J Biomol NMR (2010) 48:147–155
DOI 10.1007/s10858-010-9445-5

ARTICLE

Efficient protein production method for NMR using soluble protein tags with cold shock expression vector

Kokoro Hayashi · Chojiro Kojima

Received: 18 August 2010 / Accepted: 25 August 2010 / Published online: 16 September 2010

Abstract The *E. coli* protein expression system is one of the most useful methods employed for NMR sample preparation. However, the production of some recombinant proteins in *E. coli* is often hampered by difficulties such as low expression level and low solubility. To address these problems, a modified cold-shock expression system containing a glutathione S-transferase (GST) tag, the pCold-GST system, was investigated. The pCold-GST system successfully expressed 9 out of 10 proteins that otherwise could not be expressed using a conventional *E. coli* expression system. Here, we applied the pCold-GST system to 84 proteins and 78 proteins were successfully expressed in the soluble fraction. Three other cold-shock expression systems containing a maltose binding protein tag (pCold-MBP), protein G B1 domain tag (pCold-GB1) or thioredoxin tag (pCold-Trx) were also developed to improve the yield. Additionally, we show that a C-terminal proline tag, which is invisible in ^{1}H-^{15}N HSQC spectra, inhibits protein degradation and increases the final yield of unstable proteins. The purified proteins were amenable to NMR analyses. These data suggest that pCold expression systems combined with soluble protein tags can be utilized to improve the expression and purification of various proteins for NMR analysis.

Keywords Protein expression · *E. coli* · pCold-GST · Protein G B1 domain · Thioredoxin · Maltose-binding protein · Poly-proline tag

Background

Large amounts of protein sample are generally required for NMR analysis. *E. coli* expression systems provide many advantages for the high-level production of stable isotope-labeled proteins. However, the production of some recombinant proteins in *E. coli* is often hampered by difficulties such as low expression level and low solubility. To address these problems, a variety of methods have been developed (Makrides 1996).

One approach to overcome the difficulties is use of a protein tag system. Soluble protein tags have been used to enhance the expression level, stability and solubility of fusion proteins and to facilitate subsequent purification. Glutathione S-transferase (GST) (Smith and Johnson 1988), protein G B1 domain (GB1) (Bao et al. 2006; Zhou et al. 2001), thioredoxin (Trx) (Yasukawa et al. 1995), maltose-binding protein (MBP) (di Guan et al. 1988) and other proteins have been employed as soluble protein tags (Esposito and Chatterjee 2006).

Another approach to overcome the aforementioned difficulties concerns the development of expression vector systems. For instance, a series of cold-shock vectors, pCold I, II, III and IV, was developed in 2004 (Qing et al. 2004).

Electronic supplementary material The online version of this article (doi:10.1007/s10858-010-9445-5) contains supplementary material, which is available to authorized users.

K. Hayashi · C. Kojima
Graduate School of Biological Sciences,
Nara Institute of Science and Technology (NAIST),
8916-5 Takayama, Ikoma, Nara 630-0192, Japan

C. Kojima (✉)
Institute for Protein Research, Osaka University,
3-2 Yamadaoka, Suita, Osaka 565-0871, Japan
e-mail: kojima@protein.osaka-u.ac.jp

Present Address:
K. Hayashi
Analysis Technology Center, Fujifilm Corporation, 210
Nakanuma, Minamiashigara, Kanagawa 250-0193, Japan

In this expression vector series, the expression of target protein is under control of the *cspA* promoter and *lac* operator, thus protein expression is induced by the addition of isopropyl-*β*-D-thiogalactopyranoside (IPTG) following a temperature downshift. *cspA* mRNA has a highly efficient structure for translation initiation (Mitta et al. 1997). When a nonsense codon is introduced at the 1st, 10th or 30th position in the CspA coding region, induction of *cspA* mRNA results in the trapping of most cellular ribosomes and the inhibition of other cellular protein synthesis (Bae et al. 1997; Xia et al. 2001). This phenomenon is referred to as the 'low-temperature antibiotic effect of truncated cspA expression (LACE)' (Jiang et al. 1996). Additionally, for the pCold I vector, a translation-enhancing element (TEE) sequence is placed to enhance translation initiation (Etchegaray and Inouye 1999) and is followed by a hexahistidine (6His) tag and the factor Xa cleavage site. These techniques enable pCold vector series to induce high-yield protein production in *E. coli* at low temperature.

Notwithstanding these efforts, not all genes can be efficiently expressed in *E. coli*. To address this problem, we have developed a pCold-GST expression vector containing a GST tag sequence within the pCold I vector (Hayashi and Kojima 2008). This pCold-GST vector has been applied to 10 genes and NMR analyses of two unstable peptides. Here we report on the expression of 84 proteins using the pCold-GST vector and the development of new cold-shock vectors named pCold-GB1, pCold-Trx and pCold-MBP containing GB1, Trx and MBP tag sequences, respectively. Additionally, we report on an effective method for the preparation of unstable peptide using a C-terminal proline tag.

Materials and methods

Construction of expression plasmids

pCold-GST vector was prepared as previously described (Hayashi and Kojima 2008). pCold-GB1, pCold-Trx and pCold-MBP vectors were prepared as follows. The streptococcal GB1 gene was synthesized by two-step PCR (Ito and Wagner 2004), the Trx gene was amplified from pET32c vector (Novagen) and the MBP gene was from *E. coli* BL21. GB1, Trx or MBP gene and HRV 3C protease recognition site sequence were inserted into the *Nde*I site of pCold I vector (Takara). Following ligation, one of the two *Nde*I sites, located just before the soluble protein tag sequence, was destroyed by point mutation using QuickChange mutagenesis (Qiagen).

