Published in final edited form as:

Curr Protoc Protein Sci. 2008 May ; CHAPTER: Unit-5.21. doi:10.1002/0471140864.ps0521s52.

Recombinant protein complex expression in *E. coli*

William Selleck and Song Tan¹

Center for Gene Regulation, Department of Biochemistry & Molecular Biology, 108 Althouse Laboratory, The Pennsylvania State University, University Park, PA 16802-1014

Key terms

multisubunit protein complexes; polycistronic expression; affinity tags

Unit Introduction

It is increasingly clear that many, if not most proteins in a cell do not exist as isolated protein subunits, but instead function in multicomponent protein assemblies. Biochemical and biophysical studies of these complexes have been hampered by the difficulties associated with isolating the scarce quantities available in natural sources, while the complicated composition and organization of the complexes have challenged traditional approaches to recombinant expression in heterologous hosts such as *E. coli*. For projects requiring more material than can be purified from a natural source or for projects involving point or deletion mutations, a recombinant source is highly desirable.

The two most common approaches to produce recombinant protein complexes are to perform *in vitro* reconstitution of individually expressed and purified subunits or to implement *in vivo* reconstitution by coexpressing the subunits in an appropriate host (Fig. 1). Although *in vitro* reconstitution has been successfully used (Luger et al., 1997, Tan et al., 1996, Zalenskaya et al., 1990), the process is tedious (each subunit has to be expressed and purified, and the complex has to be further purified after reconstitution) and reconstitution yields are often low. In contrast, *in vivo* reconstitution by coexpression offers the benefits of efficiency (only one round of expression and purification) and potentially higher yields and quality of the desired complex (refolding and assembly of the complex take place in the presence of protein folding enzymes in a cellular environment). *In vivo* reconstitution has been successfully performed by coinfecting insect cells with baculoviruses expressing individual protein subunits (Tirode et al., 1999), and in bacteria from multiple plasmids (Johnston et al., 2000, McNally et al., 1988) or from specialized polycistronic plasmids (Henricksen et al., 1994, Ishiai et al., 1996, Li et al., 1997).

We have developed general polycistronic expression systems for producing protein complexes in *E. coli* (Tan, 2001, Tan et al., 2005). These systems utilize the concept of a translation cassette, comprised of the coding region with requisite START and STOP codons and preceded by translational initiation signals such as the Shine-Dalgarno sequence and translational enhancers (Tan, 2001) (Fig. 2). When transcribed into mRNA, the translation cassette contains the necessary and sufficient information for the *E. coli* translational machinery to initiate and sustain translation of the mRNA into the desired polypeptide. In the first generation pST39 system, we introduced the idea of using a modified pET3a T7-expression plasmid, pET3aTr,

1address for correspondence: Dr. Song Tan Center for Gene Regulation, Department of Biochemistry & Molecular Biology, 108 Althouse Laboratory, The Pennsylvania State University, University Park, PA 16802-1014 tel: 814-865-3355 fax: 814-863-7024 email: sxt30@psu.edu.

as the source of the translation cassette which could be subcloned using particular restriction enzyme pairs into one of four possible cassettes in the polycistronic vector pST39 (Studier et al., 1990, Tan, 2001). Since the pST39 plasmid provides transcriptional initiation and termination signals flanking the 4 cassettes, it can support the coexpression of up to 4 genes from a single plasmid. Thus, in the pST39 system, pET3aTr acts as the transfer or donor plasmid for each of the 4 possible translation cassettes in the polycistronic vector pST39. In an alternative approach, Novagen offers their Duet line of coexpression vectors that coexpress up to 2 genes each from multiple plasmids.

We have recently developed a second generation pST44 polycistronic system which improves upon the modularity of the first generation system in two distinct ways (Tan et al., 2005). Firstly, the new system now provides a separate transfer or donor plasmid for each of the 4 translation cassettes in the polycistronic vector (Fig. 2). This new design eliminates the problem of nested restriction sites inherent to the first generation polycistronic system's use of pET3aTr as the source of all 4 translation cassettes. A significant benefit of this change is that it simplifies the creation of polycistronic plasmids that express point or deletion mutants for any of the subunits. Secondly, the modular subcloning concept implemented in the first generation system to transfer translation cassettes from the transfer plasmid to the polycistronic plasmid is now extended to the process of subcloning a coding region into the transfer plasmid. In particular, modular N-terminal affinity tags are now available as single or double and cleavable or non-cleavable affinity tags, as are simple C-terminal affinity tags. We selected small (9–30 amino acids) affinity tags that we verified function efficiently as protein purification tags from *E. coli* as well as yeast, *Drosophila* and HeLa extracts (Lichty et al., 2005).

We describe here procedures to design and create polycistronic plasmids that coexpress protein complex subunits in *E. coli* using either the pST39 or pST44 expression systems. Since the subcloning steps constitute much of the effort to create a polycistronic vector, particular emphasis is placed on formulating a subcloning scheme that takes into account potential complications such as internal restriction sites. We also describe general methods to test the expression (Basic Protocol 1) and to purify the protein complex (Basic Protocol 2), as well as approaches to troubleshoot potential expression problems.

STRATEGIC PLANNING

Careful planning prior to any bench experiments usually saves much time and effort, and this is particularly true when expressing recombinant protein complexes in *E. coli* using polycistronic vectors. When one coexpresses a protein complex in *E. coli*, one frontloads effort into constructing the coexpression vector, whereas when one reconstitutes the complex from purified subunits in vitro, much of the effort is expended in the protein expression, purification and reconstitution steps (Fig. 1). In our experience, this is a significant benefit since subcloning work is almost always simpler and more reliable than performing protein expression and purification experiments. However, this also underscores the importance of carefully planning the DNA subcloning strategy to create the coexpression vectors so as to avoid time-consuming mistakes later on.

In planning a coexpression experiment, the first key decisions are which protein subunits to coexpress, which subunits to affinity tag, and to define the goals of the protein coexpression. Once these parameters have been established, one can design the subcloning strategy to create the polycistronic vector, taking into consideration internal restriction sites, the order of expression of the individual subunits, and the design of primers to perform the subcloning procedure.

Select components to coexpress in *E. coli*

Selecting the right components to coexpress may be the most important decision of the entire process. Ideally, one will know the protein components that are both necessary and sufficient to form the complex with the desired activity. The components necessary for complex formation may be known from genetic and biochemical studies in which the effects of deleting individual subunits on the purification of native complexes have been analyzed. However, defining the components sufficient to reconstitute the complex is typically more difficult. Even if powerful affinity purification methods such as TAP purification (Puig et al., 2001) are used to isolate a native complex which appears pure by silver staining, the bands which are visible may not be sufficient to reconstitute the complex. For example, additional smaller polypeptides that stain poorly may be additionally needed to form the complex. Methods to define sufficient components of a complex include reconstituting the complex *in vitro* or coexpressing the subunits in a heterologous system. The polycistronic expression systems described here can be used as both analytical tools (to determine if the selected components can form a complex) as well as preparative tools (to purify micro- and milligram amounts of complex). For example, purification of native yeast Piccolo NuA4 histone acetyltransferase complex identified 3 probable subunits: Epl1, Yng2 and Esa1. Subsequent coexpression of the 3 subunits in *E. coli* using the pST44 polycistronic system followed by purification and characterization of the resulting recombinant complex established that Epl1, Yng2 and Esa1 were necessary and sufficient components of the Piccolo NuA4 complex (Boudreault et al., 2003).

Decide which, if any, subunits to affinity tag and where (N- or C-terminus)

It may be tempting to tag all subunits to expedite purification of the complex and to track individual subunits during the expression and purification steps. The pST50Tr suite of plasmids in the pST44 purification system can facilitate this by providing 5 small N-terminal (HIS, CBP, Strep II, FLAG, HPC) and 4 C-terminal tags (HIS, Strep II, FLAG, CYD) in various combinations of single vs double and cleavable vs non-cleavable affinity tags (Tan et al., 2005) (Fig. 3a). Affinity resins and detection reagents are available for all these tags with the possible exception of the CYD (covalent yet dissociable) tag. However, one must also consider the possibility that a tag might interfere with the formation of complex, particularly if other larger protein tags such as GST (glutathione S-transferase) or MBP (maltose-binding protein) are used. Any prior knowledge of subunits or termini that can be tagged without affecting *in vivo* function or purification of the native complex will be particularly useful here. In general, we recommend tagging one subunit for simplified purification by affinity methods, but more tags can be incorporated if information is available that their presence can be accommodated without affecting formation or activity of the complex.

