Hindawi Publishing Corporation
Journal of Biomedicine and Biotechnology
Volume 2009, Article ID 591923, 8 pages
doi:10.1155/2009/591923

Research Article

A Novel Method for High-Level Production of TEV Protease by Superfolder GFP Tag

Xudong Wu, Di Wu, Zhisheng Lu, Wentao Chen, Xiaojian Hu, and Yu Ding

Department of Physiology and Biophysics, School of Life Sciences, Fudan University, Shanghai 200433, China

Correspondence should be addressed to Yu Ding, yuding@fudan.edu.cn

Received 25 August 2009; Revised 25 October 2009; Accepted 2 December 2009

Recommended by Claudio M. Soares

Because of its stringent sequence specificity, tobacco etch virus (TEV) protease is widely used to remove fusion tags from recombinant proteins. Due to the poor solubility of TEV protease, many strategies have been employed to increase the expression level of this enzyme. In our work, we introduced a novel method to produce TEV protease by using visible superfolder green fluorescent protein (*sf*GFP) as the fusion tag. The soluble production and catalytic activity of six variants of *sf*GFP-TEV was examined, and then the best variant was selected for large-scale production. After purified by Ni-NTA affinity chromatography and Q anion exchange chromatography, the best variant of *sf*GFP-TEV fusion protease was obtained with purity of over 98% and yield of over 320 mg per liter culture. The *sf*GFP-TEV had a similar catalytic activity to that of the original TEV protease. Our research showed a novel method of large-scale production of visible and functional TEV protease for structural genomics research and other applications.

1. Introduction

Nowadays, it has been a popular way to fuse target proteins with various tags to facilitate expression and purification. An efficient combination of solubility-enhancing tags, such as maltose-binding protein (MBP) [1, 2], N-Utilization substance (NusA) [3], glutathione S-transferase (GST) [4], thioredoxin (TRX) [5], trigger factor [6], and SUMO [7], will promise high-throughput expression and purification methods for many target proteins and sometimes increases their solubility. However, these fusion tags may become a drawback for further structural and functional studies [8]. Therefore, the removal of these tags is necessary in many situations. Proteases such as enterokinase, thrombin, and factor Xa [9] as well as the more specific human rhinovirus 3C protease (3CP or PreScission [10]) and tobacco etch virus (TEV) protease [11] can fulfill the task to liberate fusion tags from target proteins.

The widely used TEV protease is the 27 kDa catalytic domain of the nuclear inclusion an (NIa) protease from tobacco etch virus [12]. Among various proteases, TEV protease outstands because of its high and unique specificity. It can recognize the canonical cleavage site, ENLYFQ/G [11] and the P1' position can tolerate substitutions with small amino acids [13]. Moreover, TEV protease can be used at temperature as low as 4°C with adequate efficiency to reduce the proteolysis of the target protein. Because of these advantages, nowadays, it is used more frequently than other proteases (enterokinase, thrombin, factor Xa, and human rhinovirus 3C protease) in structural genomic research projects.

Production of TEV protease in *E. coli* has been problematic due to its low solubility. To increase its soluble production, many strategies have been addressed. First, Kapust et al. [14] designed a more stable mutant of TEV protease named S219V. van den Berg et al. [15] obtained a mutant TEV_{SH} with production of 54 mg/L culture by directed evolution. Later, Fang et al. [16] increased the production to 65 mg/L culture using chaperone coexpression and low-temperature expression methods. More recently, Blommel and Fox [17] reported a combined approach raising the production to 400 mg/L culture while Kraft et al. developed a fluorogenic substrate which was useful to determine the TEV protease's expression and folding in vivo [18].

Fluorescent protein is widely used as gene reporter and protein marker, and so forth. However, existing variants of

green fluorescent protein (GFP) often misfold when fused to other proteins. Pédelacq et al. [19] reported a robustly folded GFP called "superfolder GFP" (*sf*GFP) which could fold well regardless of the folding status or solubility of its fusion partner in *E. coli*. Furthermore, *sf*GFP fusions are more soluble than conventional GFP fusions.

In our present work, considering the high thermodynamic stability, robust folding kinetics, and solubility of *sf*GFP fusions, we tempted to fuse *sf*GFP to TEV protease hoping that *sf*GFP would increase the soluble production of TEV protease. In order to minimize the possible stereohindrance of *sf*GFP that might decrease the activity of TEV protease, we further constructed 6 variants of *sf*GFP-TEV with different linkers of various lengths and composition between *sf*GFP and TEV. Then, the catalytic activity of *sf*GFP-TEV variants was tested and compared with that of the original TEV protease without *sf*GFP tag. Finally, we obtained one variant of *sf*GFP-TEV fusion protease with soluble production of over 320 mg/L culture. Compared with the original TEV protease, this variant of *sf*GFP-TEV has similar catalytic activity and is easy for detection during expression, purification, and applications because of the presence of green fluorescence. The results of our work also present the potential of superfolder GFP to become a solubility-enhancing fusion tag with fluorescence.

2. Materials and Methods

2.1. Materials. The bacterial hosts, *E. coli* DH5α, Rosetta (DE3) pLysS, and the vector pET21a were obtained from Novagen (Madison, WI). KOD Plus polymerase and the DNA ligation kit were purchased from Toyobo (Osaka, Japan). Nucleotides, agarose gel, the DNA extraction kit, and the PCR purification kit were purchased from Roche Diagnostics (Indianapolis, IN). Primer synthesis and DNA sequence analysis were performed by Invitrogen (Shanghai, China). Restriction endonucleases were purchased from Takara (Dalian, China). The nickel-nitrilotriacetic acid (Ni-NTA) superflow matrix was obtained from Qiagen (Chatsworth, CA). Q Sepharose Fast Flow was from GE Health (Sweden). Amylose Sepharose was purchased from New England Biolabs (Hitchin, UK). Bicinchoninic acid (BCA) Protein Assay Reagent Kit was from Pierce (Rockford, IL). Imidazole, D-glucose, and D-lactose were from Sigma (St Louis, MO). All other agents are of analytic purity. PRK793-TEV expression vector was a gift from Dr. Waugh [14].

