nature biotechnology

Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies

Sara Züger[1] & Hideo Iwai[1,2]

Segmental isotopic labeling of proteins using protein ligation is a recently established *in vitro* method for incorporating isotopes into one domain or region of a protein to reduce the complexity of NMR spectra, thereby facilitating the NMR analysis of larger proteins[1–3]. Here we demonstrate that segmental isotopic labeling of proteins can be conveniently achieved in *Escherichia coli* using intein-based protein ligation. Our method is based on a dual expression system that allows sequential expression of two precursor fragments in media enriched with different isotopes. Using this *in vivo* approach, unlabeled protein tags can be incorporated into isotopically labeled target proteins to improve protein stability and solubility for study by solution NMR spectroscopy.

Segmental isotopic labeling offers a means of studying only the region of interest in proteins by NMR spectroscopy[1–3]. Unlike well-established approaches for isotopic labeling of specific amino acids[4], isotopic labeling of protein segments not only simplifies the spectral complexity by reducing the number of NMR-active atoms but also allows conventional triple resonance experiments, thereby permitting sequence-specific resonance assignments. Although NMR techniques using transverse relaxation-optimized spectroscopy[5] have extended the size limit for the observation of NMR signals to over 100 kDa[6,7], signal overlap in large proteins inherently hinders spectral analysis. Therefore, methods to reduce the signal overlap in NMR spectra, such as segmental isotopic labeling, are needed.

One method of segmental isotopic labeling relies on the *trans*-protein splicing activity of split inteins[1,3]. This approach requires purification of two or more precursor fragments with different isotopic enrichment in denatured form and refolding of these fragments *in vitro* to restore protein splicing activity for protein ligation[1,3,8]. Another approach, making use of native chemical ligation, involves ligation of a soluble protein domain containing an α-thioester with a protein (or a peptide) bearing an N-terminal cysteine, each of which are prepared separately with different isotope labels[2]. Both methods typically require laborious preparation of precursor fragments and optimization of the ligation conditions, limiting their applications[1–3,8].

In contrast to the *in vitro* methods, *in vivo* protein ligation has been used successfully in several applications, including protein cyclization[9–11] and *in vivo* fluorescence labeling[12,13]. This approach does not require individual preparation of precursor fragments before protein ligation nor additional chemical reagents[9–13]. We therefore decided to apply *in vivo* protein ligation to the segmental isotopic labeling of proteins. In particular, we focused on establishing a method for incorporating a solubility/stability enhancement tag into a labeled protein of interest, because the sequence near the splicing junction typically required for *trans*-protein splicing[10] has little importance for making fusion proteins. Fusion with bacterial proteins such as glutathione S-transferase (GST) is the most popular method to enhance solubility and stability and also allows convenient affinity purification[14]. However, such protein tags increase the complexity of the NMR spectra and must be removed before NMR analysis. Isotopic labeling of only the target protein in a fusion protein would eliminate the need to remove the solubility/stability enhancement tag. Thereby, this intein-based biosynthetic incorporation of protein tags makes it feasible to study unstable or insoluble proteins in fusion protein contexts by solution NMR.

To incorporate NMR-active isotopes into a specific protein domain in living cells, we introduced a dual vector system with two different tightly controlled inducible promoters for the individual expression of two precursor fragments for protein ligation. The bacterial culture medium was replaced during the protein expression of individual fragments to achieve segmental isotopic labeling (**Fig. 1**). Our model system consisted of the B1 domain of protein G (GB1), a chitin-binding domain (CBD) and the split intein DnaE from *Synechocystis* sp. strain PCC6803 (*Ssp* DnaE), which catalyzes *trans*-protein splicing[15]. We used two plasmids containing two different replicons (RSF1030 and ColE1) that can coexist in *Escherichia coli* (**Fig. 2**). One plasmid contains a T7/*lac* promoter for expression of the N-terminal portion of *Ssp* DnaE ($intein_N$) fused to GB1 (**Fig. 2a**). The other plasmid is designed to express the C-terminal portion of *Ssp* DnaE ($intein_C$) fused to the CBD under the tight control of the *araBAD* promoter[16] (**Fig. 2b**). The coexpression of these two fragments and the reconstitution of $intein_N$ and $intein_C$ in the bacterial cells restores the protein splicing activity of the split

[1]Department of Chemistry and [2]Department of Biochemistry, University of Saskatchewan, 110 Science Place, Saskatoon, Saskatchewan S7N 5C9, Canada. Correspondence should be addressed to H.I. (h.iwai@usask.ca or hide.iwai@gmail.com).