All 84 target gene fragments were amplified by PCR and inserted into multiple cloning sites of the plasmid vectors. The rice flowering locus T (FT) homologous protein gene was a gift from Prof. Ko Shimamoto (NAIST). The FT homologous protein gene fragment was amplified by PCR and then inserted between the *Nde*I and *Eco*RI sites of pCold-GST, pCold-GB1, pCold-Trx and pCold-MBP vectors. The human ubiquitin gene was a gift from Dr. Toshiyuki Kohno (MITILS). The ubiquitin gene was amplified by PCR and then inserted between the *Nde*I and *Eco*RI sites of pCold-GST, pCold-GB1, pCold-Trx and pCold-MBP vectors. Expression plasmid comprising human Na^+/H^+ exchanger 1 (NHE1) was constructed as previously described (Mishima et al. 2007). The *E. coli* serine chemoreceptor (Tsr) gene was a gift from Prof. Ikuro Kawagishi (Hosei University). The Tsr gene fragment was amplified by PCR and then inserted between the *Kpn*I and *Hind*III sites of pCold-GST vector. Preparation of expression plasmids for human VAMP associated protein A (VAP-A) protein and human oxisterol binding protein (OSBP) fragment were as previously described (Furuita et al. 2006, 2010).

Expression analysis

All proteins were expressed using *E. coli* Rosetta™(DE3) (Novagen) with 10 mL LB medium containing 100 μg/mL ampicillin. Each of these was grown at 37 °C until the OD_{660} was 0.6. For the pCold I and pCold-GST systems, protein expression was induced by the addition of 1 mM IPTG (Wako) following a temperature downshift from 37 to 15 °C and cells were then cultured for 16 h. For the other systems, protein expression was induced at 37 °C and cells were then cultured for 3 h at 37 °C. Harvested wet cells were resuspended in 1 mL of 50 mM Tris buffer (pH 8.0) containing 300 mM KCl and 15 mM 2-mercaptoethanol (Buffer A). The suspension was lysed by sonication and then centrifuged (6,000×*g* for 30 min at 4 °C). Target proteins were partially purified by a batch method using Glutathione Sepharose 4B (GE Healthcare) or Ni-NTA Agarose (QIAGEN). Following the addition of 100 μL of affinity resin to supernatants, unbound fractions were removed by centrifugation (3,300×*g* for 10 s at 4 °C). The affinity resin was then washed with Buffer A (1 mL × 4) to remove nonspecifically bound proteins, resuspended in 100 μL of 4 × sample buffer (40 mM Tris buffer (pH 6.8) containing 0.4% SDS, 0.4% 2-mercaptoethanol, 0.2% Bromophenol Blue and 20% glycerol) and heated at 95 °C. Whole cell lysate, insoluble fraction and soluble proteins eluted from the affinity resin were analyzed by SDS-PAGE.

Expression and purification

Non-labeled and ^{15}N-labeled ubiquitin, non-labeled NHE1 and FT homologous protein were prepared as fusion

proteins with various N-terminal soluble protein tags without tag digestion. ^{15}N-labeled FT homologous protein was prepared with tag digestion. Non-labeled OSBP and ^{15}N-labeled NHE1 and Tsr were prepared as fusion proteins with a C-terminal 6His or hexa-proline (6Pro) tag.

Expression

All proteins were expressed using *E. coli* Rosetta^{TM}(DE3). For the preparation of ^{15}N-labeled protein, cells were grown at 37 °C in 2 L M9 minimum medium containing 1 mM $MgSO_4$, 50 μM $MnCl_2$, 3.3 μM $FeCl_3$, 0.1 mM $CaCl_2$, 100 μg/mL ampicillin, 20 mg/L thymine, 20 mg/L adenosine, 20 mg/L guanosine, 20 mg/mL cytidine, 20 mg/L thiamine, 20 mg/mL biotin, 0.5 g/L $^{15}NH_4Cl$ and glucose. For ^{15}N-labeled FT homologous protein, cells were grown at 37 °C in 2 L of M9 medium containing 25 mL of LB medium. The amount of glucose was 8 g/L for NHE1 and 4 g/L for the others. When the A_{660} was 0.4–0.6, protein expression was induced by the addition of 1 mM IPTG following a temperature downshift from 37 to 15 °C and cells were then cultured overnight. For the preparation of OSBP, cells were grown at 37 °C in 2 L of LB medium. When the A_{660} was 0.8, protein expression was induced by the addition of 1 mM IPTG and cells were then cultured for 3 h at 37 °C.

Purification

Tsr and NHE1 were purified as previously described (Hayashi and Kojima 2008). Other proteins were purified as follows. Harvested cells were resuspended in 50 mM HEPES buffer (pH 8.0) containing 300 mM KCl, 0.1 mM EDTA and 15 mM 2-mercaptoethanol (buffer A). Cell suspensions were lysed by sonication and crude membranes were collected by ultracentrifugation (138,000×*g* for 30 min at 4 °C). The supernatants were then loaded onto a Glutathione Sepharose 4B or Ni-NTA Agarose column. When using the Glutathione Sepharose 4B column, the resin was washed extensively with buffer A to remove nonspecifically bound proteins and GST-tagged proteins were eluted using buffer A containing 30 mM reduced glutathione. When using the Ni-NTA Agarose column, the resin was washed extensively with buffer A containing 20 mM imidazole and Ni-tagged proteins were eluted in a stepwise fashion using buffer A containing 50, 100 and 300 mM imidazole. The N-terminal GST-tag was digested using GST-tagged human rhinovirus (HRV) 3C protease (Cordingley et al. 1990; Walker et al. 1994) supplied as PreScission protease (GE Healthcare). Gel filtration was performed using a HiLoad 16/60 Superdex 75 column (GE Healthcare) in 10 mM Bis-Tris buffer (pH 6.8) containing 50 mM KCl and 1 mM DTT.

