A Modular Polycistronic Expression System for Overexpressing Protein Complexes in *Escherichia coli*

Song Tan
Center for Gene Regulation, Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, Pennsylvania 16802-1014

Received September 14, 2000, and in revised form October 17, 2000

To facilitate studies of multicomponent protein complexes, I have developed an *Escherichia coli* expression system which coexpresses up to four polypeptides from a single plasmid. The modular nature of the system enables efficient subcloning of a gene into each of the 4 cassettes in the polycistronic expression vector. Restriction sites present in the polycistronic expression vector allow both affinity tagged and untagged complexes to be overexpressed. I demonstrate successful use of the expression system for binary and ternary complexes, including the reconstitution of the VHL-elonginC-elonginB complex in *E. coli* and purification of the complex by affinity and ion-exchange chromatography. This polycistronic expression system should provide an important alternative to in vitro reconstitution of multicomponent complexes.

Many cellular processes such as gene regulation, cell-cycle control, and metabolism require enzymatic or regulatory multicomponent protein complexes. Isolation of such complexes for biochemical and biophysical studies can be challenging, especially for regulatory complexes present in only scarce quantities. One method used to obtain micro- to milligram quantities of such complexes is to overexpress and purify recombinant versions of each component and then to reconstitute the complex in vitro. Examples of successful application of this procedure include the yeast TFIIA/TBP1/DNA complex, the nucleosome and *Escherichia coli* RNA polymerase, each of which required reconstituting four individually overexpressed and purified polypeptides (8, 19, 21). Tens of milligrams of the TFIIA/TBP/DNA complex and nucleosomes were prepared in this way for their successful crystallization and structural determination (7, 19).

Although such examples demonstrate that in vitro reconstitution of multicomponent complexes can succeed, reconstitution yields can be very low despite tedious optimization of many refolding parameters. An attractive alternative to in vitro reconstitution is in vivo reconstitution by coexpression because individual components of a complex which might not fold properly when expressed individually may fold together into a complex when coexpressed in a cellular environment. Molecular chaperones present in a cell may enhance folding of the coexpressed polypeptides into a complex by preventing aggregation of nascent unfolded polypeptides (1, 5). Coexpression also offers improved efficiency compared to in vitro reconstitution, which typically requires creating expression vectors for each component, expressing and purifying each component, reconstituting the complex and purifying the reconstituted complex from other misfolded products. In contrast, coexpression can be simpler once the coexpression plasmid(s) has/have been created because the complex reconstituted in the cell needs to be purified only once. Successful coexpression experiments have included the myosin heavy chain and light chain polypeptides which were assembled into a complex by coexpressing either from separate plasmids or from a single plasmid with separate promoters for both genes (10). Several TAFs (TBP-associated factors) which associate via histone-fold interactions have been produced in soluble form by coexpressing individual TAF polypeptides from two separate expression plasmids (2). Finally, a ternary complex of VHL-elonginC-elonginB associated with von Hippel-Lindau cancer syndrome was reconstituted in milligram quantities by coexpressing VHL and elonginB genes from a dicistronic message on one plasmid together with elonginC expressed from a separate plasmid in *E. coli* (15).

The VHL-elonginC-elonginB complex demon-

---

1 Abbreviations used: TBP, TATA-binding protein; TAF, TBP-associated factor; GST, glutathione; MBP, maltose-binding protein; TEV, tobacco etch virus; IPTG, isopropylthiogalactosidase.
strates some of the problems encountered when coexpressing more than two genes in E. coli. It is difficult to maintain more than two plasmids in E. coli even if different antibiotics are used to select the plasmids since each plasmid must have a unique but compatible origin of replication. Thus, it is problematic to coexpress more than two genes if each plasmid expresses only one gene. A general expression vector capable of coexpressing several genes at once would be useful for simultaneously overexpressing and reconstituting multicomponent complexes.

I have constructed a modular polycistronic expression system to coexpress up to four genes in E. coli from a single expression plasmid. The system is composed of the polycistronic expression vector pST39 and a transfer vector pET3aTr which facilitates subcloning of component genes into pST39. I demonstrate successful use of this polycistronic expression system to produce several binary and ternary complexes, including the efficient purification of these complexes from crude extracts by affinity chromatography by means of a fusion tag.

