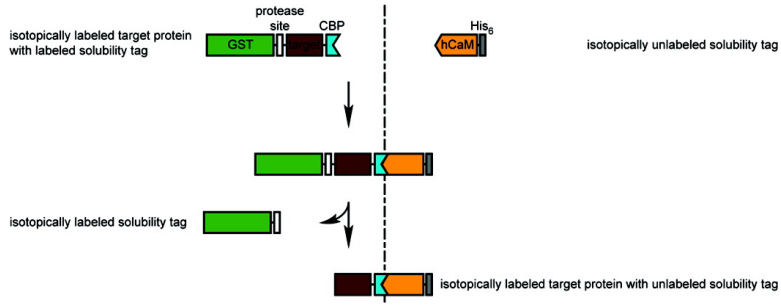
Communication

The Better Tag Remains Unseen

Florian G. Durst, Horng Der Ou, Frank Lo#hr, Volker Do#tsch, and Wesley E. Straub

J. Am. Chem. Soc., **2008**, 130 (45), 14932-14933 • DOI: 10.1021/ja806212j • Publication Date (Web): 21 October 2008



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
COMMUNICATIONS

Published on Web 10/21/2008

The Better Tag Remains Unseen

Florian G. Durst, Horng Der Ou, Frank Löhr, Volker Dötsch,* and Wesley E. Straub

Institute of Biophysical Chemistry and Center for Biomolecular Magnetic Resonance, and Cluster of Excellence Macromolecular Complexes (CEF), University of Frankfurt, Max-von-Laue Strasse 9, 60438 Frankfurt am Main, Germany

Received August 7, 2008; E-mail: vdoetsch@em.uni-frankfurt.de

With advancements in protein expression, protein purification, and data acquisition technologies, protein structure determination by nuclear magnetic resonance (NMR) is currently most restricted by the process of preparing a concentrated, nonaggregated sample. Modern cryogenic probes have considerably lowered the amount of protein needed. However, protein concentrations of at least 100 μM are still required for structural analysis.[1,2] Many proteins or isolated domains tend to aggregate or precipitate at these concentrations, thus necessitating additional measures to increase their solubility. Several approaches have addressed this problem, including buffer screens and mutational studies on the target protein.[3−5] The trial and error nature of such methods led to a preference for fusion of the onerous target protein to a well behaved protein, which acts as a solubility enhancement tag (SET).[6] The most widely adopted fusion partners, GST and MBP, facilitate high level expression, efficient purification, and elevated protein concentration, but their relatively large size could potentially complicate the spectral assignment of the protein of interest. Therefore, GST and MBP fusions are typically removed prior to the recording of NMR spectra, at which time solubility problems are often reintroduced. Smaller fusion proteins, such as protein D (95 amino acids) or protein G (B1 domain) (56 amino acids), have been pursued but still contribute a cumbersome number of signals.[6,7]

Here we present a modular system of attaching solubility enhancement tags that allows for high level expression, facile purification, and solubility enhancement with the addition of only a few resonances to the spectrum of the target protein. By utilizing two distinct expression constructs and by exploiting the tight binding of a well behaved protein to a peptide fused to the protein of interest, the system exchanges the isotopically labeled solubility enhancement tag present during expression with a nonisotopically labeled, and therefore "NMR invisible", tag.

As a proof of principle we have designed a system based on the tight binding of human calmodulin (hCaM) to the calmodulin binding peptide (CBP) (Figure 1).[8,9] The first expression construct carries the target sequence with a cleavable (PreScission protease) N-terminal GST tag and a noncleavable C-terminal CBP fusion (26 amino acids KRRWKKNFIAVSAANRFKKISSSGAL). As a model target protein, we demonstrate the solubility enhancement of the sterile alpha motif (SAM) domain from the murine transcription factor p63, which is characterized by an intrinsic low solubility.[10,11] ^{15}N-labeled GST.SAM.CBP fusion protein was expressed in *Escherichia coli* and purified using a glutathione Sepharose column, yielding 40 mg/L of pure protein. Immediately after purification, the second solubility tag was attached to GST.SAM.CBP by adding an equimolar amount of nonisotopically labeled His-tagged hCaM (hCaM.His$_6$) precharged with 4 equiv of calcium ions. The formation of the GST.SAM.CBP-hCam.His$_6$ complex ensured that a solubility tag remains attached to the target protein at all times. The isotopically labeled GST tag was proteolytically cleaved from the target with PreScission protease. Further purification was achieved by utilizing the poly histidine moiety of the hCaM.His$_6$ construct, ensuring that only a full length protein is obtained.