NMR experiments

Protein samples were dissolved in buffer comprising 95% H_2O/5% ^2H_2O. NMR spectra were acquired at 283, 288 or 303 K using an AVANCE500 NMR spectrometer with a TXI cryogenic probe (Bruker Biospin). Tsr proteins were dissolved in 10 mM citric acid buffer (pH 5.0) containing 50 mM KCl, and NHE1 proteins were dissolved in 50 mM potassium phosphate buffer (pH 6.8) containing 50 mM KCl. VAP A and OSBP were dissolved in 50 mM potassium phosphate buffer (pH 6.8) containing 100 mM KCl, 1 mM DTT and 0.1 mM EDTA. The other proteins were dissolved in 10 mM Bis-Tris buffer (pH 6.8) containing 50 mM KCl and 1 mM DTT.

Results and discussion

Application of pCold-GST vector to various proteins

The pCold-GST system has been successfully employed for the overexpression of 9 out of 10 proteins that could not otherwise be expressed using conventional *E. coli* expression systems (Hayashi and Kojima 2008). Here, the pCold-GST system was applied to 84 proteins, ranging in length from 18 to 1,161 a. a. from mammalian, plant and other organisms, and the expression level and solubility of these proteins were examined. Each expression plasmid was transformed into *E. coli* Rosetta^{TM}(DE3) and cells were then cultured overnight in 10 mL LB medium at 15 °C following induction. Cells were lysed, purified by a batch method using 100 μL of affinity resin, and samples were then analyzed by SDS-PAGE. As shown in Table 1 and the supplementary material (Tables S1 and S2), 78 out of 84 proteins were expressed in soluble fractions using the pCold-GST system. Unexpressed proteins were those derived from rice, human and yeast sources (Table S2). The length of the expressed proteins was not correlated with the expression level in the soluble fraction (Table 1). These results suggest that the pCold-GST system can be effectively employed for the expression of most proteins in the soluble fraction. Our results are remarkable considering that protein expression in the soluble fraction when using *E. coli* is problematic for many proteins.

Details concerning the expression, purification and NMR measurement of FT homologous protein (179 a. a.) are described below. The expression and affinity purification of GST-fused FT homologous protein are shown in Fig. 1a, b for the pGEX 6P-3 and pCold-GST systems, respectively. pGEX 6P-3 is one of most popular *E. coli* expression vectors employing a GST tag. The yield of the pCold-GST system was more than tenfold higher than that of pGEX 6P-3. In pCold-GST system the expression level

Table 1 Expression of various proteins in soluble fraction using pCold-GST

Target protein length (a. a.)	Number of proteins expressed in soluble fraction		
	Mammalian	Plant	Others
–50	6/6	4/4	1/1
51–150	17/17	2/2	2/2
151–300	5/8	18/18	4/4
301–999	4/5	15/16	0/1
1,000–	0/0	1/1	0/1

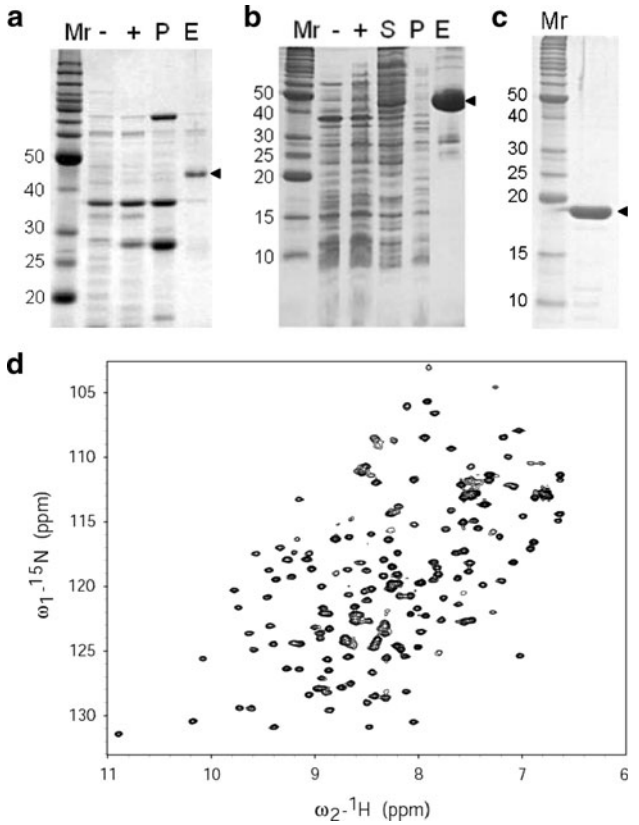

Fig. 1 **a–c** SDS-PAGE showing expression and purification of FT homologous protein. Mr, marker, −, lysate of whole cells prior to induction, +, lysate of whole cells following induction, P, pellet, S, the supernatant fraction from the lysate, E, the eluted fraction from the affinity resin. Molecular weights are shown on the *left*. The gel concentrations are 12.5, 15 and 15% for (**a–c**), respectively. **a** GST-tagged protein prepared using the pGEX 6P-3 system. **b** GST-tagged protein prepared using the pCold-GST system. **c** Tag digested purified sample. **d** ^{1}H-^{15}N HSQC spectrum of FT homologous protein

is not very high, but the solubility of the expressed GST-fused protein is high (Fig. 1b). The yield of the pGEX 6P-3 system was significantly improved for the half of examined six proteins on lowering of the induction temperature for the protein expression from 37 to 15 °C, although the yield for the rest half was not improved much (Fig. S2). FT homologous protein expressed using the pCold-GST vector was highly purified following tag digestion by HRV3C protease and gel filtration chromatography (Fig. 1c). During the purification process, non-specific cleavage and precipitation were not observed. The final sample gave a well-separated ^{1}H-^{15}N HSQC spectrum (Fig. 1d), indicating that the expressed protein possessed a folded structure. This result was also supported by the CD spectrum. Thus we can conclude that use of the pCold-GST system can result in increased yields of expressed protein while maintaining protein 3D structure.