EXPERIMENTAL PROCEDURES

DNA Constructs

The low copy number cloning vector pST35 (Table 2) was created by amplifying the ORI region of pET3a with the primers STO753 (CGGCCCGGTGGCACAACCACCTCATACTCTGATATTACGGA) and STO754 (CCGGATCCGGGCGTTAGAAAAAGATCAAAAGATCTCTTC) and subcloning this region as a BanII–RsrII fragment into pWM521 digested with the same restriction enzymes. The T7 transcription/translation cassette from pET3a was amplified using the primers STO761 (CCGAACTATCTGATATTACGGA) and STO762 (CCCTGCAAGATCGATCCGCGAATT) and STO762 (CCCTGCAAGATCGATCCGCGAATT) and then subcloned as a BanII–SpeI fragment into BanII–SpeI pST35. The resulting plasmid, pST36, is a low-copy T7 expression plasmid based on the pWM521 backbone but containing the pET3a origin of replication region. The first polycistronic expression plasmid, pST37, was constructed by ligating synthetic oligonucleotides which create the four empty cassettes into XbaI–BamHI pST36 such that the 5' XbaI site was retained, but the 3' BamHI site was eliminated. pST38 adds an additional base pair between the Stul and the BamHI sites to remove a dcm methylation site that overlaps the Stul site and prevents digestion with Stul when plasmid DNA is prepared from a dcm+ strain. The final polycistronic expression plasmid, pST39, was created in several steps. First, an XbaI–BglII fragment from pRET3a–HisTrxNYTAF17 (unpublished data) containing the translation cassette for the fusion polypeptide, HisTrxNYTAF17, with BsrGI and Apal sites just downstream of the STOP codon, was ligated into XbaI–BamHI pST38 to create pST39–HisTrxNYTAF17. The HisTrxNYTAF17 region was then removed by ligating a synthetic XbaI–BsrGI linker into XbaI–BsrGI pST39–HisTrxNYTAF17 to create pST39. Similarly, the pET3aTr–yGcn5 plasmid was constructed by amplifying a yGcn5 translation cassette with primers that append the appropriate restriction sites to either side of the cassette, and then ligating the SpeI–BglII PCR product into XbaI–BamHI pET3a (the original XbaI and BamHI sites of pET3a are eliminated in the process, leaving the XbaI and BamHI sites included in the primer as the unique XbaI and BamHI sites). pET3aTr was then created by ligating an NdeI–BamHI linker between the NdeI and BamHI sites of pET3aTr–yGcn5, thereby removing the yGcn5 coding region.

The genes for VHL, elonginB and elonginC were cloned from HeLa cDNA by PCR amplification, resulting in pST38–HisTrxNYVHL, pET3aTr–elonginB, and pET3aTr–elonginC. The dual expression vector, pST38–HisTrxNYVHL–elonginB was created by subcloning the EcoRI–HindIII elonginB cassette from pET3aTr–elonginB into EcoRI–HindIII pST38–HisTrxNYVHL. The Sad–KpnI elonginC cassette from pET3aTr–elonginC was then subcloned into pST38–HisTrxNYVHL–elonginB to create pST38–HisTrxNYVHL–elonginB–elonginC.

The genes for yTAF17 and yTAF60 amplified from yeast genomic DNA by PCR amplification were cloned into the pWM529 cloning vector (9). The internal BspEI, MluI, and SacI sites of yTAF17 and the internal HindIII site of yTAF60 were removed in a translationally silent manner by QuikChange site-directed mutagenesis (Stratagene). These genes for yTAF17 and yTAF60 were then subcloned to create pST39–HisTrxNYTAF17 and pET3aTr–yTAF60. The pST39–HisTrxNYTAF17–yTAF60 coexpression vector was constructed by subcloning the EcoRI–HindIII yTAF60 cassette from pET3aTr–yTAF60 into pST39–HisTrxNYTAF17. The coding regions of all expression constructs, and the transcriptional and translational control regions of pET3aTr and pST39 were confirmed by DNA sequencing.

