

Dual-tagging system for the affinity purification of mammalian protein complexes

Richard J. Giannone^{1,3}, W. Hayes McDonald¹, Gregory B. Hurst¹,
Ying Huang¹, Jun Wu^{1,3}, Yie Liu², and Yisong Wang¹

¹Oak Ridge National Laboratory, Oak Ridge, TN, ²National Institute on Aging,
Baltimore, MD, and ³University of Tennessee, Knoxville, TN, USA

BioTechniques 43:296-302 (September 2007)
doi 10.2144/000112550

Although affinity purification coupled with mass spectrometry (MS) provides a powerful tool to study protein-protein interactions, this strategy has encountered numerous difficulties when adapted to mammalian cells. Here we describe a Gateway®-compatible dual-tag affinity purification system that integrates regulatable expression, tetracycline motifs, and various combinations of affinity tags to facilitate the cloning, detection, and purification of bait proteins and their interacting partners. Utilizing the human telomere binding protein TRF2 as a benchmark, we demonstrate bait protein recoveries upwards of approximately 16% from as little as $1-7 \times 10^7$ cells and successfully identify known TRF2 interacting proteins, suggesting that our dual-tag affinity purification approach is a capable new tool for expanding the capacity to explore mammalian proteomic networks.

One popular method to elucidate protein-protein interactions involves the native co-purification of an affinity-tagged protein and its interacting partners, which are subsequently identified through mass spectrometry (MS) (1). Although straightforward, reproducible, and broadly used, this strategy is hampered by the efficacy of protein recoveries both in terms of sensitivity and specificity. This is especially pertinent to methodologies that use a single-step purification, in which suboptimal enrichment of the bait protein and its partners over background can lead to masking of their signals. Although improvements to MS instrumentation generally increase peptide detection sensitivities, the problem of specificity (i.e., distinguishing specific from nonspecific interacting proteins) remains. Thus ultimately, the limiting factor in the identification of specific interacting proteins lies with the purification itself.

An effort to resolve this specificity issue has been made with the introduction of the tandem affinity purification (TAP) tag. This construct consists of an immunoglobulin G (IgG) binding domain and calmodulin binding peptide domain separated by a tobacco etch virus (TEV) protease cleavage site (2). The TAP method was originally developed in yeast and has best demonstrated its utility in the systematic identification of numerous

multiprotein complexes in the yeast proteome (3). Although modifications to the original TAP methodology have been successful in examining the protein networks of mammalian cells (3–7), the strategy offers a relatively low yield of bait and specific interacting proteins (8), and the success rate usually varies on a case-by-case basis. Additionally, problems remain that are inherent to any protein tagging strategy: (i) variable exposure of the affinity tag; (ii) disruption of the bait protein's ability to fold properly; (iii) steric exclusion of interacting partners; and (iv) ectopic overexpression of the fusion protein, which can lead to complications in both the purification and identification of true interactions.

We generated five novel dual-tag purification vectors, each with a different combination of affinity tags (two per construct, varying by composition, size, and terminal location) and including either a constitutive (CMV_p) or tetracycline-regulatable promoter (Tet_p) to allow for controlled expression of the tagged bait proteins (Figure 1A). We chose Strep-Tactin binding peptide (*StrepII*-tag[®]; IBA, St. Louis, MO, USA) in most of our dual-tag combinations due to both its high binding affinity and its small (8-mer) size compared with the original streptavidin binding peptide (strep tag: 38-mer). Other novel features include a second TEV protease recognition

site to improve cleavage efficiency (data not shown) and a tetracycline motif (CCPGCC) (9) (except for C-HAtP) to easily monitor bait protein expression, purification, and localization. Moreover, all our dual-tags are constructed in Invitrogen Gateway®-compatible destination vectors, allowing for easy cloning through site-specific recombination (see the supplementary materials available online at www.BioTechniques.com).

