J Biomol NMR (2012) 52:303–313
DOI 10.1007/s10858-012-9610-0

ARTICLE

Segmental isotope labeling of proteins for NMR structural study using a protein S tag for higher expression and solubility

Hiroshi Kobayashi · G. V. T. Swapna · Kuen-Phon Wu · Yuliya Afinogenova · Kenith Conover · Binchen Mao · Gaetano T. Montelione · Masayori Inouye

Received: 10 November 2011 / Accepted: 11 January 2012 / Published online: 3 March 2012

Abstract A common obstacle to NMR studies of proteins is sample preparation. In many cases, proteins targeted for NMR studies are poorly expressed and/or expressed in insoluble forms. Here, we describe a novel approach to overcome these problems. In the protein S tag-intein (PSTI) technology, two tandem 92-residue N-terminal domains of protein S (PrS_2) from *Myxococcus xanthus* is fused at the N-terminal end of a protein to enhance its expression and solubility. Using intein technology, the isotope-labeled PrS_2-tag is replaced with non-isotope labeled PrS_2-tag, silencing the NMR signals from PrS_2-tag in isotope-filtered ^1H-detected NMR experiments. This method was applied to the *E. coli* ribosome binding factor A (RbfA), which aggregates and precipitates in the absence of a solubilization tag unless the C-terminal 25-residue segment is deleted (RbfAΔ25). Using the PrS_2-tag, full-length well-behaved RbfA samples could be successfully prepared for NMR studies. PrS_2 (non-labeled)-tagged RbfA (isotope-labeled) was produced with the use of the intein approach. The well-resolved TROSY-HSQC spectrum of full-length PrS_2-tagged RbfA superimposes with the TROSY-HSQC spectrum of RbfAΔ25, indicating that PrS_2-tag does not affect the structure of the protein to which it is fused. Using a smaller PrS-tag, consisting of a single N-terminal domain of protein S, triple resonance experiments were performed, and most of the backbone ^1H, ^{15}N and ^{13}C resonance assignments for full-length *E. coli* RbfA were determined. Analysis of these chemical shift data with the Chemical Shift Index and heteronuclear ^1H–^{15}N NOE measurements reveal the dynamic nature of the C-terminal segment of the full-length RbfA protein, which could not be inferred using the truncated RbfAΔ25 construct. CS-Rosetta calculations also demonstrate that the core structure of full-length RbfA is similar to that of the RbfAΔ25 construct.

Keywords Protein S · Intein · Protein ligation · Protein NMR · RbfA

Introduction

A major bottleneck in protein NMR structural studies involves preparation of protein samples in stable, soluble forms at relatively high concentrations, typically 0.1–0.3 mM. To improve protein solubility, various solubility enhancement tags (SETs), fused to either the N- or C-terminal ends of target proteins, have been developed (Esposito and Chatterjee 2006; Zhou and Wagner 2010). The most widely used SETs are glutathione S-transferase (GST; 26 kDa) (Smith and Johnson 1988) and maltose-binding protein (MBP; 42 kDa) (di Guan et al. 1988; Fox et al. 2001). These SETs can enhance not only protein solubility, but also protein expression. Furthermore, GST and MBP tags can be used for single-step affinity purifications. As these tags are relatively large, they generally

Electronic supplementary material The online version of this article (doi:10.1007/s10858-012-9610-0) contains supplementary material, which is available to authorized users.

H. Kobayashi · K.-P. Wu · Y. Afinogenova ·
G. T. Montelione · M. Inouye (✉)
Department of Biochemistry, Robert Wood Johnson Medical
School, Center for Advanced Biotechnology and Medicine,
679 Hoes Lane, Piscataway, NJ 08854, USA
e-mail: inouye@cabm.rutgers.edu

G. V. T. Swapna · K. Conover · B. Mao · G. T. Montelione
Department of Molecular Biology and Biochemistry,
and Northeast Structural Genomics Consortium, Rutgers,
The State University of New Jersey, 679 Hoes Lane, Piscataway,
NJ 08854, USA

have to be removed for NMR studies. However, when the tags are removed, the proteins are quite often no longer soluble. To circumvent these problems, Wagner and co-workers have developed the *Streptococcus* Protein G B1 domain (GB1) (6.2 kDa) as a SET. GB1 is a small, highly soluble domain consisting of 56 amino acid residues. Using this tag, the CIDE domain of human DNA fragment factor (DFF) 45 (GB1-DFF45) can be enhanced in solubility, yielding a well-resolved [^{1}H–^{15}N] HSQC spectrum (Zhou et al. 2001a, b). The NMR structure of the complex between GB1-DFF45 and DFF40 was subsequently also solved using this SET approach (Zhou et al. 2001a).

The shortcoming of this approach is that many extra resonances arise from the GB1 SET domain. An elegant strategy for suppressing signals from the SET is to generate a protein construct in which the SET domain is not isotope enriched, while the target protein is isotope-enriched. This can be addressed by generation of protein constructs in which the target protein, but not the SET, is isotope-enriched; the so called "invisible SET" approach. Such segmental labeling technology has been carried out previously using three different methods. The first utilized sortase A (Kobashigawa et al. 2009) to ligate an isotope-enriched target protein with an unlabeled SET. The second used human calmodulin (hCaM) as the SET to solubilize an isotope-enriched target protein containing a calmodulin binding peptide (CBP) (Durst et al. 2008). Although the sortase A and hCaM/CBP systems are both elegant and efficient, the more widely used approach for production of segmentally-labeled proteins involves intein-based protein ligation (Elleuche and Poggeler 2010; Southworth et al. 1998). The intein technology has been used successfully in NMR sample preparation, including protein-segmental labeling with isotopes (Romanelli et al. 2004; Yamazaki 1998), and to produce SET fusion proteins of otherwise insoluble proteins (Züger and Iwai 2005). Using an in vivo intein technology approach, a non-isotope labeled SET, the GB1 domain, has been fused to an insoluble chitin-binding domain (CBD) to enable solubility enhancement in the cell. The resulting fusion protein provided a well-resolved [^{1}H–^{15}N] HSQC spectrum of CBD (Züger and Iwai 2005).

While the method of Züger and Iwai (2005) for generating an "invisible SET" fusion construct is very powerful, it has two significant problems. First, if a target protein aggregates before the in vivo protein ligation, the SET cannot be efficiently fused to the target protein. In this case, the precursor which contains the SET is first expressed, and subsequently the lower-solubility target protein is produced in order to allow maximum ligation yield. However, to achieve this one has to carefully optimize the timing of the inductions and concentrations of inducers (Muona et al. 2010). Furthermore, since the target protein is expressed without a SET, this approach may not work well for poorly expressed and/or intrinsically poorly-soluble target proteins. Alternatively, a target protein can be expressed in an isotope-rich medium as a fusion protein with a SET, such as GST and MBP. In this case, however, the tag has to be removed before NMR studies, and the target protein has to be re-purified. Secondly, the use of L-arabinose for induction of a target protein in the system described by Züger and Iwai (2005) generally results in incorporation of ^{12}C-labeled amino acids generated from the arabinose into the target protein, as uniformly ^{13}C-enriched arabinose is quite expensive for the use of the inducer.