Protein Expression and Purification

Individual proteins and protein complexes were overexpressed in BL21(DE3)pLysS cells (16) containing appropriate monocistronic or polycistronic expression plasmids. Expression strains between OD600 0.4 and 0.7 are typically induced at 37 or 28°C with 0.2 mM IPTG for 2–4 h. For solubility studies, 50 ml of induced cells were harvested by centrifugation, resuspended in 10 ml of P100 buffer (50 mM sodium phosphate, pH 7.0, 100 mM NaCl, 1 mM benzamidine, and
5 mM 2-mercaptoethanol) and stored frozen at −20°C. The thawed and partially lysed sample was then sonicated and centrifuged to isolate the supernatant and pellet fractions (the pellet was resuspended in the same volume as the supernatant fraction to provide equivalent volumes for SDS–PAGE).

The VHL–elonginB–elonginC complex was purified from 1 L of BL21(DE3)pLysS cells containing the pST38–HisTrxNVHL–elonginB–elonginC polycistronic expression plasmid (induced at 37°C for 3.5 h). The soluble extract was passed over a 20 mL column of Talon cobalt-affinity resin (Clontech) equilibrated in P100 buffer (50 mM sodium phosphate, pH 7.0, 100 mM NaCl, 1 mM benzamidine, and 5 mM 2-mercaptoethanol). The column was washed with P100 buffer and P100 buffer containing 10 mM imidazole before eluting with P100 buffer containing 100 mM imidazole. This Talon pool was then incubated with the site-specific tobacco etch virus Nla protease to remove the N-terminal HisTrxN fusion tag on the VHL polypeptide. The untagged VHL–elonginB–elonginC complex was further purified by anion-exchange chromatography using a 1-mL SourceQ (Pharmacia) column. Since the HisTrxN polypeptide coelutes with the VHL–elonginB–elonginC complex by anion-exchange chromatography, the SourceQ pool was incubated with Talon resin to remove the N-terminally hexahistidine tagged HisTrxN polypeptide, yielding apparently homogeneous VHL–elonginB–elonginC complex. At least 3 mg of the complex can be isolated from 1 L of cells.

**RESULTS**

**Design Considerations**

A typical expression vector contains a promoter directing transcription of a single coding region including translational start signals (Fig. 1a). Many expression plasmids also include a transcriptional terminator to ensure efficient termination of the transcript. In contrast, an E. coli polycistronic expression vector should generate an RNA transcript containing tandem coding regions each preceded by its own translation start signal (ribosomal binding site). Thus, proteins A, B, C, and D should be coexpressed when this polycistronic expression vector is used.

**FIG. 1.** Comparison of monocistronic and polycistronic expression vectors. (a) A monocistronic expression vector typically includes transcriptional start (promoter) and stop (terminator) signals. These transcriptional control elements flank a translation cassette which contains the coding region preceded by a translational start signal. (b) A polycistronic expression vector also contains transcriptional start and stop signals, but these flank multiple translation cassettes. In this example, the transcript produced will contain four coding regions, each preceded by its own translation start signal (ribosomal binding site). Thus, proteins A, B, C, and D should be coexpressed when this polycistronic expression vector is used.
luation cassette should have a Shine-Dalgarno site, a START codon and a translational STOP codon),

(d) the translational enhancer element present upstream of the Shine-Dalgarno sequence in the pET3a vector should be maintained to ensure high levels of translation (11, 12),

(e) the gene for each polypeptide should be easily subcloned from an existing pET3a-based expression vector into the each of the four cassettes in the polycistronic expression vector using asymmetric cloning where possible,

(f) the expression vector should allow for simple subcloning of native (nonfusion) as well as tagged polypeptides to permit affinity purification of the complex,

(g) the restriction enzymes used for asymmetric sticky-end subcloning should share similar restriction conditions (temperature and buffer requirements),

(h) the expression vector should not contain dam or dcm methylation sites in the cloning sites which could interfere with restriction enzyme digestion,

(i) blunt-end restriction sites should be available for subcloning flexibility,

(j) unique priming sites should be available on either side of each of the four translation cassettes to facilitate sequencing and PCR screening of colonies,

(k) the size of the expression vector should be kept to a minimum to aid plasmid stability,

(l) the expression vector should contain few restriction sites outside of the designed subcloning sites,

(m) the expression vector should contain few restriction sites outside of the designed subcloning sites.