Define the goals of the polycistronic expression

The pST39 and pST44 polycistronic expression systems can be used in different ways depending on the design of the project. If the goal is to test coexpression of subunits for either analytical or preparative experiments where mutant or deletion variants will not be needed, then the original pET3aTr or the more flexible pST50Trc1 plasmid (with its suite of affinity tagging variants) can be used to subclone all genes. An advantage of this approach is that the resulting single plasmid can then be used as the donor of the translation cassette for each of the 4 cassettes (Tan, 2001). This is particularly useful when a variety of subcomplexes are to be generated from pool of components under investigation. However, if plasmids that express complexes with variant subunits (e.g. point or deletion mutants) are needed, the nested restriction sites within pET3aTr or pST50Trc1 will likely require multiple rounds of subcloning for each variant complex. This problem is aggravated when the variant gene replaces a gene in the 1st cassette since the polycistronic plasmid will essentially have to be created from scratch. In this case, use of the individual pST50Trc1, pST50Trc2, pST50Trc3 or pST50Trc4

plasmids is preferred so that each cassette can be replaced with a single subcloning step (pST50Trc1 is the pST50 Transfer plasmid for cassette 1) (Tan et al., 2005).

Examine gene sequences of components for internal restriction sites

Since restriction sites are used to subclone genes into the pST39 and pST44 polycistronic cassettes, the coincidental occurrence of these restriction sites internal to the genes of interest will complicate the subcloning process. For example, the presence of an internal EcoRI site in a gene will interfere with the use of EcoRI and HindIII to subclone this gene into cassette 2. It is therefore critical to examine what restriction sites are present in the genes one wishes to coexpress. We recommend using the worksheet in Table 1 to keep track of restriction sites. If one or more of the genes to be coexpressed do contain internal restriction sites, there are at least six different methods to deal with the problem.

- a. Alternate restriction sites are available for subcloning genes from the transfer plasmid (pET3aTr or the appropriate pST50Tr plasmid) into the polycistronic vectors pST39 or pST44. In particular, NheI, MfeI and BssHII yield compatible sticky ends to XbaI, EcoRI and MluI respectively (Fig. 4). BglII and NgoMIV are additionally available as alternates for BamHI and BspEI for the first generation pST39/pET3aTr system.
- b. One can select the order of subcloning to minimize the effect of the internal restriction sites. For example, if one wishes to create a pST44 polycistronic vector to express proteins A, B and C where gene C (but not genes A and B) contains an internal SacI site, consider subcloning A into cassette 1 (using the XbaI & BglII sites) followed by gene B into cassette 3 (using the SacI and KpnI sites) before subcloning gene C into cassette 2 (using the EcoRI & HindIII sites) (Fig. 4). This is an example where using individual pST50Tr plasmids (pST50Trc1 for cassette 1, pST50Trc2 for cassette 2, pST50Trc3 for cassette 3) is preferable since it avoids the problem of nested restriction sites.
- c. Partial digestion can be used to isolate the translation cassette from the transfer plasmid. In the previous example, one may be able to isolate the desired SacI-KpnI fragment from pST50Trc3-geneC by digesting first with KpnI followed by partial SacI digestion.
- d. If less than 4 genes are to be coexpressed, it may be possible to avoid a particular restriction enzyme by bypassing its use in the polycistronic vector. For example, if one or more genes of a two or three subunit complex contains an internal EcoRI site, one may consider subcloning the translation cassette for the first gene into the cassette formed by the XbaI and HindIII sites of pST39 or pST44, i.e. the union of cassettes 1 and 2 (Fig. 2, also Fig. 4a and d). In this way, one can eliminate the use of EcoRI for subcloning genes into the polycistronic expression vector.
- e. Since each of the 4 polycistronic cassettes in pST39 and pST44 contain a blunt end restriction site, it may be possible to avoid internal restriction sites by blunt-end cloning. Any gene fragment can be made blunt-ended by filling or trimming overhangs with an appropriate DNA polymerase. In the example described in item (b), gene C could be excised from pST50Trc3 using the 5' BspEI and 3' KpnI sites, not using the 5' SacI site because of the internal SacI site (Fig. 3g). That BspEI-KpnI gene C fragment could then be filled with Klenow enzyme before subcloning into the blunt end EcoRV site of the pST44 3rd cassette (Fig. 3d). This approach does not completely eliminate the problem of internal restriction sites since the genes of interest could contain internal occurrences of the blunt end sites in the 4 cassettes: StuI, SmaI, EcoRV and NruI.

- f. The final approach may be the simplest in concept: remove the internal restriction sites by site-directed mutagenesis. Although these sites are within the coding region, you can use the redundancy of the genetic code to change the DNA sequence without changing the translated amino acid sequence. The time and effort required to remove the internal restriction sites may be worthwhile if the resulting complex will be studied in depth by deletion analysis and mutagenesis experiments. For example, after we demonstrated that the yeast Piccolo NuA4 subunits Ep11, Yng2 and Esa1 could be coexpressed using the pST44 system, we removed the internal restriction sites that would otherwise complicate the creation of polycistronic plasmids with variant subunits. Once the internal restriction sites were removed, we created 20 variant plasmids for a detailed deletion analysis of each of the subunits to define the domains necessary for nucleosomal HAT activity. We were also able to rapidly create 7 point mutations in the Esa1 chromodomain, which we then used to demonstrate that a putative hydrophobic pocket in the chromodomain plays a critical role in the catalysis of nucleosome acetylation function (Selleck et al., 2005). We find QuikChange based mutagenesis (Stratagene) to be an efficient method to remove internal restriction sites by silent mutagenesis. We recommend sequencing genes modified by any *in vitro* replication such as the linear amplification employed in the QuikChange protocol or PCR in general to verify the absence of any undesirable sequence changes, even if high fidelity polymerases are employed.

Decide the order of expression of the genes along the polycistronic vector

A common concern is whether the order of expression of the coexpressed genes from a polycistronic plasmid affects expression of the complex. We tested all 6 permutations for the order of expression of the 3 Piccolo NuA4 subunits in the pST44 polycistronic plasmid and observed no detect significant differences in the expression of the complex (Tan et al., 2005). This suggests the order of expression may not be a critical parameter for polycistronic expression using the pST39 and pST44 systems. Therefore, for an initial experiment, we recommend selecting the order of expression which simplifies the subcloning scheme (see point 4(b) above).

Plan subcloning strategy

Genes are typically subcloned into the pET3aTr transfer plasmid as NdeI-BamHI fragments, following the cloning scheme of the original pET3a vector (Studier et al., 1990, Tan, 2001). Minimal provisions for affinity tags were provided in the original pET3aTr/pST39 plasmids of the first generation polycistronic system. In contrast, the pST50Tr family of transfer plasmids in the second generation polycistronic expression system introduced a modular cloning scheme which accommodates both N- and C-terminal affinity tags (Tan et al., 2005). In this modular scheme, the same BamHI-NgoMIV gene fragment can be subcloned to express untagged, cleavable or non-cleavable N-terminal tags, and C-terminal tags (Fig. 3b). No extra amino acids are added to the protein with the exception of 2 small amino acids introduced by the BamHI (Gly-Ser) or NgoMIV (Ala-Gly) restriction sites. In the case of cleavable N-terminal tags, the Gly encoded by part of the BamHI site also serves as the preferred P1' amino acid of the TEV protease site (Parks et al., 1994). Thus, using 5' BamHI and 3' NgoMIV sites allows great flexibility for adding or varying N- or C-terminal tags if an extra N-terminal Gly-Ser or C-terminal Ala-Gly can be tolerated in the protein. If even these very short fusions are undesirable, subcloning the gene as an NdeI-BsrGI fragment into the same pST50Tr plasmids will result in a protein without any non-native amino acids.