Combination of pCold vector with various soluble protein tags

GB1, Trx and MBP have been employed as soluble protein tags in a manner similar to GST. In an effort to improve yields, we combined GB1, Trx or MBP with the pCold vector, yielding pCold-GB1, pCold-Trx and pCold-MBP, respectively. These three expression vectors and pCold-GST were employed for the expression of ubiquitin, FT homologous protein and NHE1. These three proteins were expressed in 2 L of M9 medium and purified by affinity chromatography following gel filtration without tag digestion. Table 2 shows the yield of each fusion protein from 1 L of media. In this report, the fusion protein yield is used to compare the yield among the developed expression systems due to its robust reproducibility. With ubiquitin, the yield was high for all expression vectors, and the final yield of fusion protein was higher than that of the tag protein alone. These results show that ubiquitin enhances the yield of tag protein, and are consistent with previous data indicating that ubiquitin can function as a soluble protein tag (Kohno et al. 1998). With FT homologous protein, as shown in Fig. 1, the yield for pCold-GST was more than tenfold higher than that for conventional expression systems. Although the yield for pCold-Trx and pCold-GST was relatively low, that for pCold-GB1 and pCold-MBP was high. Comparing to conventional expression systems, the yield is expected to be more than 25 and 125 times higher for pCold-GB1 and pCold-MBP, respectively. NHE1 was not expressed by pCold-Trx nor by the conventional expression vector (Mishima et al. 2007; Hayashi and Kojima 2008). Use of pCold-GB1 and pCold-GST resulted in NHE1 expression although the yield was low. Use of the pCold-MBP system increased the yield of NHE1 by more than sixfold compared to pCold-GST. As demonstrated above, the combination of pCold vector with soluble protein tags, such as pCold-MBP, pCold-GB1 and pCold-GST, proved effective for protein sample preparation, and the yields observed for the pCold-MBP system were quite remarkable. It is noted that the yield after protease digestion of soluble protein tags is strongly depends on the purification protocol and the character of the target

Table 2 Comparison of yield of purified fusion protein

Sample	Target protein length (a. a.)	Yield of fusion protein (from 1 L of M9 medium)				
		pCold-I	pCold-GB1	pCold-Trx	pCold-GST	pCold-MBP
Ubiquitin	76	–	26 mg (1,460 nmol)	29 mg (1,250 nmol)	10 mg (270 nmol)	51 mg (980 nmol)
NHE1	55	Not expressed	0.17 mg (11 nmol)	Not expressed	0.25 mg (16 nmol)	2.9 mg (100 nmol)
FT homologous protein	179	1.52 mg (72 nmol)	1.75 mg (77 nmol)	0.28 mg (16 nmol)	1.2 mg (30 nmol)	24 mg (385 nmol)
Tsr	55	–	–	–	4.5 mg (144 nmol)	0.47 mg (10 nmol)
Tag only[a]	–	–	7.5 mg (917 nmol)	7.7 mg (500 nmol)	5.9 mg (247 nmol)	24 mg (705 nmol)

[a] Cells were grown in LB medium

protein, thus the pCold-MBP system may not be the best in some cases.

Figure 2 shows ^{1}H-^{15}N HSQC spectra of ubiquitin (Fig. 2a), ubiquitin fusion with soluble protein tags (GB1-ubiquitin, Trx-ubiquitin, GST-ubiquitin and MBP-ubiquitin) (Fig. 2b–e), and soluble protein tags alone (Fig. 2f–i). These samples were prepared using pCold vector in combination with soluble protein tag sequences as described above. Ubiquitin, in addition to most of the tag proteins (GB1, Trx and MBP), gave well-separated NMR spectra as has been reported. NMR analysis of GB1-ubiquitin and Trx-ubiquitin also yielded well-separated NMR spectra (Fig. 2b, c), although these were slightly broader and more complicated compared to ubiquitin (Fig. 2a). These spectral differences can be explained by the increase in molecular weight and signals derived from the tag proteins. For GST-ubiquitin, the observed NMR signals were mostly derived from ubiquitin (Fig. 2d), and most of the NMR signals usually derived from the GST tag were not observed (Fig. 2h), as previously reported (Liew et al. 2008). In this sense, GST is an NMR invisible protein tag. For MBP-ubiquitin, the spectrum was heavily overlapped and made subsequent analysis difficult (Fig. 2e). As shown above, the NMR spectra of fusion proteins were more complicated than the corresponding spectra of tag-digested proteins. To obviate this problem, the use of segmental label techniques such as sortase-mediated protein ligation (Kobashigawa et al. 2009) could be useful since NMR signals originating from the tag protein portion of fusion proteins can be eliminated without tag-digestion using this technique.

Comparison of soluble protein tags GB1, Trx, GST and MBP

In this study, four cold-shock expression systems comprising pCold-GB1, pCold-Trx, pCold-GST and pCold-MBP were used. In order to characterize these four systems, expressed fusion protein yields were compared as shown below. All tag proteins were highly expressed and purified in high yield. As shown in Table 2, the yields of GB1-ubiquitin and Trx-ubiquitin were the first and second highest, respectively, for ubiquitin. In contrast, the yields of MBP-fusion proteins were highest for NHE1 and FT homologous protein. Although ubiquitin was highly expressed using a conventional vector, NHE1 and FT homologous protein were difficult to prepare using conventional vectors. These data suggest that use of a small protein tag such as GB1 or Trx can enhance the yield of a protein which is easy to prepare by conventional methods, and that use of a large protein tag such as MBP can enhance the yield of a protein which is difficult to prepare by conventional methods.