Published online 22 May 2005; doi:10.1038/nbt1097

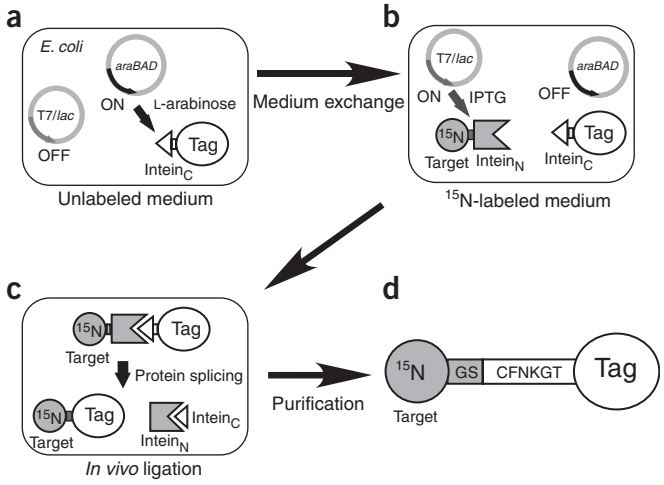

Figure 1 Biosynthetic preparation of segmentally isotope-labeled proteins using *trans*-protein splicing. (**a**) The C-terminal protein tag fused with intein$_C$ can be expressed by the addition of L-arabinose in unlabeled medium. (**b**) The expression of the N-terminal target protein fused with intein$_N$ can be subsequently induced by IPTG after the replacement of the medium with ^{15}N-labeled medium. (**c**) Protein splicing in bacterial cells produces a segmentally isotope-labeled protein by intein-mediated protein ligation. (**d**) Purified ligated product from the bacterial cells. The amino acid sequence of the linker connecting the tag and the target protein is shown.

intein, allowing self-catalytic ligation of the two protein fragments with a peptide bond without any additional cofactors or reagents (**Fig. 1**).

In vivo segmental isotopic labeling is achieved by selective expression of one of the two fragments in the medium containing the isotope (**Fig. 1**). To achieve isotopic labeling of the N-terminal segment, we grew *E. coli* cells harboring the two expression plasmids in unlabeled medium, and induced expression of the C-terminal fragment (**Figs. 1a** and **3a**). The cells were harvested by centrifugation and resuspended in ^{15}N-labeled M9 medium. Subsequently, expression of the N-terminal fragment was induced (**Figs. 1b** and **3a**). The *in vivo* ligated protein GB1-CBD was purified using the incorporated affinity tags and ion-exchange chromatography (**Fig. 3b**). The eluted fraction from the affinity chromatography contained a total of 10–15 mg of the mixture of the ligated product and the copurified precursor proteins from 1 liter of culture. After separating the ligated product and the precursor proteins by ion-exchange chromatography, we purified 1–2 mg/l of the ligated product of [^{15}N]-GB1-CBD (molecular mass 15 kDa; **Fig. 3b**, lane 2).

Protein ligation was demonstrated by the expected single peak in the spectrum of the HNCO experiment, which detects the correlation between the ^{15}N-labeled amide nitrogen of CBD and the ^{13}C-labeled carbonyl carbon of GB1 at the ligation site in [^{13}C]-GB1-[^{15}N]-CBD (**Fig. 4f**). The yield was substantially reduced compared with that of a fusion protein prepared by conventional methods without intein-mediated protein ligation. Expression levels of the individual fragments were as high as ~51 mg/l for the C-terminal fragment before the induction of the second expression, and ~35 mg/l for the N-terminal portion, as estimated by SDS-PAGE (**Fig. 3a**). The yield of the *in vivo* ligation seems to depend on three factors: degradation of the C-terminal fragment, the *in vivo* reconstitution efficiency of the split intein and the efficiency of the *trans*-protein splicing. The C-terminal fragment, which was expressed first, disappeared slowly after the second induction (**Fig. 3a**, lanes 5–8). This was presumably due to degradation of the C-terminal fragment as well as protein splicing upon reconstitution of the split intein. Degradation may be the primary factor reducing the yield. However, degradation of the C-terminal fragment was reduced when CBD was replaced with the protein tag GST (data not shown).

It was necessary to express the C-terminal fragment before the N-terminal fragment to reconstitute the split intein, because the C-terminal fragment containing both intein$_C$ and a protein tag is designed to be soluble, whereas the N-terminal fragment tends to be insoluble in the cells, depending on the fusion partner. Therefore, the solubility of intein$_N$ fused to any protein such as insoluble or membrane proteins in bacterial cells is likely to be a limiting factor of this method. The unspliced precursor proteins were always copurified with the ligated product (**Fig. 3b**), suggesting that the slow protein splicing reaction is also responsible for lowering the yield. The splicing reaction also continued during the purification (**Fig. 3b**, lane 1). Optimization of the reaction conditions, such as longer incubation of the precursor proteins *in vitro* and/or inclusion of a more efficient intein sequence, might improve the yield. Despite these limiting factors, the yield of our method is comparable to or better than that of existing *in vitro* methods[2,8]. Furthermore, our method is less labor-intensive since segmentally labeled proteins can be produced in one day.