Design primers to subclone each gene

We typically amplify the genes to be coexpressed with primers that introduce 5' BamHI sites and 3' NgoMIV sites to subclone into the appropriate pST50Tr vector (Fig. 3). For full length genes, the BamHI site is inserted just before the 2nd codon (the first codon following the START codon). The NgoMIV follows the last amino acid before the native STOP codon since the pST50Tr vectors provide the STOP codon after the affinity tag. If no C-terminal tags will be used, we attach a BsrGI site immediately following the STOP codon.

CONSTRUCTING THE POLYCISTRONIC EXPRESSION VECTOR

Once the subcloning scheme has been worked out in advance, the actual creation of the polycistronic expression vector can proceed. In general, two steps are required. First, each component gene needs to be subcloned into the appropriate transfer vector (pET3aTr or one of the pST50Tr plasmids). Then, the translation cassette for each component gene is subcloned in turn into the pST39 or pST44 polycistronic vector. Standard subcloning techniques are used to create the transfer and polycistronic plasmids.

1. Subclone each gene into pET3aTr or appropriate pST50Tr vector

After amplifying the gene with the respective primers (see step 7), the PCR product is digested with the appropriate combination of restriction enzymes (NdeI-BamHI for pET3aTr, NdeI-NgoMIV, BamHI-NgoMIV or NdeI-BsrGI for pST50Tr vectors), agarose gel purified and ligated into the digested and gel purified pST50Tr vector for the appropriate cassette position and which contains the appropriate affinity tag. Table 2 shows the scheme we use for digesting DNA with two enzymes that may or may not share compatible restriction digest buffers. The resulting transfer plasmid for that subunit should be verified by restriction digestion, and particularly if PCR was performed, sequenced through the entire coding region.

2. Subclone translation cassette from transfer vector into polycistronic vector

The translation cassette from pET3aTr or pST50Tr can be released and subcloned into the polycistronic pST39 or pST44 vector, typically using XbaI-BglII for cassette 1, EcoRI-HindIII for cassette 2, SacI-KpnI for cassette 3 and BspEI-MluI for cassette 4 (Fig. 4). These restriction enzyme pairs were selected to function well together, an important consideration since each pair of restriction sites are only 6 bp apart in pST39 or pST44 and it would be impossible to separate plasmid DNA that is digested with only one enzyme from the desired doubly-digested vector fragment. For all the enzyme combinations shown in Table 2, we generally obtain at least 5 fold more colonies when a vector + insert ligation mix is transformed compared to a vector only ligation mix.

Although the pST39 and pST44 polycistronic vectors were designed to be used in conjunction with the pET3aTr and pST50Tr transfer vectors, it is possible to bypass the transfer vectors altogether and subclone a translation cassette directly into the polycistronic expression vector. For example, the necessary translational signals, affinity tags and restriction sites can be attached to a gene by PCR amplification using appropriate primers before subcloning into the polycistronic plasmid.

SMALL SCALE EXPRESSION OF SUBUNITS AND COMPLEX (Basic Protocol

1)

Once the transfer vectors for the individual subunits and the actual polycistronic expression plasmid have been constructed, expression of the individual subunits and the protein complex can be tested in an appropriate host strain. The mechanics of conducting a polycistronic expression experiment are identical to any other *E. coli* expression experiment: the expression

plasmid is transformed into the appropriate competent *E. coli* expression strain, grown in media, induced and harvested. Reduced induction temperatures of 18°C or 28°C can be attempted instead of 37°C to increase solubility of the overexpressed protein or protein complex (Schein and Noteborn, 1988). Since pST39 and pST44 are T7 based expression vectors, the common *E. coli* host strains for T7 based expression are typically used. These host strains include BL21(DE3), BL21(DE3)pLysS and their CodonPlus (Stratagene) and Rosetta (Novagen) variants which provide additional tRNAs to assist expression of genes containing rarely used *E. coli* codons (see troubleshooting step 1 below). Since the pET3aTr and pST50Tr transfer plasmids are also bona fide T7 expression plasmids, one can verify the integrity of the translation cassette by testing the expression of the corresponding subunit from the appropriate transfer plasmid, a simple and useful control experiment. We usually perform small scale expression on a 100 ml scale in 500 ml shaker flasks, but smaller scale experiments can also be performed. In the protocol below, we induce the BL21(DE3)pLysS cells with IPTG, but other induction schemes such as the auto-induction scheme developed by William Studier can be used.

Materials

Competent *E. coli* host strain such as BL21(DE3)pLysS

TYE plates (see recipe)

2xTY media (see recipe)

0.2 M IPTG solution (see recipe)

1x SDS sample buffer (unit 10.1)

P300-EDTA buffer (see recipe)

T100 buffer (see recipe)

liquid nitrogen (for freezing cell suspension)

shaking incubator such as New Brunswick model C24 incubator

spectrophotometer to measure optical density at 600 nm

microcentrifuge

tabletop or floor centrifuge with rotor which accepts 250 ml centrifuge bottles, such as

Heraeus Megafuge1.0

Additional reagents and equipment for preparing and analyzing samples for polyacrylamide gel analysis (UNIT 5.2)

1. Transform expression plasmid into appropriate competent *E. coli* expression strain, such as BL21(DE3)pLysS. For example, incubate 1 to 100 ng of plasmid DNA with 100 µl of competent BL21(DE3)pLysS cells on ice for 15 minutes, heat shock at 42°C for 30 seconds, place on ice for 10 seconds, add 500 µl 2xTY media, grow in a shaking incubator at 37°C for 15–30 minutes, then plate 300 µl of the transformation mix onto a TYE plate containing 100 µg/ml ampicillin and 25 µg/ml chloramphenicol. You should obtain 50 to 1000 colonies after incubating the transformation plate at 37°C for 12 to 16 hours.

To avoid complications due to unintended induction of cultures grown in complex media like 2xTY or LB, we always use freshly transformed cells for our expression experiments with BL21(DE3) strains in growth media such as 2xTY or LB, and we generally use these colonies within 36 hours of the overnight incubation. If you obtain inconsistent expression results using freshly transformed cells grown in 2xTY or LB media, you may wish to try the Studier defined auto-induction media which greatly reduces unintended induction of IPTG-inducible cultures by controlling for the presence of the inducer lactose (Studier, 2005).

2. Inoculate 100 ml of 2xTY with 50 µg/ml ampicillin and 25 µg/ml chloramphenicol with 3 colonies (0.5 – 1.0 mm diameter) of the expression colony from the fresh transformation plate. Grow in shaking incubator at appropriate temperature (typically 37°C) and 200 rpm.
3. When the OD₆₀₀ is between 0.5 and 0.9 (blanked against 2xTY media), induce the culture by adding 100 µl 0.2 M IPTG. This usually requires 4 to 6 hours at 37°C.
4. Transfer 500 µl of the uninduced culture (i.e. cells used for the optical density measurement) to a 1.5 ml microfuge tube. Centrifuge in a microfuge for 1 minute. Aspirate off the supernatant and resuspend the pelleted cells in 100 µl PGLB.
5. At hourly intervals up to 4 hours after inducing the culture with IPTG, transfer 250 µl of the induced culture to a 1.5 ml microfuge tube. Centrifuge in a microfuge for 1 minute, aspirate off the supernatant and resuspend the pelleted induced cells in 100 µl PGLB.
6. At 3 hours after induction, transfer 50 ml of the culture to a 50 ml Falcon tube and centrifuge at 4,000 g for 10 minutes in a tabletop centrifuge at room temperature (not 4°C). Pour off the supernatant, and resuspend the cell pellet in 10 ml appropriate buffer. Flash freeze in liquid nitrogen and store at –20°C.