We selected human telomeric repeat binding factor 2 (TRF2) to evaluate our dual-tagging system. TRF2 is a key telomere binding protein that functions to stabilize the t-loop configuration, a structure that both protects the chromosome end from being recognized as damaged DNA and represses telomere elongation by telomerase (10). Several telomere-associated and DNA damage repair proteins are known to interact with TRF2 (10,11) and thus provides both an effective means to assess the efficacy of our tagging system and an opportunity to gather potentially novel insight regarding TRF2 function.

As shown in Figure 1, B and F, all TRF2 fusions produced proteins of anticipated size (or very nearly so). Moreover, the level of Tet_p -driven TRF2 fusion protein expression was tetracycline concentration-dependent (Figure 1C). This demonstrates the capability to modulate fusion protein expression, potentially overcoming problems encountered in overexpression systems. For example, not only can expression be adjusted to near physiological level, but also, bait proteins that would otherwise impede cellular growth and/or viability can be repressed until specific experimental conditions are met. This feature greatly expands the tags' applicability.

The CCPGCC motif featured in four out of five dual-tags allows for the visualization of the fusion protein in both live cells and cellular lysates using Lumio™ (Invitrogen, Carlsbad, CA, USA), a conditionally fluorescent, membrane-permeable compound based on the fluorescein arsenical hairpin (FIAsH) reagent (9,12). The expected co-localization of TRF2-C-StH with the telomere in fixed cells (Figure 1D) and its similar punctate staining pattern

Benchmarks

in live cells (Figure 1E) indicate that the tag does not interfere with TRF2's subcellular localization. This useful feature provides a means to (i) rapidly infer bait protein function following tagging based on proper localization; (ii) assess transfection efficiency; (iii) confirm putative interacting partners by co-localization; or (iv) monitor the purification progress directly by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Figure 1F). Interestingly, the tetracysteine motif can also function as an affinity tag when paired with FLAsH-conjugated agarose beads (12), providing yet another means by which the bait-complex could be purified.

To evaluate the purification efficiency of our tagging system, two plates (10 cm²) of 293T cells were transfected with TRF2-C-StH. Cells (approximately 10⁷) were collected 48 h posttransfection, lysed, and tagged TRF2-purified (see supplementary materials). Western blot analysis was used to monitor the recovery at each purification step. Quantitative analysis of the density of each band against the percentage of total volumes loaded in each sample reveals that the final eluates contain approximately 6% of the TRF2-C-StH fusion protein present in the lysate (Figure 2A).

To assess bait recovery with regards to downstream MS analysis, each TRF2 construct (Figure 1A) was transfected into 293T cells and purified as described in the supplementary materials. In addition, several controls were included to identify proteins that nonspecifically bind to the affinity beads, the tag itself, and to proteins in general. Each sample eluate was trichloroacetic acid (TCA)-precipitated, digested, and analyzed by two-dimensional liquid chromatography tandem mass spectrometry (2-D LC-MS/MS). Resultant tandem mass spectra were searched against the human International Protein Index (IPI) database v3.05 using DBDigger (13). Data from multiple pull-downs of each TRF2 fusion and controls were filtered, organized, and compared using DTASelect and Contrast (14). Expectedly, TRF2 was identified in all pull-downs (search parameters and filter criteria can be found in the