The pST39 polycistronic expression system is composed of two plasmids: pST39 and pET3aTr. pST39 is the expression vector containing four empty cassettes, while pET3aTr is a monocistronic expression vector that simplifies subcloning of component genes into the pST39 polycistronic expression vector.

Construction of the pST39 polycistronic expression plasmid was aided by the modular design of the pET3a expression vector on which pST39 is based (18). pET3a contains a translation cassette composed of T7 gene 10 translation initiation signals, including the translational enhancer and Shine-Dalgarno sequence, positioned before cloning sites for the gene to be overexpressed. This translation cassette is flanked by an upstream T7 p10 promoter and a downstream Tø transcriptional terminator (Fig. 2a). Since the translational cassette contains all the information necessary for its transcript to be translated, a polycistronic version of pET3a need only have several tandem copies of the translation cassette placed between the existing transcriptional start and stop signals of pET3a. However, to fulfill criteria (k) and (l), a small synthetic plasmid, pWM521, was used as the starting point for construction of the polycistronic expression vector instead of the pBR322 backbone of pET3a. pWM521 is a 2000-bp plasmid with similar properties to pUC9, on which it is conceptually based, except that pWM521 contains only half of the restriction sites found in pUC plasmids (9). The high copy number pWM521 plasmid was converted into a low copy number version, pST35, by replacing the mutant ColE1 origin of replication of pUC9 with the normal, low copy number ColE1 origin of replication of pBR322 as found in pET3a. The pST35 plasmid thus fulfills criteria (k), (l), and (m) and provides the backbone for the pST39 polycistronic expression plasmid.

Constructing a polycistronic expression vector in pST39 requires creating and subcloning up to four different translation cassettes. Clearly each translation cassette in pST39 needs to be flanked by unique subcloning sites. If a gene were already subcloned into pET3a, PCR amplification of that translation cassette with primers that incorporate the appropriate restriction sites would allow transfer of the translation cassette into pST39. However, any in vitro replication such as PCR amplification carries the risk of introducing undesired mutations due to reduced proof-reading activity. To circumvent such concerns, a variant of pET3a, pET3aTr (for pET3a transfer), was constructed which facilitates the transfer of any gene into pST39. pET3aTr contains appropriate restriction sites flanking the pET3a translation cassette (Figs. 2b and 3a). If the gene is already subcloned into pET3a, the entire translation cassette containing the gene can be subcloned into pET3aTr (using NdeI–BamHI or XbaI–BamHI sites, for example). Alternatively, the gene can be subcloned directly into pET3aTr. The translation cassette containing the gene of interest can then be excised from pET3aTr with restriction sites which allow subsequent subcloning into each of the cassettes of the polycistronic expression vector, pST39. Besides acting as a source of translational cassettes for subcloning into pST39, pET3aTr is itself a monocistronic expression vector since all the transcriptional and translational elements in pET3a remain in pET3aTr.

pST39 contains four subcloning cassettes, each of which consists of a 5’ subcloning site (XbaI, EcoRI, Sad and BspE1 in the four cassettes), a 3’ subcloning site (BamHI, HindIII, KpnI, MluI), and a blunt restriction site (StuI, Smal, EcoRV, NruI) between these 5’ and 3’ subcloning sites (Figs. 2c and 3b). The translation cassette for geneX in pET3aTr-geneX can be subcloned into the second cassette, EcoRI and HindIII would usually be used, while Sad–KpnI and
BspEI–MluI would be used for the third and fourth cassettes, respectively (Fig. 4). These restriction enzyme pairs were selected to be compatible in the appropriate restriction enzyme buffer of New England Biolabs, and to favor simple subcloning with low number background colonies which result from incomplete digestion and consequent religation of the vector fragment (criteria (e) and (g)).

One potential problem is that the genes of interest, particularly larger genes, might contain internal restriction sites for these subcloning sites. Therefore, alternate restriction sites with compatible ends are provided in pET3aTr wherever possible. These alternate restriction sites are NheI (XbaI), MfeI (EcoRI), NgoMIV (BspEI), BglII (BamHI), and BssHII (MluI) where the restriction site being replaced is shown in parentheses (Figs. 2b and 3a). Thus, if geneX contains an internal XbaI site, it can still be excised from pET3aTr-geneX with NheI and BamHI and subcloned into the XbaI–BamHI pST39 vector since XbaI and NheI share the same sticky ends (assuming that geneX does not also contain an NheI site). For other situations where restriction sites internal to geneX prevent use of the sticky-ended subcloning sites, blunt end restriction sites (StuI, SamI, EcoRV, and NruI) are available to allow the subcloning of almost any fragment by blunt-end ligation (criteria (i)). Alternatively, site-directed mutagenesis can be employed to eliminate undesirable internal restriction sites in geneX.

