Protein Expression

Prokaryotic Expression

pET System Tutorial

The premier E. coli expression system

The pET System is the most powerful system yet developed for the cloning and expression of recombinant proteins in *E. coli*. Based on the T7 promoter-driven system originally developed by Studier and colleagues (1–3), Novagen's pET System has been used to express thousands of different proteins. Please refer to pET System Overview earlier in this chapter for the basic features of the pET System. This section describes some of the unique characteristics of the system in greater detail.

Control Over Basal Expression Levels

The pET System provides six possible vector-host combinations that enable tuning of basal expression levels to optimize target gene expression (2). These options are necessary because no single strategy or condition is suitable for every target protein.

Host Strains

After plasmids are established in a non-expression host, they are most often transformed into a host bearing the T7 RNA polymerase gene (λDE3 lysogen) for expression of target proteins. Figure 1 illustrates in schematic form the host and vector elements available for control of T7 RNA polymerase levels and the subsequent transcription of a target gene in a pET vector. In λDE3 lysogens, the T7 RNA polymerase gene is under the control of the ***lacUV5*** promoter. This allows some degree of transcription in the uninduced state and in the absence of further controls is suitable for expression of many genes whose products have innocuous effects on host cell growth. For more stringent control, hosts carrying either pLysS or pLysE are available. The pLys plasmids encode T7 lysozyme, which is a natural inhibitor of T7 RNA polymerase, and thus reduces its ability to transcribe target genes in uninduced cells. pLysS hosts produce low amounts of T7 lysozyme, while pLysE hosts produce much more enzyme and, therefore, represent the most stringent control available in λDE3 lysogens (4).

Thirteen different host strains

Thirteen different host strains are available as λDE3 lysogens (see Table 1). The most widely used hosts are BL21 and its derivatives, which have the advantage of being deficient in both ***lon*** (5) and ***ompT*** proteases. The B834 strain is a methionine auxotroph and, therefore, enables high specific activity labeling of target proteins with ^{35}S-methionine or selenomethionine (6). The BLR strain is a ***recA***$^-$ derivative that improves plasmid monomer yields and may help stabilize target plasmids containing repetitive sequences. Two thioredoxin reductase (***trxB***) mutant strains (AD494, BL21***trxB***) are available that facilitate disulfide bond formation in the ***E. coli*** cytoplasm (7). The Origami™, Origami B