This sample will be used to check the solubility of the overexpressed protein or protein complex and/or to perform small scale affinity purification. Although the optimal time of induction may not be known, we find that 3 hours is often an acceptable period of induction for overexpression experiments at 37°C or 28°C. Note that the BL21(DE3) pLysS cells should be centrifuged at room temperature and not 4°C to avoid cell lysis at low temperatures due to the presence of T7 lysozyme in the cells.

The cells should be resuspended in a buffer appropriate for the downstream purification steps. For example, for protein complexes with a histidine-tagged subunit, we use P300-EDTA buffer, whereas for samples without an affinity tag, we use T100 buffer. Cells can be resuspended using a 10 ml plastic or glass pipette and a pipette bulb or mechanical pipet-aid filler. Resuspended cells can be stored at –20°C or –80°C for at least 6 months.

7. Analyze the time point samples (0, 1, 2, 3, 4 h) by SDS-PAGE. You will hopefully observe induction of the desired polypeptide(s) by Coomassie Blue staining, but you may need to employ more sensitive detection methods such as Western blotting with an appropriate antibody if the expression levels are low.

SMALL SCALE PURIFICATION OF TAGGED SUBUNITS AND COMPLEX

(Basic Protocol 2)

If an affinity tag such as the HIS peptide has been incorporated into one of the coexpressed subunit, a quick and simple small scale affinity purification can often confirm the coexpression and reconstitution of the complex. This is particularly helpful when the level of expression of complex subunits is not high enough to detect easily using Coomassie Blue staining of the

crude extract or if antibodies are not available for the subunits. The following procedure which employs a small scale batch purification over the cobalt-based Talon resin (Clontech) has been successfully used to purify complexes containing HIS-tagged subunits. The protocol can also be adapted to use a variety of immobilized metal affinity chromatography resins or coated plates that are commercially available.

Materials

Affinity chromatography resin, such as Talon metal affinity resin (Clontech)

P300-EDTA (see recipe)

TYE plates (see recipe)

2xTY media (see recipe)

0.2 M IPTG solution (see recipe)

Bio-Spin disposable spin columns (BioRad catalog #732–6008)

sonicator (Any sonicator capable of disrupting *E. coli* cells will suffice. We use a Branson S-450D 400 watt probe sonicator with a 1/2" probe) microcentrifuge

tabletop centrifuge

Additional reagents and equipment for preparing and analyzing samples for polyacrylamide gel analysis (UNIT 5.2)

1. Transfer 1 ml of resuspended Talon resin to a 15 ml Falcon tube. Add 10 ml of deionized or MilliQ water. Mix by inversion several times. Spin in a tabletop centrifuge at 700 g for 2 minutes at room temperature to sediment the resin.
2. Pour off (discard) the supernatant, add 10 ml of P300-EDTA and mix by inversion several times. Spin in a tabletop centrifuge at 700 g for 2 minutes at room temperature. Discard this supernatant. If you will not use the resin in the next 10 minutes, cap the tube so that the resin does not dry out.
3. Prepare soluble cellular extract of the induction sample prepared in step 6 of the "small scale expression of subunits and complexes" procedure above. For example, BL21 (DE3)pLysS cells properly resuspended in buffer should lyse by simply freeze-thawing the cell suspension, but the released nucleic acid will create a viscous extract. We prefer to additionally sonicate the extract to shear the nucleic acid as well as to increase the efficiency of lysis. We use a Branson sonicator with a 1/2" probe head, 20 x 0.5 sec pulses at 30% full power for 10 ml extract delivered in 2 rounds of 10 x 0.5 sec pulses with a 30 sec incubation on ice in between. Centrifuge the sonicated extract at 15,000 g for 3 minutes in either a superspeed centrifuge or alternatively, in a microfuge after aliquoting into 6–8 1.5 ml Eppendorf tubes. The supernatant of this centrifugation will constitute the cleared extract.

The centrifugation and subsequent steps of this procedure can be performed at room temperature, but will work equally well at 4°C if you are concerned about the stability of your protein complex. If you prefer not to sonicate the extract, you can add 20 µg/ml DNaseI and 1 mM MgCl₂ final concentration to enzymatically reduce the viscosity of the extract.

4. Transfer the cleared extract (supernatant of previous centrifugation) to the 15 ml Falcon tube containing the pre-equilibrated 0.5 ml of Talon resin (from step 2). Mix gently for 20 minutes.
5. Sediment the Talon resin in a tabletop centrifuge at 700 g for 5 minutes. Transfer supernatant (Talon flow through) to a new 15 ml Falcon tube. Wash the resin with 10 ml P300 buffer twice using the same centrifugation conditions to sediment the resin between washes.

Although the P300 washes generally do not contain much of the tagged protein, it is prudent to save these washes until you have assayed the imidazole eluted fractions of step 6.

6. Resuspend the washed resin (~0.5 ml) in 1 ml P300-EDTA and transfer this resin suspension to a BioRad Bio-Spin column. Allow the buffer to drain from the resin.
7. Elute bound proteins with four 0.5 ml aliquots of P300 buffer + 100 mM imidazole using gravity feed. Collect each 0.5 ml elution into a separate 1.5 ml Eppendorf tube.

You should observe darkening of the resin to a richer mauve color when the imidazole is in contact with the resin. The first fraction typically contains little protein since 0.5 ml of elution buffer is needed to displace the 0.5 ml of buffer in the resin. Most of the eluted proteins typically elute in fraction 2, with decreasing amounts of protein in fractions 3 and 4. However, if the volume of resin in the column is less than 0.5 ml, you may observe significant eluted protein in fraction 1.

8. Analyze samples by SDS-PAGE followed by Coomassie Blue or silver staining, or if necessary by Western blot analysis using an antibody raised against an affinity tag or one of the proteins of interest. Sample results for the purification of two or three subunit complexes can be found in Figures 5 and 6. The eluted protein can usually be stored on ice for a day or two, or for longer periods at -20°C after addition of glycerol to a final concentration of 20%.

REAGENTS AND SOLUTIONS

Antibiotic stock solutions (1000x concentrated for liquid culture work)

Ampicillin: 50 g/liter in deionized or distilled water. Filter through 0.22 μm filter. Store ≤ 3 months at -20°C .

Chloramphenicol: 25 g/liter in ethanol (95% or 100% ethanol). Store ≤ 1 year at -20°C .

Isopropyl- β -thiogalactopyranoside (IPTG)

0.2 M stock solution in deionized or distilled water. Filter through 0.22 μm filter. Store ≤ 1 year at -20°C .

TYE plate medium

Per liter:

10 g Bacto-tryptone

5 g yeast extract

8 g NaCl

15 g agar

deionized or distilled water to 1 liter

Sterilize 30 min at 121°C

Allow media to cool to about 60°C before adding antibiotics (typically 100 mg/ml ampicillin and/or 25 mg/ml chloramphenicol). Media that is too hot when poured into plates will produce excessive condensation, which if not dried off completely, will increase the probability of plate contamination during storage. Pour about 25 ml of medium to each of about 40 sterile disposable petri dishes, stacking the plates on top of each other to reduce condensation. Allow the medium to set and then dry overnight at room temperature. The dried plates can be stored at 4°C in the plastic sleeve bag used to ship the petri dishes. Plates containing antibiotic plates should be used within 1–2 months.

2xTY medium

Per liter:

16 g Bacto-tryptone

10 g yeast extract

5 g NaCl

deionized or distilled water to 1 liter

Dispense 100 ml into 500 ml Erlenmeyer flasks (preferably thick-walled to reduce breakage).

Plug with foam plug and then aluminum foil.

Sterilize 30 min at 121°C

Add antibiotics just before use (typically 50 mg/ml ampicillin and/or 25 mg/ml chloramphenicol)

P300-EDTA buffer	
50 mM sodium phosphate pH 7.0	3.05 ml 0.5 M Na ₂ HPO ₄
	1.95 ml 0.5 M NaH ₂ PO ₄
100 mM NaCl	1 ml 5 M NaCl
1 mM benzamidine	50 µl 1 M benzamidine
5 mM 2-mercaptoethanol	17 µl stock (14.4 M) 2-mercaptoethanol
	MilliQ water to 50 ml

Benzamidine is a general protease inhibitor that is both water soluble and stable, in contrast to protease inhibitors like PMSF.