2.2. Construction of sfGFP-TEV and TEV Expression Vectors. We have previously reconstructed an expression vector, designated pT7His, which contained the N-terminal His_{10} and C-terminal His_6 tags from the vector pET21a. The detailed vector construction procedure was similar to that of pT7470 with N-terminal His_6 and C-terminal His_6 tags [20]. We optimized the codon usage of superfolder GFP's cDNA by referring to its amino acid sequence [19]. The whole gene synthesis of superfolder GFP was accomplished by 2 rounds PCR with 18 central primers listed in Table 1, one 5′ primer 5′-GATATACATATGAGCAAAGGCGAAGAA-3′ and one 3′ primer 5′-GCCGGATCCGCCCCCGGAACC-CCCTCCGTTATTGTTATTCTTGTACAGCTCGTCCAT-3′. Considering that the C-terminal poly (R) in PRK793-TEV would decrease the solubility of TEV protease [17], we replaced the poly (R) with residue E to construct the plasmid TEV 238Δ by PCR with primers 5′-GGGGGTAGCGGC-GGTGGCAGCGGCGGAGAAAGCTTGTTTAAG-3′ and 5′-TTACTCGAGTCATTCATTCATGAGTTGAGTCGC-3′. We have constructed 6 recombinant *sf*GFP-TEV fusion proteins with different linkers. The linker region of *sf*GFP-TEV-His_6 Nd1–6 was listed in Table 2. The plasmid TEV 238Δ was also used as the PCR template to produce the control TEV protease. The PCR product was incorporated into the expression vector MBP-LTL-His_6 [21]. The final expression vector MBP-LTL-TEV-His_6 which produced TEV-His_6 (MBP tag was self-cleaved during expression) was employed as a control in further experiments.

2.3. Expression of sfGFP-TEV, TEV-His_6, and MBP-EGFP. The expression vectors mentioned above were transformed into *E. coli* strain Rosetta (DE3) pLysS. After the colony had grown overnight at 37°C in 5 mL of LB medium with 100 μg/mL ampicillin, 0.5 mL of the bacterial suspension was transferred into a 2L flask containing 250 mL autoinduction medium. (For 1 liter culture, we used 4 flasks to ensure the sufficient oxygen supply). The autoinduction medium was prepared as studier's original protocol [22]. Standard stock solutions include 20*P (1 M Na_2HPO_4, 1 M KH_2PO_4, and 0.5 M $(NH_4)_2SO_4$), 50*M (1.25 M Na_2HPO_4, 1.25 M KH_2PO_4, 2.5 M NH_4Cl, and 0.25 M Na_2SO_4), and 50*5052 (25% glycerol, 2.5% glucose, and 10% D-lactose); the working autoinduction medium was assembled by adding sterile concentrated stock solutions into sterile water. When the cells had grown (250 rpm) at 37°C to an optical density at 600 nm (OD600) of 0.6 (around 3 hours), the cells were cooled to 19°C and shaken at 250 rpm for 20 hours. Finally, the cells were collected by centrifugation at 6,000 ×g for 20 minutes and stored at −80°C. In order to reflect the real-time expression level of *sf*GFP-TEV, the induced *E. coli* cells in the autoinduction medium were collected at 0, 2, 4, 6, 8, 10, 12, 14, and 16 hours, respectively. The fluorescence of 100 μL *E. coli* cells in the 96-well plates was recorded by DTX 880 multimode detector (Beckman) using bottom reading method with 485 nm excitation filter and 535 nm emission filter.

2.4. Purification of sfGFP-TEV and TEV-His_6. The *sf*GFP-TEV-His_6 Nd1–6 recombinant proteins were all first purified by Ni-NTA affinity chromatography. The frozen cell pellet was thawed and resuspended in Buffer A (50 mM Tris–HCl [pH 8.0], 150 mM NaCl, 10% [v/v] glycerol, 20 mM imidazole). Then, the cells were lysed by sonication on ice and the lysate was cleared by two-round 20-minute centrifugation at 20,000 ×g. The retained supernatant was loaded onto a Ni-NTA Superflow column which was pre-equilibrated with Buffer A. After loading, the Ni-NTA column was washed

Table 1: Primers for whole gene synthesis of superfolder GFP.

No. and length	Oligo Sequences (5′ → 3′)
S1F 59	ATGAGCAAAGGCGAAGAACTGTTTACCGGCGTGGTGCCGATTCTGGTTGAACTGGATGG
S2R 59	GCCTTCGCCGCGCACGCTAAACTTATGGCCATTAACATCGCCATCCAGTTCAACCAGAA
S3F 59	TAGCGTGCGCGGCGAAGGCGAAGGCGATGCGACCAACGGCAAACTGACCCTGAAGTTTA
S4R 59	AGGGTCGGCCACGGCACCGGCAGTTTGCCGGTGGTGCAAATAAACTTCAGGGTCAGTTT
S5F 59	CGGTGCCGTGGCCGACCCTGGTGACCACCCTGACCTATGGCGTGCAGTGCTTTAGCCGC
S6R 59	CGGACTTAAAGAAATCATGGCGTTTCATGTGATCCGGATAGCGGCTAAAGCACTGCACG
S7F 59	CATGATTTCTTTAAGTCCGCGATGCCGGAAGGCTATGTTCAGGAACGCACCATTAGCTT
S8R 59	TTTCACTTCCGCGCGGGTCTTATAGGTGCCATCATCTTTAAAGCTAATGGTGCGTTCCT
S9F 59	GACCCGCGCGGAAGTGAAATTTGAAGGCGATACCCTGGTGAACCGCATTGAACTGAAAG
S10R 59	TTATGGCCCAGAATGTTGCCATCTTCTTTGAAGTCAATGCCTTTCAGTTCAATGCGGTT
S11F 59	GCAACATTCTGGGCCATAAACTGGAGTACAATTTCAACAGCCATAACGTGTATATTACC
S12R 59	TTTTGAAATTCGCTTTGATGCCGTTCTTCTGTTTGTCCGCGGTAATATACACGTTATGG
S13F 59	ATCAAAGCGAATTTCAAAATCCGCCATAATGTGGAAGATGGCAGCGTGCAGCTGGCGGA
S14R 59	CACCGGGCCATCGCCGATCGGGGTATTCTGCTGATAGTGATCCGCCAGCTGCACGCTGC
S15F 59	GATCGGCGATGGCCCGGTGCTGCTGCCGGACAATCACTACCTGAGCACCCAGTCCGTGC
S16R 54	CATGTGATCGCGTTTCTCATTCGGATCTTTGCTCAGCACGGACTGGGTGCTCAG
S17F 54	TGAGAAACGCGATCACATGGTTCTGCTGGAGTTTGTGACCGCGGCGGGTATCAC
S18R 44	CTTGTACAGCTCGTCCATGCCATGAGTGATACCCGCCGCGGTCA