These four cold-shock expression systems were further investigated including the comparison with the original pCold-I system. For pCold-GST, the expression level of five out of six proteins is significantly improved comparing to pCold-I as described (Hayashi and Kojima 2008). The yield of NHE1 was significantly improved using pCold-GB1, pCold-GST and pCold-MBP, comparing to pCold-I (Fig. S3 and Table 2). The yield of FT homologous protein was significantly improved for all of examined cold-shock expression systems comparing to pGEX 6P-3 (Figs. 1 and S4), and pCold-MBP comparing to pCold-I (Table 2), but not for other systems (Table 2). The yield of the pCold-GST system was much higher than pCold-MBP for Tsr (Table 2). As described in our previous report, the yield of the degradative protein such as Tsr is significantly improved using pCold-GST (Hayashi and Kojima 2008).

As demonstrated, GST, GB1 and MBP can be utilized for the expression of target protein. GST-fused protein is especially useful for NMR sample preparation since most of the NMR signals from GST tag are not observable (Fig. 2d, h). Moreover, the GST tag is useful for protein purification given availability of the highly efficient resin, glutathione Sepharose. Hence, use of a GST tag will be the first choice (Hayashi and Kojima 2008). The thermal stability of GST was much lower than that of MBP and GB1 (Table S3), while the yield of MBP-fused protein was the highest (Table 2). Thus, use of an MBP tag will be the second choice.

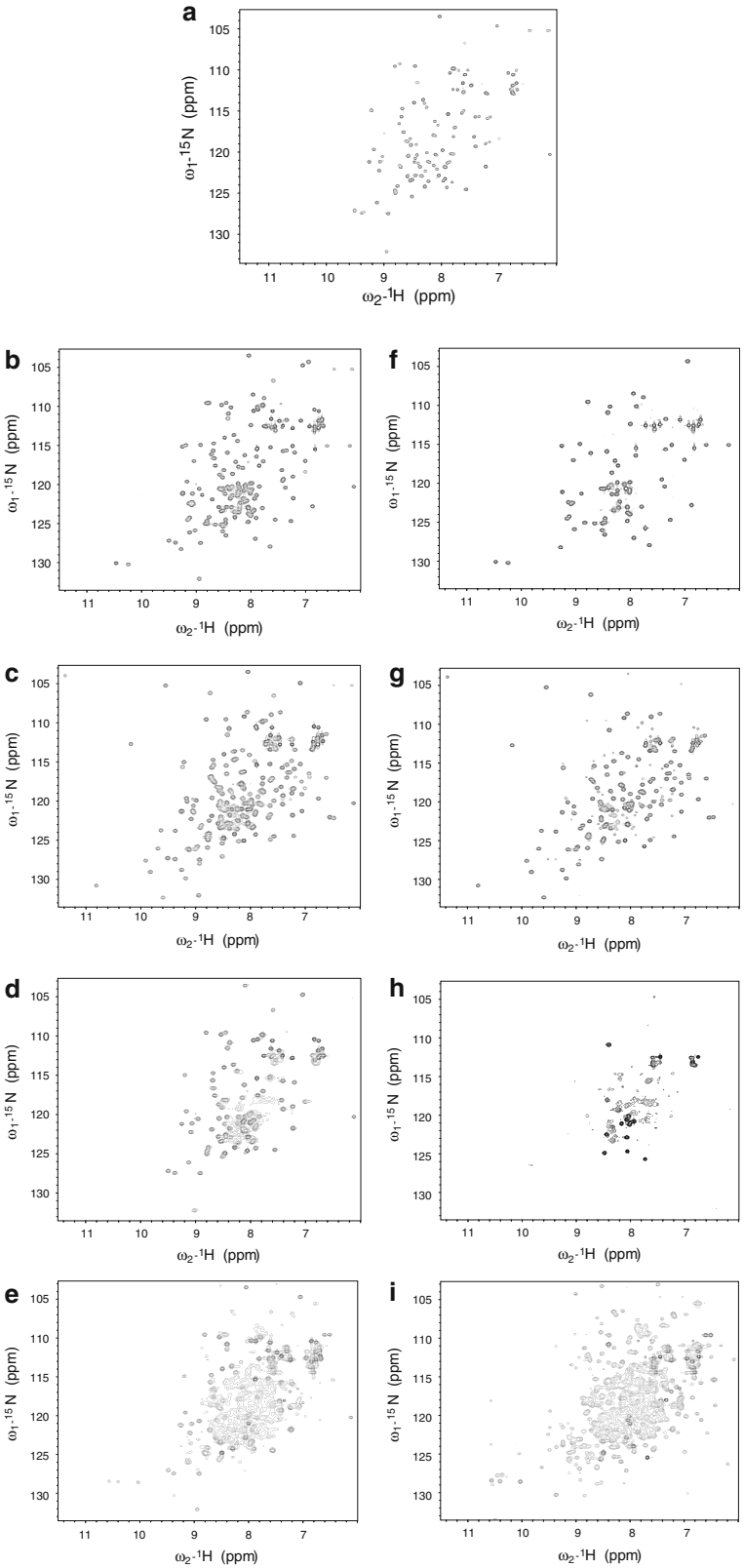

Fig. 2 ^{1}H-^{15}N HSQC spectra of ubiquitin (**a**), ubiquitin fusion proteins with soluble protein tags (**b–e**), soluble tag proteins alone (**f–i**). **b** GB1-ubiquitin, **c** Trx-ubiquitin, **d** GST-ubiquitin, **e** MBP-ubiquitin, **f** GB1, **g** Trx, **h** GST and **i** MBP

About soluble protein tags, many of them (Esposito and Chatterjee 2006) were not employed in this report. Moreover, we were focusing on NMR sample preparation using pCold expression system, thus GST tag may not be the first choice for different purposes and different systems. For example, Trx tag is reported to be the best for protein

expression among MBP, GST, Trx, NUS A, ubiquitin and SUMO tags (Marblestone et al. 2006), although Trx tag is not good for the cold-shock expression system.