P300-EDTA buffer + 100 mM imidazole	
P300-EDTA buffer	20 ml P300-EDTA buffer
100 mM imidazole	0.136 imidazole

T100 buffer	
20 mM Tris-Cl pH 8.0	1 ml 1 M Tris-Cl pH 8.0
100 mM NaCl	1 ml 5 M NaCl
0.5 mM EDTA, Na ₂	50 µl 0.5 M EDTA, Na ₂
1 mM benzamidine	50 µl 1 M benzamidine
10 mM 2-mercaptoethanol	35 µl stock (14.4 M) 2-mercaptoethanol
	MilliQ water to 50 ml

COMMENTARY

Background Information

The pST39 and pST44 polycistronic expression system enable multicomponent protein complexes to be coexpressed and assembled in *E. coli*. The modular nature of the original pST39 system has been extended in the second generation pST44 system to facilitate the inclusion of both N-terminal and C-terminal affinity tags for simplified purification of the recombinant complexes. The complexes that have been successfully expressed using either the pST39 or pST44 systems include:

- a. the von Hippel-Lindau VHL/elonginB/elonginC tumor suppressor complex (3 subunits) (Tan, 2001),
- b. several TFIID associated histone-fold complexes: scTAF9/6, scTAF4/12, scTAF11/13 (each 2 subunits). The scTAF9/6 and scTAF4/12 were further reconstituted *in vitro* into the scTAF9/6/4/12 histone-fold octamer (Selleck et al., 2001, Tan, 2001),
- c. the SAGA associated histone-fold complex of scTAF12/Ada1 (2 subunits), which was then reconstituted with scTAF9/6 *in vitro* into the scTAF9/6/12/Ada1 SAGA histone-fold subcomplex (unpublished data),
- d. the catalytic Ada2/Ada3/Gcn5 (3 subunits) and the Ada2/Gcn5 (2 subunits) histone-acetyltransferase subcomplexes of yeast SAGA coactivator complex (Balasubramanian et al., 2002, Boyer et al., 2002),
- e. the catalytic Piccolo NuA4 histone acetyltransferase subcomplex of yeast NuA4 coactivator complex (3 subunits: Epl1, Yng2 and Esa1) (Boudreault et al., 2003), as well as 27 variant complexes for deletion analysis of each subunit and mutational analysis of the Esa1 chromodomain function (Selleck et al., 2005),
- f. the catalytic human Piccolo histone acetyltransferase subcomplex of human Tip60 complex (3 subunits: Epc1, Ing3 and Tip60) (Doyon et al., 2004),
- g. the Aurora kinase/INCENP chromosomal passenger complex (2 subunits) (B. Hnatkovich and S.T., unpublished data),
- h. the RNA-binding Nrd1-Nab3 heterodimer involved in premature transcriptional termination (2 subunits) (Arigo et al., 2006, Carroll et al., 2007)
- i. the Asf1 histone chaperone bound to a histone H3/H4 heterodimer (3 subunits) (English et al., 2006),
- j. the ESCRT-I (3 or 4 subunits) and ESCRT-II (3 subunit) endosomal trafficking complexes (Hierro et al., 2004, Kostelansky et al., 2007, Kostelansky et al., 2006),
- k. the CENP-O class CENP-O, CENP-P, CENP-Q and CENP-50 kinetochore proteins (4 subunits) (Okada et al., 2006), and (personal communication, T. Fukagawa)
- l. the NDC-80 (4 subunits) and the MIS-12 (4 subunits) components of the microtubule-binding core of the kinetochore (Cheeseman et al., 2006). In addition, the 110 kDa KNL-1 protein copurified with the MIS-12 complex when coexpressed with that complex, and the KNL-1 protein, MIS-12 complex and NDC-80 complex could be reconstituted *in vitro* into the 9 protein KMN network.

The crystal structure of the ESCRT-II endosomal trafficking complex indicates that the 3 subunits form a complex with a stoichiometry of 2:1:1. This demonstrates that the pST39 polycistronic expression system is capable of successfully reconstituting complexes with stoichiometries more complicated than 1:1.

Although the pST39 and pST44 polycistronic expression systems were designed to coexpress protein subunits for recombinant protein complex production, they could be used for other potential applications. For example, it should be possible to coexpress a post-translational modification enzyme with its protein substrate to produce the appropriately modified protein. Such an approach has been used to produce recombinant phosphorylated cyclin-dependent kinase protein using two -co-transformed monocistronic plasmids. (Russo et al., 1996). Similarly, proteins that require particular protein chaperones for efficient folding can be coexpressed with the appropriate chaperone (Yasukawa et al., 1995). An alternative application would be to use the polycistronic expression vector to coexpress two different types of macromolecules. For instance, it should be possible to combine a translation cassette with a transcription cassette to coexpress an RNA-binding protein with its cognate RNA-binding site. This might be particularly useful for multicomponent protein-RNA complex where the proteins and RNA components are intimately cofolded together.

A particular application that we are currently exploring is to combine *in vivo* reconstitution with *in vitro* reconstitution to produce even more complicated assemblies. For instance, we have prepared an 8 component, 320 kDa HAT enzyme/nucleosome complex containing the Piccolo Nua4 histone acetyltransferase complex (3 subunit, MW = 120 kDa for a particular truncated complex) and a recombinant nucleosome core particle (4 histone protein subunits and a 147 bp DNA, MW = 200 kDa) (Selleck et al., 2005). The Piccolo NuA4 complex was produced by polycistronic expression in *E. coli*, but the recombinant nucleosome core particle was assembled *in vitro* from individually expressed and purified components. Another example of combining *in vitro* and *in vivo* reconstitution to assemble a large multicomponent complex is found in the exciting report of the biochemical reconstitution of the 9 subunit microtubule-binding unit of the kinetochore, the KMN network, from 2 subcomplexes separately coexpressed in *E. coli* using the pST39 system (Cheeseman et al., 2006).

It is worth mentioning that the limit of 4 coexpressed subunits in the pST39 and pST44 polycistronic expression systems is arbitrary and can be easily lifted by increasing the number of cassettes on a single plasmid. In addition, one can create a second or even a third compatible polycistronic expression vector to further increase the number of subunits that will be coexpressed in *E. coli*. In this way, it may be possible to reconstitute complexes containing 8, 12 or even more subunits by polycistronic expression. Such systems would be greatly welcome in many fields where interesting multisubunit complexes being discovered and characterized each week.

Critical Parameters and Troubleshooting

1. No or very low expression of complex subunit(s) detected—One benefit of the modular design of the pST39 and pST44 polycistronic system using a transfer plasmid as the source of the translation cassette is that the transfer plasmid is itself a monocistronic expression vector that can express the appropriate subunit. Thus, if a subunit of the desired complex expresses poorly or not at all in the polycistronic expression vector, the expression of that subunit alone can be checked in its transfer plasmid. This feature can help one to isolate the cause of poor expression. One potential cause is errors such as point deletions or insertions in the coding region possibly introduced during PCR steps used to subclone the gene into the transfer or polycistronic vectors. If poor or no expression of a particular subunit is detected, we recommend confirming the integrity of the entire coding region by DNA sequencing the transfer or polycistronic expression vector using the provided sequencing priming sites (Fig. 4) and/or custom internal primers.