In order to overcome the shortcomings of existing SET methods described above, we have developed an alternative "invisible SET" approach using protein S, a major spore coat protein from *Myxococcus xanthus*, a gram-negative developmental bacterium (Inouye et al. 1979). Previously, we have shown that when human proteins are tagged with tandem repeated N-terminal domains of protein S (PrS$_2$), their expression and solubility are significantly enhanced (Kobayashi et al. 2009). Moreover, PrS$_2$-tagged *E. coli* OmpR was found to fully retain its function as a transcription regulator, indicating that the PrS$_2$-tag does not interfere with the structure and function of the target protein that is fused to the tag (Harlocker et al. 1995; Kobayashi et al. 2009). Using this simple and novel approach (outlined in Fig. 1a, b), referred here as the Protein S Tag-Intein (PSTI) technology, we demonstrate highly efficient segmental protein labeling, in which an isotope-labeled target protein is fused to a non-isotope labeled protein S SET for ^{1}H-detected heteronuclear-filtered NMR studies.

We have applied the PSTI technology to the *E. coli* ribosome binding factor A, RbfA. The full-length RbfA protein (133 residues) slowly precipitates in solution precluding NMR studies, which to date have been restricted to a truncated construct lacking the C-terminal 25 residues (RbfAΔ25) (Huang et al. 2003; Swapna et al. 2001). Although RbfAΔ25 retains its biological function of suppressing growth defects at low temperature, the growth-suppression efficiency of RbfAΔ25 is significantly lower than that of full-length RbfA. Structural studies on full-length RbfA are thus important to fully elucidate structure–function relationships in RbfA. Using the PSTI technology we are now able to produce isotope-labeled full length RbfA and achieve almost complete backbone resonance assignments. Analysis of these chemical shift data with both TALOS+ (Shen et al. 2009) and CS-Rosetta (Shen et al. 2008), together with ^{1}H–^{15}N hetNOE measurements, demonstrate that the full-length RbfA protein structure is similar to that of RbfAΔ25, and that the C-terminal 25-residue region of the full-length protein is largely disordered in the absence of any binding partners.

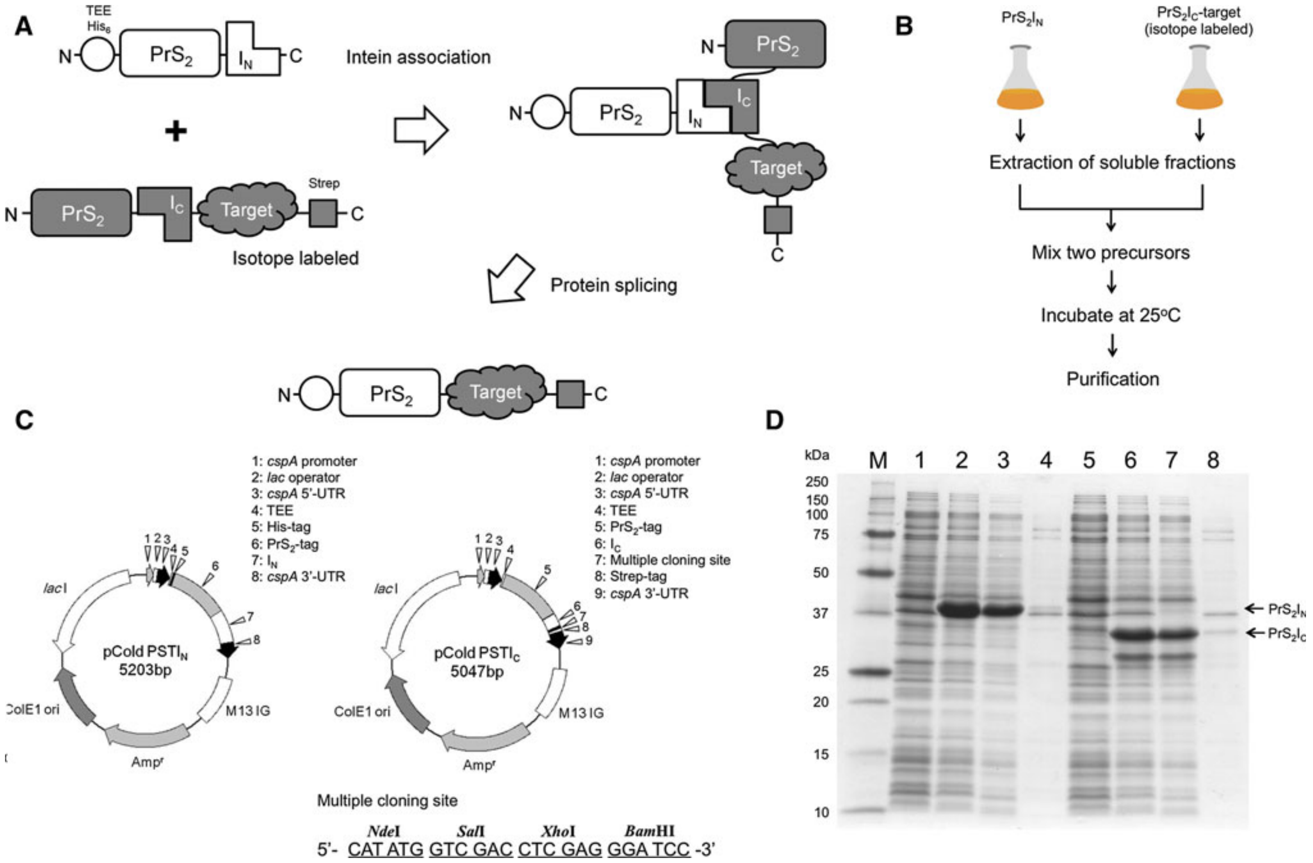

Fig. 1 A novel technology for segmental labeling combining PST and intein technology. **a** PSTI technology for segmental isotope labeling was designed by combining PST (Kobayashi et al. 2009) and intein (Elleuche and Poggeler 2010) technologies. *White* and *gray* are indicating non-labeled and isotope labeled proteins, respectively. **b** Flowchart of the simple protocol for preparation of segmentally-labeled fusion proteins using PSTI technology. **c** Structure of the pCold-PSTI$_N$ and pCold-PSTI$_C$ vectors used for PSTI technology. The sequence of multiple cloning site of PSTI$_C$ is shown, along with the corresponding restriction enzyme sites. **d** Efficiency of expression and solubility of PrS$_2$I$_N$ and PrS$_2$I$_C$. Cell cultures of PrS$_2$I$_N$ and PrS$_2$I$_C$ at 0 and 16 h after induction with 1 mM IPTG at 15°C were collected and analyzed by SDS-PAGE. The solubility of two precursors was examined by SDS-PAGE with the cultures at 16 h after induction. *Lanes: 1*, 0 h of PrS$_2$I$_N$; *2*, 16 h of PrS$_2$I$_N$; *3*, soluble fraction of PrS$_2$I$_N$; *4*, insoluble fraction of PrS$_2$I$_N$; *5*, 0 h of PrS$_2$I$_C$; *6*,16 h of PrS$_2$I$_C$; *7*, soluble fraction of PrS$_2$I$_C$; and *8*, insoluble fraction of PrS$_2$I$_C$. PrS$_2$, PrS$_2$-tag; *I$_N$* intein$_N$ domain, *I$_C$* intein$_C$ domain, *M13 IG* intergenic region of M13 bacteriophage, *UTR* untranslated region, *M* molecular weight marker

