Identification of Novel Protein-Protein Interactions Using A Versatile Mammalian Tandem Affinity Purification Expression System*

Matthew Knuesel†, Yong Wan§, Zhan Xiao¶, Eric Holinger‡, Nick Lowe‡, Wei Wang‡, and Xuedong Liu‡

Identification of protein-protein interactions is essential for elucidating the biochemical mechanism of signal transduction. Purification and identification of individual proteins in mammalian cells have been difficult, however, due to the sheer complexity of protein mixtures obtained from cellular extracts. Recently, a tandem affinity purification (TAP) method has been developed as a tool that allows rapid purification of native protein complexes expressed at their natural level in engineered yeast cells. To adapt this method to mammalian cells, we have created a TAP tag retroviral expression vector to allow stable expression of the TAP-tagged protein at close to physiological levels. To demonstrate the utility of this vector, we have fused a TAP tag, consisting of a protein A tag, a cleavage site for the tobacco etch virus (TEV) protease, and the FLAG epitope, to the N terminus of human SMAD3 and SMAD4. We have stably expressed these proteins in mammalian cells at desirable levels by retroviral gene transfer and purified native SMAD3 protein complexes from cell lysates. The combination of two different affinity tags greatly reduced the number of nonspecific proteins in the mixture. We have identified HSP70 as a specific interacting protein of SMAD3. We demonstrated that SMAD3, but not SMAD1, binds HSP70 in vivo, validating the TAP purification approach. This method is applicable to virtually any protein and provides an efficient way to purify unknown proteins to homogeneity from the complex mixtures found in mammalian cell lysates in preparation for identification by mass spectrometry. Molecular & Cellular Proteomics 2:1225–1233, 2003.

The transmission of a wide range of biological signals depends upon direct physical interaction of specific cellular proteins (1). A more thorough understanding of the precise mechanisms of these signal transduction pathways relies on the identification and biochemical characterization of the specific molecules involved. Identification and characterization of protein complexes in mammalian cells have been challenging partly due to the complexity of protein mixtures in cell extracts. Definitive identification of components present in physiologically relevant protein complexes often requires considerable amounts of proteins purified to homogeneity by multistep chromatography. Classical biochemical purification methods rely heavily on the biophysical properties of a given protein. Because each protein is unique, designing a purification protocol for a given protein is often empirical. The complexity of the mammalian genome and lack of a generic protein complex purification protocol hamper the progress in elucidating the functional interactions between cellular proteins. There is a critical need for developing widely applicable and efficient purification procedures for isolating protein complexes from mammalian cells.

Among a plethora of protein purification techniques, affinity purification appears to be the most efficient and gentle discriminatory separation technique for the retrieval of protein complexes. In particular, well characterized small affinity tags such as FLAG, His6, or glutathione S-transferase have been widely used in affinity purification of recombinant proteins or protein complexes. The conditions used for purification of proteins fused with these tags are often straightforward, robust, and reproducible. Two generic affinity purification procedures utilizing two consecutive affinity purification steps have been developed and have proven to be highly effective in the identification of protein complexes in yeast (2, 3). Rigaut et al. described the tandem affinity purification (TAP)1 method (3). Two affinity tags, protein A and calmodulin-binding peptide (CBP), separated by a TEV protease cleavage site were