Poor expression can also be observed if the gene contains codons that are poorly utilized in *E. coli*. For example, the Arg tRNA which recognizes the AGA and AGG codons is present in limiting quantities in *E. coli* strains such as BL21(DE3)pLysS commonly used for T7-based

expression. These codons are more frequent in eukaryotic genes and dramatic improvements in expressing certain eukaryotic genes have been observed when an *E. coli* host strain is supplemented with additional Arg tRNA^{AGA/AGG} (Brinkmann et al., 1989). Strains such as CodonPlus and Rosetta which provide several potentially limiting tRNAs on a pACYC-based plasmid compatible with the pBR322/pUC-based pST39 and pST44 polycistronic plasmids are commercially available and widely used. One significant advantage of polycistronic expression plasmids such as pST39 and pST44 over multiple cotransformed monocistronic plasmids is compatibility with such host strains engineered to handle poor codon usage. Since these host strains express the otherwise limiting tRNAs on a plasmid, that plasmid must contain an origin of replication and antibiotic resistance marker compatible and distinct from the plasmids which coexpress the complex subunits. Most expression plasmids in common use contain either the pMB1 (pBR322/pUC ColE1 compatibility group) or the p15A (pACYC) origins of replication. If one pBR322-based and one pACYC-based expression plasmid are used to coexpress a two subunit complex, the CodonPlus or Rosetta strains cannot be used since these strains already contain a pACYC-based plasmid to express the supplementary tRNA. In contrast, no such conflict exists for the polycistronic pST39 or pST44 plasmids which would coexpress both subunits from a single pBR322/pUC-based plasmid. Furthermore, this plasmid compatibility problem is compounded when more than two subunits need to be coexpressed using coexisting monocistronic vectors

2. One or more overexpressed protein is insoluble—We have observed multiple instances where an individual protein is expressed as insoluble polypeptide on its own, but becomes soluble when coexpressed with its partner subunits (Balasubramanian et al., 2002, Boudreault et al., 2003, Selleck et al., 2001). This suggests that some proteins may not fold on their own, but rather cofold with their partner subunits into complexes. It is therefore possible to observe that a soluble complex is formed by coexpression, but that the excess of one subunit is found in inclusion bodies. For example, soluble VHL/elonginB/elonginC complex was expressed using the pST39 polycistronic vector, but excess VHL and elonginC polypeptide were found in the insoluble phase of the crude extract (Tan, 2001). However, if the majority of complex subunits are expressed as insoluble polypeptides, one may wish to test conditions that favor expression of soluble proteins. Such conditions include inducing the cells at temperatures lower than 37°C (Schein and Noteborn, 1988) (we often try 28°C or 18°C) or growing the cells in special media such as one containing betaine and a high concentration of sorbitol (Blackwell and Horgan, 1991).

If altering growth or induction conditions does not improve the situation, one should consider if any incorporated tags might interfere with complex formation. The modular design of the second generation pST44 system facilitates the incorporation of affinity tags at different termini or on different subunits. If a subcomplex of a larger protein assembly is being coexpressed, one may also wish to consider the possibility that the current subcomplex contains regions that normally interact with subunits of the larger parent assembly. Such interacting regions might aggregate in the absence of their cognate subunits and it may be possible to express a soluble and functional subcomplex when these regions are removed (Fribourg et al., 2001).

One possible outcome of the expression experiment is that the individual subunits are expressed but do not reconstitute into a complex. This suggests that the subunits are not sufficient for complex formation. The missing component(s) may be additional protein subunits, other macromolecular subunits such as RNA or DNA, some post-translational modification not present in *E. coli* or even a critical small molecule cofactor. It is also possible that a particular protein chaperone might be necessary for the folding of the complex from its component subunits. If information is available as to the role of potential missing partner subunits, RNA, post-translational modifications or protein chaperones, it may be possible to exploit the polycistronic expression vector to coexpress the appropriate additional subunits or enzymes.

3. The complex can be expressed and purified, but smaller polypeptides remain after purification—

It is not unusual to observe that a coexpressed complex of full length components contains smaller polypeptides despite multiple chromatographic purification steps, particularly if the complex is a subcomplex from a larger assembly (see point 2 above). These contaminating fragments could result from limited proteolysis during purification, limited proteolysis in the host strain during induction or translation from cryptic initiation sites. Figs. 5 and 6 provide examples of one step affinity purifications where many of the contaminating smaller molecular weight bands likely correspond to limited proteolysis products. Use of protease inhibitor cocktails may alleviate limited proteolysis during purification, but other approaches are necessary for the other two potential problems. Limited proteolysis in the cell could suggest that the complex contains disordered regions that are particularly susceptible to proteolysis. If one needs to isolate the unproteolyzed complex, affinity tags can be used to eliminate truncated products, particularly if one knows the nature of the truncated proteins. For example, if a contaminating band has been identified by N-terminal sequence and/or mass spectrometric analysis to correspond to a truncation missing the N-terminus of a subunit, one may wish to attach an N-terminal tag to ensure that complex containing that subunit with an intact N-terminus can be isolated. We used a similar approach to purify recombinant human Tip60 catalytic HAT subcomplex of Epc1/Ing3/Tip60 from presumed degradation products (Doyon et al., 2004). An alternate approach would be to eliminate the potentially disordered regions by preparing recombinant truncations. The second generation pST44 polycistronic system is well-suited for this purpose. We have observed the overall expression yield to increase for certain complexes when extraneous regions were removed, presumably because these more optimized complexes are more proteolytically resistant (unpublished data).

4. The purified complex appears to contain a mixture of different subcomplexes

—Interacting proteins may form obligate complexes where the individual subunits do not normally exist as well-behaved entities on their own (see troubleshooting item 2 above), or as nonobligate complexes where complexes and individual subunits can coexist. When isolating nonobligate complexes, one may observe mixtures of the desired complex and either individual subunits or subcomplexes. For example, polycistronic expression of a ternary $\alpha\beta\gamma$ complex might also produce $\alpha\beta$, $\alpha\gamma$ or $\beta\gamma$ binary subcomplexes as well as individual α , β or γ subunits. If the subcomplexes possess sufficiently different chemical properties, it may be possible to separate them from the desired full complex by conventional chromatography methods. Strategically placed affinity tags may also be used to facilitate the purification of the desired complex. For example, if a different affinity tag is available on each of the three subunits, it should be possible to use three successive affinity purification steps to isolate complex containing all three subunits. The suite of pST50Tr plasmids with N- and C-term affinity tags can simplify the construction of polycistronic plasmids which express appropriately tagged subunits for such purposes.

Anticipated Results

If all goes well, each of the subunits in the polycistronic vector will be expressed in *E. coli* and associate with its partner subunits as judged by copurification in the small scale affinity purification experiment. Suggested ways of tackling results different from this anticipated outcome are discussed in the previous Critical Parameters and Troubleshooting section.

Time Considerations

Although the actual protein expression experiment can be completed in two days, and the small scale affinity purification in an additional day, significantly more time will be required to construct the polycistronic expression vector. The time needed for the subcloning steps will depend on the number of subunits to be coexpressed and the complexity of the subcloning

procedure, typically determined by the size of each subunit gene and the absence or presence of internal restriction sites in each gene. If none of the genes for a 4 subunit complex contain internal restriction sites used for the subcloning the genes into the transfer or polycistronic plasmids, it is possible to create the polycistronic plasmid for the 4 subunit complex in 2–3 weeks assuming all subcloning steps are successful on the first attempt. We have selected the restriction enzyme pairs for each subcloning step to improve the chances of subcloning success.

In general, if the vector and insert restriction fragments are properly gel purified, one should obtain good subcloning success with the pST39 or pST44 systems. The time needed to create the polycistronic plasmid will be longer if internal restriction sites need to be removed by site-directed mutagenesis.