To permit affinity purification of coexpressed proteins which form a complex, cassette 1 of pST39 includes additional restriction sites to simplify subcloning of constructs containing fusion tags. For example, N-terminal glutathione S-transferase (GST), maltose-binding protein (MBP), hexahistidine, or thioredoxin tags can be introduced into cassette 1 as XbaI–BamHI translational cassettes (or NdeI–BamHI fragments containing only the coding region if a pST39 expression vector containing the XbaI–NdeI translational signals is already available) (Figs. 2b and 2c). In this case, the
gene to be fused to the N-terminal tag can be subcloned into cassette 1 as a BamHI–BsrGI or BamHI–ApaI fragment (Figs. 2c and 3b).

The pST39/pET3aTr combination has been designed to allow one transfer vector, the pET3aTr construct, to act as the source of the insert no matter which of the four cassette is to be used. This modularity allows great flexibility in the order of expression of the (up to) four polypeptides. However, this design also requires that the genes be subcloned into the four cassettes in the order cassette 1 cassette 2 cassette 3 cassette 4 since each successive translation cassette from pET3aTr will contain the restriction sites used for the preceding cassette. Thus, if a SacI–KpnI translation cassette from pET3aTr-gene1 is subcloned into pST39 first (into cassette 3), the XbaI–BamHI sites of cassette 1 or the EcoRI–HindIII sites of cassette 2 cannot be used since all these sites are present within the SacI–KpnI insert in cassette 3. In such situations where such out-of-order subcloning might be desirable, it may be possible to insert the appropriate restriction sites using PCR amplification with primers which incorporate those sites.

To facilitate sequencing and PCR screening of colonies, unique sequencing and PCR priming sites flank each translation cassette (Fig. 2c). Commonly used sequencing/PCR primers available in many laboratories were selected for this purpose. The oligonucleotide sequences which anneal to these sites are presented in Table 1.

To facilitate sequencing and PCR screening of colonies, unique sequencing and PCR priming sites flank each translation cassette (Fig. 2c). Commonly used sequencing/PCR primers available in many laboratories were selected for this purpose. The oligonucleotide sequences which anneal to these sites are presented in Table 1.

Use of the Polycistrionic Expression Vector

I have used the pST39 expression vector and the closely related pST38 vector, which lacks the Apal and BsrGI sites just downstream of the first cassette (Table 2), to express two different eukaryotic complexes in E. coli. I overexpressed the VHL–elonginC–elonginB complex since Stebbins et al. showed that this complex can reconstitute in vivo when coexpressed, albeit from two different expression vectors (15). Aside from changing the fusion tag from GST to HisTrxN (hexahistidine-tagged E. coli thioredoxin followed by tobacco etch virus (TEV) N1a protease site to allow site-specific cleavage of the fusion polypeptide), the polypeptides expressed were the same as Stebbins et al. I also tested the expression of the intermediate construct, pST38–HisTrxNVHL–elonginB. Figure 5a shows that polypeptides with sizes corresponding to HisTrxNVHL and elonginB are induced in cells containing the pST38–HisTrxNVHL–elonginB expression plasmid. However, both these polypeptides fail to fold properly and are found in the insoluble part of the cell extract. However, when the pST38–HisTrxNVHL–elonginB–elonginC plasmid is used instead, all three polypeptides with sizes corresponding to HisTrxNVHL, elonginB, and elonginC are induced and significant quantities of all three polypeptides are found in the soluble portion of the cell extract (Fig. 5b, lanes 1–4). ElonginB appears to be expressed in limiting amounts since excess HisTrxNVHL and elonginC are found in
the insoluble part of the extract (Fig. 5b, lane 3). The ternary HisTrxNVHL–elonginC–elonginB can be purified from the crude extract by metal affinity chromatography over Talon resin (Fig. 5b, lanes 4 and 5). Subsequent treatment with the site-specific TEV N1a protease and ion-exchange/metal affinity chromatography produces purified VHL–elonginC–elonginB complex (Fig. 5b, lanes 6 and 7).