1 The abbreviations used are: TAP, tandem affinity purification; IRES, internal ribosome entry site; GFP, green fluorescent protein; TBS, Tris-buffered saline; MS/MS, tandem mass spectrometry; MCS, multiple cloning site; CBP, calmodulin-binding peptide; TEV, tobacco etch virus; MALDI, matrix-assisted laser desorption ionization; BMP, bone morphogenetic protein; TGF-β, transforming growth factor-β; Cdk, cyclin-dependent kinase.
fused to the protein of interest. The TAP-tagged protein was expressed in yeast cells at close to physiological concentration to form a complex with endogenous components. Extracts prepared from cells expressing the TAP-tagged protein are subjected to two successive purification steps. The first purification step involved binding of the protein A tag to an IgG column and eluting bound materials by incubation with TEV protease. This elution method significantly reduced the contaminating proteins bound nonspecifically to the column. The eluates that contained protein complexes were further purified by incubation with calmodulin-coated beads to remove TEV protease and contaminants. Highly purified protein complexes were eluted by adjusting the concentration of calcium ions with chelating reagents. The retrieved protein complexes were then subjected to tryptic digestion, and identities of proteins in the complexes were determined by tandem mass spectrometry. A similar technique using three different tags (CBP, His6, and hemagglutinin) called MAFT for “multiple affinity purification” has also been developed and used for purification of cyclin-Cdk complexes in yeast (4). Therefore, successive steps of affinity purification can produce an extremely pure preparation of the protein of interest with associated proteins in yeast under standard conditions. These technological advances, coupled with sensitive mass spectrometry analysis, have allowed systematic study of the functional organization of close to 600 protein complexes in the yeast Saccharomyces cerevisiae (5, 6).

The application of the tandem affinity tag purification approach in mammalian cells has not been fully explored (7). In yeast, it is relatively easy to replace the gene encoding the endogenous protein by a tagged recombinant version through homologous recombination. The expression levels of the tagged protein often track the endogenous protein. Such a feat is not as easily accomplished in cultured cells. To develop a generic protein purification scheme that is applicable to the identification of protein complexes in mammalian cells, we incorporated the TAP tag into a bicistronic retroviral expression vector. We have shown previously that a bicistronic retroviral vector offers an efficient way to stably express the exogenous proteins at controllable levels (8). Here we demonstrated that a tandem affinity tag consisting of the IgG binding domain and a FLAG tag separated by a TEV cleavage site fused to the N terminus of SMAD3 allows for the rapid purification of SMAD3 and its associated proteins from mammalian cells. We have detected a novel interaction between SMAD3 and HSP70 by sequencing proteins copurified with SMAD3 using mass spectrometry. Interaction between SMAD3 and HSP70 was confirmed by a communoprecipitation experiment and appears to be very specific as no association can be detected between HSP70 and SMAD1, a protein that is highly homologous to SMAD3.
Construction of a Bicistronic Retroviral TAP Tag Vector—The tandem affinity purification method has been proven to be a powerful generic protein purification method for functional proteomics studies in yeast (7). To develop a comparable methodology to allow purification of protein complexes from mammalian cultured cells, we constructed a versatile retroviral TAP vector. As shown in Fig. 1, the TAP cassette, which contains two copies of the IgG binding domain of protein A of *Staphylococcus aureus*, two tandem copies of the TEV protease recognition sequence, and a FLAG tag, was inserted into the pMX-IRES-GFP vector (8). The original TAP tag used in yeast consists of two IgG binding domains, one TEV cleavage site, and the calmodulin-binding peptide (3, 11). We opted to use FLAG instead of CBP because there are a significant number of proteins in mammalian cells that bind calmodulin beads in a calcium-dependent manner. For example, SMAD3, the protein used in the current study, binds calmodulin beads avidly (12–14). We introduced an extra copy of the TEV cleavage sequence into the tag to improve the TEV cleavage efficiency and the release of the protein complex from IgG beads. Spacer sequences were also placed between tags to reduce interference between tags in successive purification steps. Finally, the cassette was placed under the control of the SP6 promoter, which allows us to easily check for correct fusion between the protein of interest and the TAP tag by in vitro transcription/translation in rabbit reticulocyte lysates.

The bicistronic retroviral vector was chosen to express the TAP-tagged protein in mammalian cells. Retroviral gene transfer allows stable expression of exogenous proteins in a variety of mammalian cell lines within 2 days of infection. We have shown previously that a bicistronic retroviral vector like pMX-IRES-GFP enables us to generate mammalian cells stably expressing multiple genes at predetermined levels (8). This feature is important because protein complex purification often requires expression of the protein of interest at close to its endogenous levels. The quantitative selection feature of the bicistronic vector gives us the option to tune the levels of expression of the target protein in mammalian cells.