References

- Arigo JT, Carroll KL, Ames JM, Corden JL. Regulation of yeast NRD1 expression by premature transcription termination. *Mol Cell* 2006;21:641–51. [PubMed: 16507362]
- Balasubramanian R, Pray-Grant MG, Selleck W, Grant PA, Tan S. Role of the Ada2 and Ada3 transcriptional coactivators in histone acetylation. *J Biol Chem* 2002;277:7989–7995. [PubMed: 11773077]
- Barrios A, Selleck W, Hnatkovich B, Kramer R, Sermwittayawong D, Tan S. Expression and purification of recombinant yeast Ada2/Ada3/Gcn5 and Piccolo NuA4 histone acetyltransferase complexes. *Methods* 2007;41:271–7. [PubMed: 17309836]
- Blackwell JR, Horgan R. A novel strategy for production of a highly expressed recombinant protein in an active form. *FEBS Lett* 1991;295:10–12. [PubMed: 1765138]
- Boudreault AA, Cronier D, Selleck W, Lacoste N, Utley RT, Allard S, Savard J, Lane WS, Tan S, Cote J. Yeast enhancer of polycomb defines global Esa1-dependent acetylation of chromatin. *Genes Dev* 2003;17:1415–1428. [PubMed: 12782659]
- Boyer LA, Langer MR, Crowley KA, Tan S, Denu JM, Peterson CL. Essential role for the SANT domain in the functioning of multiple chromatin remodeling enzymes. *Mol Cell* 2002;10:935–42. [PubMed: 12419236]
- Brinkmann U, Mattes RE, Buckel P. High-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the dnaY gene product. *Gene* 1989;85:109–114. [PubMed: 2515992]
- Carroll KL, Ghirlando R, Ames JM, Corden JL. Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. *Rna* 2007;13:361–73. [PubMed: 17237360]
- Cheeseman IM, Chappie JS, Wilson-Kubalek EM, Desai A. The conserved KMN network constitutes the core microtubule-binding site of the kinetochore. *Cell* 2006;127:983–97. [PubMed: 17129783]
- Doyon Y, Selleck W, Lane WS, Tan S, Cote J. Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans. *Mol Cell Biol* 2004;24:1884–1896. [PubMed: 14966270]
- English CM, Adkins MW, Carson JJ, Churchill ME, Tyler JK. Structural basis for the histone chaperone activity of Asf1. *Cell* 2006;127:495–508. [PubMed: 17081973]
- Fribourg S, Romier C, Werten S, Gangloff YG, Poterszman A, Moras D. Dissecting the interaction network of multiprotein complexes by pairwise coexpression of subunits in *E. coli*. *J Mol Biol* 2001;306:363–373. [PubMed: 11237605]
- Henricksen LA, Umbricht CB, Wold MS. Recombinant replication protein A: expression, complex formation, and functional characterization. *J Biol Chem* 1994;269:11121–32. [PubMed: 8157639]
- Hierro A, Sun J, Rusnak AS, Kim J, Prag G, Emr SD, Hurley JH. Structure of the ESCRT-II endosomal trafficking complex. *Nature* 2004;431:221–225. [PubMed: 15329733]
- Ishiai M, Sanchez JP, Amin AA, Murakami Y, Hurwitz J. Purification, gene cloning, and reconstitution of the heterotrimeric single-stranded DNA-binding protein from *Schizosaccharomyces pombe*. *J Biol Chem* 1996;271:20868–78. [PubMed: 8702843]
- Johnston K, Clements A, Venkataramani RN, Trievel RC, Marmorstein R. Coexpression of proteins in bacteria using T7-based expression plasmids: expression of heteromeric cell-cycle and transcriptional regulatory complexes. *Protein Expr Purif* 2000;20:435–443. [PubMed: 11087683]

- Kostelansky MS, Schluter C, Tam YY, Lee S, Ghirlando R, Beach B, Conibear E, Hurley JH. Molecular architecture and functional model of the complete yeast ESCRT-I heterotetramer. *Cell* 2007;129:485–98. [PubMed: 17442384]
- Kostelansky MS, Sun J, Lee S, Kim J, Ghirlando R, Hierro A, Emr SD, Hurley JH. Structural and functional organization of the ESCRT-I trafficking complex. *Cell* 2006;125:113–26. [PubMed: 16615894]
- Li C, Schwabe JW, Banayo E, Evans RM. Coexpression of nuclear receptor partners increases their solubility and biological activities. *Proc Natl Acad Sci U S A* 1997;94:2278–83. [PubMed: 9122185]
- Lichty JJ, Malecki JL, Agnew HD, Michelson-Horowitz DJ, Tan S. Comparison of affinity tags for protein purification. *Protein Expr Purif* 2005;41:98–105. [PubMed: 15802226]
- Luger K, Rechsteiner TJ, Flaus AJ, Wayne MM, Richmond TJ. Characterization of nucleosome core particles containing histone proteins made in bacteria. *J Mol Biol* 1997;272:301–311. [PubMed: 9325091]
- McNally EM, Goodwin EB, Spudich JA, Leinwand LA. Coexpression and assembly of myosin heavy chain and myosin light chain in *Escherichia coli*. *Proc Natl Acad Sci U S A* 1988;85:7270–3. [PubMed: 3050991]
- Okada M I, Cheeseman M, Hori T, Okawa K, McLeod IX, Yates JR 3rd, Desai A, Fukagawa T. The CENP-H-I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres. *Nat Cell Biol* 2006;8:446–57. [PubMed: 16622420]
- Parks TD, Leuther KK, Howard ED, Johnston SA, Dougherty WG. Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase. *Anal Biochem* 1994;216:413–417. [PubMed: 8179197]
- Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, Wilm M, Seraphin B. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* 2001;24:218–229. [PubMed: 11403571]
- Russo AA, Jeffrey PD, Pavletich NP. Structural basis of cyclin-dependent kinase activation by phosphorylation. *Nat Struct Biol* 1996;3:696–700. [PubMed: 8756328]
- Schein CH, Niteborn MH. Formation of soluble recombinant proteins in *Escherichia coli* is favored by lower growth temperature. *Bio/Technology* 1988;6:291–294.
- Selleck W, Fortin I, Sernwittayawong D, Cote J, Tan S. The *Saccharomyces cerevisiae* Piccolo NuA4 histone acetyltransferase complex requires the Enhancer of Polycomb A domain and chromodomain to acetylate nucleosomes. *Mol Cell Biol* 2005;25:5535–42. [PubMed: 15964809]
- Selleck W, Howley R, Fang Q, Podolny V, Fried MG, Buratowski S, Tan S. A histone fold TAF octamer within the yeast TFIID transcriptional coactivator. *Nat Struct Biol* 2001;8:695–700. [PubMed: 11473260]
- Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 2005;41:207–34. [PubMed: 15915565]
- Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* 1990;185:60–89. [PubMed: 2199796]
- Tan S. A modular polycistronic expression system for overexpressing protein complexes in *Escherichia coli*. *Protein Expr Purif* 2001;21:224–234. [PubMed: 11162410]
- Tan S, Hunziker Y, Sargent DF, Richmond TJ. Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature* 1996;381:127–151. [PubMed: 8610010]
- Tan S, Kern RC, Selleck W. The pST44 polycistronic expression system for producing protein complexes in *E. coli*. *Protein Expr Purif* 2005;40:385–395. [PubMed: 15766881]
- Tirode F, Busso D, Coin F, Egly JM. Reconstitution of the transcription factor TFIIF: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Mol Cell* 1999;3:87–95. [PubMed: 10024882]
- Yasukawa T, Kanei-Ishii C, Maekawa T, Fujimoto J, Yamamoto T, Ishii S. Increase of solubility of foreign proteins in *Escherichia coli* by coproduction of the bacterial thioredoxin. *J Biol Chem* 1995;270:25328–31. [PubMed: 7592692]
- Zalenskaya K, Lee J, Gujuluva CN, Shin YK, Slutsky M, Goldfarb A. Recombinant RNA polymerase: inducible overexpression, purification and assembly of *Escherichia coli* rpo gene products. *Gene* 1990;89:7–12. [PubMed: 2197183]

Figure 1. Schematic representation of *in vitro* and *in vivo* reconstitution methods to produce a 4 subunit recombinant protein complex. (a) With *in vitro* reconstitution, each of the 4 subunits is expressed and purified individually before combining to form the complex, which must be further purified from partial or misfolded complexes. (b) In contrast, *in vivo* reconstitution requires only a single expression step and one round of purification steps to produce purified the 4 subunit complex.

Figure 2.