We also tested the *in vivo* incorporation of a protein tag into other proteins (**Fig. 3b**, lane 3). The prion-inducing domain of yeast Sup35 spontaneously aggregates to form filaments[17], and therefore it is difficult to characterize this domain by solution NMR without making a fusion protein. The N-terminal residues 1–61 of yeast Sup35 protein[18] were ligated *in vivo* with the solubility enhancement tag GB1 at the C-terminus[19]. Lane 3 in **Figure 3b** shows the eluate from the affinity chromatography step and indicates that Sup35(1–61) ligated to the fusion tag (GB1) can be produced by *in vivo* protein ligation. Hence, our strategy is applicable for incorporating protein tags into aggregation-prone proteins.

Next, we investigated the selectivity of biosynthetic segmental isotopic labeling. **Figure 4** illustrates the segmental isotopic enrichment of our model system. The [^{1}H, ^{15}N]-heteronuclear single-quantum coherence (HSQC) spectrum of [^{15}N]-GB1-[^{15}N]-CBD, of which both N- and C-terminal fragments were expressed and ligated in ^{15}N-enriched M9 medium, illustrates the uniform ^{15}N-labeling of GB1-CBD (**Fig. 4a**). We could identify all the expected peaks of GB1 (ref. 20) and CBD[21] in this spectrum, confirming *in vivo* protein ligation. With the segmentally ^{15}N-labeled samples, we expect to observe only the N-terminal fragment (78 amino acids; **Fig. 4c**) or

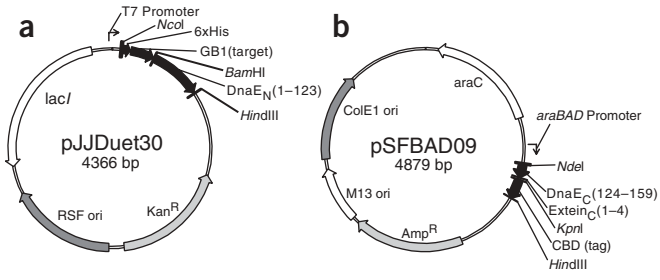

Figure 2 Structure of the two plasmids for *in vivo* segmental isotopic labeling. (**a**) Schematic map of pJJDuet30 for the expression of the N-terminal portion, derived from pRSFDuet-1 (Novagen). (**b**) Schematic map of pSFBAD09 for the expression of the C-terminal portion. This vector is derived from pBAD. See Methods for details.

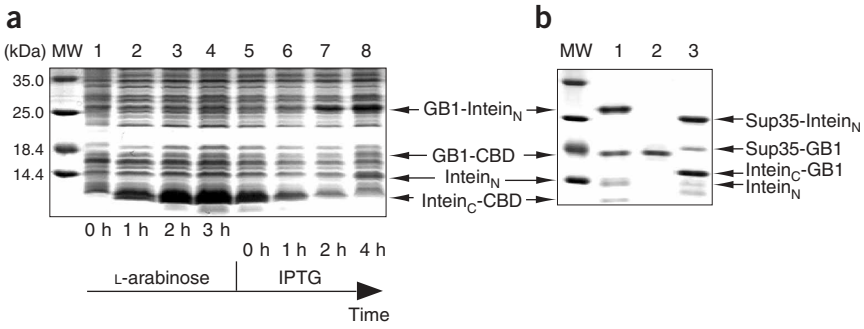

Figure 3 SDS-PAGE analysis of *in vivo* protein ligation. (**a**) SDS-PAGE analysis of the expression of GB1-Intein$_N$ and Intein$_C$-CBD. The 18% SDS gels were stained with Coomassie Blue. Lanes 1–8 are total cell lysates prepared by boiling in SDS loading buffer. Lane 1, before induction with 0.2% L-arabinose; lanes 2–4, 1, 2, and 3 h after the induction with L-arabinose, respectively; lane 5, before induction with 1 mM IPTG; lanes 6–8, 1, 2, and 4 h after the induction with 1 mM IPTG, respectively. GB1-Intein$_N$, GB1-CBD, Intein$_N$, and Intein$_C$-CBD are indicated by arrows. (**b**) SDS-PAGE analysis of the purification of the *in vivo* ligated proteins of GB1-CBD and Sup35-GB1. Lanes 1–2, purification of GB1-CBD. Lane 1, before loading to anion-exchange column after the IgG-Sepharose chromatography and dialysis; lane 2, elution from the anion-exchange chromatography. Lane 3, purification of Sup35-GB1, illustrating the elution from IgG chromatography. Sup35-intein$_N$, Sup35-GB1, Intein$_N$, and Intein$_C$-GB1 are indicated by arrows.

