GFP_{UV}, renatured FATT-fused proteins can be completely cleaved by factor Xa to liberate the untagged target proteins, which are finally purified on the second anion exchange column (lane 4). Monodisperse behavior under gel filtration chromatography (data not shown) as well as the results of the functional assay (see next section) strongly suggest that all proteins assume native conformations.

Functional integrity of the extracellular proteins produced by FATT-fusion strategy

MeCPA is known to exhibit relatively low selectivity toward the carboxy terminal amino acid residue of its substrates as compared to other carboxypeptidases,[10] 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).[11] Using the FATT-fusion strategy, we successfully obtained >10 mg of soluble MeCPA precursor (proMeCPA) from 1 L of bacterial culture. The ~40 kDa proMeCPA obtained after FATT cleavage by factor Xa was further incubated with chymotrypsin to remove the N-terminal autoinhibiting propeptide, resulting in an ~30-kDa active enzyme [Fig. 5(A) lanes 1 and 2]. The active MeCPA was able to remove the C-terminal hexahistidine tag from a model substrate protein at enzyme/substrate ratios higher than 1:1000, as confirmed by the electrophoretic mobility shift of the substrate [Fig. 5(A)] and the disappearance of the antipoly-His immunostaining (data not shown). The reaction was completely inhibited by the chelater 1,10-phenanthroline [Fig. 5(A), lane 8], corroborating the previous report that MeCPA is a Zn protease.[12] Activity measurement using the chromogenic synthetic substrate hippuryl-L-phenylalanine indicated that activated MeCPA had specific activity of ~15 units/mg protein, which is about 1/6 of that of commercially available bovine pancreas CPA.

Cutinase is a hydrolytic enzyme that degrades cutin, the polyester compound found in higher plants, and its strong esterolytic activity has been exploited in various industrial applications.[13] As cutinase is a secreted extracellular enzyme, its heterologous expression in *E. coli* has relied on production in the periplasmic space directed by bacterial secretion signals.[13] By using the FATT-fusion strategy, we could extract a large quantity of cutinase from the bacterial cytosol, and obtained soluble enzyme after the oxidative refolding described above [Fig. 4(B)]. When the esterase activity was evaluated using the chromogenic substrate *p*-nitrophenyl butylate, we confirmed that the purified enzyme possessed high activity with a k_{cat}/K_m value of 1849 $mM^{-1}\ s^{-1}$ [Fig. 5(B)], which was even higher than that obtained for the native enzyme produced in bacterial periplasm (1159 $mM^{-1}\ s^{-1}$).[14]

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 (antipeptide tag),[7] A5201A (anti-human nicastrin),[15] and SG/19 (anti-human β1 integrin).[16] In general, either periplasmic 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,[17] and 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 tag, 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 very similar to that obtained with the same scFv prepared from solubilized inclusion bodies by a conventional refolding method [Fig. 5(C)].[18] Furthermore, the A5201A and SG/19 scFvs prepared using the FATT-fusion strategy could detect their native antigens in enzyme-linked immunosorbent assay (ELISA)- and immunofluorescence-based assays, respectively [Fig. 5(D,E)]. As the functional refolding of the latter two scFvs from solubilized 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.

The benefits of FATT-tag fusion

In this study, we found that the unique properties of the APP hyperacidic region can be harnessed by fusing it to a target protein, enabling (1) near-perfect soluble expression in bacterial cytosol, (2) low-cost purification on an anion-exchange resin, and (3) a single-protocol renaturing/refolding method from the crude bacterial cytosol for many extracellular proteins that harbor disulfide bridges necessary for their structural and functional integrities. The idea of fusing a highly acidic sequence as a purification handle is not unprecedented; Hober and coworkers[19] reported the design of a negatively charged fusion partner aimed at facilitating anion exchange