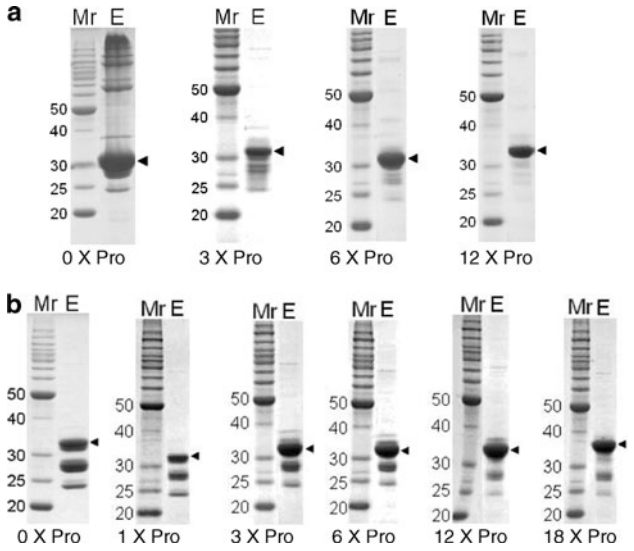

Fig. 3 SDS-PAGE showing the affinity purification of NHE1 (**a**) and Tsr (**b**) with varied C-terminal proline lengths. Mr, marker, E, the eluted fraction from the affinity resin. Molecular weights and the number of proline residue are shown on the *left* and *bottom*, respectively

Sample preparation of unstable protein using C-terminal poly-histidine and poly-proline tags

We previously reported on the expression and purification of unstable protein with a C-terminal 6His tag using the pCold-GST system (Hayashi and Kojima 2008). The C-terminal 6His tag was used to separate target protein from degraded product, and the reduction in yield was unavoidable due to the degradation. In this study, a C-terminal poly-proline tag was used to inhibit degradation during the expression and purification of two unstable protein fragments 55 a. a. in length. Figure 3 shows the results of the protein expression. The number of C-terminal proline residues varied from 0 to 18 for Tsr, and from 0 to 12 for NHE1. In both cases, protein degradation was partially prevented. The minimum number of C-terminal proline residues found to inhibit degradation was about 3–6. Therefore, the expression and purification of Tsr and NHE1 containing a C-terminal hexa-proline (6Pro) tag was performed.

NHE1 and Tsr each containing a C-terminal 6Pro or 6His tag were expressed using pCold-GST. GST fusion proteins were expressed in 2 L of M9 medium and purified by affinity chromatography. Final samples were purified by gel filtration following tag digestion. The final yield of C-terminal 6Pro tagged NHE1 was 0.75 mg (125 nmol) per 1 L of M9 medium, whereas the final yield of C-terminal 6His tagged NHE1 was 0.06 mg (10 nmol). ^{1}H-^{15}N HSQC spectra of