the C-terminal fragment (57 amino acids; **Fig. 4b**). **Figures 4b** and **4c** demonstrate that either the N-terminal or the C-terminal segments of GB1-CBD can be isotopically enriched with ^{15}N by controlling the expression and the isotope content of the culture media. Unfortunately, the selectivity of segmental isotopic labeling was reduced because of isotope-scrambling. In the [^{1}H, ^{15}N]-HSQC spectrum of the C-terminal-labeled sample GB1-[^{15}N]-CBD, we observed strong signals at the position of the N-terminal portion (GB1), which correspond to on average ~25% of the signal intensities of the C-terminal portion (CBD) (**Fig. 4b**, circles, and **d**, middle). On the other hand, only about 7% of the signal intensities of the N-terminal portion (GB1) were observed at the position of the C-terminal portion (CBD) in the spectrum of [^{15}N]-GB1-CBD (**Fig. 4c**, circles, and **d**, bottom). We were able to reduce the 'isotope scrambling' observed for the N-terminal segmental isotopic labeling to ~3% or less by introducing a washing step during the medium exchange to remove residual L-arabinose (**Fig. 4e**). This allowed isotopic enrichment of the N-terminal portion with high selectivity (>97%). The selectivity for C-terminal isotope labeling might be improved by replacing the CBD with a more stable protein domain, because the degradation of the C-terminal fragment could be reduced by replacing the CBD with GST.