I have also used the pST39 expression vector to coexpress the yeast TAF17/TAF60 complex in E. coli. Yeast TAF17 and TAF60 are homologous to Drosophila TAF42 and TAF62, respectively, and these two Drosophila TAFs have been shown to form a heterotetrameric complex via histone–fold interactions (20). When the fusion HisTrxN–yTAF17 is expressed on its own in a T7 expression system, a very high level of overexpression is observed, but essentially all of the protein is insoluble when expressed at either 37 or 28°C (Fig. 6a, lanes 1–6). Similar results are observed for nonfusion expression of yTAF17 (data

**Table 1**

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>T7</td>
<td>TAATACGACTCTATAGGG</td>
<td>20</td>
</tr>
<tr>
<td>M13 univ</td>
<td>TGTTAAAACGACGCCG</td>
<td>18</td>
</tr>
<tr>
<td>M13−47</td>
<td>CGCCAGGTTTTCCGTC</td>
<td>19</td>
</tr>
<tr>
<td>M13 rev−48</td>
<td>AGCGGATAACAATTTCACA</td>
<td>19</td>
</tr>
<tr>
<td>STO720</td>
<td>GCTAGTTATTGGCTAGCGG</td>
<td>19</td>
</tr>
<tr>
<td>STO767</td>
<td>ACTGGCCGTCGTTTACA</td>
<td>18</td>
</tr>
<tr>
<td>STO768</td>
<td>GACTGGGAAAACCTTCGGG</td>
<td>19</td>
</tr>
<tr>
<td>STO769</td>
<td>TGTGAAATTATTCTCCGCT</td>
<td>19</td>
</tr>
</tbody>
</table>
It is likely that this very high level of expression is possible because yTAF17 fails to fold on its own and is protected from cellular proteases when sequestered in inclusion bodies. However, when E. coli cells containing pST39–HisTrxNyTAF17–yTAF60 are induced, modest expression of 32- and 60-kDa polypeptides that correspond to HisTrxN–yTAF17 and yTAF60 is detected, and both of these polypeptides are found predominantly in the soluble extract when expressed at either 37 and 28°C (some coexpressed HisTrxN–yTAF17 remains in the pellet fraction when induction is performed at 37°C) (Fig. 6a, lanes 8–15). These two polypeptides also coelute from a metal affinity resin, and the complex can be further purified by ion-exchange chromatography (Fig. 6b). Although much less HisTrxN–yTAF17 is produced when coexpressed with yTAF60 compared to on its own (presumably because the complex is soluble and excess complex is either degraded in the cell or is toxic to the cell), the coexpressed HisTrxNyTAF17–yTAF60 complex can be purified in milligram quantities. Furthermore, this purified material elutes in a single peak by size exclusion chromatography with an elution volume consistent with a heterotetrameric complex (data not shown). Thus, coexpression appears to produce a properly folded yTAF17–yTAF60 complex.

**DISCUSSION**

In vivo reconstitution by coexpression is based on the premise that individual polypeptides of a multicompo-
nent complex might self-assemble properly in a heterologous expression system. Since certain complexes might require chaperones, cofactors, or particularly oxidizing or reducing environments, this premise will not always be valid. However, the variety of complexes successfully coexpressed to date suggests that many complexes may self-assemble in E. coli (3, 4, 6, 13, 14). The pST39 expression system provides an efficient way of constructing polycistronic expression vectors for in vivo reconstitution of up to four polypeptides.