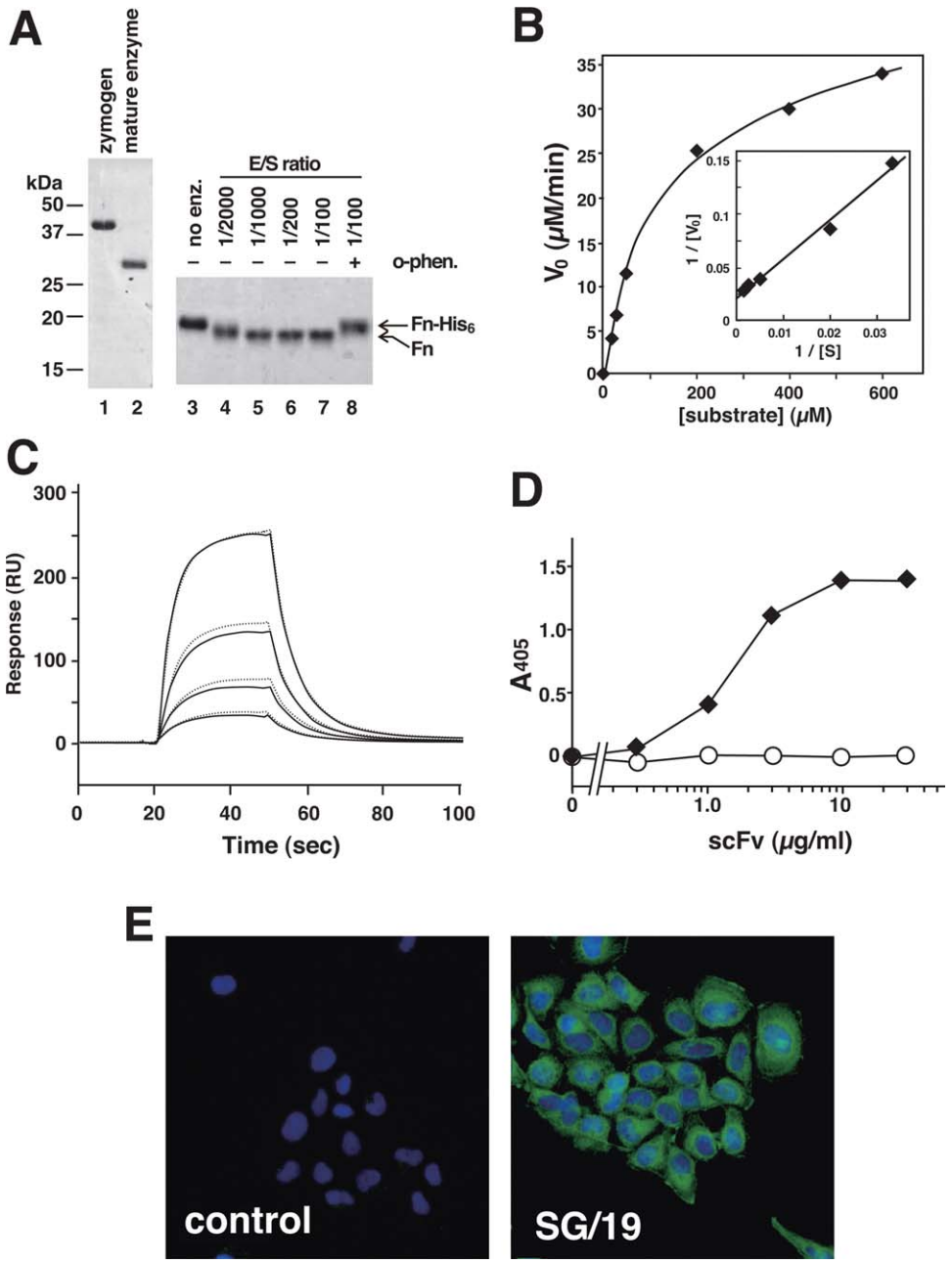

Figure 5. Biological activities of extracellular proteins prepared with FATT-assisted refolding. (A) Carboxypeptidase activity of refolded MeCPA. The purified MeCPA was in the inactive zymogen form, which migrates as a 40-kDa protein (Lane 1) and can be activated by treatment with chymotrypsin to yield the 30-kDa active enzyme (Lane 2). The C-terminal peptide trimming activity of the activated MeCPA was assessed by observing a digestion-dependent band shift for the model substrate (Fn-His_6) after incubation at varying enzyme/substrate ratio (wt/wt) in the absence (Lanes 3–7) or presence (Lane 8) of 2.5 m*M* 1,10-phenanthroline (*o*-phen). (B) Enzymatic activity of refolded cutinase. Release of *p*-nitrophenol by 2 n*M* enzyme was monitored spectrophotometrically at 400 nm ($\varepsilon = 15{,}400\ M^{-1}\ cm^{-1}$) and the initial velocity was plotted against substrate concentration. The kinetic numbers were derived from the double reciprocal plot (inset). (C) Binding kinetics of refolded P20.1 scFv toward antigen peptide assessed by surface plasmon resonance. P20.1 scFv was produced either using FATT-assisted soluble expression/refolding (solid lines) or via conventional insoluble expression/refolding method (dotted lines) and serially diluted solutions (62.5, 125, 250, and 500 n*M*) were flowed over a sensor chip surface bearing the "P4" antigen peptide (GYPGQV) for 30 s. (D) ELISA format binding assay for the refolded antinicastrin A5201A scFv. C-terminally Myc-tagged A5201A (◆) or P20.1 (○) scFvs were allowed to interact with antigen-coated wells and were detected by anti-Myc antibody. (E) 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.

purification. In this case, however, the tag was a structured protein and applied only to the soluble proteins, and no attempts were made to remove the tag after the purification. We are also given a wide variety of options when it comes to the soluble expression in the bacterial cytosol, from commercially available fusion systems to condition optimization for the bacterial growth, induction, and coexpression with accessory proteins.[20] Still, success depends greatly on the properties of particular target protein and one must test each condition empirically. Among the 21 proteins tested so far that are completely insoluble after bacterial expression without any tags, 20 showed either complete or greatly increased recovery in the soluble fraction when using the FATT-fusion strategy.