Materials and methods

Plasmid construction

A schematic figure of the strategy and constructs used in implementing PSTI technology is presented in Fig. 1a. The overall simple protocol is summarized in Fig. 1b. The concept is to fuse, using intein technology, an unlabeled PrS$_2$I$_N$ (PrSI$_N$) construct, which includes the N-terminal region of an intein (I$_N$) fused to a PrS$_2$ (PrS) SET, with an isotope-enriched construct, PrS$_2$I$_C$-target, which includes the PrS$_2$ SET fused to the C-terminal region of an intein (I$_C$) and the target protein. These constructs are expressed using the pCold expression system, which provides protein expression induction at low temperatures (Qing et al. 2004). Each plasmid was constructed so as to exclude the nucleotide sequence ‘ACA’, which is cleaved by *E. coli* MazF, allowing us to use the PSTI technology together with Single Protein Production (SPP) system (Suzuki et al. 2005). For construction of pCold-PSTI$_N$ and pCold-PSTI$_C$, ACA-less PrS$_2$-tag, ACA-less DnaE$_N$ (*Nostoc punctiforme*: I$_N$) and ACA-less DnaE$_C$ (*Synechocystis sp*: I$_C$) (Lockless and Muir 2009) genes were synthesized by GenScript Inc. The ACA-less PrS$_2$-tag was inserted into pCold IV (sp-4) (Takara Bio), resulting in ACA-less pCold-PST. For pCold-PSTI$_N$, Translation Enhancing Element (TEE consisting of MNHKV) (Etchegaray and Inouye 1999) and a His$_6$-tag were inserted into N-terminal end of ACA-less pCold-PST, and the synthesized gene coding for I$_N$ was inserted using *Kpn*I and *Bam*HI sites in the multiple cloning site of pCold-PST (Kobayashi et al. 2009). For pCold-PSTI$_C$, the synthesized gene fragment coding for I$_C$

was inserted using *Kpn*I and *Nde*I sites, and a streptavidin-binding tag (Strep-tag, including the polypeptide sequence WSHPQFEK) was inserted at C-terminal end by site directed mutagenesis. Expression vector pCold-SPSTI$_N$ was constructed by replacing PrS$_2$-tag in pCold-PSTI$_N$ with a single N-terminal domain of protein S (PrS). Schematic vector maps and complete amino acid sequences of pCold-PSTI$_N$ and pCold-PSTI$_C$ are shown in Fig. 1c and Supplementary Figure S1, respectively.

Protein expression

Escherichia coli BL21 cells were transformed with pCold-PSTI$_N$ or pCold-SPSTI$_N$ and the transformed cells were grown at 37°C in M9-casamino acid medium to an optical density at 600 nm of 0.5 units. The cultures were cooled down with ice-cold water for 5 min and incubated at 15°C for 45 min with shaking, and IPTG was added to the final concentration of 1 mM to induce protein expression. The protein was expressed at 15°C for 16 h.

The transformed *E. coli* BL21 cells with pCold-PSTI$_C$-RbfA or pCold-PSTI$_C$-RbfAΔ25 (coding for a construct in which the C-terminal 25 residues are deleted) were grown at 37°C in M9-minimum medium to O.D$_{600}$ = 0.5 units. The cells were harvested by centrifugation and resuspended in M9-minimum medium containing $^{15}NH_4Cl$ for PrS$_2$I$_C$-RbfAΔ25 or $^{15}NH_4Cl$ and ^{13}C-glucose for PrS$_2$I$_C$-RbfA. IPTG was then added (1 mM) and incubated at 15°C for 16 h to induce protein expression.

Protein splicing and purification

The harvested cells were resuspended in 30 ml of splicing buffer (50 mM Tris, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 7.2) (Lockless and Muir 2009), washed once, and then disrupted using a French press. The cell debris and insoluble proteins were removed by centrifugation at 17,000×*g* for 20 min at 4°C, and the supernatant was further centrifuged at 85,000×*g* for 1 h at 4°C to remove the membrane fraction. The soluble fractions of PrS$_2$I$_N$ (or PrSI$_N$) and PrS$_2$I$_C$-RbfAΔ25 (or PrS$_2$I$_C$-RbfA) were mixed at 1 to 1 molar ratio, and incubated at 25°C for 16 h without shaking, to allow intein ligation. The spliced product was purified with Ni-NTA Agarose and Strep-Tactin Super flow Plus (QIAGEN) according to vendor-supplied protocols. Next, the resultant proteins were loaded onto a gel-filtration column (Superdex 200, GE Healthcare) and eluted with NMR buffer (25 mM NaOAc, 150 mM NaCl, 10 mM β-mercaptoethanol, pH 4.0). SDS-PAGE and MALDI-TOF mass spectrometry were performed to verify the purity (>90%) and molecular weight.

RbfAΔ25 was expressed and purified using protocols described previously (Swapna et al. 2001).

NMR spectroscopy

All NMR samples were prepared at 0.4–0.5 mM in NMR buffer at pH 4.0, containing 10% ^2H_2O, in 5-mm Shigemi NMR tubes. All NMR experiments were performed at 35°C on a Bruker Avance 800 MHz NMR system equipped with a 5-mm cryogenic triple-resonance NMR probe. 2D [^{1}H–^{15}N] HSQC spectra with TROSY detection were acquired for RbfAΔ25, PrS$_2$-RbfAΔ25 (^{15}N), PrS$_2$-RbfA (^{15}N, ^{13}C) and PrS-RbfA (^{15}N, ^{13}C), and 2D [^{1}H–^{13}C] HSQC, 3D HNCO, 3D HNCA, 3D HNCACB and 3D HNcoCACB triple resonance experiments were acquired for PrS$_2$-RbfA (^{15}N, ^{13}C) and PrS-RbfA (^{15}N, ^{13}C). NMRPipe (Delaglio et al. 1995), Sparky (Goddard and Kneller, University of California, San Francisco) and AutoAssign software (Moseley et al. 2001; Zimmerman et al. 1997) were used for processing, peak picking and analysis of the resonance assignments. Automated backbone resonance assignments for ^{15}N, ^{1}H, ^{13}C', $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ resonances were extended and in some cases verified manually. The assignments were validated using the Assignment Validation Software (AVS) (Moseley et al. 2004). Resonance assignments for full-length RbfA have been deposited in the BioMagResDB (BMRB ID: 18147).

Generation of backbone secondary structure and 3D model of RbfA using TALOS+ and CS-Rosetta

The backbone secondary structure of RbfA was generated by using the TALOS+ program (Shen et al. 2009). The standard CS-Rosetta protocol (Shen et al. 2008) was used to generate a 3D models of full-length RbfA from backbone chemical shift data. In these calculations, the RbfAΔ25 (PDB ID; 1KKG) structure was excluded from the fragment generation process, and 10,000 decoys were generated.

Results

Construction of expression system for PrS$_2$I$_N$ and PrS$_2$I$_C$-target

PSTI technology was designed on the basis of the intein technology taking advantage of PrS$_2$-tag to enhance the level of expression and solubility of target proteins. As illustrated in the schematic shown in Fig. 1a, two precursors, PrS$_2$I$_N$ (non-labeled) and PrS$_2$I$_C$-target (isotope labeled) which are both enhanced in their expression and solubility by PrS$_2$-tag, are mixed together to initiate the intein reaction. After formation of an intermediate, protein splicing occurs. This results in a segmentally-labeled protein, consisting of a PrS$_2$ (non-labeled) segment and a target protein (isotope labeled) segment, including 14 extra isotope-labeled amino acid residues derived from cloning site, streptavidin-binding purification tag (Strep-tag), and junction amino acids (Lockless and Muir 2009).

To establish the PSTI technology, expression vectors for two precursors, pCold-PSTI$_N$ and pCold-PSTI$_C$, were first constructed using pCold (sp-4) vectors (Suzuki et al. 2006), as shown in Fig. 1c. The pCold-PSTI$_N$ vector consists of TEE which enhances translation in *E. coli* (polypeptide sequence MNHKV) (Etchegaray and Inouye 1999), His$_6$-tag, PrS$_2$-tag and I$_N$. The pCold-PSTI$_C$ vector consists of TEE, PrS$_2$-tag, I$_C$, a fragment containing multiple cloning sites and a Strep-tag. Both expression vectors were constructed using nucleotide ACA-less genes so that the proteins can be selectively expressed using the Single Protein Production (SPP) system (Suzuki et al. 2005).