continued on next page

Figure 1. Control elements of the pET System

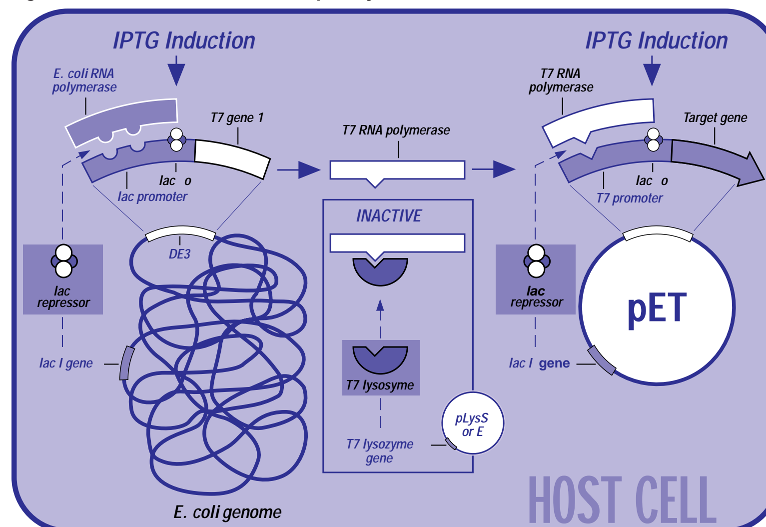


Table 1. pET System Host Strains

Strain	Derivation	Key Feature(s)	Antibiotic Resistance	Available as Competent Cells
AD494 AD494(DE3) AD494(DE3)pLysS	K-12	*trxB* mutant; facilitates cytoplasmic disulfide bond formation	Kan Kan Kan + Cam	yes yes yes
BL21 BL21(DE3) BL21(DE3)pLysS BL21(DE3)pLysE	B834	Lacks *lon* and *ompT* proteases	none none Cam Cam	yes yes yes no
BL21*trxB*(DE3) BL21*trxB*(DE3)pLysS	BL21	BL21 *trxB* mutant; facilitates cytoplasmic disulfide bond formation	Kan Kan + Cam	yes yes
BLR BLR(DE3) BLR(DE3)pLysS	BL21	BL21 *recA* mutant; stabilizes tandem repeats	Tet Tet Tet + Cam	yes yes yes
B834 B834(DE3) B834(DE3)pLysS	B strain	met auxotroph; ^{35}S-met and selenomethionine labeling	none none Cam	no yes yes
HMS174 HMS174(DE3) HMS174(DE3)pLysS HMS174(DE3)pLysE	K-12	*recA* mutant Rif resistance	Rif Rif Rif + Cam Rif + Cam	yes yes yes no
NovaBlue NovaBlue(DE3)	K-12	*recA, endA, lacI^q* cloning, plasmid preps	Tet Tet	yes yes
Origami™ Origami(DE3) Origami B(DE3)pLysS	K-12	*trxB/gor* mutant, greatly facilitates cytoplasmic disulfide bond formation	Kan + Tet Kan + Tet Kan + Tet + Cam	yes yes yes
Origami B Origami(DE3) Origami B(DE3)pLysS	Tuner	BL21 *lacY* deletion, *trxB/gor* mutant, greatly facilitates cytoplasmic disulfide bond formation; allows precise control with IPTG	Kan + Tet Kan + Tet Kan + Tet + Cam Kan + Tet + Cam	yes yes yes yes
Rosetta™ Rosetta(DE3) Rosetta(DE3)pLysS	Tuner	Enhances expression of proteins having codons rarely used in *E. coli*, *lacY* deletion mutant	Cam Cam Cam	yes yes yes
Rosetta-gami™ Rosetta-gami(DE3) Rosetta-gami(DE3)pLysS	Origami	Enhances expression of proteins having codons rarely used in *E. coli*, *trxB/gor* mutant	Kan + Tet + Cam Kan + Tet + Cam Kan + Tet + Cam	yes yes yes
RosettaBlue™ RosettaBlue(DE3) RosettaBlue(DE3)pLysS	NovaBlue	Enhances expression of proteins having codons rarely used in *E. coli*, *recA, endA, lacI^q*	Tet + Cam Tet + Cam Tet + Cam	yes yes yes
Tuner™ Tuner(DE3) Tuner(DE3)pLysS	BL21	BL21 *lacY* deletion mutant allows precise control with IPTG	none none Cam	yes yes yes

pET System Tutorial *continued*

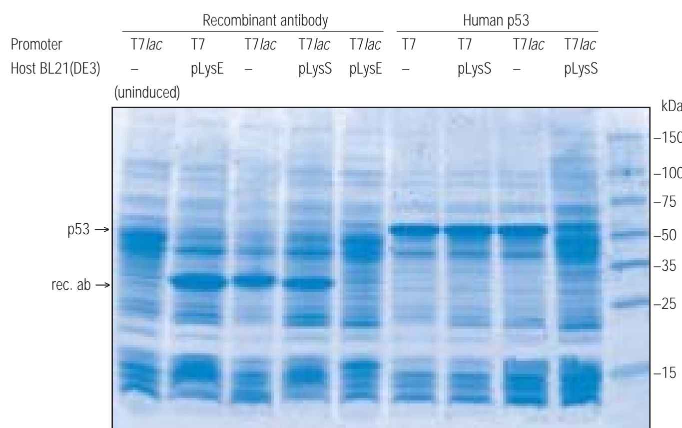


Figure 2. Effect of vector/host combination on expression levels of two proteins

The indicated cell cultures were grown at 37°C to OD_{600} of approximately 0.8 and expression induced with 1 mM IPTG for 2.5 h. Total cell protein samples were run along with Novagen's Perfect Protein™ Markers on a 4–20% SDS polyacrylamide gradient gel followed by staining with Coomassie blue. Vectors used were pET-20b(+) and pET-22b(+) for the recombinant antibody and pET-23b(+) and pET-21b(+) for p53.

and Rosetta-gami™ strains are double mutants of ***trxB/gor***, which are the key enzymes in both major reductive pathways (8). These hosts thus represent a significant advantage for the formation of properly folded disulfide-containing proteins. The Rosetta™, Rosetta-gami, and RosettaBlue™ strains supply the tRNAs for six codons used only rarely in *E. coli*, which alleviates poor expression caused by incompatible codon usage in some eukaryotic genes (9). The K-12 derivative strains HMS174, NovaBlue, and RosettaBlue are ***recA***$^-$, like BLR. These strains may stabilize certain target genes whose products may cause the loss of the DE3 prophage. NovaBlue and RosettaBlue are potentially useful as stringent hosts due to the presence of the high affinity ***lacI***^{q} repressor encoded by the F episome. In addition, Novagen offers the λDE3 Lysogenization Kit for making new expression hosts with other genetic backgrounds. An alternative for expressing extremely toxic genes or preparing a new λDE3 lysogen is to provide T7 RNA polymerase by infection with λCE6. Although not as convenient as inducing a λDE3 lysogen with IPTG, this strategy may be preferred for certain applications.