with Buffer A with 40 mM imidazole. The column was equilibrated again with Buffer A and then eluted with Buffer B (50 mM Tris–HCl [pH 8.0], 150 mM NaCl, 10% [v/v] glycerol, 500 mM imidazole). The fluorescence of 100 μL Ni-NTA purification sample was also recorded by DTX 880 multimode detector (Beckman) using top reading method with 485 nm excitation filter and 535 nm emission filter.

The collected elution from Ni-NTA affinity chromatography was immediately diluted with 5 volume QA (50 mM Tris–HCl [pH 8.0], 10% [v/v] glycerol). The dilution was loaded onto a Q Sepharose Fast Flow column pre-equilibrated with QA. After washed with QA, the protein was eluted with a linear 0–0.9 M NaCl gradient by automatically mixing QA and QB (50 mM Tris–HCl [pH 8.0], 1 M NaCl, 10% [v/v] glycerol). Fractions were analyzed by SDS-PAGE and quantified by BandScan 4.30 (Glyko) and were pooled based on their purity and concentration. The concentration of pooled protein sample from Q anion exchange chromatography and elution sample from Ni-NTA affinity chromatography was determined by BCA method according to the reagent kit protocol (Pierce). The pooled protein was dialyzed in dialysis buffer (50 mM Tris–HCl [pH 8.0], 150 mM NaCl, 0.5 mM EDTA, 10% [v/v] glycerol) at 4°C and then diluted with storage buffer (50 mM Tris–HCl [pH 8.0], 150 mM NaCl, 0.5 mM EDTA, 80% [v/v] glycerol) to a protein concentration of ~2 mg/mL in 40% glycerol. The purified protein was finally stored at −20°C.

TEV-His_6 was purified by Ni-NTA affinity chromatography using the similar methods described above. The purified protein was dialyzed in dialysis buffer and diluted with storage buffer to a protein concentration of ~1 mg/mL in 40% glycerol. The purified TEV-His_6 was stored at −20°C.

2.5. Purification of TEV Protease Substrate MBP-EGFP. For the purification of MBP-EGFP, the cell pellet was thawed and resuspended in Amylose A buffer (50 mM Tris–HCl [pH 8.0], 150 mM NaCl, 10% [v/v] glycerol). After sonication, the supernatant was retained by two-round 20-minute centrifugation at 20,000 ×g and then loaded onto the Amylose Sepharose Column pre-equilibrated with Amylose A buffer. MBP-EGFP was eluted with Amylose B buffer (50 mM Tris–HCl [pH 8.0], 150 mM NaCl, 10% [v/v] glycerol, 20 mM maltose). The purified protein sample was analyzed by SDS-PAGE and quantified by BandScan 4.30 (Glyko). The purified MBP-EGFP was dialyzed in dialysis buffer (150 mM NaCl, 10% [v/v] glycerol, 1 mM EDTA) at 4°C and then stored at 4°C.

2.6. Activity Assay of sfGFP-TEV and TEV-His_6. The catalytic activity of *sf*GFP-TEV-His_6 Nd1–6 and TEV-His_6 was determined by cleaving the substrate MBP-EGFP which contained a TEV cleavage site between MBP and EGFP. Prior to activity assay, the protein concentration of *sf*GFP-TEV-His_6 Nd1–6, TEV-His_6, and MBP-EGFP was determined by BCA method according to the reagent kit protocol (Pierce). The time course assay was conducted at 17°C for a given incubation time (0, 5, 10, 20, 40, 60, 90, 120, 180, and 240 minutes, respectively). The mass ratio of substrate to enzyme (calculated by the mass of effective TEV protease) is 100 : 1. At any given time, the reaction was stopped by adding 3×loading buffer (150 mM Tris-HCl [pH 6.8], 300 mM DTT, 6% [w/v] SDS, 0.06% [w/v] bromophenol blue, 30% [v/v] glycerol). The samples were boiled at 95°C for 3 minutes and then loaded onto 12% SDS-PAGE gel for electrophoresis. After visualized by staining with Coomassie G-250, the gel was quantified by BandScan 4.30 (Glyko) to establish the

Table 2: Linker region of *sf*GFP-TEV-His_6 Nd1–6.

Construction Code	Abbreviation	Linker length (aa)	Anticipated linker composition
Sf GFP-TEV-His_6 Nd1	Nd1	2	...THG/GS/RD...
Sf GFP-TEV-His_6 Nd2	Nd2	5	...THG/GSKGP/RD...
Sf GFP-TEV-His_6 Nd3	Nd3	8	...THG/NNPGSKGP/RD...
Sf GFP-TEV-His_6 Nd6	Nd6	8	...THG/GSNLFKGP/RD...
Sf GFP-TEV-His_6 Nd4	Nd4	11	...THG/NNPGSNLFKGP/RD...
Sf GFP-TEV-His_6 Nd5	Nd5	14	...THG/MDPNNPGSNLFKGP/RD...