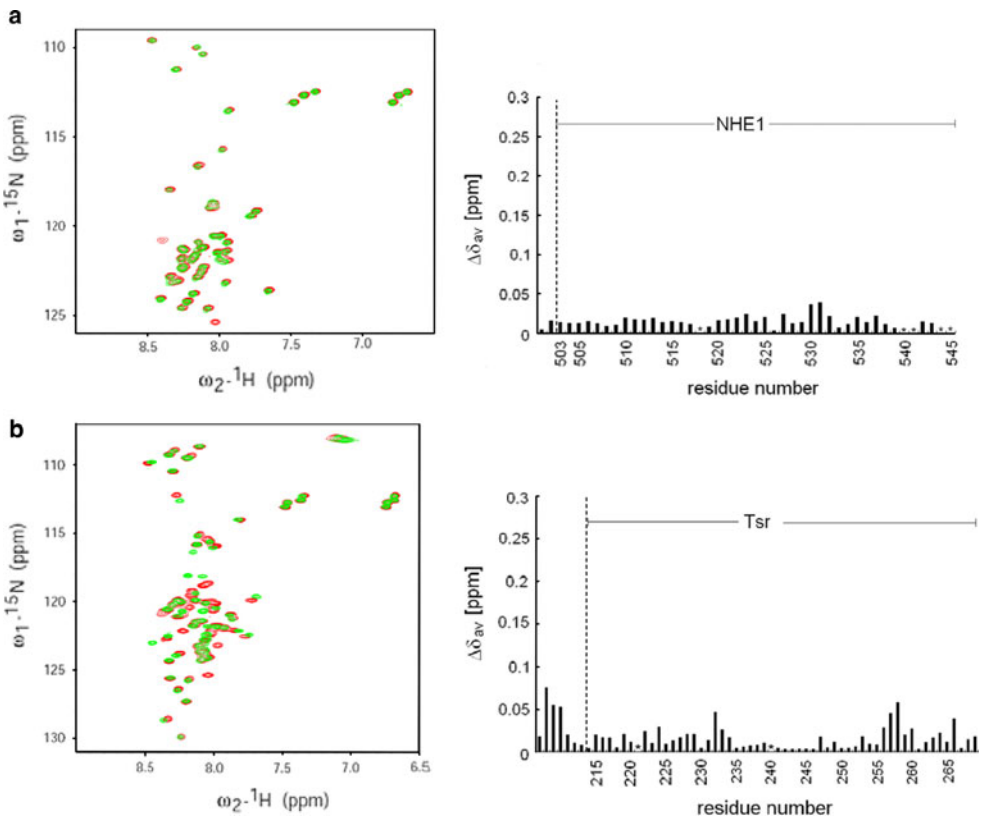

Fig. 4 ^{1}H-^{15}N HSQC spectra of C-terminal 6His (*red*) and 6Pro (*green*) tagged proteins (*left*), and the combined ^{1}H and ^{15}N chemical shift differences ($\Delta\delta_{av} = \{\Delta\delta(^{1}H)^2 + [\Delta\delta(^{15}N)/5]^2\}^{1/2}$) between 6His and 6Pro tagged proteins (*right*). Asterisks indicate unassigned residues. **a** NHE1. **b** Tsr

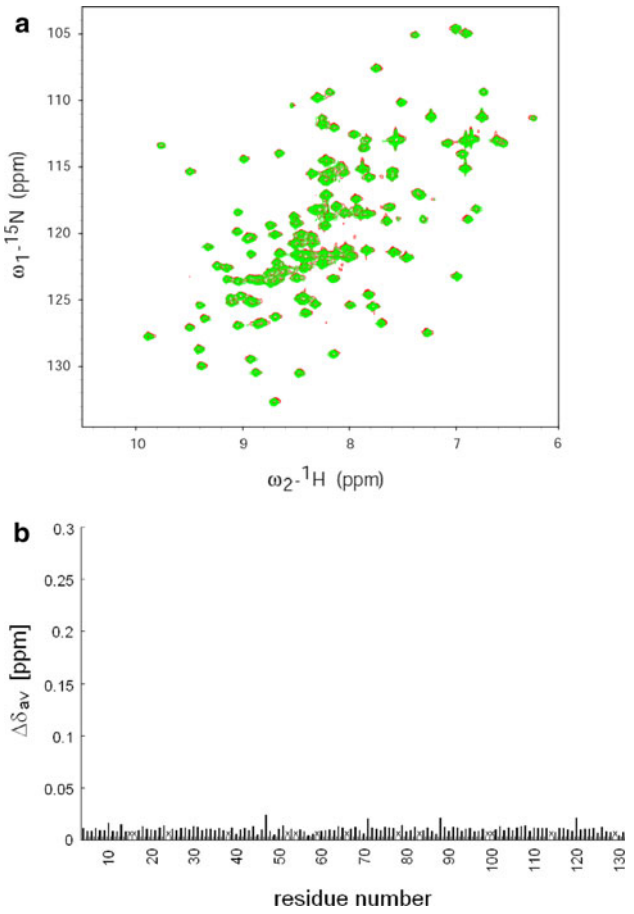

Fig. 5 **a** ^{1}H-^{15}N HSQC spectra of VAP-A protein complexed with C-terminal 6His (*red*) and 6Pro (*green*) tagged OSBP proteins. **b** Combined ^{1}H and ^{15}N chemical shift differences ($\Delta\delta_{av} = \{\Delta\delta(^{1}H)^2 + [\Delta\delta(^{15}N)/5]^2\}^{1/2}$). Crosses indicate unassigned residues

NHE1 and Tsr are shown in Fig. 4a, b for 6Pro and 6His tagged proteins in green and red, respectively. The chemical shift differences between 6Pro and 6His tagged proteins were small for both protein fragments, although several N-H peaks showed relatively large differences for Tsr.

In an effort to determine the influence of the C-terminal 6Pro tag on protein-protein interactions, the interaction between a VAP-A and 6Pro tagged OSBP fragment was examined. VAP-A specifically binds OSBP fragment containing the FFAT motif, and NMR signals of the complex have been reported (Furuita et al. 2006, 2010). Two complexes were prepared using ^{15}N-labeled VAP-A and non-labeled 6His or 6Pro tagged OSBP. ^{1}H-^{15}N HSQC spectra of complexes are shown in Fig. 5 for 6His and 6Pro tagged samples in red and green, respectively, and no significant chemical shift difference was observed. These data suggest that C-terminal 6Pro tagged sample forms an identical complex as the 6His tagged sample, and no effect will be expected on protein-protein interactions.

As demonstrated above, the use of a C-terminal poly-proline tag is favorable in the preparation of degradative proteins. To the best of our knowledge, the stability enhancement by the addition of a C-terminal poly-proline tag has never been reported for the protein preparation, and the inhibition mechanism against the protein degradation is not known. Two possible inhibition mechanisms are considered at this moment. First, the non-specific peptidases may not prefer the proline residue in the C-terminal position. For example, peptidase D of *E. coli* cleaves dipeptide substrates non-specifically, but the proline residue is not accepted in the C-terminal position of the substrates (Schroeder et al. 1994). Second, C-terminal poly-proline residues may form a particular kind of conformation, such as polyproline II helix (Williamson 1994). Such structure may stabilize the C-terminal region of proteins and inhibit the degradation. Therefore the use of the C-terminal poly-proline tag will not inhibit endopeptidases, and in fact it could not inhibit proteases completely as shown in Fig. 3.

Conclusion

Here, we reported on the development of an *E. coli* protein expression system for NMR that utilized a variety of cold-shock expression vectors such as pCold-GST, pCold-GB1, pCold-Trx and pCold-MBP. All systems were successful in obtaining a large amount of fusion proteins. As demonstrated, the pCold-GST system is the first choice given fusion protein NMR character and ease of purification. The pCold-MBP system is the second choice given the high yield and thermal stability of the fusion protein. Use of a C-terminal 6Pro tag was successful in the preparation of ordinarily unstable proteins given that protein degradation present during the expression and purification steps was inhibited. Furthermore, the proline residue was not observed in any of the ^{1}H-^{15}N HSQC spectra. The presence of a C-terminal 6Pro tag did not result in any marked influence on the structure or function of the fusion proteins. Thus, the combination of pCold-GST or pCold-MBP systems with the C-terminal 6Pro tag should prove useful for the preparation of most protein samples for NMR analysis.