The solubility-enhancing effect from FATT fusion can be rationalized by its highly hydrophilic and charged nature, but that does not necessarily explain its ability to facilitate efficient refolding of denatured proteins in the presence of unrelated proteins in the bacterial cytosol. As it is generally thought that (re)folding and aggregation are competing reactions,[4] conditions that disfavor aggregation would in effect promote protein folding. In solution at neutral pH, the 14-kDa FATT tag polypeptide behaves as a particle with an effective exclusion volume corresponding to an 80-kDa globular protein [Fig. 2(C)]. A similar observation was reported for the negatively charged carbohydrate polymer hyaluronan, where a 12-kDa hyaluronan molecule had a hydrodynamic radius similar to 67-kDa albumin.[21] It is, therefore, predicted that the FATT tag portion occupies large average volume, "shielding" the aggregation-prone folding intermediates from intimate association with neighboring molecules.

Since the beneficial properties of FATT tag reported here are unlikely to be sequence-specific, it should be possible to redesign the tag by incorporating additional functional motifs or shuffling the amino acid sequence to obtain multiple custom-made FATT tags 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 efficiently as the N-terminally tagged versions (data not shown). Another intriguing possibility is the potential use of FATT tag as a way to improve/modify the bioactivity of recombinant proteins. The biophysical properties of FATT tag suggest that its fusion may have a similar effect to chemical modification with polyethylene glycol, which often improves *in vivo* stability and lowers the toxicity of the target protein. Combined with the fact that the tag sequence is derived from an abundant human protein and is hence minimally immunogenic, this unique fusion strategy may find various useful applications in the field of biotechnology that are not limited to the ones reported here.