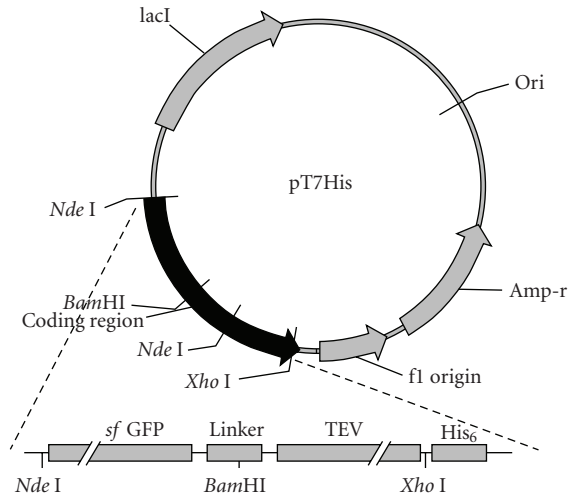

Figure 1: Maps of expression vector for *sf*GFP-TEV-His_6 Nd1–6 used in our work. The vector map was created by Vector NTI software. All variants of *sf*GFP-TEV-His_6 use the same vector pT7His derived from pET21a. The coding region in black will produce *sf*GFP-TEV-His_6 Nd1–6.

time-course curve. The reaction condition was 75 mM NaCl, 0.5 mM EDTA, 25 mM Tris-HCl 8.0, and 10% [v/v] glycerol.

3. Results and Discussion

3.1. Construction of Expression Vector for sfGFP-TEV and TEV. In order to maximize the expression level of the recombinant *sf*GFP-TEV proteases, we first synthesized the *sf*GFP gene according to the synonymous codon choice which is optimal for the *Escherichia coli* translational system. Figure 1 shows the vector map we used for high-level expression of *sf*GFP-TEV-His_6 Nd1–6. The *sf*GFP-TEV coding sequence was cloned to the pET derived vector pT7His which possesses the strong bacteriophage T7 promoter, ensuring the high level expression of target protein. Considering that the linker between *sf*GFP and TEV might have effects on the stability and catalytic activity of fusion protease, we constructed 6 variants of *sf*GFP-TEV-His_6 with different linkers. The linker here is defined as the peptide between C-terminus of *sf*GFP "THG" and N-terminus of TEV "RDYNP." The composition of different linkers with lengths varying from 2 to 14aa could be referred to Table 2. We also incorporated a small peptide "GGG" at the C-terminus of TEV; so the C-terminuses of *sf*GFP-TEV-His_6 Nd1–6 and TEV-His_6 are all "LMNEGGGLEHHHHHH." Our first attempt of *sf*GFP-TEV vector construction did not include the GGG small peptide between TEV and C-terminus His6 tag. However, during the Ni-NTA purification step, more than 70% expressed fusion protein did not bind with the Ni-NTA resin (data not shown). Perhaps the steric structure of TEV hindered His6 tag from binding with Ni-NTA resin. So we added the flexible GGG peptide between TEV and His6 tag. Almost all of the new version fusion protein can bind with Ni-NTA in the buffer containing relatively high concentration (20 mM) of imidazole.

3.2. Fusion of sfGFP to TEV Greatly Increases the Soluble Production of TEV Protease. After autoinduction, *sf*GFP-TEV-His_6 Nd1–6 were all purified by Ni-NTA affinity chromatography and Q anion exchange chromatography. After purification, there was an obvious main band around the molecular weight of 53 kDa (Figures 2(a) and 2(b)). Table 3 summarizes the purification results from 1-L culture medium. According to Bandscan software analysis, all variants of fusion protease were obtained with over 96% purity. Among them, Nd2, Nd4, and Nd5 were purified with over 98% purity. With the fusion of *sf*GFP, all variants could be purified by two-step chromatography with soluble production of over 200 mg. In particular, we could obtain around 320 mg of *sf*GFP-TEV Nd2 from 1-L culture medium. Because the molecular weight of *sf*GFP-TEV Nd2 and TEV-His_6 was 53.8 kDa and 28.8 kDa, respectively, 320 mg/L of *sf*GFP-TEV Nd2's effective TEV composition was close to 171 mg/L ($320 * 28.8/53.8 = 171$) of TEV-His_6. We also constructed the control expression vector for TEV protease without any tags, but there was almost no detectable TEV protease expressed under the same induction condition (data not shown). Therefore, the fusion of *sf*GFP to TEV significantly increases the soluble production of TEV protease.

3.3. Purification of TEV-His_6 and MBP-EGFP. We have also expressed and purified TEV-His_6 as control and MBP-EGFP as TEV protease's substrate. During expression, the MBP tag of MBP-LTL-TEV-His_6 would be cleaved and then TEV-His_6 was released. By Ni-NTA affinity chromatography and further dialysis, about 140 mg of TEV-His_6 could be obtained from 1-L culture medium with around 98% purity. The