Generation of Stable Mammalian Cell Lines Expressing TAP-tagged SMAD3 and SMAD4—We tested the feasibility of the TAP method in mammalian cells by TAP tagging the low abundance signaling molecules SMAD3 and SMAD4. SMADs are downstream effectors of activated TGF-β receptors (15). Upon TGF-β stimulation, SMAD2, -3, and -4 form a heterologenic complex that translocates to the nucleus and regulates transcription of TGF-β-responsive genes. The exact composition of the TGF-β–activated SMAD signaling complex is largely unknown (15). We constructed TAP-SMAD3 and TAP-SMAD4 expression vectors and prepared recombinant retroviruses to infect mink lung epithelial PE2S cells, human breast cancer MDA-MB468 cells, human colon cancer SW480 cells, and mouse myoblast C2C12 cells. We have been able to obtain cell lines expressing TAP-SMAD3 and -SMAD4 in all the cell lines we have tested so far. We have shown previously that the levels of expression of the gene downstream of the IRES (in this case GFP) and the expression levels of the gene cloned upstream of the IRES are highly correlated in stably infected target cells (8). Sorting cells based on the GFP expression levels, we were able to obtain cells either overexpressing SMAD3 or expressing SMAD3 or SMAD4 at close to endogenous levels. For example, TAP-SMAD3 was overexpressed in mink lung cells (Fig. 2D). We were not able to assess the levels of overexpression of SMAD3 in mink lung cells due to the lack of an antibody that recognizes both
endogenous mink SMAD3 and recombinant human SMAD3.

In MDA-MB468 cells, we were able to obtain cell lines that express SMAD3 at levels 5-fold over the endogenous SMAD3 (data not shown). In C2C12 mouse myoblast cells, we were able to express TAP-SMAD4 at the levels comparable with endogenous expression levels as revealed by Western blot analysis (Fig. 2).

Epitope tagging sometimes alters the biochemical properties of the target protein (16). To determine whether cell lines expressing TAP-tagged SMAD3 display any significant alterations in TGF-β signaling, we measured the PAI-1 luciferase reporter gene induction in parental mink lung cells (PE25) and cells stably expressing TAP-SMAD3 in response to TGF-β. PE25 is a cell line derived from MuLV1 (CCL64) by stable integration of a TGF-β-responsive luciferase reporter gene (PE2.1-lux) into the genome (17). As shown in Fig. 2C, there is an 18-fold increase in luciferase activity in response to TGF-β stimulation in the wild type cells (PE25); similar inductions were observed in cell lines stably expressing TAP-SMAD3 except that the basal levels of luciferase activity are slightly lower than in the wild type cells. This could be a result of SMAD3 overexpression in these cells. This result suggests that TAP-SMAD3 does not adversely affect TGF-β signaling in these cells.

**Fig. 2.** Generation of mammalian stable cell lines expressing TAP-tagged SMAD3. **A,** mink lung cells were infected with the pRAV-FLAG-SMAD3 recombinant virus. Two days after infection, GFP-positive cells were sorted using a MoFlo cell sorter (DakoCytomation). **R1,** gate for non-infected cells; **R2,** gate for infected cells. **B,** schematic diagram of the structure of recombinant TAP-tagged SMAD3. **UTR,** untranslated region. **C,** effect of overexpression of TAP-SMAD3 on transcriptional activation of the TGF-β-responsive PAI-1 promoter (PE2.1) (17) in the presence or absence of 100 pm TGF-β. The data shown represent three averaged experiments, each using 100,000 cells/well and normalized to protein concentrations measured by Bradford assay. **D,** Western blot analysis of TAP-SMAD3 expression using an anti-human SMAD3 antibody (Santa Cruz Biotechnology). One hundred μg of total cell lysates were loaded in each lane. Endogenous mink SMAD3 is not recognized by this antibody, and only the recombinant TAP-SMAD3 was detected. **Con,** control. **E,** Western blot analysis of TAP-SMAD4 and endogenous SMAD4 expression in C2C12 myoblast cells. One hundred μg of cell lysates were subjected to SDS-PAGE, transferred to nitrocellulose, and blotted with a monoclonal anti-SMAD4 (B8, Santa Cruz Biotechnology) antibody. **endo,** endogenous.