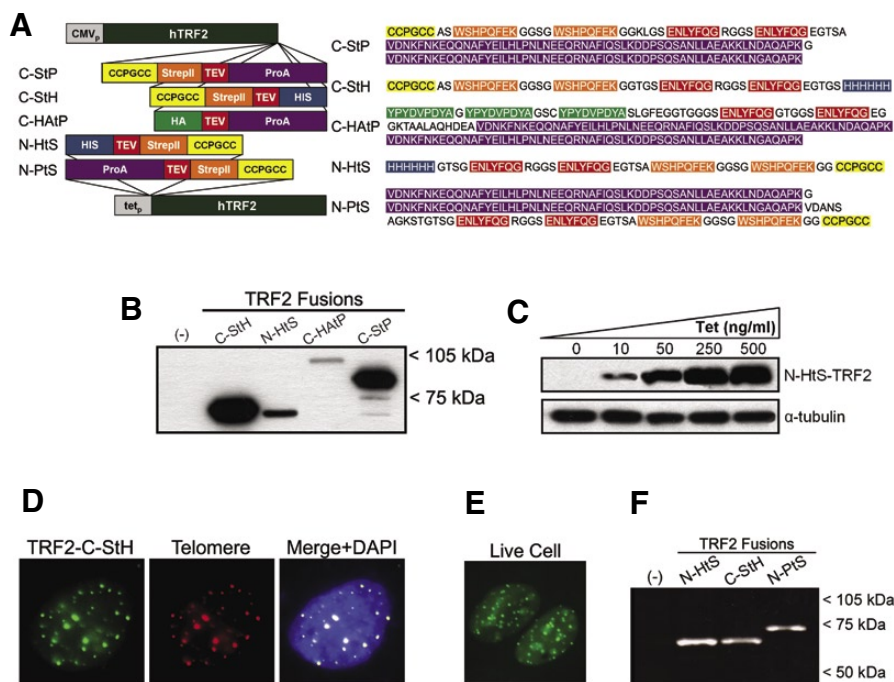


Figure 1. Features and validation of the dual-tags. (A) Dual-tags, their relative positions in TRF2 fusions, and their amino acid sequence. C, C'-terminal; N, N'-terminal; CCPGCC, a tetracysteine motif; S, Strep-Tactin binding peptides (*StrepII*-tag); t, tobacco etch virus (TEV) protease cleavage site; P, immunoglobulin G (IgG) binding domain of protein A from *Staphylococcus aureus* (ProA tag); H, 6 \times histidine (His tag); HA, influenza hemagglutinin epitope (HA) tag. Colors: yellow, tetracysteine motif; orange, *StrepII*-tag; purple, ProA tag; red, TEV cleavage site; blue, His tag; green, HA tag. (B) Western blot analysis (anti-TRF2) depicting the expression and anticipated size of representative fusions 24 h after transfection of respective telomeric repeat binding factor 2 (TRF2) fusion constructs into 293T cells. (C) Modulation of fusion protein expression via the Tet-regulatable promoter, visualized by Western blot analysis (anti-Strep tag, anti- α -tubulin). (D) Co-localization of transfected TRF2-C-StH with telomeres in U2OS cells after fixation. TRF2-C-StH (green, Lumio), telomeres [red, telomeric fluorescence in situ hybridization (FISH)], DNA [blue, 4',6-diamidino-2-phenylindole (DAPI)], and co-localization (yellow, merged panel). (E) Live cell image of TRF2-C-StH punctate staining on telomeres revealed by Lumio. (F) Lumio detection of TRF2 fusion protein expression and size, separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). TRF2, telomeric repeat binding factor 2.

supplementary materials; MS data are available upon request).

Using the control-filtered MS data, we identified the major proteins previously shown to associate with the TRF2 complex, such as telomeric repeat binding factor 2-interacting protein 1 (RAP1), TRF1-interacting nuclear protein 2 (TIN2), TINT1/PTOP/PIP1 (TPP1), ataxia telangiectasia mutated (ATM), Werner syndrome ATP-dependent helicase (WRN), Bloom syndrome protein (BLM), Ku antigen 70 (Ku70), Ku antigen 80 (Ku80), and poly(ADP-ribose) polymerase 1 (PARP1) (10,11,15) (see Supplementary Table S1) along with potentially novel TRF2-interacting proteins (manuscript in preparation). RAP1, a relatively low-abundant protein, was confidently identified in TRF2 pull-down samples

from as little as approximately 10⁷ adherent cells. In fact, all of the known TRF2-associated proteins identified in this study originated from samples containing no more than 7 \times 10⁷ cells. Taken together, these data demonstrate the efficacy and sensitivity of our dual-tag purification system.

Comparing each of the dual-tags, we found that all five generated sufficient recovery for MS analysis, yielding the TRF2 bait protein, known TRF2-interacting protein(s), and several new candidate TRF2-associated partners. In this study, we mainly focused on the His/*StrepII*-based tags (Figure 1A), as they produced the best TRF2 sequence coverage, the largest number of MS/MS spectra assignable to TRF2 peptides, and identified the most known TRF2-interacting proteins (Supplementary

Benchmarks

Table S1), relative to the other dual-tags. However, this is not to suggest that these tags will always outperform the rest.