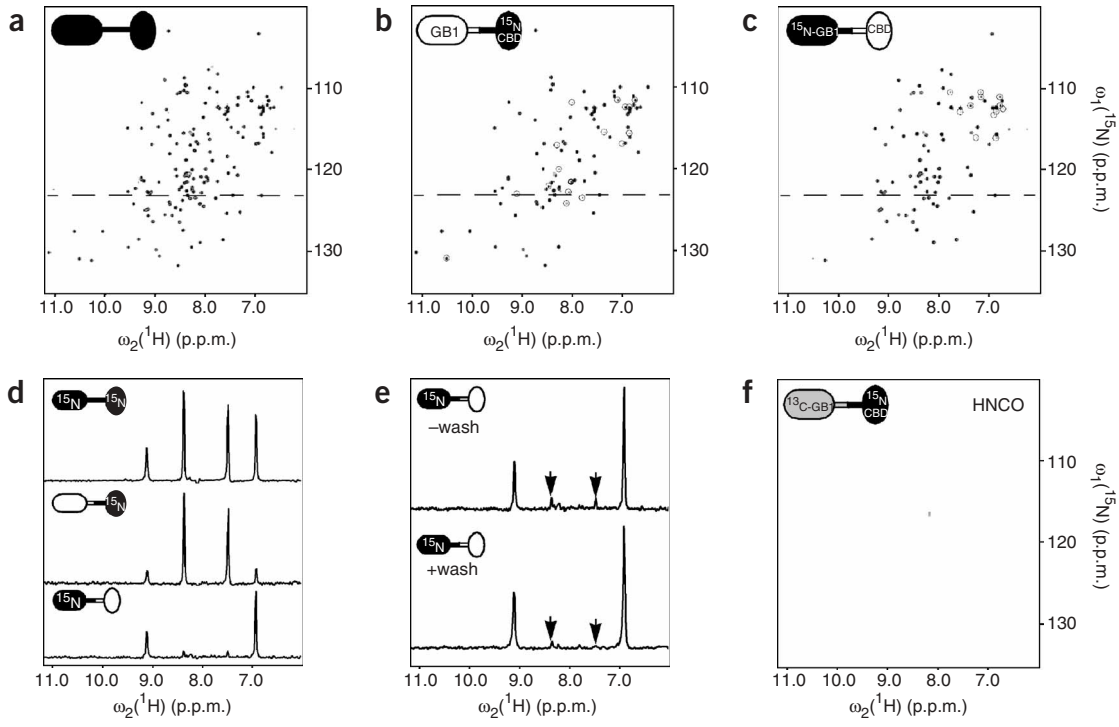

Figure 4 NMR spectra of the fully and segmentally isotope-labeled proteins. (**a**) [^{1}H, ^{15}N]-HSQC spectrum of uniformly ^{15}N-labeled GB1-CBD, [^{15}N]-GB1-[^{15}N]-CBD produced by *in vivo* protein ligation in ^{15}N-labeled M9 medium. (**b**) [^{1}H, ^{15}N]-HSQC spectrum of GB1-[^{15}N]-CBD. Circles indicate the signals from GB1 owing to isotope scrambling. (**c**) [^{1}H, ^{15}N]-HSQC spectrum of [^{13}C, ^{15}N]-GB1-CBD. The signals from CBD are marked by circles. (**d**) Cross sections along ω_2 at the dashed lines taken from the [^{1}H, ^{15}N]-HSQC spectra of [^{15}N]-GB1-[^{15}N]-CBD (top), GB1-[^{15}N]-CBD (middle), and [^{15}N]-GB1-CBD (bottom). (**e**) Cross sections along ω_2 at the dashed lines taken from the [^{1}H, ^{15}N]-HSQC spectra of [^{15}N]-GB1-CBD without washing step (upper) and with washing step (lower). (**f**) The first plane of the HNCO spectrum of [^{13}C]-GB1-[^{15}N]-CBD.

We chose the *araBAD* promoter to modulate the expression level of the C-terminal fragment to optimize *in vivo* protein ligation. However, a promoter such as *araBAD* may complicate the ^{15}N, ^{13}C double labeling of proteins required for triple resonance experiments, because the L-arabinose (2 g/l) used as the inducer can serve as a carbon source, unless ^{13}C-labeled L-arabinose is used. Although glucose is known to repress the protein expression under the *araBAD* promoter[16,22], the repression effect had no influence on the expression of the C-terminal part when a high concentration of L-arabinose was used together with M9 medium (**Fig. 3**, lanes 1–4). A high concentration of unlabeled L-arabinose could also dilute ^{13}C-labeled glucose for the C-terminal labeling. Therefore, we tested only N-terminal segmental isotopic labeling for ^{13}C, ^{15}N double labeling. In this case, we were successfully able to produce ^{13}C, ^{15}N doubly labeled protein with high selectivity using $^{13}C_6$-D-glucose and ^{15}N ammonium chloride as the sole carbon and nitrogen sources. The backbone resonance assignments of the GB1 portion in GB1-CBD were successfully performed using the conventional triple resonance assignment procedure as illustrated in the HNCACB spectrum of [^{13}C, ^{15}N]-GB1-CBD (**Supplementary Fig. 1** online). There was no interference of the NMR signals from the C-terminal CBD in backbone resonance assignment, indicating that our approach can incorporate an unlabeled protein tag at the C-terminal end of a labeled protein for NMR analysis.

In this article we have described a method for intein-based biosynthetic segmental labeling. The method may be generally useful for incorporating unlabeled solubility/stability enhancement tags into proteins that are insoluble in bacterial overexpression systems, which account for more than 50% of the proteins in a genome[23]. *In vivo* segmental isotopic labeling of soluble multi-domain proteins is also feasible, provided that the linker sequence introduced by protein ligation is not crucial for structural studies. It is anticipated that the method could be extended to a three-fragment ligation by introducing another inducible promoter and a split intein. Finally, our method may offer a tool to study proteins and their conformational changes in living cells by in-cell NMR[24].