High Stringency T7*lac* Promoter

In addition to the choice of three basic expression stringencies at the host level, the pET system provides two different stringency options at the level of the T7 promoter itself: the "plain" T7 promoter and the T7*lac* promoter (8; also shown in Figure 1). The T7*lac* promoter contains a 25 bp *lac* operator sequence immediately downstream from the 17 bp promoter region. Binding of the *lac* repressor at this site effectively reduces transcription by T7 RNA polymerase, thus providing a second *lacI*-based mechanism (besides the repression at *lacUV5*) to suppress basal expression in λDE3 lysogens. pET plasmids with the T7*lac* promoter also carry their own copy of *lacI* to ensure that enough repressor is made to titrate all available operator sites.

In practice, it is usually worthwhile to test several different vector/host combinations to obtain the best possible yield of protein in its desired form (10). Figure 2 illustrates dramatic differences in the expression of two target proteins with various combinations.

Control Over Induced Expression Levels

In many cases the expression of optimal levels of active, soluble protein depends on host cell background, culture conditions, and vector configuration; often the conditions for highest activity of a target protein do not correlate with conditions that produce the highest mass of target protein. In addition to offering variable stringency based on vector/host combinations that provide control over basal expression of T7 RNA polymerase, the pET System offers precise control over target protein expression based on inducer (IPTG) concentration, made possible by the *lacY* mutation in the Tuner™, Rosetta, and Origami B host strains.

continued on next page

pET System Advantages

Powerful

- Highest expression levels, tightest control over basal expression
- Choice of vectors and host strains to control basal and induced expression levels
- Precise control of induced expression with IPTG in Tuner™ and Rosetta™ hosts
- Rare codons supplied by Rosetta hosts
- Origami™ strains for enhanced disulfide bond formation in the cytoplasm

Versatile

- Choice of N-terminal and C-terminal fusion tags for detection, purification and localization
- Expanded multiple cloning sites in all three reading frames
- f1 origin of replication for mutagenesis and sequencing

Rapid

- *E. coli*-based system for rapid results
- Convenient restriction sites for subcloning from other vectors
- Choice of methods for one-step purification of target proteins without antibodies

Complete

- Variety of system configurations plus many supporting products

1 2 3 4 5 6 7 8 9 10 11 Appendices Indices

Protein Expression

Prokaryotic Expression

pET System Tutorial *continued*

Choosing a pET Vector

A wide variety of pET vectors is available. All are derived from pBR322 and vary in leader sequences, expression signals, fusion tags, relevant restriction sites, and other features. There are two major categories of pET plasmids known as transcription vectors and translation vectors:

- The transcription vectors [including pET-21(+), pET-23(+), and pET-24(+)] express target RNA but do not provide translation signals. They are useful for expressing proteins from target genes that carry their own bacterial translation signals. ***Note that the transcription vectors can be identified by lack of a letter suffix after the name.***
- The translation vectors contain efficient translation initiation signals and are designed for protein expression. Most contain cloning sites in reading frames a, b, or c that correspond to the GGA, GAT, or ATC triplet of the *Bam*H I site, respectively.

Primary Considerations

Choosing a pET vector for expression usually involves a combination of factors. Consider the following three primary factors:

- The application intended for the expressed protein
- Specific information known about the expressed protein
- Cloning strategy

Applications for proteins expressed in pET vectors vary widely. For example, analytical amounts of a target protein may be needed for activity studies, screening and characterizing mutants, screening for ligand interactions, and antigen preparation. Large amounts of active protein may be required for structural studies, use as a reagent, or affinity matrix preparation. Any number of vectors may be suitable for expression of analytical amounts of protein for screening or antigen preparation, yet only one combination of vector, host strain, and culture conditions may work best for large scale purification. If a high yield of active protein is needed on a continual basis, it is worth testing a matrix of vector, host, and culture combinations to find the optimal result.