In an attempt to increase bait protein recovery, we modified a freeze/thaw lysis protocol (see the supplementary materials) to keep the cell lysis as concentrated as possible and were able to triple the recovery of TRF2 to approximately 16% in the final eluates (Figure 2B), equivalent to approximately 200 pmol TRF2 from the lysate based on a dot-blot analysis (Figure 2C). In addition, using the freeze/thaw method, pulldown experiments ($n = 4$) identified, on average, most (five out of nine) known TRF2-interacting proteins listed in Supplementary Table S1, suggesting a greater enrichment of known TRF2-interacting proteins per pulldown compared with those not processed by freeze/thaw (average two out of nine; $n = 4$). Moreover, according to the total, nonredundant protein counts reported by DTASelect, the total number of proteins found in the pulldown experiments without freeze/thaw ($n = 4$) was 435 ± 56 , compared with 388 ± 114 with freeze/thaw ($n = 4$). Therefore, the enrichment of specific interacting proteins in freeze/thaw samples was not the result of increased total protein counts. These data perhaps indicate that keeping the initial lysis concentrated may enhance the recovery of specific TRF2 interacting proteins by ensuring minimal perturbation of the natural equilibrium that exists between interacting partners.

In this study, we have developed a dual-tagging system that offers versatile features to address various issues related to the difficulty of identifying interacting proteins in mammalian cells. With several available dual-tag constructs, this Gateway-compatible system provides enormous flexibility to rapidly modify tag composition, terminal location, and bait protein expression, while at the same time providing a convenient means to visualize the bait protein in vivo and in vitro.

ACKNOWLEDGMENTS

We acknowledge Dr. Lea Harrington for kindly providing the pcDNA-FLAG-

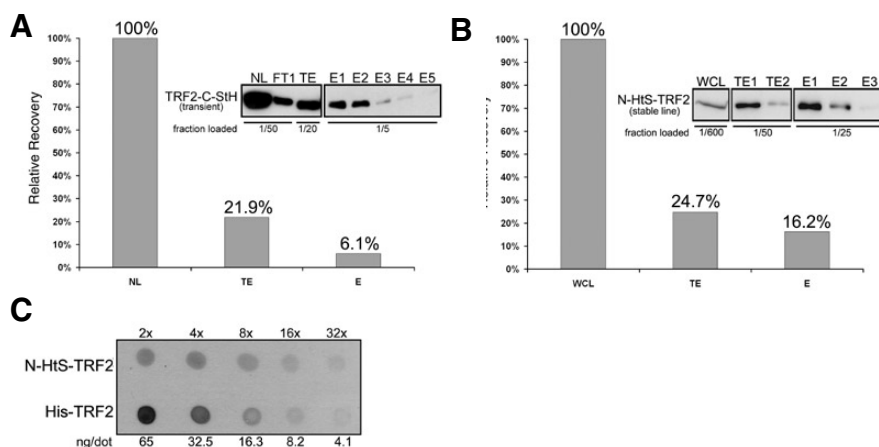


Figure 2. Efficient purification of dual-tagged telomeric repeat binding factor 2 (TRF2). (A) Western blot analysis (anti-Strep tag) depicting the complete purification of transiently expressed TRF2-C-StH. Bar graph representing the estimated recovery of bait protein relative to input from the lysate. (B) Western blot analysis and graph showing the complete purification and estimated recovery of a stably expressing N-HtS-TRF2 fusion protein using freeze/thaw lysis. NL, nuclear lysate; WCL, whole-cell lysate; FT1, flow-through of lysate from Ni-NTA resin; TE, tobacco etch virus (TEV) protease-mediated elution from Ni-NTA resin; E, eluates from Strep-Tactin resin. (C) Dot blot analysis [anti-His tag (H-15; Upstate)] comparing the amount of N-HtS-TRF2 in the freeze/thaw-processed whole-cell lysate to a His-purified TRF2 of known concentration, both serially diluted (1:2).

hTRF2 plasmid for this study. The authors acknowledge the support of the Laboratory Directed Research and Development (LDRD) Program of Oak Ridge National Laboratory, and the Office of Biological and Environmental Research, U.S. Department of Energy, under Contract DE-AC05-00OR22725 with UT-Battelle, LLC, and the DOE Genomics:GTL grants, respectively. Y.L. acknowledges the support by the Intramural Research Program of the National Institutes of Health/National Institute on Aging.