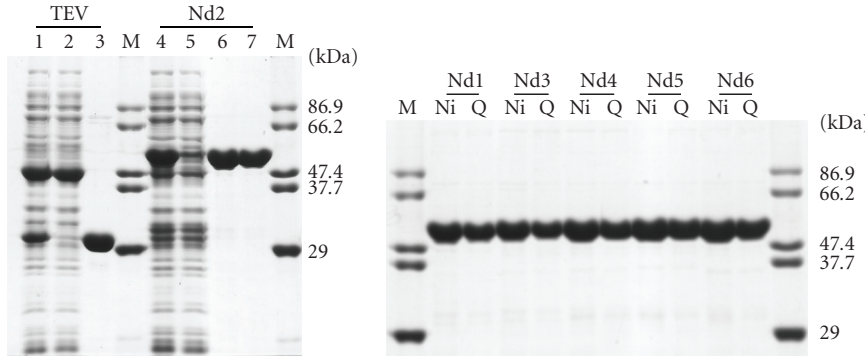

Figure 2: SDS-PAGE analysis of the expression and purification of *sf*GFP-TEV-His_6 Nd1–6 and TEV-His_6. (a) Lanes 1–3 are the result for TEV-His_6: Lane 1: autoinduced whole-bacterial lysate; Lane 2: flow through from Ni-NTA affinity chromatography; Lane 3: the purified TEV-His_6 eluted from Ni-NTA affinity chromatography. Lanes 4–7 are the result for *sf*GFP-TEV-His_6 Nd2: Lane 4: autoinduced whole-bacterial lysate; Lane 5: flow through from Ni-NTA affinity chromatography; Lane 6: the purified *sf*GFP-TEV-His_6 Nd2 eluted from Ni-NTA affinity chromatography; Lane 7: the pooled sample of purified *sf*GFP-TEV-His_6 Nd2 eluted from Q anion exchange chromatography. (b) The results for *sf*GFP-TEV-His_6 Nd1 and Nd3–6. "Ni" represents the results of the purified protein eluted from Ni-NTA affinity chromatography; "Q" represents the pooled sample of purified protein eluted from Q anion exchange chromatography. "M" in (a) and (b) represents the protein marker.

Table 3: Purification results of *sf*GFP-TEV-His_6 Nd1–6 and TEV-His_6 collected from 1-L of expression culture medium.

Abbreviation	Theoretical MW (kDa)	Total protein after Ni-NTA (mg)	Purity	Total protein after High Q (mg)	Purity	Total activity (μmol/h)	Specific activity (μmol/h/mg)
Nd1	53.5	244 ± 1	90%	221 ± 2	96%	7.3 ± 0.2	0.067 ± 0.002
Nd2	53.8	334 ± 2	94%	323 ± 1	98%	13.7 ± 0.3	0.085 ± 0.002
Nd3	54.1	220 ± 4	93%	211 ± 3	97%	8.0 ± 0.1	0.076 ± 0.001
Nd4	54.5	304 ± 6	93%	293 ± 4	98%	11.9 ± 0.5	0.082 ± 0.003
Nd5	55.0	299 ± 1	93%	288 ± 2	98%	11.3 ± 0.6	0.079 ± 0.005
Nd6	54.2	308 ± 5	93%	300 ± 4	96%	12.0 ± 0.5	0.080 ± 0.003
TEV	28.8	140 ± 3	98%	—	—	12.6 ± 0.6	0.090 ± 0.005

electrophoresis results show that molecular weight of TEV-His_6 is around 29 kDa (Figure 2(a)). The substrate MBP-EGFP could also be purified with over 95% purity by Amylose affinity chromatography.

3.4. Cleavage Activity Assay of sfGFP-TEV and TEV. The cleavage activity assay of *sf*GFP-TEV-His_6 Nd1–6 and TEV-His_6 could be determined by cleaving the substrate MBP-EGFP at the cleavage site "ENLYFQ/G" between MBP and EGFP. By SDS-PAGE, the remaining MBP-EGFP could be separated sufficiently with released MBP and EGFP (Figure 3(a)). After we set the quantity of MBP-EGFP at 0 min as 100%, the time course curve could be plotted by quantitatively analyzing the digested MBP-EGFP at the given time. Figure 3(b) shows the time course curve of *sf*GFP-TEV-His_6 Nd1–6, TEV-His_6, and 2% TEV-His_6. Compared with the time course curve of TEV-His_6, we found that *sf*GFP-TEV-His_6 Nd1–6 had different degrees of loss of catalytic activity. Among them, Nd2 had the closest curve to TEV-His_6. Ranking the cleavage rate at 60 minutes, the second highest ranked Nd2 could digest around 66% substrate, which retained about 95% catalytic activity of TEV-His_6. Moreover, TEV-His_6 and all variants of *sf*GFP-TEV-His_6 except Nd1 could efficiently cleave over 98% substrate after incubation for 4 hours at 17°C. However, the control 2% TEV-His_6 could only cleave less than 7% substrate under the same condition (Figure 3(c)). In conclusion, *sf*GFP-TEV-His_6 Nd2 retained the most of catalytic activity among all variants.

Fusion tags are widely used to facilitate protein expression and purification. However, due to its drawback in structural and functional studies, these tags always need to be removed by various proteases. TEV protease is an ideal protease receiving most attention, thanks to its high specificity as well as toleration of a wide range of temperatures and presence of detergents [23]. One bottleneck for TEV protease is low soluble production due to its poor solubility. Researchers have tried many strategies including in silico design [24], direct-evolution [15], or coexpression with chaperone to increase its soluble production. These efforts have raised the production from 1 mg/L to 65 mg/L culture [16]. More recently, Blommel and Fox reported a