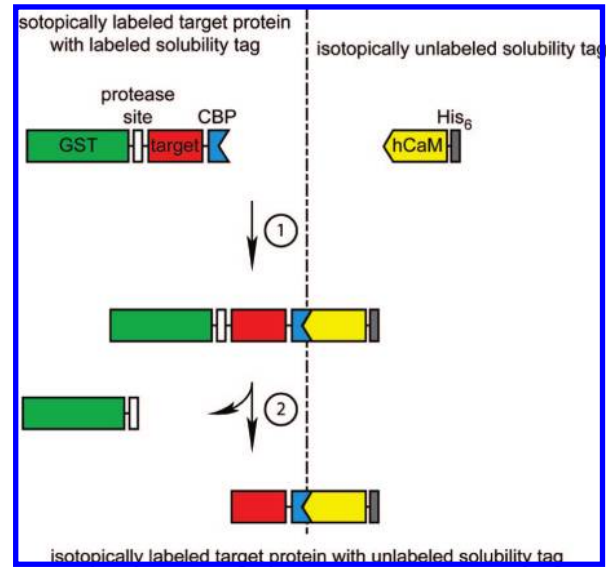


Figure 1. Schematic representation for attaching an "NMR invisible" solubility enhancement tag. The target protein is expressed as an isotopically labeled fusion protein. Addition of nonisotopically labeled His-tagged calmodulin (hCaM.His$_6$) leads to calcium dependent complex formation with the calmodulin binding peptide at the C-terminus of the target sequence (Step 1). Removal of the labeled GST tag by protease cleavage leaves the target protein with an unlabeled, i.e., "NMR invisible," solubility tag (Step 2). A similar calmodulin based system has been used for the measurement of residual dipolar couplings.[12]

In preparation of NMR experiments, pure SAM.CBP-hCaM.His$_6$ could be concentrated beyond a protein concentration of 2 mM, whereas SAM.CBP, which was not complexed with hCaM.His$_6$, precipitated at 100 μM. The calmodulin tag also prolonged the sample lifetime. After 7 days, the sample of SAM.CBP had decreased from 100 μM to 35 μM due to precipitation, whereas SAM.CBP-hCaM.His$_6$ maintained a solubility of 2 mM.

To investigate the effect of the CBP-tag on the spectral quality of the target protein we compared [^{15}N,^{1}H]-TROSY-HSQC spectra of untagged SAM, SAM.CBP, and SAM.CBP-hCaM.His$_6$. Final protein concentrations used for NMR experiments, determined from UV absorbance at 280 nm, were 100 μM for SAM and SAM.CBP and 1 mM for SAM.CBP-hCaM.His$_6$.

Comparison of the spectra of SAM and SAM.CBP shows that the addition of the tag does not influence the conformation of the protein since chemical shift differences in the NMR resonances of

10.1021/ja806212j CCC: $40.75

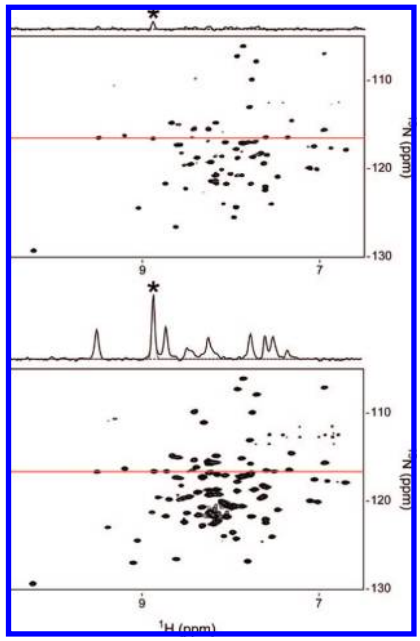


Figure 2. Comparison of [^{15}N,^{1}H]-TROSY-HSQC spectra of SAM (A) and SAM.CBP-hCaM.His$_6$ (B). The spectrum in (A) was measured with 4 scans per increment at a concentration of 100 μM, and the spectrum in (B) with 2 scans per increment and at a concentration of 1 mM. Cross sections shown above each spectrum were taken from the position indicated by the red line. Asterisks mark the position of the amide proton of D524 (numbering according to full length TAp63α). Both spectra were measured on a Bruker Avance spectrometer operating at a ^{1}H frequency of 600 MHz and equipped with a cryogenic probe.

the SAM domain can only be detected for the last amino acid. Of the 26 amino acids and the 2 amino acids used as a linker between the end of the SAM domain and the CBP tag (Ser Arg), only 3 are visible as additional peaks. This relatively low number can be explained with the unfolded nature of the tag in the absence of calmodulin that facilitates fast exchange of the amide protons with water. The spectral quality of the SAM.CBP spectrum is worse than the quality of the SAM spectrum, and SAM.CBP tends to aggregate even more than SAM alone, most likely due to the presence of an extra stretch of 28 unfolded amino acids. This shows that the CBP tag by itself does not lead to higher solubility and rather worsens the situation (Supporting Information).