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

REFERENCES

- Aebersold, R. and M. Mann. 2003. Mass spectrometry-based proteomics. *Nature* 422:198-207.
- Rigaut, G., A. Shevchenko, B. Rutz, M. Wilm, M. Mann, and B. Seraphin. 1999. A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* 17:1030-1032.
- Gavin, A.C., M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J.M. Rick, et al. 2002. Functional organization of the yeast proteome by

- systematic analysis of protein complexes. *Nature* 415:141-147.
- Bouwmeester, T., A. Bauch, H. Ruffner, P.O. Angrand, G. Bergamini, K. Croughton, C. Cruciat, D. Eberhard, et al. 2004. A physical and functional map of the human TNF- α /NF- κ B signal transduction pathway. *Nat. Cell Biol.* 6:97-105.
- Cox, D.M., M. Du, X. Guo, K.W. Siu, and J.C. McDermott. 2002. Tandem affinity purification of protein complexes from mammalian cells. *BioTechniques* 33:267-270.
- Westermarck, J., C. Weiss, R. Saffrich, J. Kast, A.M. Musti, M. Wessely, W. Ansorge, B. Seraphin, et al. 2002. The DEXD/H-box RNA helicase RHII/Gu is a co-factor for c-Jun-activated transcription. *EMBO J.* 21:451-460.
- Drakas, R., M. Prisco, and R. Baserga. 2005. A modified tandem affinity purification tag technique for the purification of protein complexes in mammalian cells. *Proteomics* 5:132-137.
- Burckstummer, T., K.L. Bennett, A. Preradovic, G. Schutze, O. Hantschel, G. Superti-Furga, and A. Bauch. 2006. An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. *Nat. Methods* 3:1013-1019.
- Griffin, B.A., S.R. Adams, and R.Y. Tsien. 1998. Specific covalent labeling of recombinant protein molecules inside live cells. *Science* 281:269-272.
- de Lange, T. 2005. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.* 19:2100-2110.
- d'Adda di Fagagna, F., S.H. Teo, and S.P. Jackson. 2004. Functional links between telomeres and proteins of the DNA-damage response. *Genes Dev.* 18:1781-1799.
- Thorn, K.S., N. Naber, M. Matuska, R.D. Vale, and R. Cooke. 2000. A novel method of affinity-purifying proteins using a bis-arsenical fluorescein. *Protein Sci.* 9:213-217.

Benchmarks

13. **Tabb, D.L., C. Narasimhan, M.B. Strader, and R.L. Hettich.** 2005. DBDigger: reorganized proteomic database identification that improves flexibility and speed. *Anal. Chem.* 77:2464-2474.
14. **Tabb, D.L., W.H. McDonald, and J.R. Yates 3rd.** 2002. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* 1:21-26.
15. **Gomez, M., J. Wu, V. Schreiber, J. Dunlap, F. Dantzer, Y. Wang, and Y. Liu.** 2006. PARP1 is a TRF2-associated poly(ADP-ribose)polymerase and protects eroded telomeres. *Mol. Biol. Cell* 17:1686-1696.

Received 2 March 2007; accepted 13 July 2007.

Address correspondence to Yisong Wang, Biosciences Division Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. e-mail: ywa@ornl.gov; or Yie Liu, Laboratory of Molecular Gerontology, Gerontology Research Center, National Institute on Aging, 5600 Nathan Shock Drive, Baltimore, MD 21224-6825, USA. e-mail: liuyie@grc.nia.nih.gov

To purchase reprints of this article, contact: Reprints@BioTechniques.com