After construction of these vectors, we examined the efficiency of the expression level and solubility. As shown in Fig. 1d, we observed dramatic induction of both PrS$_2$I$_N$ and PrS$_2$I$_C$ (lanes 2 and 6). Importantly, both precursors were fully soluble (lane 3 and 7) as they both include a PrS$_2$ SET.

Production of segmentally-labeled PrS$_2$-RbfAΔ25 (^{15}N)

Next, we examined if the PrS$_2$-tag influences the structure of the fused protein using heteronuclear NMR spectroscopy. Previously, we have demonstrated that the PrS$_2$-tag does not affect the function of fused *E. coli* OmpR (Harlocker et al. 1995; Kobayashi et al. 2009). Based on these results, we anticipated that the PrS$_2$-tag and a target protein fused to the PrS$_2$-tag would fold independently and that NMR frequencies of the tagged protein should be identical to those obtained for the target protein by itself. To test this, we used *E. coli* RbfAΔ25, consisting of 108 residues excluding the 25 C-terminal residues of full-length RbfA as a control system. The resonance assignments (Swapna et al. 2001) and NMR-derived 3D structure (Huang et al. 2003) of RbfAΔ25 have been described previously.

SDS-PAGE data for the soluble fractions of PrS$_2$I$_N$ and PrS$_2$I$_C$-RbfAΔ25 are shown in lane 1 and 2, Fig. 2a. As shown in lane 1 and 2, both the PrS$_2$I$_N$ and PrS$_2$I$_C$-RbfAΔ25 precursors are highly soluble (see Supplementary Table S1). Precursors of intein which contain split intein domains tend to be insoluble because of their unstructured forms (Otomo et al. 1999; Yamazaki 1998) and denaturation/refolding steps are often required to form the functional intein structure. However, using the PSTI technology, the PrS$_2$-tag imparts a significant solubilizing effect on the two precursors which contain the split intein domains. Thus they can be used for protein splicing reaction without the need for denaturation and refolding steps, simplifying the protocol (Fig. 1b) for preparing "invisible SET" protein samples.

The two precursors were next mixed in a 1:1 molar ratio to initiate protein *trans*-splicing reaction (lane 3 in Fig. 2a). After a 16 h incubation at 25°C, a spliced product, PrS$_2$-tag (non-labeled)-RbfAΔ25 (^{15}N), appeared (lane 4 in Fig. 2a). Notably, the density of the PrS$_2$I$_N$ and the PrS$_2$I$_C$-RbfAΔ25

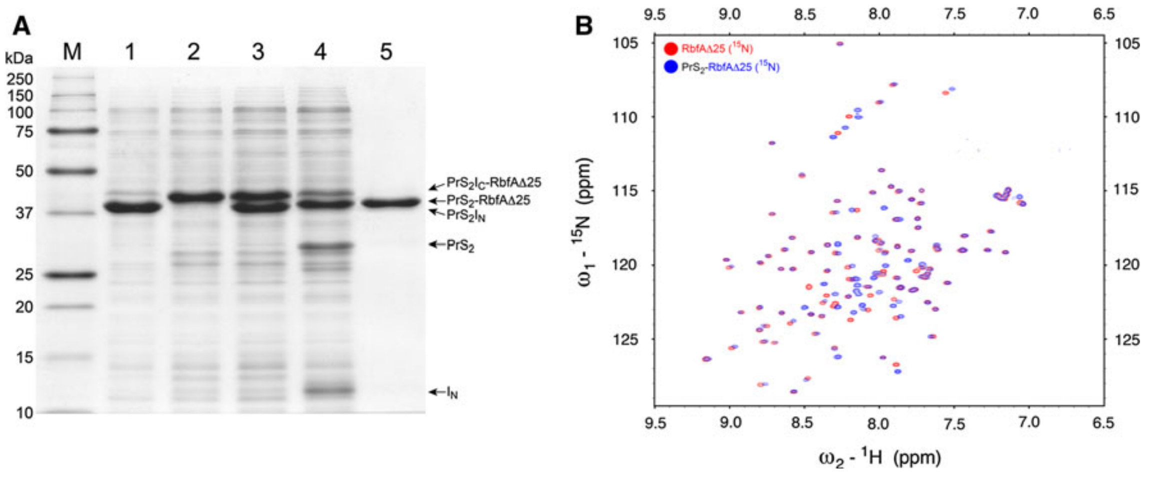

Fig. 2 In vitro protein ligation using PSTI and [^{1}H–^{15}N]-TROSY HSQC of segmentally-labeled PrS$_2$-RbfAΔ25 (^{15}N). **a** SDS-PAGE analysis of protein ligation of PrS$_2$I$_N$ and PrS$_2$I$_C$-RbfAΔ25. Protein ligation was performed for 16 h at 25°C, and the efficiency of the reaction was examined by SDS-PAGE. In addition to spliced product (PrS$_2$-RbfAΔ25) and both precursors, some by-product, cleaved PrS$_2$-tag and I$_N$ at 29 and 12 kDa are indicated by *arrows* (*lane 4*). *M* molecular weight marker; *Lanes: 1*, soluble fraction of PrS$_2$I$_N$; *2*, soluble fraction of PrS$_2$I$_C$-RbfAΔ25; *3*, 0 h of protein ligation; *4*, 16 h of protein ligation; *5*, purified PrS$_2$-RbfAΔ25 (^{15}N). **b** [^{1}H–^{15}N]-TROSY HSQC spectrum of PrS$_2$-RbfAΔ25 (^{15}N) and comparison with RbfAΔ25. The [^{1}H–^{15}N]-TROSY HSQC spectrum of segmentally-labeled PrS$_2$-RbfAΔ25 (^{15}N) (*blue peaks*) on the spectrum RbfAΔ25 (*red peaks*). Both spectra were recorded at 800 MHz and temperature of 35°C

bands are significantly reduced after the splicing reaction (compare lane 3 with lane 4). In contrast to the conventional approach, in which the two precursors are separately purified and then mixed for splicing reaction (Muona et al. 2010), using the PSTI technology, we observe a highly efficient splicing reaction by simply mixing the two cell lysates without prior purification of the precursors. Subsequently, the final splicing product was purified, as described in Materials and methods section, and analyzed by SDS-PAGE (lane 5 in Fig. 2a). The molecular weight of the product was estimated to be 35 kDa, consistent with its theoretical value (Supplementary Figure S2).

[^{1}H–^{15}N]-TROSY HSQC spectra were used to compare the differences between purified segmentally-labeled PrS$_2$-RbfAΔ25 (^{15}N) (blue peaks) and the non-tagged RbfAΔ25 protein (red peaks) (Fig. 2b). In the overlaid [^{1}H–^{15}N]-TROSY HSQC, most peaks observed from PrS$_2$-RbfAΔ25 (^{15}N) superimpose well with those of RbfAΔ25, although a small number of peaks exhibit small chemical shift perturbations. Most of the chemical shift differences between RbfAΔ25 and PrS$_2$-RbfAΔ25 are quite small, except in the 5-residue segment at C-terminal end adjacent to the Strep-tag (see Supplementary Figure S3). These results indicate that we successfully generated segmentally ^{15}N-labeled RbfAΔ25 in the fusion protein, and that all the signals from PrS$_2$-tag were silenced. Furthermore, we conclude from HSQC data that PrS$_2$-tag and RbfAΔ25 have folded independently without interfering with each other, and that the structure of RbfAΔ25 in the final fusion product shows no major conformational changes compared to RbfAΔ25 alone. The extra peaks observed in Fig. 2b arise from the linker region including the junction amino acids between PrS$_2$-tag and RbfAΔ25 (SGVH), the restriction enzyme site for cloning (GS for *Bam*HI site), and the Strep-tag (WSHPQFEK) added at the C-terminal end of the fusion protein for purification purpose.