Materials and Methods

Construction of expression vectors

The FATT-fusion expression cassette was obtained by polymerase chain reaction (PCR) amplifying the DNA segment corresponding the 97-residue hyperacidic region from human APP695 (a gift from K. Yoshikawa). With additional extension PCR cycles, an *Nde*I site and a FLAG epitope sequence (DYKDDDDV) were inserted at the 5′ end, while a "TARGET tag (P4 × 3)" sequence[18] followed by a factor Xa cleavage site (IEGR/GIP) and *Bam*HI site was inserted at the 3′ end. This cassette was then ligated into a pET11b vector (Novagen) following *Nde*I and *Bam*HI digestion, resulting in the plasmid pFATT3. DNA fragments coding for various target proteins, flanked by either *Bam*HI or *Bgl*II sites, were PCR-amplified from respective complementary DNA and cloned in-frame into the *Bam*HI site of the pFATT3. 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; *Fusarium solani* cutinase (synthetic DNA), residues 32–230; Rpt1, residues 1–467; Rpt2, residues 1–437; Rpt3, residues 1–428; Rpt4, residues 1–437; Rpt5, residues 1–434; and Rpt6, residues 1–405. All scFv constructs contained the V_H and V_L regions of their respective monoclonal antibody intervened by a $(GGGGS)_3$ linker and followed by a hexahistidine tag and a Myc tag.

Production and purification of FATT-tagged proteins

Expression plasmids were transformed into the *E. coli* strain BL21(DE3) cells and cultured overnight at 37°C. The overnight culture was diluted into 1 L of fresh LB medium. The culture was grown at 37°C until the OD600 reached 0.5 and then was induced with 1 m*M* isopropyl β-D−1-thiogalactopyranoside (IPTG), followed by an additional 4 h growth at 37°C. The cells were then harvested and suspended in 50 m*M* Tris, 150 m*M* NaCl, pH 8.0 [Tris-buffered saline (TBS)], and lysed using ultrasonic disruptor UD-201 (TOMY) in the presence of protease inhibitor cocktail containing leupeptin, pepstatin A, and phenylmethylsulfonyl fluoride (PMSF). The cell lysate was incubated for 30 min at room temperature (RT) with nuclease (Benzonase, Novagen) to partially digest genomic DNA, and centrifuged for 10 min at 15,000*g*. For purification of FATT-tagged proteins from the cytosol, the soluble fraction of cell lysate was directly loaded onto Q-sepharose FF (GE healthcare). The column was then washed with five column volumes of wash buffer (50 m*M* Tris, 400 m*M* NaCl, pH 8.0), followed by the same volume of

elution buffer (50 m*M* Tris, 1 *M* NaCl, pH 8.0). The eluted sample was dialyzed against TBS and treated with 2 units/mL of factor Xa (Novagen) for 16 h at 4°C to remove the tag. Tag-free target proteins were separated from cleaved FATT-tag and the remaining FATT-tagged protein by being passed through a column of Q-sepharose FF. Further purification was achieved by size exclusion chromatography on a Superdex 200 (GE healthcare) column equilibrated with TBS.

Oxidative refolding of FATT-tagged extracellular proteins directly from bacterial lysates

For the production of natively folded extracellular proteins with disulfide bonds, various proteins tagged with FATT-tag were expressed in the bacterial cytosol and then subjected to the "direct refolding" method coupled with one-step concentration by anion exchange resin as shown in Figure 4. Briefly, all the proteins in the cytosolic fraction were denatured by adding 6 *M* guanidine hydrochloride (Gdn-HCl), filtered (0.45 μm), and then serially diluted with a buffer containing 0.4 *M* arginine hydrochloride and 375 μ*M* oxidized glutathione over 20 h at 4°C with gentle stirring to achieve 100-fold dilution from the original sample. The renatured FATT-tagged proteins could then be concentrated in one step by capturing with and eluting from Q-sepharose FF anion exchange resin. Soluble but unstructured FATT-tagged proteins did not bind well to the resin prior to the refolding process, an observation for which we have no good explanation. Proteins eluted from the resin with a buffer containing 500 m*M* NaCl were pooled and dialyzed against TBS, then subjected to factor Xa treatment and further purification as described in the previous section.

Western blot analysis

The total cell lysates from the induced bacterial cultures were separated on SDS-PAGE using 5–20% gradient gels and transferred onto a PVDF membrane. After blocking with Blocking-One reagent (Nacalai Tesque), the membranes were incubated first with primary antibodies P20.1 IgG or anti-FLAG polyclonal antibody (Sigma Aldrich), then with horseradish peroxidase-conjugated secondary antibodies against mouse or rabbit immunoglobulin, respectively (Sigma Aldrich). Blots were visualized by ECL Plus reagent (GE-Healthcare) and recorded in an ImageQuant LAS4000 mini system (GE healthcare).

