Journal of Biomolecular NMR, **20:** 11–14, 2001.
KLUWER/ESCOM

A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins

Pei Zhou[a], Alexey A. Lugovskoy[a,b] & Gerhard Wagner[a,b,c,*]
[a]*Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, U.S.A.*
[b]*Committee on Higher Degrees in Biophysics, Harvard University, Cambridge, MA 02138, U.S.A.*
[c]*MIT/Harvard Center for Magnetic Resonance, Francis Bitter Magnet Laboratory, Cambridge, MA 02139, U.S.A.*

Received 18 December 2000; Accepted 5 February 2001

Key words: protein G, protein stabilization, protein tag

Abstract

Protein-fusion constructs have been used with great success for enhancing expression of soluble recombinant protein and as tags for affinity purification. Unfortunately the most popular tags, such as GST and MBP, are large, which hinders direct NMR studies of the fusion proteins. Cleavage of the fusion proteins often re-introduces problems with solubility and stability. Here we describe the use of N-terminally fused protein G (B1 domain) as a non-cleavable solubility-enhancement tag (SET) for structure determination of a dimeric protein complex. The SET enhances the solubility and stability of the fusion product dramatically while not interacting directly with the protein of interest. This approach can be used for structural characterization of poorly behaving protein systems, and would be especially useful for structural genomics studies.

Introduction

Probably the most serious problem for determining protein structures by NMR is to prepare well-behaving protein samples. After an interesting protein is identified, it is necessary to overproduce the protein and to find conditions under which the expressed protein is stable and soluble at concentrations at least in the 100-μM range. Efficiently making well behaving protein is of particular interest for structural genomics efforts, for which structure determination of representative protein folds at low cost is crucial. The major challenge in this approach is a robust preparation of the sample, as only 25% of overproduced proteins are biochemically stable and suitable for structural studies (Christendat et al., 2000).

Multiple approaches have been proposed to address this problem. Buffer conditions screening and introduction of point mutations in the protein of interest (Huang et al., 1996; Bagby et al., 1997) have been useful in some systems. However, these methods are largely guided by trial and error, which makes them unsuitable for high throughput studies where extensive screening for large numbers of proteins is prohibitively costly.

Protein tags have been used extensively to enhance expression, stability and solubility of fusion proteins and to facilitate their purification. However, these tags have to be removed for structural studies. For X-ray structure determination, the high mobility of the protein tag, which is often independent of the protein of interest, interferes with crystallization and structure determination. Although independent mobility is less of a concern in NMR spectroscopy, the size of most common protein tags, such as GST or MBP, is too large to make the structural characterization of a fusion protein by NMR possible.

Since it is only the size limitation that hinders the use of protein tags in NMR studies, we decided to use the 56 amino acid protein G B1 domain. It is a highly stable and soluble molecule, and the com-

*To whom correspondence should be addressed. E-mail: wagner@hms.harvard.edu

plete assignment of the chemical shifts has been reported (Gronenborn et al., 1991). An approach related to what we report here has been used successfully to screen whether bacterially expressed protein was folded, without going through time-consuming purification procedures (Huth et al., 1997). The authors reported that an N-terminal fusion with protein G B1 could be used to accelerate the initial assessment of protein NMR projects such that, in a matter of days, the solubility and stability of a protein can be determined. Here we used a non-removable protein G B1 tag to solubilize and stabilize the NMR samples during the process of structure determination. Using this system, we were able to obtain a significant improvement in both solubility and stability of the heterodimeric complex between regulatory domains of human DNA fragmentation factor 40 (DFF40) and human DNA fragmentation factor 45 (DFF45) CIDE domains. This allowed us to determine the structure of this complex. The SET approach presented here can provide a robust and straightforward way to produce biochemically well-behaving NMR samples for structure determination of proteins that are insufficiently soluble and stable by themselves.

Materials and methods

Design of a chimeric protein containing protein G B1 domain and DFF45 CIDE domain

The chimeric protein containing protein G B1 domain and DFF45 CIDE domain was generated by two-step PCR using three primers: (1) 5′- GGA GAT ATA CAT ATG CAG TAC AAG CTT ATC CTG -3′; (2) 5′- TAG AGT CCG GAT CTC GCC AGA TTC GGT TAC CGT GAA GGT TTT -3′; (3) 5′- GCA GCC GGA TCC TCA ATC TGA ATC TGA ATT GTT GTA TGC CCA 3′. First a PCR reaction was carried out using primer 1 and primer 2 encoding residues 1–56 of the protein G B1 domain and the first six residues of DFF45 CIDE domain (S12–L17). A second PCR reaction was carried out using the PCR product from the previous reaction and the third primer to obtain the final DNA insert containing a chimeric protein (protein G B1 M1–E56 and DFF45-CIDE S12–D100), which we call gbDFF45-CIDE in the later discussions. This insert was cloned in a pET30 a(+) vector between the Nde I site and BamHI site and the fusion protein was overproduced in *Escherichia coli* BL21(DE3) cell line.