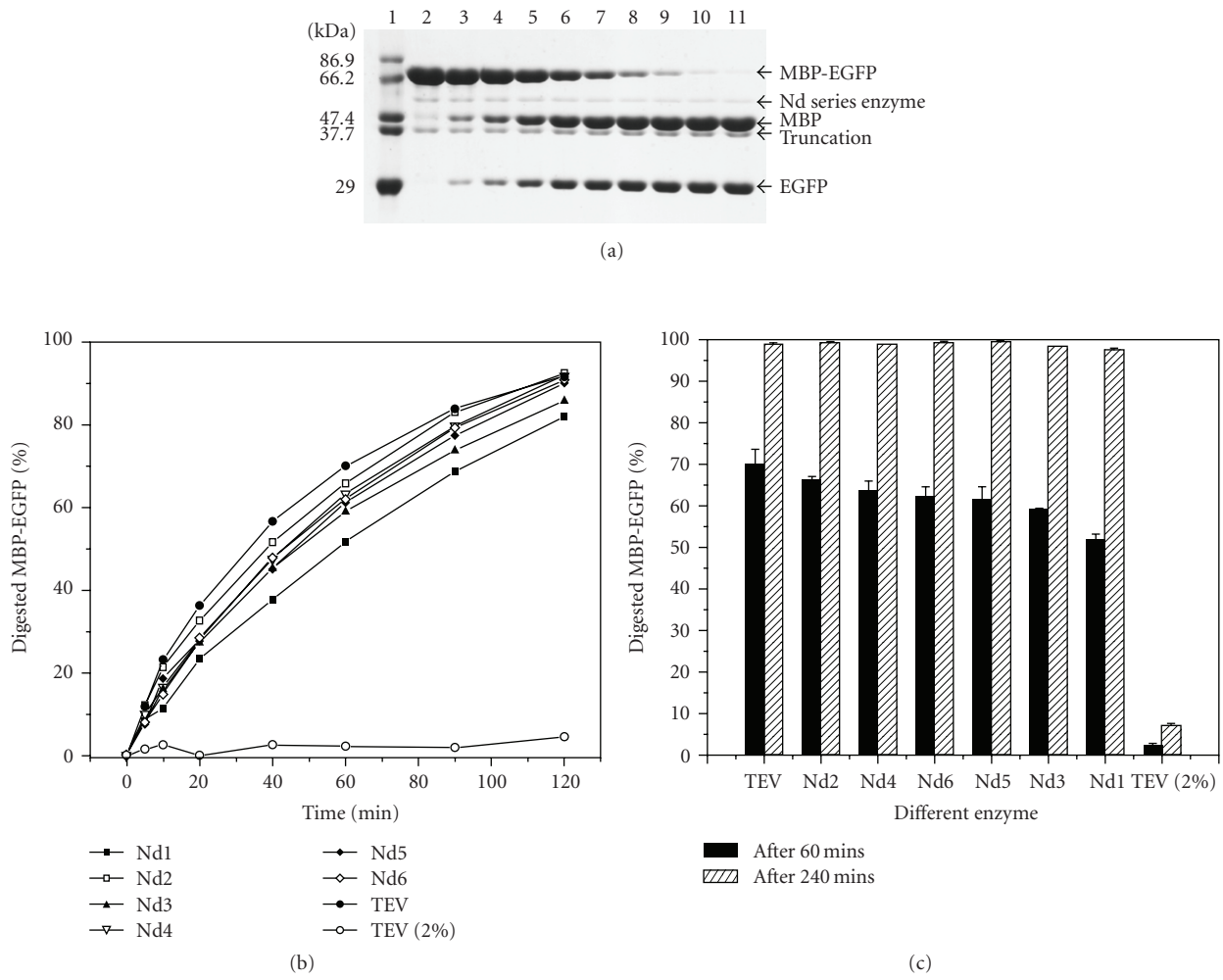

Figure 3: SDS-PAGE and time course analysis of catalytic activity on MBP-EGFP of *sf*GFP-TEV-His$_6$ and TEV-His$_6$. (a) One representative of SDS-PAGE analysis. 100 μg MBP-EGFP was incubated with 2 μg *sf*GFP-TEV-His$_6$ Nd2. Lane 1 is the protein marker. Lanes 2–11 represent different incubation time (0, 5, 10, 20, 40, 60, 90, 120, 180, 240 minutes, resp.). (b) The time course curve analysis. The time course curve was plotted according to the quantitative analysis of the SDS-PAGE by Bandscan 4.30. Each data point was the average of three independent tests. (c) Bar representation based on time course curve. The black bars represent the percentage of the digested MBP-EGFP after incubation for 60 minutes. The striped bars represent the percentage after incubation for 240 minutes. The ranking of the bar is based on the sorting result of cleavage efficiency after incubation for 60 minutes. The height of the bar is the average of three independent tests with standard error on top of the bar.

production of 400 mg/L culture by optimizing each step in expression, and purification [17]. However, the whole process of expression, purification and characterization of recombinant TEV protease was not visible to naked eye. Our attempts to express recombinant TEV protease fused with commonly used GST, TRX, and NusA tags all failed (data not shown). GST and TRX fused TEV proteases were most in the inclusion body and NusA fusion strategy gave less than 50% full length fusion protein.

In this paper, we introduce a novel method to increase the soluble production of TEV protease by fusing *sf*GFP to TEV protease. The results show that the production of *sf*GFP-TEV-His$_6$ Nd2 fusion protease reached 320 mg/L culture. Thanks to *sf*GFP's high folding kinetics, thermodynamic stability, *sf*GFP might work as a platform for the folding of TEV protease to prevent the formation of inclusion body. Compared with MBP which brings high metabolic burden for the host, *sf*GFP is much smaller and has fluorescence easy for detection. Figure 4 showed that the expression (Figure 4(a)) and purification (Figure 4(b)) procedure of *sf*GFP-tagged fusion protein can be monitored and quantified real-time by the fluorescence emitted from *sf*GFP, thus greatly simplified the procedure of *sf*GFP tagged target proteins expression and purification. We suggest that *sf*GFP could be employed as a colored solubility-enhancing tag for other small proteins with poor soluble production.