Assessment of biological activities of extracellular proteins

The carboxypeptidase activity of refolded MeCPA was assessed by the ability to remove C-terminal His-tag attached to a model substrate protein Fn10, a 94-residue fragment derived from the 10th FNIII repeat region of human fibronectin.[22] Purified proMeCPA was incubated with a chymotrypsin at a molar ratio of 1/83 for 1 h at RT to release the propeptide. After the inactivation of chymotrypsin by 1 m*M* PMSF, the activated MeCPA was mixed with the Fn10-His_6 at various enzyme:substrate ratios and incubated for 4 h at RT, followed by SDS-PAGE analysis on a 15% polyacrylamide gel. The esterase activity of cutinase was evaluated using the chromogenic substrate *p*-nitrophenylbutylate as described previously.[14] P20.1 scFv-antigen interactions were measured using BIACORE 2000 as described previously.[18] Binding of anti-human nicastrin A5201A scFv to the native antigen was assessed by an ELISA as follows. Microtiter plates (Nunc) were coated with recombinant nicastrin ectodomain fragment,[15] blocked with 1% bovine serum albumin (BSA), and incubated with varying concentrations of scFv for 1 h at RT. The bound scFvs were probed with anti-Myc polyclonal antibody (MBL) followed by peroxidase-conjugated secondary antibody. The antigen-binding activity of the purified scFv of anti-human β1 integrin SG/19[16] was evaluated by indirect immunofluorescence microscopy on β1-expressing HelaS3 cells. Briefly, HelaS3 cells were seeded onto 18-mm glass coverslips coated with poly-L-lysine. The cells were fixed for 15 min at RT in phosphate-buffered saline (PBS) containing 4% formaldehyde. After blocking with 1% BSA in PBS, they were then incubated with 2 μg/mL SG/19 scFv or control P20.1 scFv at RT for 1 h. After washing three times with PBS, cells were incubated with 1 μg/mL Alexa 488-conjugated anti-Myc IgG (MBL) and Hoechst33342 (Invitrogen) for 2 h at RT. 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).

Acknowledgments

We thank K. Tamura-Kawakami, E. Mihara, and Takuro Yasuda for their excellent technical support, M. Nakano for the preparation of the manuscript, and S. Thompson for critically editing the manuscript.

References

1. Gräslund S, Nordlund P, Weigelt J, Hallberg BM, Bray J, Gileadi O, Knapp S, Oppermann U, Arrowsmith C, Hui R, Ming J, dhe-Paganon S, Park HW, Savchenko A, Yee A, Edwards A, Vincentelli R, Cambillau C, Kim R, Kim SH, Rao Z, Shi Y, Terwilliger TC, Kim CY, Hung LW, Waldo GS, Peleg Y, Albeck S, Unger T, Dym O, Prilusky J, Sussman JL, Stevens RC, Lesley SA, Wilson IA, Joachimiak A, Collart F, Dementieva I, Donnelly MI, Eschenfeldt WH, Kim Y, Stols L, Wu R, Zhou M, Burley SK, Emtage JS, Sauder JM, Thompson D, Bain K, Luz J, Gheyi T, Zhang F, Atwell S, Almo SC, Bonanno JB, Fiser A, Swaminathan S, Studier FW, Chance MR, Sali A, Acton TB, Xiao R, Zhao L, Ma LC, Hunt JF, Tong L, Cunningham K, Inouye M, Anderson