The utility of the pST39 polycistronic expression system has been demonstrated with the successful in vivo reconstitution of the ternary VHL–elonginB–elonginC and the binary yeast TAF17/60 complexes. The VHL–elonginB–elonginC experiment also showed that significantly different levels of overexpression were possible for different genes in the same polycistronic vector since higher levels of VHL and elonginC expression are detected than for elonginB (Fig. 5b, lane 2). This does not simply reflect incomplete transcription by RNA polymerase or a failure of the ribosome to reinitiate translation since elonginC is further downstream of the promoter than elonginB and yet appears to be better expressed than elonginB (compare Fig. 5a, lane 1, and Fig. 5b, lane 2). It is worth noting that the polypeptides in excess (VHL and elonginC) were found in the insoluble part of an extract. This is consistent with the crystal structure, which suggests that the

FIG. 6. Expression of HisTrxN–yTAF17 and coexpression of HisTrxN–yTAF17–yTAF60. (a) Expression and solubility of HisTrxN–yTAF17 alone or coexpressed with yTAF60. The expression of HisTrxN–yTAF17 in BL21(DE3)pLysS cells is shown in lanes 1–3 (37°C) and 4–6 (28°C), while the coexpression of HisTrxN–yTAF17 with yTAF60 is shown in lanes 8–11 (37°C) and 12–15 (28°C). Whole cell extracts of uninduced samples are shown in lanes 8 and 12, while whole cell extracts of induced samples are shown in lanes 1, 4, 9, and 12. The supernatant fractions after sonication are shown in lanes 2, 5, 10, and 14, while the pellet fractions after sonication are shown in lanes 3, 6, 11, and 15. Molecular weight markers are shown in lane 7. The position of the fusion HisTrxN–yTAF17 and the nonfusion yTAF60 are indicated by arrows. (b) Purification of the yTAF17–60 complex. Total cell extracts of BL21(DE3)pLysS cells containing pST39–HisTrxNyTAF17–yTAF60 before and after induction at 37°C are shown in lanes 1 and 2, respectively. The cobalt affinity chromatography pool before and after digestion with TEV Nla protease are shown in lanes 3 and 4, respectively. The final purified yTAF17/60 complex after anion-exchange chromatography is shown in lane 5.
VHL–elonginB–elonginC complex is an obligatory complex (i.e., one in which the individual components cannot fold independently) with predominantly hydrophobic interactions between the subunits. In general, many polypeptides can be expressed only in an insoluble form in E. coli and a significant proportion of these might similarly be components of obligatory complexes. If true, this would also suggest that soluble overexpression might be possible if such polypeptides were coexpressed with their cognate partner polypeptides.

One major advantage of the pST39 polycistronic expression system over previous polycistronic plasmids developed ad hoc to express particular pairs or groups of polypeptides is the modular nature of this system. Once each component gene has been subcloned into pET3aTr, it is simple to construct a polycistronic expression vector. It is also simple to alter the order in which the genes are expressed along the vector, an important consideration since the order of expression may play a role in how efficiently a complex is constituted in E. coli (W. Selleck and S. Tan, unpublished observations). The modular nature of the system will also assist in dissecting the interactions between components of large multicomponent enzyme or regulatory protein complexes. Different combinations of component polypeptides can be coexpressed by constructing appropriate polycistronic expression vectors. Since the same pET3aTr vector provides the translation cassette for each of the four positions in pST39, constructing a new expression vector to coexpress a new combination of polypeptides can be performed quickly and easily once the individual pET3aTr vectors have been created. For example, we have used the pST39 system to examine how the SAGA transcriptional coactivator components Ada2, Ada3, and Gcn5 interact to form a catalytic subcomplex (Balasubramanian, R., et al., submitted). This modular design using a transfer vector in combination with a polycistronic expression vector should be applicable to other bacterial or eukaryotic expression system that permits polycistronic expression.

One limitation of the pST39 polycistronic expression system is the maximum of four genes that can be coexpressed together. This limitation was imposed arbitrarily and can be lifted by incorporating additional restriction sites in the transfer and polycistronic plasmids. However, this might not be the most efficient method of overexpressing many more than four genes in E. coli, especially if the genes are large since a very long transcript would have to be produced by T7 RNA polymerase. I am currently investigating the use of a second polycistronic expression containing a compatible origin of replication and which itself could coexpress up to four genes, to increase the number of coexpressed genes to 8. Expression systems like these may permit us to reconstitute sufficient quantities of complicated protein complexes for biochemical and biophysical studies.

ACKNOWLEDGMENTS

I thank Will Selleck for preparing Fig. 6b. I am also very grateful to the Tan Laboratory and the Penn State gene regulation community as well as Arne Wörn, Peter Hackel, and Tim Richmond for helpful discussions.

REFERENCES