**Table:**

<table>
<thead>
<tr>
<th>Cells</th>
<th>Mink Lung Cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGF-β</td>
<td>Con</td>
</tr>
<tr>
<td><strong>97 kDa</strong></td>
<td>-</td>
</tr>
<tr>
<td><strong>66 kDa</strong></td>
<td>-</td>
</tr>
<tr>
<td><strong>45 kDa</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>TAP-Smad3</strong></td>
<td>3</td>
</tr>
</tbody>
</table>

**D.** Western blot analysis of TAP-Smad4 expression in C2C12 myoblast cells. One hundred μg of cell lysates were subjected to SDS-PAGE, transferred to nitrocellulose, and blotted with a monoclonal anti-SMAD4 (B8, Santa Cruz Biotechnology) antibody. **endo,** endogenous.
**Purification of the SMAD3 Complex**—To test whether a tandem affinity tag purification method works in our system, we first chose to purify the SMAD3 protein complex from mink lung cells overexpressing TAP-SMAD3. A half-billion mink lung cells grown on 150-mm culture dishes were treated with TGF-β for 1 h and were harvested by flash freezing with liquid nitrogen. Protein extracts were prepared from these cells and incubated with IgG-Sepharose 6 Fast Flow beads. The IgG beads were recovered by brief centrifugation, washed three times with TBS buffer, and incubated with IgG-Sepharose 6 Fast Flow beads. The IgG-Sepharose beads were eluted by incubation with 1 mg/ml FLAG peptide three times with TBS buffer, and the IgG-Sepharose beads (Sigma) for 1 h at 4 °C. After washing the beads three times with TBS buffer, proteins bound to the M2 beads were eluted by incubation with 1 mg/ml FLAG peptide three times for 30 min each at 4 °C. We found that high concentrations (at least 1 mg/ml) of FLAG peptide were necessary to achieve efficient elution of protein bound to the M2 beads. Second step purification with FLAG beads further reduces the complexity of protein mixture. The 50-kDa SMAD3 is the major protein in the eluates from M2 beads, suggesting the two-step tandem purification procedure can efficiently purify TAP-SMAD3 and its associated proteins from cultured mammalian cells.

**Identification of Copurifying Proteins by Mass Spectrometry**—To identify the proteins that were copurified with TAP-SMAD3, protein mixtures obtained from two successive affinity purification steps were fractionated on a 4–20% SDS-PAGE gradient gel and visualized by silver staining. Visible bands were excised and subjected to tryptic digestion. A MALDI quadrupole time-of-flight instrument (ABI/MDS Sciex QSTAR) was used for identification of proteins excised after in-gel digestion. Mass accuracy and resolution better than 0.1 Da in mass spectrometry and MS/MS mode were achieved on this orthogonal MALDI hybrid tandem mass spectrometer. The mass spectrometry spectrum acquired for the prominent 55-kDa protein band on the oMALDITM QSTAR indicated that this protein was SMAD3 after searching the peptide mass fingerprint using MASCOT and Prospector search engines. The peptide coverage with SMAD3 was over 42%. Three ions were selected for MS/MS analysis, and peptide sequence information unequivocally identified this protein as SMAD3 (data not shown). Similar analysis enabled us to identify the 45-kDa protein as IgG and the 30-kDa band as TEV protease. The presence of IgG is probably due to leaching of the IgG beads used in the first step of purification. The peptide mass fingerprint of the band with an apparent mass of 70 kDa suggests this protein may be HSP70 (Fig. 4). The coverage of the peptide mass map is about 36% identical to mink HSP70 in the database (25 matching peptides). The precursor ion at m/z 1199.69 was isolated, and a tandem mass spectrum was acquired. As shown in Fig. 5, the peptide sequence DAGTIA-GLNVLR was matched to mink HSP70 unambiguously after database searching using the MASCOT program. Attempts have been made to identify other visible bands on the gel, but so far we have not been able to unambiguously match data on other bands.