S, Janjua H, Shastry R, Ho CK, Wang D, Wang H, Jiang M, Montelione GT, Stuart DI, Owens RJ, Daenke S, Schütz A, Heinemann U, Yokoyama S, Büssow K, Gunsalus KC (2008) Protein production and purification. Nat Methods 5:135–146.
2. Terpe K (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol 60:523–533.
3. Lichty JJ, Malecki JL, Agnew HD, Michelson-Horowitz DJ, Tan S (2005) Comparison of affinity tags for protein purification. Protein Expr Purif 41:98–105.
4. Eiberle MK, Jungbauer A (2010) Technical refolding of proteins: do we have freedom to operate? Biotechnol J 5:547–559.
5. Reinhard C, Hebert SS, De Strooper B (2005) The amyloid-beta precursor protein: integrating structure with biological function. EMBO J 24:3996–4006.
6. Gralle M, Ferreira ST (2007) Structure and functions of the human amyloid precursor protein: the whole is more than the sum of its parts. Prog Neurobiol 82:11–32.
7. 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. J Proteomics 73:1777–1785.
8. Takeuchi J, Tamura T (2004) Recombinant ATPases of the yeast 26S proteasome activate protein degradation by the 20S proteasome. FEBS Lett 565:39–42.
9. Chow MK, Amin AA, Fulton KF, Fernando T, Kamau L, Batty C, Louca M, Ho S, Whisstock JC, Bottomley SP, Buckle AM (2006) The REFOLD database: a tool for the optimization of protein expression and refolding. Nucleic Acids Res 34:D207–D212.
10. Austin BP, Tozser J, Bagossi P, Tropea JE, Waugh DS (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 Expr Purif 77:53–61.
11. Austin BP, Waugh DS (2012) Isolation of Metarhizium anisopliae carboxypeptidase A with native disulfide bonds from the cytosol of Escherichia coli BL21(DE3). Protein Expr Purif 82:116–124.
12. Joshi L, St Leger RJ (1999) Cloning, expression, and substrate specificity of MeCPA, a zinc carboxypeptidase that is secreted into infected tissues by the fungal entomopathogen Metarhizium anisopliae. J Biol Chem 274:9803–9811.
13. Carvalho CM, Aires-Barros MR, Cabral JM (1999) Cutinase: from molecular level to bioprocess development. Biotechnol Bioeng 66:17–34.
14. Goncalves AM, Serro AP, Aires-Barros MR, Cabral JM (2000) Effects of ionic surfactants used in reversed micelles on cutinase activity and stability. Biochim Biophys Acta 1480:92–106.
15. Hayashi I, Takatori S, Urano Y, Miyake Y, Takagi J, Sakata-Yanagimoto M, Iwanari H, Osawa S, Morohashi Y, Li T, Wong PC, Chiba S, Kodama T, Hamakubo T, Tomita T, Iwatsubo T (2012) Neutralization of the gamma-secretase activity by monoclonal antibody against extracellular domain of nicastrin. Oncogene 31:787–798.
16. Nagae M, Re S, Mihara E, Nogi T, Sugita Y, Takagi J (2012) Crystal structure of alpha5beta1 integrin ectodomain: atomic details of the fibronectin receptor. J Cell Biol 197:131–140.
17. Sun W, Xie J, Lin H, Mi S, Li Z, Hua F, Hu Z (2012) A combined strategy improves the solubility of aggregation-prone single-chain variable fragment antibodies. Protein Expr Purif 83:21–29.
18. Nogi T, Sangawa T, Tabata S, Nagae M, Tamura-Kawakami K, Beppu A, Hattori M, Yasui N, Takagi J (2008) Novel affinity tag system using structurally defined antibody–tag interaction: application to single-step protein purification. Protein Sci 17:2120–2126.
19. Hedhammar M, Graslund T, Uhlen M, Hober S (2004) Negatively charged purification tags for selective anion-exchange recovery. Protein Eng Des Sel 17:779–786.
20. Waugh DS (2005) Making the most of affinity tags. Trends Biotechnol 23:316–320.
21. Ohlson M, Sorensson J, Lindstrom K, Blom AM, Fries E, Haraldsson B (2001) Effects of filtration rate on the glomerular barrier and clearance of four differently shaped molecules. Am J Physiol Renal Physiol 281:F103–F113.
22. Takagi J, Erickson HP, Springer TA (2001) C-terminal opening mimics "inside–out" activation of integrin a5b1. Nature Struct Biol 8:412–416.