Overexpression and purification of gbDFF45 CIDE (12–100), DFF45 CIDE (12–100) and their complex with DFF40 CIDE (1–80)

The DFF45 CIDE domain (12–100) was cloned into pGEX6P2 vector. The CIDE domain of DFF40 1–80 was sub-cloned into pET30 a(+) with the His_6-tag fused at the C-terminus. Cells transformed with GST-fused DFF45 CIDE or gbDFF45 CIDE were grown at 37 °C and induced with 1 mM isopropyl-D-thiogalactoside at 20 °C in M9-minimal media supplemented with ^{15}N-NH_4Cl (1 g/L) for production of ^{15}N labeled protein. Unlabeled DFF40 CIDE was obtained in a similar way except for growing cells in LB-media. The cell pellets of ^{15}N-labeled GST-DFF45 CIDE and ^{15}N-gbDFF45 CIDE were mixed with unlabeled DFF40 CIDE prior to sonication. The complexes between GST-DFF45/DFF40 CIDE complex and gbDFF45/DFF40 CIDE complex are purified by Ni^{2+} NTA affinity chromatography using the manufacturer's protocol (Qiagen). GST was removed by cleavage with Prescission protease (Amersham Pharmacia) at 4 °C for 2 h. Purified DFF40/45 CIDE complexes were exchanged into NMR buffer containing 20 mM phosphate, 50 mM NaCl, 5 mM DTT in H_2O/D_2O (9/1).

A protein G tagged CIDE/CIDE complex has superior biochemical behaviour and displays higher quality NMR spectra

The quality of the ^1H-^{15}N heteronuclear single quantum coherence (HSQC) spectrum is a sensitive measure of the biochemical behaviour of the protein in solution. We used such spectra to examine the solution behavior of the N-terminal domain of DFF45. Figure 1A shows that this domain has very little dispersion of its cross peaks, indicating that it is primarily unfolded. When adding unlabeled N-terminal domain of DFF40 (1–80) the dispersion of the HSQC spectrum increases dramatically (Figure 1B), indicating that the protein folds upon binding DFF40. However, the complex has very low solubility and precipitates within days. This changed dramatically when we used the SET approach. When ^{15}N gbDFF45 was used to form the complex with DFF40 the quality of the HSQC spectrum increased dramatically (Figure 1C). In order to quantitatively compare the properties of the complex, an HSQC spectrum of the untagged complex was compared with that of a complex where DFF45 was fused with the SET, recorded under exactly the same experimental conditions. The superior quality of the SET complex is obvious. Furthermore, addition of the

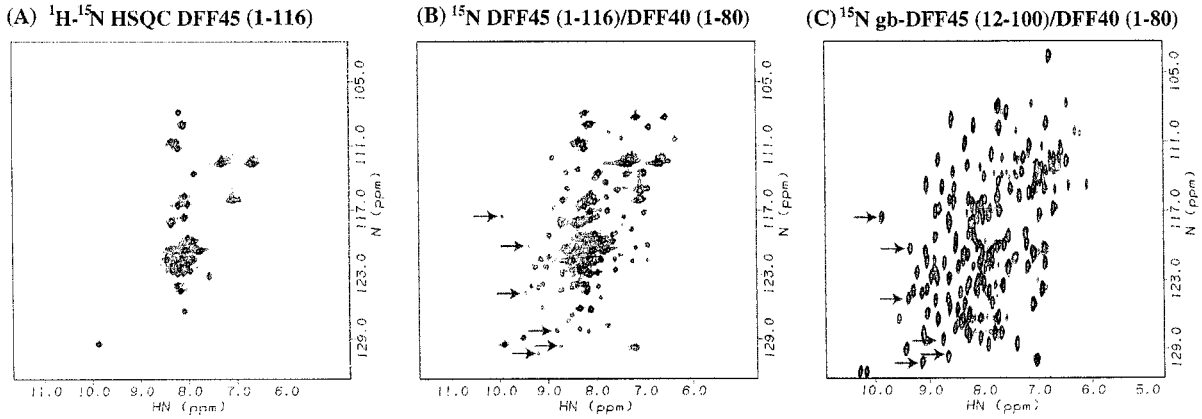

Figure 1. (A) ^{15}N-HSQC spectrum of the free ^{15}N-labeled DFF45 NTD (1–116). (B) ^{15}N-HSQC spectrum of the ^{15}N-labeled DFF45 NTD (1–116) in complex with unlabeled DFF40 NTD (1–80). Arrows indicate distinct resonances of folded DFF45. (C) ^{15}N-HSQC spectrum of the ^{15}N-labeled chimeric gbDFF45 (12–100) in complex with unlabeled DFF40 NTD (1–80). Arrows indicate distinct resonances of folded DFF45.