Application of the PSTI technology to full-length RbfA

We next attempted to apply the PSTI technology to a protein whose structure cannot otherwise be studied by NMR due to its poor solubility and tendency to aggregate under solution NMR conditions. Full-length RbfA has limited solubility, and when prepared for NMR studies it aggregates within a few hours at room temperature and precipitates. For these reasons, our earlier attempts to determine resonance assignment for full-length RbfA using triple resonance NMR (TR-NMR) experiments were unsuccessful (Swapna et al. 2001).

To test if the PrS$_2$-tag is capable of providing a stable, soluble, full-length RbfA sample suitable for NMR experiments, the *rbfA* gene was cloned into pCold-PSTI$_C$ and expressed in M9-minimum medium supplemented with ^{15}NH$_4$Cl and ^{13}C-glucose. The soluble fractions of PrS$_2$I$_N$ (lane 1 in Fig. 3a) and PrS$_2$I$_C$-RbfA (lane 2 in Fig. 3a) (Supplementary Table S1) were mixed, and then subjected to the protein splicing reaction to generate segmentally-labeled PrS$_2$-RbfA (lane 4 in Fig. 3a). The resulting purified PrS$_2$ (non-labeled)-RbfA (uniformly ^{15}N- and ^{13}C-enriched) (lane 5 in Fig. 3a) was highly stable, even at concentrations higher than 0.5 mM (Supplementary Table

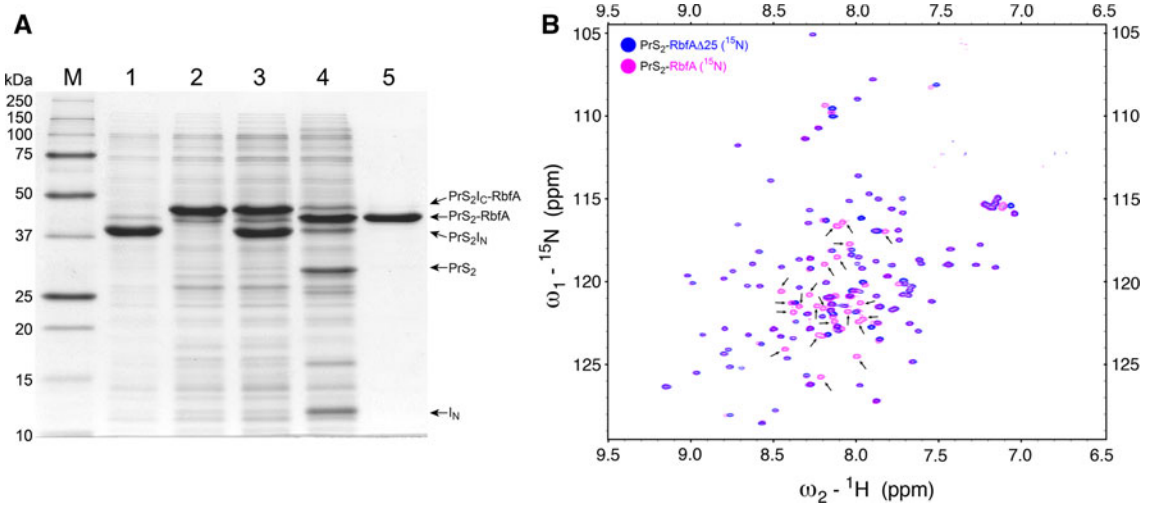

Fig. 3 In vitro protein ligation using PSTI and [^{1}H–^{15}N]-TROSY HSQC of segmentally-labeled PrS$_2$-RbfA (^{15}N, ^{13}C). **a** SDS-PAGE analysis of protein ligation of PrS$_2$I$_N$ and PrS$_2$I$_C$-RbfA. Protein ligation was performed at 25°C and the efficiency of the reaction was examined by SDS-PAGE. *M* molecular weight marker; *Lanes: 1*, soluble fraction of PrS$_2$I$_N$; *2*, soluble fraction of PrS$_2$I$_C$-RbfA; *3*, 0 h of protein ligation; *4*, 16 h of protein ligation; *5*, purified PrS$_2$-RbfA (^{15}N, ^{13}C). **b** [^{1}H–^{15}N] HSQC spectrum of PrS$_2$-RbfA (^{15}N, ^{13}C) and comparison with PrS$_2$-RbfAΔ25 (^{15}N). The [^{1}H–^{15}N]-TROSY HSQC of PrS$_2$-RbfA (^{15}N, ^{13}C) (*magenta peaks*) was overlaid with PrS$_2$-RbfAΔ25 (^{15}N) (*blue peaks*). *New peaks*, including those subsequently assigned to residues in the 25-residue C-terminal segment of RbfA, are indicated by *arrows*

S1). Remarkably, the resulting fusion protein remained soluble at ~0.5 mM concentration for more than one month at room temperature (data not shown).

[^{1}H–^{15}N]-TROSY HSQC spectrum of purified segmentally-labeled PrS$_2$-RbfA (^{15}N, ^{13}C) shown in Fig. 3b (magenta peaks) has very similar spectral features compared with the PrS$_2$-RbfAΔ25 (^{15}N) fusion construct (blue peaks). The [^{1}H–^{15}N]-TROSY HSQC spectrum of PrS$_2$-RbfA (^{15}N, ^{13}C) exhibits well-dispersed peaks, and the C-terminal 25-residues do not severely affect the chemical shift values in the N-terminal 108 residues, indicating that there is no significant perturbation of the 3D structure of RbfAΔ25. Importantly this spectrum exhibits some additional peaks (arrows in Fig. 3b), which were subsequently assigned to the C-terminal 25-residue segment of full-length RbfA.

The single PrS-tag improves the NMR properties of the fusion protein compared with PrS$_2$-tag

In order to compare the differences in secondary and tertiary structures between RbfA and RbfAΔ25, we used segmentally-labeled PrS$_2$-RbfA (^{15}N, ^{13}C) to collect TR-NMR spectra required for determining backbone resonance assignments. However, we observed that most peaks in these 3D TR-NMR experiments recorded on PrS$_2$-RbfA (^{15}N, ^{13}C) are weak, despite the relatively high protein concentration (0.4–0.5 mM). We suspect that the large molecular weight of the fused PrS$_2$-RbfA (38 kDa), together with any transient interactions that may occur between PrS$_2$-tag and RbfA, contribute to broadening the resonance line widths. In order to overcome this line-broadening problem, we first added a longer, flexible linker, consisting of ten consecutive Glycine residues, between the PrS$_2$-tag and RbfA. However, this strategy did not provide a significant improvement in the TR-NMR data (data not shown).

We next attempted to shorten the PrS$_2$-tag to include only a single N-terminal domain (NTD) of protein S (PrS). While the two tandem repeated NTDs of protein S from *M. xanthus* allow a one-step affinity purification through the PrS$_2$-tag using myxospores, the single PrS domain is not capable of binding to the myxospores (Kobayashi et al. 2009). However, the single PrS-tag is still capable of both high expression and solubility enhancement of the HAMP domain of *E. coli* EnvZ (Kishii et al. 2007). Therefore, we tested the smaller single-domain PrS-tag as a SET for RbfA.