METHODS

Plasmid construction. The 123 amino acids of the N-terminal portion of *Ssp* DnaE (intein$_N$) was amplified from pMSNd(16ex)[25] (a gift from Henry Paulus) using the oligonucleotides TT01 (5′-TCGGGATCCTGCCTCAGTTTTGGTAC-3′) and TT02 (5′-TGCCAAGCTTTATTTATTTAATTG-3′). The gene of the B1 domain of the IgG-binding protein G contains a single mutation A2Q (a gift from Stephen F. Marino) and was amplified using the oligonucleotides SK002 (5′-TTTCTAGATTAGGATCCTTCCGTTACGGTG-3′) and SK012 (5′-TCCT TACATATGCAGTACAAACTTATCCTG-3′). Both genes were ligated using the *Bam*HI site and cloned into pRSFDuet-1 (Novagen) using the *Nde*I and *Hin*dIII sites, which allows expression of the resulting GB1-Intein$_N$ fusion with the N-terminal hexahistidine tag. The resulting plasmid pJJDuet30 contains a T7/*lac* promoter and the fusion gene, of which expression can be induced by isopropyl β-D-thiogalactoside (IPTG). This plasmid also contains the kanamycin resistance gene (*Kan*^{R}), the RSF origin of replication (RSF ori) and the *lac*I gene (**Fig. 2a**). The gene of the C-terminal portion of *Ssp* DnaE (intein$_C$) synthesized by PCR, which will be described elsewhere, was amplified by the following primers SZ01 (5′-TGAATTTCATATGGTTAAAGTTATCG-3′) and SZ02 (5′-TTGGGTACCTTTGTTAAAACAGTTGGC-3′). The amplified gene codes 36 amino acids of intein$_C$ and the additional four amino acids CFNK of the natural flanking extein sequence of DNA polymerase IIIα. The gene of CBD was amplified from pTYB4 (New England Biolabs) using the primers PE73_CBD5 (5′-CGGAATTCGGTACCACAAATCCTGGTGTATCCG-3′) and PE74_CBD3 (5′-AAGGCCTTAAGCTTATTGAAGCTGCCACAAGGC-3′). The plasmid pSFBAD09 was constructed by introducing the genes of intein$_C$ and CBD with the *Kpn*I, *Nde*I and *Hin*dIII sites into the expression vector of a variant of the pBAD vector (Invitrogen) that lacks an additional *Nde*I site in the plasmid (**Fig. 2b**). The plasmid pSFBAD09 also contains the ColE1 origin of replication (ColE1 ori), the phage M13 origin of replication (M13 ori), the gene of araC (araC) and the ampicillin resistance gene (*Amp*^{R}). The plasmid pSZBAD1PG was derived from pSFBAD09 by replacing the gene of CBD with the gene of GB1. This plasmid expresses a fusion protein of intein$_C$ and GB1. The plasmid for the expression of Sup35(1–61) and intein$_N$ was derived from pJJDuet30 by replacing hexahistidine-GB1 by Sup35(1–61). The gene of Sup35(1–61) was amplified from the plasmid pET-Sup35NM[26] (a gift from Hideki Taguchi) using the two primers SZ044 (5′-CCTACCATGGGATCG GATTCAAAC-3′) and SZ037 (5′-CTGGATCCTTGATAGCCACC-3′) and cloned into pJJDuet30 using the *Nco*I and *Bam*HI restriction sites resulting in the plasmid pSZDuet30. The DNA sequences of the vectors pJJDuet30 and pSFBAD09 were deposited under DDBJ/EMBL/GenBank accession no. DQ016035 and DQ016036, respectively.

Expression of [^{15}N]-GB1-[^{15}N]-CBD. *E. coli* ER2566 (New England Biolabs) was transformed by the plasmids pJJDuet30 and pSFBAD09 for the protein expression. The cells were grown at 37 °C in 1 liter of M9 minimal medium supplemented with $^{15}NH_4Cl$ (0.5 g/l) as the sole nitrogen source, 50 μg/ml ampicillin, 50 μg/ml carbenicillin and 50 μg/ml kanamycin. The proteins were first induced with 0.2% (wt/vol) L-arabinose at an OD_{600} of 0.4–0.6 for 30 min followed by the addition of a final concentration of 1 mM IPTG and incubation for another 4 h. The cells were harvested and frozen at −80 °C for further purification.

Expression of [^{13}C, ^{15}N]-GB1-CBD. The cells harboring the two plasmids were initially grown in 0.75 liter Luria-Bertani (LB) medium containing 100 μg/ml ampicillin and 50 μg/ml kanamycin at 37 °C to an OD_{600} of 0.4–0.6 and induced with 0.2% (wt/vol) L-arabinose for 3 h. Subsequently, the cells were spun down for 20 min at 850*g* and resuspended in 1.25 liter M9 medium with $^{15}NH_4Cl$ (0.5 g/l) and [$^{13}C_6$]D-glucose (2 g/l) as the sole nitrogen and carbon sources, supplemented with 50 μg/ml carbenicillin, 50 μg/ml ampicillin and 50 μg/ml kanamycin. As an optional washing step, the cells were resuspended with 100 ml M9 medium to remove residual arabinose. Then the cells were spun down for another 20 min at 850*g* and resuspended in M9 medium. The culture was shaken for 15 min before the second induction to adjust the cells to the new environment. Expression of the N-terminal fragment was induced with a final concentration of 1 mM IPTG for another 4 h. The cells were harvested and frozen at −80 °C.

Expression of GB1-[^{15}N]-CBD. The cells harboring the two plasmids for GB1-CBD were grown in 0.75 liter M9 supplemented with $^{15}NH_4Cl$ (0.5 g/l) as the sole nitrogen source and the antibiotics at 37 °C to an OD_{600} of 0.4–0.6, induced with 0.2% (wt/vol) L-arabinose for 3 h. The second induction with IPTG was carried out in 1 liter of unlabeled M9 medium supplemented as described above.