SET increased the solubility of the complex threefold (from 0.2 mM to 0.6 mM). The stability of the sample increased approximately sixfold (from 5 days to $>$ 30 days at 23 °C).

A substantial problem with the use of a fusion protein for NMR studies is an increase in spectral complexity. However, in our case the attachment of protein G B1 tag caused only little spectral overlap. In addition, the resonance frequencies of protein G B1 tag are very similar to those of free protein G B1 domain, and thus can be quickly identified.

The protein G B1 tag does not interact with the CIDE/CIDE complex

A common concern about the use of fusion proteins for structural studies is that the protein tag may interfere with the physical properties of the protein of interest. This seems to be a particular problem when the protein tag is bigger than a protein of interest. To examine this possibility we carefully examined the ^{15}N and ^{13}C NOESY spectra of the gbCIDE/CIDE complex. Despite a careful examination, we did not observe any interdomain NOEs between the CIDE/CIDE complex and the attached protein G B1 tag, indicating that the latter is not packing against either of the CIDE domains. This observation is further supported by a distinct relaxation behavior and narrow linewidths of the protein G B1 tag resonances, compared to those of the CIDE/CIDE complex (data not shown). To analyze the spectra of the complex we used TROSY-type spectra (Pervushin et al., 1997; Salzmann et al., 1999) and found them to be especially beneficial for this situation. The intensities of the resonances of the slowly tumbling CIDE/CIDE complex were significantly enhanced compared to those of rapidly tumbling protein G B1 tag. We attribute the distinct NMR properties of protein G B1 tag to its high acidic/basic nature, which causes it to be solvent accessible, rather than to pack against the CIDE/CIDE complex.

Application in structural genomics

Structural genomics, focusing on the rapid structure determination of all protein folds, has recently become a prominent area of structural biology. However, experiments on *Methanobacterium thermoautrophicum* showed that only a limited number of proteins are suitable for either NMR or X-ray analysis. Up to 75% of the targets of interest need further optimization on a case-specific basis, which hampers the high throughput technology (Christendat et al., 2000). Protein fusions with tags, such as GST and MBP, usually display better solubility and stability. However, these tags need to be cleaved off, and the biochemical behaviour of the protein changes following this removal, often resulting in decreased solubility of the protein of interest.

Our approach of creating a chimeric protein by selecting the highly charged, soluble, but yet small protein G B1 domain as a tag may prove useful as a general way to approach this problem. It would be especially suited for NMR studies of protein domains with molecular weight below 30 kDa.

Conclusions

In this paper we describe the use of protein G B1 domain as a SET for the DFF45 CIDE/DFF40 CIDE complex. This approach led to a significant improvement in solubility and stability of the sample, which enabled the detailed structural characterization of this complex system (Zhou et al., manuscript submitted). Although this approach has only been tested for a single protein system, and the general applicability has yet to be demonstrated, we believe that this method can make the difference to enable structural studies of poorly behaving proteins and may be particularly beneficial for structural genomics studies.

Acknowledgements

This work was supported by NIH (GM 47467) and NSF (MCB9316938).

References

Bagby, S., Tong, K.I., Liu, D., Alattia, J.R. and Ikura, M. (1997) *J. Biomol. NMR*, **10**, 279–282.

Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Savchenko, A., Cort, J.R., Booth, V., Mackereth, C.D., Saridakis, V., Ekiel, I., Kozlov, G., Maxwell, K.L., Wu, N., McIntosh, L.P., Gehring, K., Kennedy, M.A., Davidson, A.R., Pai, E.F., Gerstein, M., Edwards, A.M. and Arrowsmith, C.H. (2000) *Nat. Struct. Biol.*, **7**, 903–909.

Gronenborn, A.M., Filpula, D.R., Essig, N.Z., Achari, A., Whitlow, M., Wingfield, P.T. and Clore, G.M. (1991) *Science*, **253**, 657–661.

Huang, B., Eberstadt, M., Olejniczak, E.T., Meadows, R.P. and Fesik, S.W. (1996) *Nature*, **384**, 638–641.

Huth, J.R., Bewley, C.A., Jackson, B.M., Hinnebusch, A.G., Clore, G.M. and Gronenborn, A.M. (1997) *Protein Sci.*, **6**, 2359–2364.

Pervushin, K., Riek, R., Wider, G. and Wüthrich, K. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 12366–12371.

Salzmann, M., Wider, G., Pervushin, K., Senn, H. and Wüthrich, K. (1999) *J. Am. Chem. Soc.*, **121**, 844–848.

Zhou, P., Lugovskoy, A.A., McCarty, J., Li, P. and Wagner, G., Manuscript submitted.