Catalytic activity is another important factor to be examined in our work. We constructed 6 variants of *sf*GFP-TEV-His$_6$ in all. The catalytic tests show that *sf*GFP-TEV-His$_6$ Nd2 with a linker of only five residues "GSKGP" has the closest catalytic activity to TEV-His$_6$. After one-hour incubation at 17°C, over 65% MBP-EGFP could be cleaved

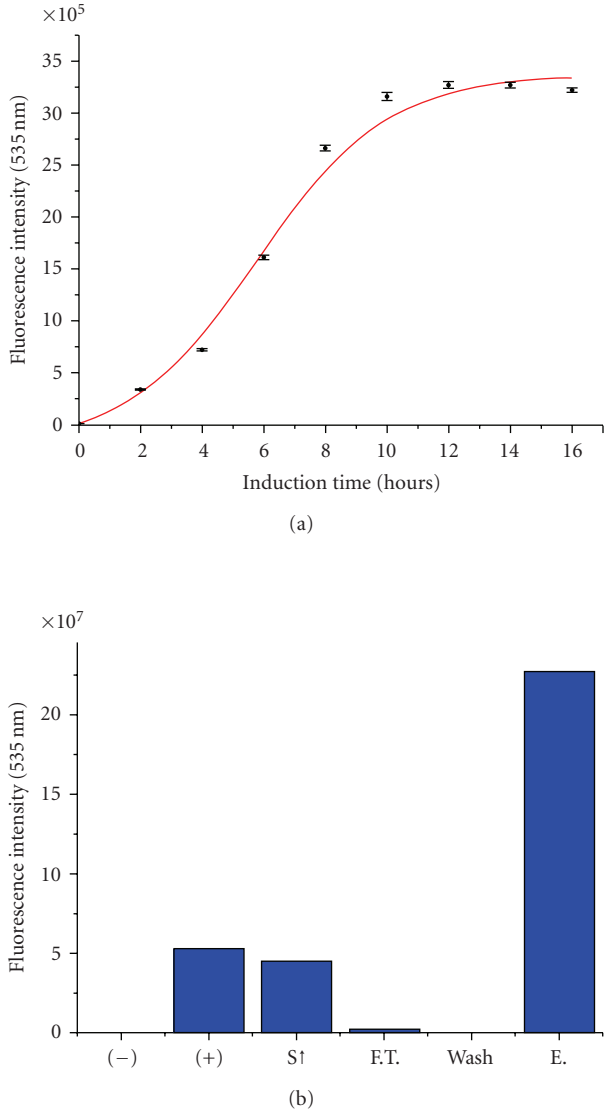

Figure 4: The quantification of *sf*GFP fluorescence during *sf*GFP-TEV-His_6 Nd2 expression and purification. (a) The fluorescence intensity versus time curve of *sf*GFP-TEV-His_6 Nd2 by autoinduction at 19°C. The fluorescence of 100 μL cultured *E. coli* cells in the autoinduction medium was collected using bottom reading method with 485 nm excitation filter and 535 nm emission filter. The height of the bar is the average of three independent tests with standard error on top of the bar. (b) One representative *sf*GFP fluorescence quantification during Ni-NTA purification. The fluorescence was collected using top reading method with 485 nm excitation filter and 535 nm emission filter. (−): uninduced cell lysate; (+): induced whole cell lysate; S↑: supernatant of the sonication; F.T.: flow through fraction from Ni-NTA chromatography; Wash: wash fraction from Ni-NTA chromatography; E.: eluated fraction by imidazole from Ni-NTA chromatography.

by *sf*GFP-TEV-His_6 Nd2 while TEV-His_6 cleaved around 70% substrate (Figure 3). In contrast, *sf*GFP-TEV-His_6 Nd1 has the lowest specific activity, which might be explained by the importance of three residues "KGP" on the correct folding and stability of TEV protease.

When preserved in 4°C for a long time, TEV-His_6 was not stable and would precipitate and completely lose the catalytic activity within one week (data not shown). However, *sf*GFP-TEV-His_6 Nd2 would not precipitate for more than one month and still retained about 60% catalytic activity, which showed much higher stability than original TEV-His_6. The *sf*GFP tag not only increased the solubility of the target protein during expression and purification but also increased its stability. Though the increase of effective TEV protease yield of *sf*GFP-TEV was only ~22% (from 140 mg to 171 mg TEV-His_6 per liter culture), during the long time cleavage experiment, the increased stability of *sf*GFP-TEV significantly outrun the original TEV protease widely used. This feature is vital because structural genomics required large-scale production of tag-free target proteins by TEV protease. Besides, the fluorescence characteristic of *sf*GFP tag provided an accurate, visible, and high-throughput measurement to quantify the fused target protein. The trace existence of *sf*GFP tagged TEV can be sensitive and easily detected by spectrofluorometer. By detecting the *sf*GFP fluorescence intensity, we can also accurately quantify the recombinant *sf*GFP-TEV protease. Like the original TEV protease, the His_6 tag of *sf*GFP-TEV makes it very easy to remove *sf*GFP-TEV from cleaved target protein by Ni-NTA chromatography after cleavage experiment.

In nature, evolution has shown its power of merging different domains to create a novel enzyme with great property. With rational design, we can also take advantage of available proteins to improve the property of certain enzymes. Our research showed that *sf*GFP tag significantly improved the solubility, expression level, and stability of TEV protease, which is important for the large-scale production of functional TEV protease used in structural genomics research.

Abbreviations

BCA: Bicinchoninic acid
GFP: Green fluorescent protein
GST: Glutathione S-transferase
Ni-NTA: Nickel-nitrilotriacetic acid
*sf*GFP: Superfolder green fluorescent protein
TEV: Tobacco etch virus protease
TRX: Thioredoxin.

Acknowledgments

The authors especially thank Professor Zhihong Zhang for the critical reviews, additions, and useful suggestions on this manuscript drafts. The work was supported by the Grants (no. 30600107, 30500113, and 30670499) from the National Natural Science Foundation of China, Shanghai Leading Academic Discipline Project (Project no. B111), National Talent Training Fund in Basic Research of China (no. J0630643), and Xi Yuan Scholar Program (2008).

References

[1] C. Di Guan, P. Li, P. D. Riggs, and H. Inouye, "Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein," *Gene*, vol. 67, no. 1, pp. 21–30, 1988.