Protein purification of GB1-CBD. The cells were lysed with B-PER II bacterial protein extraction solution (Pierce) and spun down for 20 min at 35,000*g*. The lysate was loaded on a HiTrap Chelating HP column (Amersham Biosciences) charged with Ni^{2+} ions. The hexahistidine tagged proteins and proteins associating with the His-tagged proteins were eluted with 50 mM sodium phosphate (pH 8.0), 300 mM NaCl and 250 mM imidazole. The eluted fraction was then loaded on an IgG-Sepharose (Amersham Biosciences)-packed column and washed with 50 mM HEPES (pH 7.2), 150 mM NaCl. The proteins bound to the resin were eluted with 50 mM glycine (pH 3.0) and immediately neutralized with 1 M Tris (pH 9.0) followed by overnight dialysis against 20 mM Tris (pH 8.0), 1 mM EDTA, 0.1 mM phenyl methyl sulfonyl fluoride. The dialyzed protein solution was loaded on a HiTrap Q FF anion-exchange column (Amersham Biosciences) and the flow-through containing the pure ligated protein was collected. The solution was then concentrated and the buffer exchanged to 20 mM sodium phosphate (pH 6.5), 0.01% (wt/vol) sodium azide, 10% (vol/vol) D_2O for NMR measurements.

Expression of Sup35-GB1. *E. coli* ER2566 bearing the two plasmids pSZBAD1PG and pSZDuet30 were grown at 37 °C in 0.25 liter of LB medium containing 100 μg/ml ampicillin and 50 μg/ml kanamycin. The proteins were

first induced with 0.2% (wt/vol) L-arabinose at an OD_{600} of 0.4–0.6 for 30 min followed by the addition of a final concentration of 1 mM IPTG and incubation for another 4 h. The cells were harvested and frozen at −80 °C for further purification. Purification of Sup35-GB1 was carried out with B-PER II bacterial protein extraction solution and IgG chromatography.

NMR spectroscopy. All NMR measurements were performed on a Bruker AV-600 equipped with a triple resonance cryoprobe at the Saskatchewan Structural Science Center. The water signals were suppressed by using the watergate sequence[27]. The spectra were recorded with ~0.2–0.4 mM samples at 25 °C either in 500 µl or 250 µl volumes. A total of 256 (ω_1) × 1,536 (ω_2) data points with $t_{1,max}$ = 47.8 ms, $t_{2,max}$ = 91.7 ms were recorded for the [^{1}H, ^{15}N]-HSQC experiment. For the HNCACB experiment[28], a total of 80 (ω_1) × 40 (ω_2) × 1,024 (ω_3) with $t_{1,max}$ = 4.6 ms, $t_{2,max}$ = 9.1 ms, $t_{3,max}$ = 65.9 ms were recorded resulting in the total recording time of 6 h. A total of 64 (ω_2) × 1536 (ω_1) with $t_{1,max}$ = 14.6 ms, $t_{2,max}$ = 91.7 ms were recorded for the 2D-[^{1}H, ^{15}N]-HNCO experiment. For data processing and spectra analysis, the programs PROSA[29] and XEASY[30] were used.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

We thank Henry Paulus, Hideki Taguchi and Stephen F. Marino for providing the plasmids containing *Ssp* DnaE intein, Sup35NM and GB1, respectively. We also thank Pui Hang Tam, Jennifer Jin, Sarah Figley for their technical assistance and the Saskatchewan Structural Science Center for the use of NMR instruments. H.I. thanks Peter Güntert for his continuous support. This study is supported by the National Science and Engineering Research Council of Canada (298493-4) and the Biomolecular Structure Teaching and Research Program at the University of Saskatchewan.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 19 November 2004; accepted 5 April 2005
Published online at http://www.nature.com/naturebiotechnology/