**Specific Interaction between SMAD3 and HSP70**—To determine whether HSP70 is truly associated with SMAD3 inside cells and that the interaction between SMAD3 and HSP70 is specific, we transfected FLAG-tagged SMAD3 or its related protein SMAD1 into HeLa cells. SMAD1 shares 65% homology with SMAD3 but is a downstream signaling mediator of BMP signaling. Both SMAD1 and SMAD3 are pathway-restricted receptor-regulated SMADs because they are the substrates of different type I receptors in the TGF-β superfamily of cytokines. The levels of SMAD3 and SMAD1 expression in transfected cells were determined by immunoprecipitation-Western blot analysis. Expression of SMAD3 is slightly higher than SMAD1 (Fig. 6, lane 1 versus lane 4). Cotransfection of respective constitutively active type I receptors results in a slight decrease in SMAD3 and SMAD1 levels (Fig. 6, lane 3 versus lane 5). To determine whether SMAD3 and SMAD1 associate with HSP70, lysates of SMAD3- and SMAD1-transfected cells were immunoprecipitated with FLAG antibody, and the presence of HSP70 in the immunoprecipitates was determined by Western blot analysis with an anti-HSP70 an-
tibody. Significant amounts of HSP70 were found to associate with SMAD3 (Fig. 6, middle panel, lane 1). The association between SMAD3 and HSP70 does not depend on the presence of SMAD4 nor constitutively active TGF-β type I receptor, suggesting this association is ligand-independent. In contrast, no association was found between SMAD1 and HSP70 under any of the three circumstances. This result suggests that SMAD3 is complexed with HSP70 and that this association is highly specific.

To determine whether association between SMAD3 and HSP70 depends on the subcellular localization of SMAD3, communoprecipitation experiments were performed with two SMAD3 mutants that are defective in nuclear localization in HeLa cells (18, 19). Association between HSP70 and SMAD3 with a defective nuclear localization signal was as robust as the association between HSP70 and wild type SMAD3 (Fig. 7). This experiment indicated that nuclear localization is not a prerequisite for HSP70 and SMAD3 association, suggesting that the association must occur in the cytosol.

**DISCUSSION**

Nearly every major cellular process is carried out by assemblies of large complexes of proteins (1). Elucidation of the composition of these protein complexes has increasingly become a focus of proteomic exploration. High throughput exploration of yeast protein complexes has been made possible by the development of the tandem affinity tag purification method. Application of a similar approach in mammalian cells has lagged behind in part due to the lack of a high throughput gene transfer approach and an effective system for controlling the levels of expression of the exogenous gene. Here we describe a tandem affinity tag purification system that is suitable for such studies in mammalian cells. We demonstrate that such a system is very effective in purifying the tagged protein and its associated protein complex from cultured mammalian cells. We have definitively identified at least one unsuspected novel interaction between SMAD3 and HSP70 using mass spectrometry and validated the specificity of this interaction in vivo. It is our expectation that this effective
The tandem affinity tag described here has a different configuration than any previous version (11). The original TAP tag developed for the yeast system consists of two IgG binding domains, a TEV protease cleavage site, and the calmodulin-binding peptide (3). Two tandem TEV cleavage sites were introduced into our system. This manipulation increased the efficiency of TEV cleavage and reduced the amount of TEV protease that is required for an exhaustive digestion, hence decreasing the potential exogenous proteins introduced in the mixture for mass spectrometry analysis. We also substituted the CBP with a FLAG epitope tag because there are considerable numbers of proteins in mammalian cells that bind calmodulin beads with high affinity. The SMAD3 protein studied here displays avid affinity for calmodulin beads. This property can be advantageous in this situation in that we can introduce an extra step of purification using calmodulin beads. Compared with the CBP tag, FLAG is relatively small and has a relative high affinity for M2 beads. With high concentrations of FLAG peptide (>1 mg/ml), it is a straightforward process to elute binding proteins off the column. The disadvantage of using FLAG affinity medium is that the final elution contains excess amounts of FLAG peptide that is incompatible with liquid chromatography-MS/MS sequence application. An extra purification step using a gel filtration column is necessary to remove the small FLAG peptide for liquid chromatography-MS/MS.