Creation of a polycistronic expression vector. The pST50Trc1–4 plasmids provide the translation cassettes for the 4 possible positions in the pST44 polycistronic expression vector. Specific pairs of restriction enzymes are used to subclone the translation cassettes from the pST50Trc1–4 plasmids into pST44 (XbaI-BglII for cassette 1, EcoRI-HindIII for cassette 2, SacI-KpnI for cassette 3, BspEI-MluI for cassette 4). The XbaI-BglII translation cassette from pST50Trc1 plasmid is detailed to highlight the translational enhancer (ϵ) and the Shine-Dalgarno sequence (SD) which precede the coding region (black bar bracketed by NdeI and BsrGI sites). Restriction sites and coding regions for translation cassettes 1, 2, 3, 4 are shown in blue, green, yellow and red respectively. Transcriptional promoter and termination signals on pST50Trc1–4 are present but not shown. All restriction sites shown are unique. Plasmids are not drawn to scale.

Figure 3.

Modular subcloning scheme facilitates incorporation of a wide range of affinity tag combinations. (a) Suite of affinity tags available for the pST50Trc1–4 transfer plasmids. 5 N-terminal, short affinity tags are available as single noncleavable (first column), single cleavable (second column), or double cleavable tags (third column). In addition, 4 C-terminal, noncleavable affinity tags are provided (fourth column). These 14 affinity tag combinations are available for each of pST50Trc1, pST50Trc2, pST50Trc3 and pST50Trc4 plasmids. The names and the sequences of the 6 different affinity tags are provided on the right. (b) Modular subcloning scheme for subcloning genes into pST50Trc1–4 plasmids. The same BamHI–NgoMIV coding region can be subcloned into an appropriate pST50Trc1–4 plasmid to express an untagged subunit (top row), an N-terminal, non-cleavable tagged subunit (second row), a N-terminal, cleavable tagged subunit (third row) or a C-terminal, non-cleavable tagged subunit (bottom row). Use of this subcloning scheme adds an N-terminal Gly-Ser (from the GGATCC BamHI site) and a C-terminal Ala-Gly (from the GCCGGC NgoMIV site) to the native protein. Other subcloning schemes are also possible including NdeI–BsrGI (for untagged subunits with no non-native amino acids), BamHI–BsrGI (for N-terminal tagged subunits only) and NdeI–NgoMIV (for C-terminal tagged subunits only). The scissors icon represents the tobacco etch virus (TEV) protease recognition site or the TEV protease itself. The horizontal black bar 5' of the NdeI site in each plasmid represents the translational initiation signals such as the translational enhancer and Shine-Dalgarno sequence.

Figure 4.

Plasmid maps for the vectors in the pST39 and pST44 polycistronic expression systems. The location of the T7 ϕ 10 transcription promoter and terminator are shown for each of the plasmids, while the translational enhancer (ϵ) and Shine-Dalgarno (SD) sites are shown for the transfer plasmids pET3aTr and pST50Trc1–4. The restriction sites which define the 4 polycistronic cassettes of pST39 and pST44 are shown in bold, as are the sites on the corresponding transfer plasmids. A specific plasmid is available for each cassette in the pST50Trc1–4 transfer plasmids. For example, pST50Trc1 would typically be used as the donor plasmid for subcloning the XbaI-BglII translation cassette into cassette 1 of the pST44 polycistronic expression vector. However, the same flanking restriction sites (shown in gray) for subcloning the translation cassette as EcoRI-HindIII (for cassette 2), SacI-KpnI (for cassette 3) and BspEI-MluI (for cassette 4) fragments are also available in pST50Trc1. This allows pST50Trc1 to function as the donor plasmid for all 4 cassettes if desired. All occurrences of each of the selected restriction sites in the plasmids are shown. Priming sites for sequencing or PCR primers are depicted as labeled arrows, with their sequences shown in (c). The β -lactamase gene present in each of the plasmids confers ampicillin resistance (not shown).

Figure 5.

Sample small scale purification of a two subunit complex. The HIS-tagged *A. thaliana* Gcn5 deletion construct was coexpressed with an Ada2 deletion construct. Equivalent volumes of the whole cell extract (after sonication), extract pellet, extract supernatant and Talon flow through (what did not bind to the Talon resin) are shown in lanes 1 through 4 respectively. In this experiment, 5 ml of the extract supernatant was incubated with 0.5 ml of washed Talon metal affinity resin. Note that cell lysis of this extract was incomplete given that significant amounts of all proteins were found in the pellet fraction. Since most *E. coli* proteins are soluble, one normally expects relatively little protein in the pellet fraction. Also note that the tagged Gcn5 and untagged Ada2 proteins bound efficiently to the Talon resin as judged by the relative absence of these proteins in the Talon flow through fraction. Elution using three 0.5 ml fractions of solution containing 100 mM imidazole releases the bound Ada2/Gcn5 complex, predominantly in fraction 2 (lanes 6 through 8). The asterisk marks a known *E. coli* contaminant, EF-Tu, found in many of our preparations of HIS-tagged protein complexes purified over Talon resin (Barrios et al., 2007). Molecular weight markers are shown in lane 8, with their molecular weights in kilodaltons shown to the right of the gel.

Figure 6.

Sample small scale purification of a three subunit complex. A HIS-tagged *S. cerevisiae* Epl1 deletion construct was coexpressed with full length Esa1 and a Yng2 deletion construct from the same species to produce the 3 subunit Piccolo NuA4 complex. Equivalent volumes of the whole cell extract (after sonication), extract pellet, extract supernatant and Talon flow through (what did not bind to the Talon resin) are shown in lanes 1 through 4 respectively. 5 ml of the extract supernatant was incubated with 0.5 ml of Talon metal affinity resin. In this example, efficient cell lysis occurred since relatively little protein was present in the pellet fraction. Although less obvious than in the example in Fig. 5, the three coexpressed subunits bound efficiently to the Talon resin as judged by the relative absence of these proteins in the Talon flow through fraction. Elution using three 0.5 ml fractions of solution containing 100 mM imidazole releases the bound Epl1/Yng2/Esa1 Piccolo NuA4 complex, predominantly in fraction 2 (lanes 6 through 8). The asterisk marks the EFTu contaminating *E. coli* protein (Barrios et al., 2007). Molecular weight markers are shown in lane 8, with their molecular weights in kilodaltons shown to the right of the gel.

Table 1

Polycistronic expression vector worksheet

		component	1	2	3	4
		gene name				
pST50TrecL-4		nonfusion 5'	NdeI			
		fusion 5'	BamHI			
		fusion 3'	NgoMIV			
		nonfusion 3'	BsrGI			
pST44 cassette 1		5' site	XbaI			
		alt 5' site	NheI			
		blunt site	SmaI			
		3' site	BglII			
pST44 cassette 2		5' site	EcoRI			
		alt 5' site	MfeI			
		blunt site	SmaI			
		3' site	HindIII			
pST44 cassette 3		5' site	SacI			
		blunt site	EcoRV			
		3' site	KpnI			
pST44 cassette 4		5' site	BspEI			
		blunt site	NruI			
		3' site	MluI			
		alt 3' site	BssHII			

Table 2

Recommended double digestion conditions (for New England Biolabs enzymes)

enzymes to use	digest 1	digest 2
NdeI & BamHI	digest with NdeI & BamHI in NEBuffer BamHI for 2 hours	
NdeI & NgoMIV	digest with NdeI & NgoMIV in NEBuffer 4 for 2 hours	
NdeI & BsrGI	digest with NdeI & BsrGI in NEBuffer 2 for 2 hours	
BamHI & NgoMIV	digest with NgoMIV in NEBuffer 4 for 90 min.	add 150 mM NaCl and BamHI for 90 min.
BamHI & BsrGI	digest with BsrGI in NEBuffer 2 for 90 min.	add 100 mM NaCl and BamHI for 90 min.
NgoMIV & BsrGI	digest with NgoMIV & BsrGI in NEBuffer 4 for 2 hours	
XbaI & BglII	digest with XbaI in NEBuffer 2 for 90 min.	add 50 mM NaCl and BglII for 90 min.
EcoRI & HindIII	digest with EcoRI & HindIII in NEBuffer EcoRI for 2 hours	
SacI & KpnI	digest with SacI & KpnI in NEBuffer 1 for 2 hours	
BspEI & MluI	digest with BspEI & MluI in NEBuffer 3 for 2 hours	