1. Yamazaki, T. *et al.* Segmental isotope labeling for protein NMR using peptide splicing. *J. Am. Chem. Soc.* **120**, 5591–5592 (1998).
2. Xu, R., Ayers, B., Cowburn, D. & Muir, T.W. Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies. *Proc. Natl. Acad. Sci. USA* **96**, 388–393 (1999).
3. Otomo, T., Ito, N., Kyogoku, Y. & Yamazaki, T. NMR observation of selected segments in a larger protein: central-segment isotope labeling through intein-mediated ligation. *Biochemistry* **38**, 16040–16044 (1999).
4. McIntosh, L.P. & Dahlquist, F.W. Biosynthetic incorporation of 15N and 13C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. *Q. Rev. Biophys.* **23**, 1–38 (1990).
5. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. USA* **94**, 12366–12371 (1997).
6. Yu, H. Extending the size limit of protein nuclear magnetic resonance. *Proc. Natl. Acad. Sci. USA* **96**, 332–334 (1999).
7. Wider, G. & Wüthrich, K. NMR spectroscopy of large molecules and multimolecular assemblies in solution. *Curr. Opin. Struct. Biol.* **9**, 594–601 (1999).
8. Otomo, T., Teruya, K., Uegaki, K., Yamazaki, T. & Kyogoku, Y. Improved segmental isotope labeling of proteins and application to a larger protein. *J. Biomol. NMR* **14**, 105–114 (1999).
9. Scott, C.P., Abel-Santos, E., Wall, M., Wahnon, D.C. & Benkovic, S.J. Production of cyclic peptides and proteins *in vivo*. *Proc. Natl. Acad. Sci. USA* **96**, 13638–13643 (1999).
10. Evans, T.C., Jr. *et al.* Protein *trans*-splicing and cyclization by a naturally split intein from the *dnaE* gene of *Synechocystis* species PCC6803. *J. Biol. Chem.* **275**, 9091–9094 (2000).
11. Iwai, H., Lingel, A. & Plückthun, A. Cyclic green fluorescent protein produced *in vivo* using an artificially split PI-*Pfu*I intein from *Pyrococcus furiosus*. *J. Biol. Chem.* **276**, 16548–16554 (2001).
12. Ozawa, T., Nogami, S., Sato, M., Ohya, Y. & Umezawa, Y. A fluorescent indicator for detecting protein-protein interactions *in vivo* based on protein splicing. *Anal. Chem.* **72**, 5151–5157 (2000).
13. Giriat, I. & Muir, T.W. Protein semi-synthesis in living cells. *J. Am. Chem. Soc.* **125**, 7180–7181 (2003).
14. Makrides, S.C. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev.* **60**, 512–538 (1996).
15. Wu, H., Hu, Z.M. & Liu, X.Q. Protein *trans*-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803. *Proc. Natl. Acad. Sci. USA* **95**, 9226–9231 (1998).
16. Guzman, L.M., Belin, D., Carson, M.J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).
17. King, C.Y. *et al.* Prion-inducing domain 2–114 of yeast Sup35 protein transforms *in vitro* into amyloid-like filaments. *Proc. Natl. Acad. Sci. USA* **94**, 6618–6622 (1997).
18. King, C.Y. & Diaz-Avalos, R. Protein-only transmission of three yeast prion strains. *Nature* **428**, 319–323 (2004).
19. Zhou, P., Lugovskoy, A.A. & Wagner, G. A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins. *J. Biomol. NMR* **20**, 11–14 (2001).
20. Orban, J., Alexander, P., Bryan, P. & Khare, D. Assessment of stability differences in the protein-G B1 and B2 domains from hydrogen-deuterium exchange - comparison with calorimetric data. *Biochemistry* **34**, 15291–15300 (1995).
21. Ikegami, T. *et al.* Solution structure of the chitin-binding domain of *Bacillus circulans* WL-12 chitinase A1. *J. Biol. Chem.* **275**, 13654–13661 (2000).
22. Boomershine, W.P., Raj, M.L., Gopalan, V. & Foster, M.P. Preparation of uniformly labeled NMR samples in *Escherichia coli* under the tight control of the *araBAD* promoter: expression of an archaeal homolog of the RNase P Rpp29 protein. *Protein Expr. Purif.* **28**, 246–251 (2003).
23. Christendat, D. *et al.* Structural proteomics of an archaeon. *Nat. Struct. Biol.* **7**, 903–909 (2000).
24. Serber, Z. & Dötsch, V. In-cell NMR spectroscopy. *Biochemistry* **40**, 14317–14323 (2001).
25. Mills, K.V. & Paulus, H. Reversible inhibition of protein splicing by zinc ion. *J. Biol. Chem.* **276**, 10832–10838 (2001).
26. Inoue, Y., Kishimoto, A., Hirao, J., Yoshida, M. & Taguchi, H. Strong growth polarity of yeast prion fiber revealed by single fiber imaging. *J. Biol. Chem.* **276**, 35227–35230 (2001).
27. Piotto, M., Saudek, V. & Sklenar, V. Gradient-tailored excitation for single-quantum NMR-spectroscopy of aqueous-solutions. *J. Biomol. NMR* **2**, 661–665 (1992).
28. Wittekind, M. & Mueller, L. HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha-carbon and beta-carbon resonances in proteins. *J. Magn. Reson. B* **101**, 201–205 (1993).
29. Güntert, P., Dötsch, V., Wider, G. & Wüthrich, K. Processing of multidimensional NMR data with the new software PROSA. *J. Biomol. NMR* **2**, 619–629 (1992).
30. Bartels, C., Xia, T.H., Billeter, M., Güntert, P. & Wüthrich, K. The program XEASY for computer-supported NMR spectral-analysis of biological macromolecules. *J. Biomol. NMR* **6**, 1–10 (1995).