Acknowledgments The authors thank Ms. Hiroko Kinoshita, Akiko Yasuba and Momoko Yoneyama for their technical assistance. This work was supported by grants from MEXT through the Global COE Program, Target Proteins Research Program and Grants in Aid for Scientific Research, and from the Suntory Institute for Bioorganic Research.

References

Bae W, Jones PG, Inouye M (1997) CspA, the major cold shock protein of Escherichia coli, negatively regulates its own gene expression. J Bacteriol 179:7081–7088

Bao WJ, Gao YG, Chang YG, Zhang TY, Lin XJ, Yan XZ, Hu HY (2006) Highly efficient expression and purification system of small-size protein domains in Escherichia coli for biochemical characterization. Protein Expr Purif 47:599–606

Cordingley MG, Callahan PL, Sardana VV, Garsky VM, Colonno RJ (1990) Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro. J Biol Chem 265:9062–9065

di Guan C, Li P, Riggs PD, Inouye H (1988) Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein. Gene 67:21–30

Esposito D, Chatterjee DK (2006) Enhancement of soluble protein expression through the use of fusion tags. Curr Opin Biotechnol 17:353–358

Etchegaray JP, Inouye M (1999) Translational enhancement by an element downstream of the initiation codon in Escherichia coli. J Biol Chem 274:10079–10085

Furuita K, Mishima M, Kojima C (2006) ^{1}H, ^{13}C and ^{15}N resonance assignments of the VAP-A: OSBP complex. J Biomol NMR 36(Suppl 1):69

Furuita K, Jee J, Fukada H, Mishima M, Kojima C (2010) Electrostatic interaction between oxysterol-binding protein and VAMP-associated protein A revealed by NMR and mutagenesis studies. J Biol Chem 285:12961–12970

Hayashi K, Kojima C (2008) pCold-GST vector: a novel cold-shock vector containing GST tag for soluble protein production. Protein Expr Purif 62:120–127

Ito T, Wagner G (2004) Using codon optimization, chaperone co-expression, and rational mutagenesis for production and NMR assignments of human eIF2 alpha. J Biomol NMR 28:357–367

Jiang W, Fang L, Inouye M (1996) Complete growth inhibition of Escherichia coli by ribosome trapping with truncated cspA mRNA at low temperature. Genes Cells 1:965–976

Kobashigawa Y, Kumeta H, Ogura K, Inagaki F (2009) Attachment of an NMR-invisible solubility enhancement tag using a sortase-mediated protein ligation method. J Biomol NMR 43:145–150

Kohno T, Kusunoki H, Sato K, Wakamatsu K (1998) A new general method for the biosynthesis of stable isotope-enriched peptides using a decahistidine-tagged ubiquitin fusion system: an application to the production of mastoparan-X uniformly enriched with ^{15}N and ^{15}N/^{13}C. J Biomol NMR 12:109–121

Liew CK, Gamsjaeger R, Mansfield RE, Mackay JP (2008) NMR spectroscopy as a tool for the rapid assessment of the conformation of GST-fusion proteins. Protein Sci 17:1630–1635

Makrides SC (1996) Strategies for achieving high-level expression of genes in Escherichia coli. Microbiol Rev 60:512–538

Marblestone JG, Edavettal SC, Lim Y, Lim P, Zuo X, Butt TR (2006) Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. Protein Sci 15:182–189

Mishima M, Wakabayashi S, Kojima C (2007) Solution structure of the cytoplasmic region of Na^{+}/H^{+} exchanger 1 complexed with essential cofactor calcineurin B homologous protein 1. J Biol Chem 282:2741–2751

Mitta M, Fang L, Inouye M (1997) Deletion analysis of cspA of Escherichia coli: requirement of the AT-rich UP element for cspA transcription and the downstream box in the coding region for its cold shock induction. Mol Microbiol 26:321–335

Qing G, Ma LC, Khorchid A, Swapna GV, Mal TK, Takayama MM, Xia B, Phadtare S, Ke H, Acton T, Montelione GT, Ikura M, Inouye M (2004) Cold-shock induced high-yield protein production in Escherichia coli. Nat Biotechnol 22:877–882

Schroeder U, Henrich B, Fink J, Plapp R (1994) Peptidase D of *Escherichia coli* K-12, a metallopeptidase of low substrate specificity. FEMS Microbiol Lett 123:153–160

Smith DB, Johnson KS (1988) Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. Gene 67:31–40

Walker PA, Leong LE, Ng PW, Tan SH, Waller S, Murphy D, Porter AG (1994) Efficient and rapid affinity purification of proteins using recombinant fusion proteases. Biotechnology (NY) 12:601–605

Williamson MP (1994) The structure and function of proline-rich regions in proteins. Biochem J 297:249–260

Xia B, Etchegaray JP, Inouye M (2001) Nonsense mutations in cspA cause ribosome trapping leading to complete growth inhibition and cell death at low temperature in Escherichia coli. J Biol Chem 276:35581–35588

Yasukawa T, Kanei-Ishii C, Maekawa T, Fujimoto J, Yamamoto T, Ishii S (1995) Increase of solubility of foreign proteins in Escherichia coli by coproduction of the bacterial thioredoxin. J Biol Chem 270:25328–25331

Zhou P, Lugovskoy AA, Wagner G (2001) A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins. J Biomol NMR 20:11–14