A pCold-SPSTI$_N$ expression vector, which encodes the PrS and I$_N$ domain (PrSI$_N$), was expressed in a soluble form and used to produce PrS-RbfA (Supplementary Figure S4). PrS-RbfA has a molecular weight of 28 kDa, compared to 38 kDa for the PrS$_2$-RbfA construct. The smaller fusion protein, PrS-RbfA (^{15}N, ^{13}C), is stable in solution (see Supplementary Table S1), with no evidence of precipitation at about 0.5 mM protein concentration after over more than one month at room temperature. Its [^{1}H–^{15}N]-TROSY HSQC spectrum is nearly identical to that of PrS$_2$-RbfA (^{15}N, ^{13}C) (Supplementary Figure S5).

PrS-RbfA (^{15}N, ^{13}C) provides much better TR-NMR data than PrS$_2$-RbfA (^{15}N, ^{13}C) (Supplementary Figure S6). These TR-NMR data were adequate for determining an extensive set of backbone resonance assignment; N: 96.0%, H: 96.0%, C$^\alpha$: 98.6%, C$^\beta$: 97.0% and C': 78.0%, respectively. A summary of the intra-residue and sequential connections used to determine these resonance assignments is presented in Supplementary Figure S7 and the assigned [^{1}H–^{15}N]-TROSY HSQC spectrum is shown in Fig. 4a. The newly detected peaks assigned to the 25-residue C-terminal segment have backbone amide H^N resonance frequencies ranging from 8 ppm to 8.5 ppm (red letters in Fig. 4a), indicating that this C-terminal segment has a helical or disordered structure in full-length RbfA.

The secondary structure of full-length RbfA, summarized in Fig. 4b, was next determined from these backbone chemical shift data using TALOS+ (Shen et al. 2009). The full-length RbfA protein (black bar graph) has similar secondary structure to that observed in the 3D solution NMR structure of RbfAΔ25 (red and blue bars) (Huang et al. 2003). These backbone chemical shift data in the 25-residue C-terminal end of RbfA (residue 109-133) indicate that it is a largely disordered polypeptide segment.

In order to better compare full-length RbfA with RbfAΔ25, and to better characterize the conformation in the 25-residue C-terminal segment, backbone [^{1}H–^{15}N] heteronuclear NOE (HetNOE) data were also obtained using segmentally-labeled PrS-RbfA (^{15}N, ^{13}C) (Fig. 4c). Comparing these HetNOE data for full-length PrS-solubilized RbfA with those of RbfAΔ25, we observed that each backbone N–H site of RbfA has a HetNOE value similar of the corresponding site in RbfAΔ25. However, the ^{1}H–^{15}N NOE values in the 25-residue C-terminal segment of full-length are small or negative, indicating that this 25-residue segment is largely disordered in the full-length construct.

Based on experimental chemical shifts including ^{13}C$^\alpha$, ^{13}C$^\beta$, CO, ^{15}N and ^{1}HN, a chemical-shift-directed structure prediction of the full-length RbfA structure was performed using CS-Rosetta (Shen et al. 2008). These results are summarized in Fig. 4d, f, and in Supplementary Figure S8 and Supplementary Table S2. The best-scored decoy from a total of 10,000 calculated conformers is shown in Fig. 4d. This model is very similar to the published solution NMR structure of RbfAΔ25 (Fig. 4e). Importantly, the full-length RbfA structure has a helix-kink-helix motif in residue segment 74–76, a typical structural feature of the RbfA family (Huang et al. 2003). Residue Ala75, which is well conserved at this kink region is also localized in the kink position of full-length RbfA (red colored residue in both panels).