Another important consideration in designing experiments to identify the protein complexes that are associated with the protein of interest is the choice of cell line. Many of the studies discussed here employ a mink lung epithelial cell line. The mink genome and proteome have not been fully sequenced. Fortunately, the mink HSP70 sequence is deposited in the GenBankTM database. Some of the MS/MS spectra we obtained could not be readily assigned based on human and mouse protein sequences. This may have hampered our efforts to fully assign all visible bands on the gel unambiguously. Future studies will be focusing on characterizing protein complexes in cell lines from organisms such as humans and mice, whose genomes have been fully sequenced and annotated. Successful identification of an unsuspected interaction between SMAD3 and HSP70 provides an example that this strategy works even under a less optimal situation.

We have validated the interaction between SMAD3 and HSP70 in transfection assays. However, the functional significance of this interaction remains to be elucidated. A variety of transcription factors, steroid hormone receptors, and protein kinases involved in signal transduction are found to associate with HSP70 (20, 21). Genetic studies in yeast strongly support...
NLS mutants of subcellular localization of SMAD3. Two FLAG-tagged SMAD3 mutants (mut) that are defective in nuclear localization signal (NLS) and unable to translocate to the nucleus upon TGF-β stimulation were transfected into HeLa cells. The nature of these two mutations was described previously (18). Interactions between these two SMAD3 mutants and endogenous HSP70 were determined by coimmunoprecipitation assay as described in Fig. 6. Lane 1, FLAG-SMAD3 transfected with TAP expression vector; lane 3, FLAG-SMAD3 cotransfected with SMAD4 and constitutively active type I receptor (T204D); lane 4, FLAG-SMAD3 cotransfected with SMAD4 and constitutively active BMP type I receptor (T204D); lane 5, FLAG-SMAD1 cotransfected with SMAD4; lane 6, FLAG-SMAD1 cotransfected with SMAD4 and constitutively active BMP type I receptor. IP, immunoprecipitation; WB, Western blot; F, FLAG.

Expression of the TAP-tagged exogenous protein in our system is under the control of the retroviral long terminal repeat promoter. Because the provirus integrates at many sites in the genome, the expression levels of exogenous proteins in a given cell are also influenced by the local chromatin structure. In addition, as there is no block for superinfection, multiple retroviral integrations can occur in the same cells and engender additional variability in the system. By using quantitative selection markers such as GFP, we can take advantage of this inherent variability in the system using a fluorescence-activated cell sorter to obtain cells expressing exogenous proteins at a desirable level. However, this system has its own limitations. Unlike the TAP method in yeast where one can replace the endogenous copy of the gene of interest with a TAP-tagged version under the control of the native promoter in the haploid strain to achieve close to its natural expression levels, endogenous proteins are still present in our system and may compete for incorporation into the protein complexes, thus reducing the efficiency of protein complex purification. A possible solution to this problem could be to use cell lines that have homozygous deletions of the gene of interest obtained by a genetic knockout approach or naturally occurring tumor cells containing a homozygous deletion of the gene of interest. Alternatively, knockdown expression of the endogenous gene by small interfering RNA could alleviate the competition problem. The feasibility of the latter solution has been recently demonstrated in cultured *Drosophila* cells (23).

The application of this generic tandem affinity tag purification method is not limited to identifying novel protein-protein interactions. It is conceivable that rapid purification of tagged exogenous proteins from mammalian cultured cells will facilitate analysis and identification of post-translational modifications of ectopically expressed proteins under various physiological states. Thus, this technology will potentially prove to be a valuable tool in the broad field of proteomic exploration.

**Acknowledgments**—We thank Natalie Ahn and Katheryn Resing for help with mass spectrometry and numerous suggestions during the course of this study. We thank Kristen Bjorkman for critically reading the manuscript and Chang-Zhang Chen for technical help.

* This work was supported by grants from the Elsa U. Pardee Foundation and by Grant CA95527-01 from the National Institutes of
Health (to X. L.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

To whom correspondence should be addressed. Tel.: 303-735-6161; Fax: 303-735-6161; E-mail: Xuedong.Liu@Colorado.edu.

REFERENCES