[2] K. D. Pryor and B. Leiting, "High-level expression of soluble protein in *Escherichia coli* using a His6-tag and maltose-binding-protein double-affinity fusion system," *Protein Expression and Purification*, vol. 10, no. 3, pp. 309–319, 1997.

[3] G. D. Davis, C. Elisee, D. M. Mewham, and R. G. Harrison, "New fusion protein systems designed to give soluble expression in *Escherichia coli*," *Biotechnology and Bioengineering*, vol. 65, no. 4, pp. 382–388, 1999.

[4] D. B. Smith and K. S. Johnson, "Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase," *Gene*, vol. 67, no. 1, pp. 31–40, 1988.

[5] E. R. LaVallie, E. A. DiBlasio, S. Kovacic, K. L. Grant, P. F. Schendel, and J. M. McCoy, "A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm," *Nature Biotechnology*, vol. 11, no. 2, pp. 187–193, 1993.

[6] O. Kandror, M. Sherman, M. Rhode, and A. L. Goldberg, "Trigger factor is involved in GroEL-dependent protein degradation in *Escherichia coli* and promotes binding of GroEL to unfolded proteins," *EMBO Journal*, vol. 14, no. 23, pp. 6021–6027, 1995.

[7] J. G. Marblestone, S. C. Edavettal, Y. Lim, P. Lim, X. Zuo, and T. R. Butt, "Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO," *Protein Science*, vol. 15, no. 1, pp. 182–189, 2006.

[8] M. H. Bucher, A. G. Evdokimov, and D. S. Waugh, "Differential effects of short affinity tags on the crystallization of *Pyrococcus furiosus* maltodextrin-binding protein," *Acta Crystallographica Section D*, vol. 58, no. 3, pp. 392–397, 2002.

[9] R. J. Jenny, K. G. Mann, and R. L. Lundblad, "A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa," *Protein Expression and Purification*, vol. 31, no. 1, pp. 1–11, 2003.

[10] M. G. Cordingley, P. L. Callahan, V. V. Sardana, V. M. Garsky, and R. J. Colonno, "Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro," *Journal of Biological Chemistry*, vol. 265, no. 16, pp. 9062–9065, 1990.

[11] T. D. Parks, K. K. Leuther, E. D. Howard, S. A. Johnston, and W. G. Dougherty, "Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase," *Analytical Biochemistry*, vol. 216, no. 2, pp. 413–417, 1994.

[12] J.-A. Daros, M. C. Schaad, and J. C. Carrington, "Functional analysis of the interaction between VPg-proteinase (NIa) and RNA polymerase (NIb) of tobacco etch potyvirus, using conditional and suppressor mutants," *Journal of Virology*, vol. 73, no. 10, pp. 8732–8740, 1999.

[13] R. B. Kapust, J. Tözsér, T. D. Copeland, and D. S. Waugh, "The P1′ specificity of tobacco etch virus protease," *Biochemical and Biophysical Research Communications*, vol. 294, no. 5, pp. 949–955, 2002.

[14] R. B. Kapust, J. Tözsér, J. D. Fox, et al., "Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency," *Protein Engineering*, vol. 14, no. 12, pp. 993–1000, 2001.

[15] S. van den Berg, P.-Å. Löfdahl, T. Härd, and H. Berglund, "Improved solubility of TEV protease by directed evolution," *Journal of Biotechnology*, vol. 121, no. 3, pp. 291–298, 2006.

[16] L. Fang, K.-Z. Jia, Y.-L. Tang, D.-Y. Ma, M. Yu, and Z.-C. Hua, "An improved strategy for high-level production of TEV protease in *Escherichia coli* and its purification and characterization," *Protein Expression and Purification*, vol. 51, no. 1, pp. 102–109, 2007.

[17] P. G. Blommel and B. G. Fox, "A combined approach to improving large-scale production of tobacco etch virus protease," *Protein Expression and Purification*, vol. 55, no. 1, pp. 53–68, 2007.

[18] M. Kraft, D. Radke, G. D. Wieland, P. F. Zipfel, and U. Horn, "A fluorogenic substrate as quantitative in vivo reporter to determine protein expression and folding of tobacco etch virus protease in *Escherichia coli*," *Protein Expression and Purification*, vol. 52, no. 2, pp. 478–484, 2007.

[19] J.-D. Pédelacq, S. Cabantous, T. Tran, T. C. Terwilliger, and G. S. Waldo, "Engineering and characterization of a superfolder green fluorescent protein," *Nature Biotechnology*, vol. 24, no. 1, pp. 79–88, 2006.

[20] Y. Ding, W. Jiang, Y. Su, H. Zhou, and Z. Zhang, "Expression and purification of recombinant cytoplasmic domain of human erythrocyte band 3 with hexahistidine tag or chitin-binding tag in *Escherichia coli*," *Protein Expression and Purification*, vol. 34, no. 2, pp. 167–175, 2004.

[21] Y. Wu, Y. Zhou, J. Song, X. Hu, Y. Ding, and Z. Zhang, "Using green and red fluorescent proteins to teach protein expression, purification, and crystallization," *Biochemistry and Molecular Biology Education*, vol. 36, no. 1, pp. 43–54, 2008.

[22] F. W. Studier, "Protein production by auto-induction in high density shaking cultures," *Protein Expression and Purification*, vol. 41, no. 1, pp. 207–234, 2005.

[23] A. K. Mohanty, C. R. Simmons, and M. C. Wiener, "Inhibition of tobacco etch virus protease activity by detergents," *Protein Expression and Purification*, vol. 27, no. 1, pp. 109–114, 2003.

[24] L. D. Cabrita, D. Gilis, A. L. Robertson, Y. Dehouck, M. Rooman, and S. P. Bottomley, "Enhancing the stability and solubility of TEV protease using in silico design," *Protein Science*, vol. 16, no. 11, pp. 2360–2367, 2007.