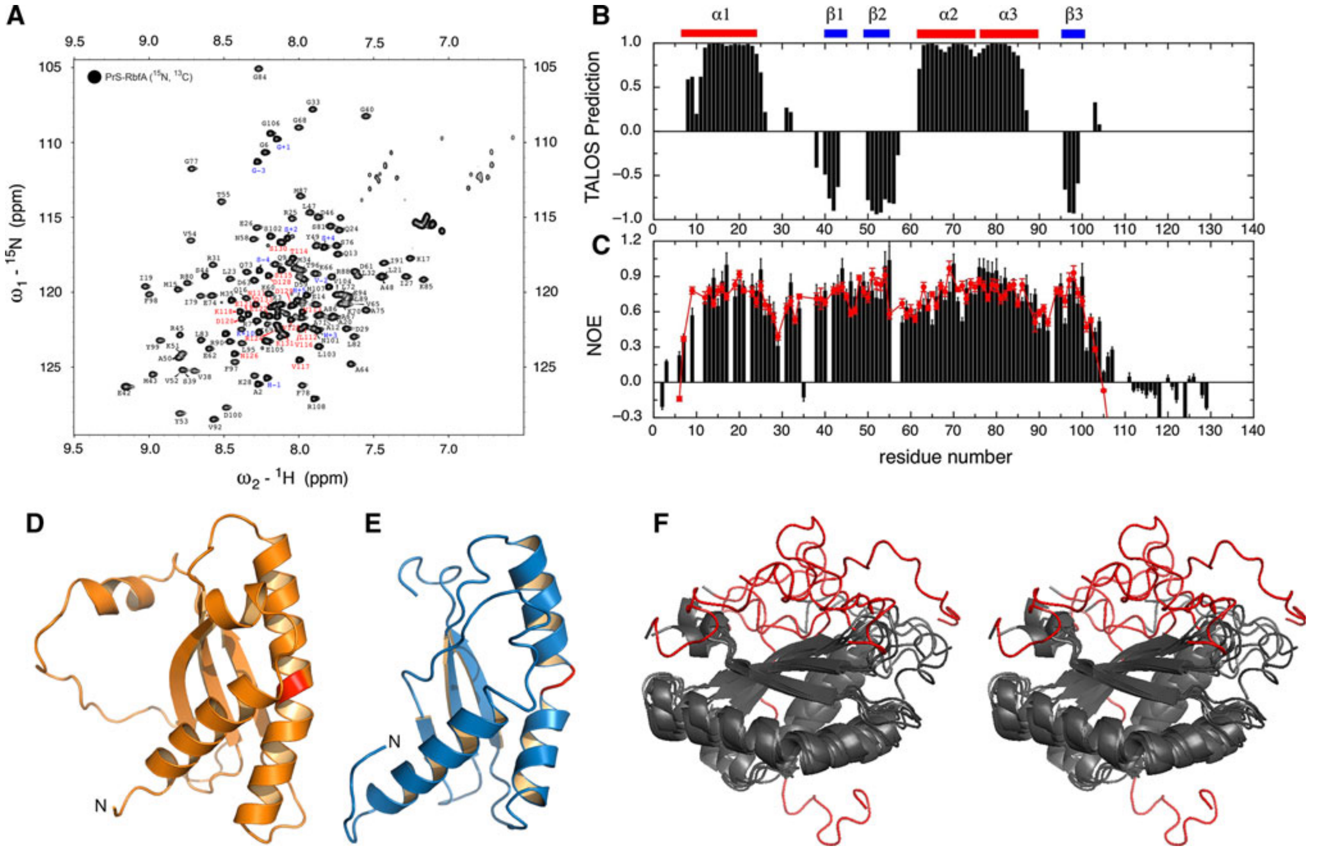

Fig. 4 [^{1}H–^{15}N]-TROSY HSQC of full-length *E. coli* RbfA with sequence-specific assignments. **a** [^{1}H–^{15}N]-TROSY HSQC spectrum of segmentally-labeled PrS-RbfA (^{15}N, ^{13}C) recorded at 800 MHz and temperature of 35°C. The peaks indicated by *red letters* are assigned to residues in the 25-residue C-terminal segment of full-length RbfA. The peaks from a linker at N-terminal end of RbfA are labeled with *minus* (from −1 to −4; *blue letters*), and the peaks from restriction enzyme site and Strep-tag at C-terminal are labeled with plus (from +1 to +10; *blue letters*). **b** Characterization of the secondary structure of *E. coli* RbfA using TALOS+ (Shen et al. 2009). The secondary structure elements of RbfAΔ25 (*red*, α-helix and *blue*, *β*-strand *bars*) determined by solution NMR (Huang et al. 2003) are indicated as a guide along the *top* of the figure. **c** Variation of heteronuclear [^{1}H–^{15}N] NOE at 600 MHz of PrS-solubilized full-length RbfA with PrS-tag (*black histograms*) and RbfAΔ25 (*red line graph*) at 35°C. Samples were prepared at 0.2 mM protein concentration. **d** CS-Rosetta model of full-length *E. coli* RbfA. The *ribbon diagram* shows the 3D structure of full-length PrS-solubilized RbfA determined using backbone chemical shift data with CS-Rosetta (Shen et al. 2008). The α-helix in the C-terminal region was observed in only some of the Rosetta models, and may possibly be transiently present in the solution structure. Residue Ala75 is labeled by red color. **e** A *ribbon diagram* of the solution NMR structure RbfAΔ25 determined using extensive NMR data (PDB ID; 1KKG) (Huang et al. 2003). Residue Ala75 is labeled by *red color*. **f** The stereo view of overlay of the five CS-Rosetta decoys of full-length RbfA with lowest rescored Rosetta energy

Superimposition of the five lowest rescaled energy CS-Rosetta structures, shown in Fig. 4f, demonstrates that (1) the low energy CS-Rosetta structures are very similar to each other throughout most of the structure, indicating good convergence of the calculations; (2) the structure derived from CS-Rosetta is similar to that reported for truncated RbfAΔ25, indicating that the PrS-tag does not change the structure, and (3) the C-terminal 25 residues do not converge to a unique structure in the CS-Rosetta calculations, supporting the view that this region is not well-ordered. Some (but not all) of the low energy CS-Rosetta models, like the one shown in Fig. 4d, predict some possible tendency to form an α-helix in residues 105–112 of the C-terminal segment. Based on these CS-Rosetta calculations, the ^{1}H–^{15}N hetNOE data, and the backbone chemical shifts themselves, we conclude that the 25-residue C-terminal segment of full-length RbfA is largely disordered in the absence of binding partners. This novel structural information could not be determined for the full-length RbfA protein prior to our use of PSTI technology.

Discussion

The PSTI technology has been developed for NMR protein sample preparation by combining the PrS_2 SET with intein methods for segmental labeling. The combined technology is well suited for preparing NMR samples of proteins which exhibit difficulties in their expression and/or solubility. In this report we applied the technology to full-length RbfA, which cannot otherwise be prepared in a stable soluble form. Using PSTI technology we successfully produced segmentally-labeled PrS_2-RbfA. The PrS_2-tag significantly enhanced the solubility and stability of full-length RbfA (Supplementary Table S1), allowing the collection of some NMR spectra. Importantly, the 2D [^{1}H–^{15}N]-TROSY HSQC spectrum of PrS_2-RbfAΔ25 is almost identical to that of RbfAΔ25 (Fig. 2b), indicating that PrS_2-tag does not influence the structure of the fused target protein. By then switching to a smaller protein tag, PrS (10 kDa), good quality TR-NMR spectra could be recorded on the full-length RbfA. Using these spectra, we determined nearly complete backbone ^{1}H, ^{15}N and ^{13}C resonance assignments (Fig. 4a). Based on these chemical shift data, full-length RbfA was shown to share a similar secondary and a tertiary structure with RbfAΔ25, except for the C-terminal 25-residue segment of RbfA. This result is consistent with our previous observation that the RbfAΔ25 largely retains the RbfA function to suppress growth defects at lower temperature, suggesting that the truncation does not result in major conformational changes in the protein structure (Huang et al. 2003; Swapna et al. 2001). Our results also indicate that the C-terminal 25-residue segment is largely disordered, with perhaps some tendency to form a transient α-helix in the segment of residues 105–112. This is the first example using the PSTI technology for the assignment and structural analysis of a protein which is not stable with respect to slow precipitation without a protein S tag.

Compared with other "invisible solubility enhancement tag" technologies (reviewed by Zhou and Wagner 2010), the PSTI technology has the following advantages. In conventional invisible SET systems, additional procedures, such as protease digestion and purification steps, are required to remove the isotope-labeled protein tag which is used for enhancement of the expression and solubility of the target protein (Durst et al. 2008; Kobashigawa et al. 2009; Zhou and Wagner 2010). However, protease digestion sometimes is problematic as it causes non-specific cleavages. For example, the PrS_2-tag itself is cleaved by Tobacco Etch Virus protease (TEV) even though the PrS_2-tag does not contain the canonical cleavage sequences (polypeptide sequence E-x-x-Y-x-Q-G) (Kobayashi et al. 2009). Moreover, in some cases the efficiency of protease digestion is low (Esposito and Chatterjee 2006). Therefore trial and error is sometimes required to identify a proper protease compatible with both the non-labeled protein tag and the target protein. Furthermore, the additional purification steps cause reduction in the yield of the final product.

The PSTI technology is designed to generate a segmental labeled protein by replacement of isotope-labeled PrS_2-tag with non-isotope labeled PrS_2 (PrS)-tag (Fig. 1a). Therefore it does not require the removal of the labeled SET by proteolysis and additional purification steps. Protease digestion and additional purification steps are also not required when using other in vivo protein ligation methods (Züger and Iwai 2005). However, in the method, precursor containing a labeled target protein is expressed without a SET, and the timing of the inductions and the concentration of inducers critically influences the final protein yield, requiring preliminary experiments to optimize conditions (Muona et al. 2010). In addition, this method cannot be used for poorly expressed proteins, since the target protein is expressed without an expression enhancement tag. Furthermore, the use of uniformly labeled ^{13}C-arabinose is quite expensive (US $2,400 for a 1-l culture containing 0.2% of arabinose; Omicon Biochemicals, Inc.).

On the other hand, using PSTI technology, both the unlabeled PrS_2I_N ($PrSI_N$) and the isotope-enriched PrS_2I_C-target protein precursors are linked to a SET, which enhances their expression and solubility (Fig. 1d, Supplementary Table S1). Thus, the PSTI technology is more robust for producing poorly soluble and/or poorly expressed SET-tagged target proteins. Also, as every construct is designed to utilize the SPP system using ACA-less sequences and pCold (sp-4) vectors (Suzuki et al. 2006, 2005), *E.coli* cultures used for producing the isotope-enriched segments can be condensed 20–50 fold without affecting the protein yield (Suzuki et al. 2006). This allows significant cost savings, particularly when using expensive isotope-labeled materials such as stereo-array isotope-labeled amino acids (SAIL) (Kainosho et al. 2006). Using the PSTI technology together with the 20-fold condensed SPP system, we have successfully obtained samples of PrS_2-RbfA (^{15}N) providing good quality [^{1}H–^{15}N]-TROSY HSQC spectra (Supplementary Figure S9).