The binding of calmodulin to SAM.CBP induces structure within the tag, causing 23 additional signals to become observable (Supporting Information). Differences in chemical shifts caused by the addition of calmodulin are detectable for the last amino acid, as well as some minor changes for some amino acids of the C-terminal helix. In case the addition of calmodulin leads to significant differences in chemical shifts of the protein of interest, the linker length between the protein and the CBP tag could be increased.

Although the "invisible" calmodulin solubility tag does not contribute signals to the target protein, it does increase its effective size. The larger mass translates into a slightly increased line width. However, the average line broadening for amide proton resonances of the SAM domain was 2 Hz, and therefore, the gain in sensitivity achieved through higher solubility outweighs this small increase in peak width.

The most comparable efforts toward an "NMR invisble" SET have utilized the intein ligation system to ligate an isotpocially labeled target protein with a nonisopically labeled solubility tag.[13] The intein systems are elegant, and the final product possesses no additional resonances due to added tags. The system is, however, experimentally more complicated since it entails enzymatic splicing of amide bonds, a process with a highly variable yield.

In conclusion, we have developed a modular and post-translational system of adding a protein solubility enhancement tag based on calmodulin and the calmodulin binding peptide. The system effectively enhances solubility, and because one component is expressed separately in a nonisotopically labeled medium, a minimal number of resonances are added to the target spectra. In principle, the solubility of the system can be easily further enhanced without increasing the number of resonances by fusing another protein (for example, protein G (B1 domain), protein D or even MBP) to hCaM. This might be necessary for proteins with very low intrinsic solubility. Such a fusion might also be necessary for modifying the pI of the entire Protein.CBP-hCaM.His$_6$ construct (pI of CBP-hCaM.His$_6$: 4.8). Moreover our system is not restricted to the interaction of hCaM and CBP but can be replaced by any protein−peptide pair that has a submicromolar binding constant. One very tempting alternative could be the use of PDZ domains, which tightly bind to very short peptide sequences.[14] Such noncalcium dependent solubility enhancement tags would be important for proteins that are affected themselves by the presence of calcium ions.

Acknowledgment. This work was supported by the Center for Biomolecular Magnetic Resonance at the University Frankfurt (BMRZ), EU-Grant EPISTEM (LSHB-CT-019067), and the Cluster of Excellence Frankfurt (Macromolecular Complexes).

Supporting Information Available: [^{15}N,^{1}H]-TROSY-HSQC spectra of SAM, SAM.CBP, and SAM.CBP-hCaM.His$_6$. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Hajduk, P. J.; Gerfin, T.; Boehlen, J. M.; Haberli, M.; Marek, D.; Fesik, S. W. *J. Med. Chem.* **1999**, *42*, 2315–2317.
(2) Serber, Z.; Richter, C.; Moskau, D.; Bohlen, J.-M.; Gerfin, T.; Marek, D.; Haberli, M.; Baselgia, L.; Laukien, F.; Stern, A. S.; Hoch, J. C.; Dotsch, V. *J. Am. Chem. Soc.* **2000**, *122*, 3554–3555.
(3) Bagby, S.; Tong, K. I.; Liu, D.; Alattia, J. R.; Ikura, M. *J. Biomol. NMR* **1997**, *10*, 279–282.
(4) Huang, B.; Eberstadt, M.; Olejniczak, E. T.; Meadows, R. P.; Fesik, S. W. *Nature* **1996**, *384*, 638–641.
(5) Lepre, C. A.; Moore, J. M. *J. Biomol. NMR* **1998**, *12*, 493–499.
(6) Zhou, P.; Lugovskoy, A. A.; Wagner, G. *J. Biomol. NMR* **2001**, *20*, 11–14.
(7) Forrer, P.; Jaussi, R. *Gene* **1998**, *224*, 45–52.
(8) Shifman, J. M.; Mayo, S. L. *J. Mol. Biol.* **2002**, *323*, 417–423.
(9) Zheng, C. F.; Simcox, T.; Xu, L.; Vaillancourt, P. *Gene* **1997**, *186*, 55–60.
(10) Cicero, D. O.; Falconi, M.; Candi, E.; Mele, S.; Cadot, B.; Di Venere, A.; Rufini, S.; Melino, G.; Desideri, A. *Cell Biochem. Biophys.* **2006**, *44*, 475–489.
(11) Yang, A.; Kaghad, M.; Wang, Y.; Gillett, E.; Fleming, M. D.; Dotsch, V.; Andrews, N. C.; Caput, D.; McKeon, F. *Mol. Cell* **1998**, *2*, 305–316.
(12) Feeny, J.; Birdsall, B.; Bradbury, A. F.; Biekofsky, R. R.; Bayley, P. M. *J. Biomol. NMR* **2001**, *21*, 41–48.
(13) Zuger, S.; Iwai, H. *Nat. Biotechnol.* **2005**, *23*, 736–740.
(14) Harris, B. Z.; Lau, F. W.; Fujii, N.; Guy, R. K.; Lim, W. A. *Biochemistry* **2003**, *42*, 2797–2805.

JA806212J