Using PSTI technology, a single unit of PrS (10 kDa) was found to be most useful for providing good quality TR-NMR spectra (Supplementary Figure S6). Larger protein tags such as PrS_2 (20 kDa), GST (26 kDa) or MBP (42 kDa) appear to cause broadening of the resonance line widths, precluding TR-NMR studies. The advantage of using a smaller SET, GB1 (6.2 kDa), for NMR studies has been documented previously (Huth et al. 1997). When GB1 is fused with DFF45, the fusion protein provides high-quality well resolved [^{1}H–^{15}N] HSQC, spectra suitable for structural studies (Zhou et al. 2001b). Importantly, such GB1-tagged proteins also provide good quality spectra in TR-NMR studies, allowing complete structure determination of fused proteins (Kobashigawa et al. 2009; Zhou et al. 2001a). However, GB1 is only infrequently used as a purification tag.

While the GB1 "invisible SET" is a powerful system for NMR sample preparation, a unique feature of using the

PrS_2/PrS SETs is that they have similar pI's (4.61 and 4.60), similar solubilizing properties, and other similar biophysical properties. This allows applications in which the PrS_2-tag is first explored, and later is replaced with the smaller PrS-tag for preparing "invisible SET" samples for NMR studies; in this case both the PrS and PrS_2 fusion constructs are expected to have similar solubility properties. The PrS_2-tagged construct is well-suited for purification applications using myxospores, while the PrS-tagged construct is most suitable for "invisible SET" applications in protein NMR studies. In conclusion, the PSTI technology provides an alternative 'invisible SET' approach, which can be generated using a simple and robust protocol, that will be generally useful for protein sample production and NMR studies.

Acknowledgements We thank Tom W. Muir (Rockefeller University) and Yohei Miyanoiri (Nagoya University) for helpful discussions. This work was supported in part by the National Institutes of General Medical Science Protein Structure Initiative (PSI-Biology) program, grant U54 GM 094597 (to G. T. M).

References

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6(3):277–293

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(1):21–30

Durst FG, Ou HD, Lohr F, Dotsch V, Straub WE (2008) The better tag remains unseen. J Am Chem Soc 130(45):14932–14933

Elleuche S, Poggeler S (2010) Inteins, valuable genetic elements in molecular biology and biotechnology. Appl Microbiol Biotechnol 87(2):479–489

Esposito D, Chatterjee DK (2006) Enhancement of soluble protein expression through the use of fusion tags. Curr Opin Biotechnol 17(4):353–358

Etchegaray JP, Inouye M (1999) Translational enhancement by an element downstream of the initiation codon in *Escherichia coli*. J Biol Chem 274(15):10079–10085

Fox JD, Kapust RB, Waugh DS (2001) Single amino acid substitutions on the surface of *Escherichia coli* maltose-binding protein can have a profound impact on the solubility of fusion proteins. Protein Sci 10(3):622–630

Harlocker SL, Bergstrom L, Inouye M (1995) Tandem binding of six OmpR proteins to the ompF upstream regulatory sequence of *Escherichia coli*. J Biol Chem 270(45):26849–26856

Huang YJ, Swapna GV, Rajan PK, Ke H, Xia B, Shukla K, Inouye M, Montelione GT (2003) Solution NMR structure of ribosome-binding factor A (RbfA), a cold-shock adaptation protein from *Escherichia coli*. J Mol Biol 327(2):521–536

Huth JR, Bewley CA, Jackson BM, Hinnebusch AG, Clore GM, Gronenborn AM (1997) Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR. Protein Sci 6(11):2359–2364

Inouye M, Inouye S, Zusman DR (1979) Biosynthesis and self-assembly of protein S, a development-specific protein of *Myxococcus xanthus*. Proc Natl Acad Sci USA 76(1):209–213

Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Mei Ono A, Guntert P (2006) Optimal isotope labelling for NMR protein structure determinations. Nature 440(7080):52–57

Kishii R, Falzon L, Yoshida T, Kobayashi H, Inouye M (2007) Structural and functional studies of the HAMP domain of EnvZ, an osmosensing transmembrane histidine kinase in *Escherichia coli*. J Biol Chem 282(36):26401–26408

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(3):145–150

Kobayashi H, Yoshida T, Inouye M (2009) Significant enhanced expression and solubility of human proteins in *Escherichia coli* by fusion with protein S from *Myxococcus xanthus*. Appl Environ Microbiol 75(16):5356–5362

Lockless SW, Muir TW (2009) Traceless protein splicing utilizing evolved split inteins. Proc Natl Acad Sci USA 106(27): 10999–11004

Moseley HNB, Monleon D, Montelione GT (2001) Automatic determination of protein backbone resonance assignments from triple resonance nuclear magnetic resonance data. Nucl Magn Reson Biol Macromol 339(Pt B):91–108

Moseley HN, Sahota G, Montelione GT (2004) Assignment validation software suite for the evaluation and presentation of protein resonance assignment data. J Biomol NMR 28(4):341–355

Muona M, Aranko AS, Raulinaitis V, Iwai H (2010) Segmental isotopic labeling of multi-domain and fusion proteins by protein trans-splicing in vivo and in vitro. Nat Protoc 5(3):574–587

Otomo T, Teruya K, Uegaki K, Yamazaki T, Kyogoku Y (1999) Improved segmental isotope labeling of proteins and application to a larger protein. J Biomol NMR 14(2):105–114

Qing G, Ma LC, Khorchid A, Swapna GV, Mal TK, Takayama MM, Xia B, Phadtare S, Ke H, Acton T et al (2004) Cold-shock induced high-yield protein production in *Escherichia coli*. Nat Biotechnol 22(7):877–882

Romanelli A, Shekhtman A, Cowburn D, Muir TW (2004) Semisynthesis of a segmental isotopically labeled protein splicing precursor: NMR evidence for an unusual peptide bond at the N-extein-intein junction. Proc Natl Acad Sci USA 101(17):6397–6402

Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu G, Eletsky A, Wu Y, Singarapu KK, Lemak A et al (2008) Consistent blind protein structure generation from NMR chemical shift data. Proc Natl Acad Sci USA 105(12):4685–4690

Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44(4):213–223

Smith DB, Johnson KS (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. Gene 67(1):31–40

Southworth MW, Adam E, Panne D, Byer R, Kautz R, Perler FB (1998) Control of protein splicing by intein fragment reassembly. EMBO J 17(4):918–926

Suzuki M, Zhang J, Liu M, Woychik NA, Inouye M (2005) Single protein production in living cells facilitated by an mRNA interferase. Mol Cell 18(2):253–261

Suzuki M, Roy R, Zheng H, Woychik N, Inouye M (2006) Bacterial bioreactors for high yield production of recombinant protein. J Biol Chem 281(49):37559–37565

Swapna GV, Shukla K, Huang YJ, Ke H, Xia B, Inouye M, Montelione GT (2001) Resonance assignments for cold-shock protein ribosome-binding factor A (RbfA) from *Escherichia coli*. J Biomol NMR 21(4):389–390

Yamazaki T (1998) Segmental Isotope Labeling for Protein NMR Using Peptiide Splicing. J Am Chem Soc 120:2

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

Zhou P, Lugovskoy AA, McCarty JS, Li P, Wagner G (2001a) Solution structure of DFF40 and DFF45 N-terminal domain complex and mutual chaperone activity of DFF40 and DFF45. Proc Natl Acad Sci USA 98(11):6051–6055

Zhou P, Lugovskoy AA, Wagner G (2001b) A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins. J Biomol NMR 20(1):11–14

Zimmerman DE, Kulikowski CA, Huang Y, Feng W, Tashiro M, Shimotakahara S, Chien C, Powers R, Montelione GT (1997) Automated analysis of protein NMR assignments using methods from artificial intelligence. J Mol Biol 269(4):592–610

Züger S, Iwai H (2005) Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies. Nat Biotechnol 23(6):736–740