Any information available about the target protein may help determine the choice of vector. For example, some proteins require no extraneous sequence on one or both termini for activity. Most pET vectors enable cloning of unfused sequences; however, expression levels may be affected if a particular translation initiation sequence is not efficiently utilized in *E. coli*. In these cases, an alternative is to construct a fusion protein with efficiently expressed amino terminal sequences (available with many pET vectors) and then remove the fusion partner following purification by digestion with a site-specific protease. The LIC (ligation-independent cloning) strategy is especially useful for this approach, because the cloning procedure enables the removal of all amino terminal vector-encoded sequences with either enterokinase or Factor Xa .

Cloning strategies can affect the choice of vector due to the need for restriction site and reading frame compatibilities. Because many of the pET vectors share common restriction site configurations, it is usually possible to clone a target gene into several vectors with a single preparation of the insert. Different considerations apply when using PCR cloning strategies. The LIC Vector Kits are recommended for this purpose, and enable the preparation of inserts by PCR and eliminate the need for restriction digestion of vector or insert. The Appendix contains sequences of cloning and expression regions for the pET vectors.

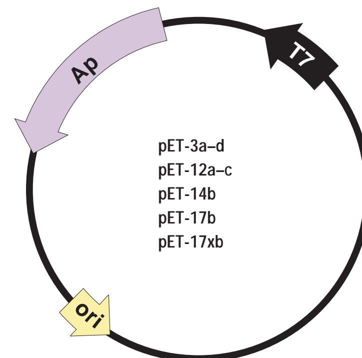
Solubility and Cellular Localization

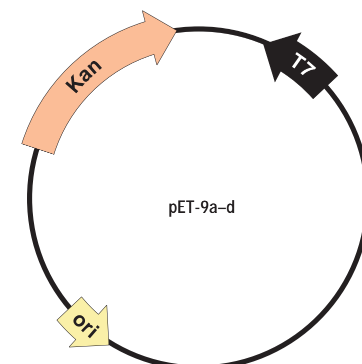
Once you have considered your application and cloning strategy, a good starting point for any expression project is to determine the cellular localization and solubility of the target protein. In many applications, it is desirable to express proteins in their soluble, active form.

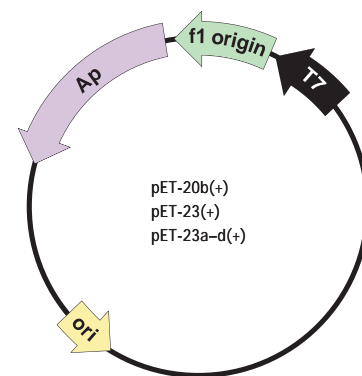
Solubility of a particular target protein is determined by a variety of factors, including the individual protein sequence. In most cases, solubility is not an all-or-none phenomenon; the vector, host, and culture conditions can be used to increase or decrease the proportion of soluble and insoluble forms obtained. The choice of vector and expression host can significantly increase the activity and amount of target protein present in the soluble fraction. A vector can enhance solubility and/or folding in three ways: 1, provide for fusion to a polypeptide that itself is highly soluble (e.g., GST, thioredoxin, NusA), 2, provide for fusion to an enzyme that catalyzes disulfide bond formation (e.g., thioredoxin, DsbA, DsbC), or 3, provide a signal sequence for translocation into the periplasmic space. When using vectors designed for cytoplasmic expression, folding can be improved in hosts that are permissive for the formation of disulfide bonds in the cytoplasm. The thioredoxin reductase (*trxB*) mutation has been shown to allow the formation of disulfide bonds in the *E. coli* cytoplasm, which is further enhanced by the additional mutation in the glutathione reductase (*gor*) gene in Origami™ and Rosetta-gami™ hosts (8). Induction at lower temperatures (15°–25°C) can also increase the proportion of soluble target proteins.

The pET-43.1 and pET-32 vectors incorporate fusion tags specifically designed to enhance the solubility of target proteins in the *E. coli*

continued on next page







pET vector backbones: "plain" T7 promoter vectors

The pET vectors shown here and on the facing page are grouped according to the functional elements present on the plasmid backbones. Features of the T7 cloning and expression regions (indicated here as "T7" and "T7*lac*") are shown on the following pages and in the Appendix. Note that the pET vector sequences are numbered by the pBR322 convention, so that the T7 expression region is reversed on these maps and in the published sequences.

pET System Tutorial *continued*

cytoplasm. These vectors are also compatible with *trxB* mutant hosts AD494 and BL21*trxB*, and with the *trxB*/*gor* mutant Origami™, Origami B, and Rosetta-gami™ strains. The pET-43.1a–c(+) vectors incorporate the 495 aa solubility-promoting NusA (Nus•Tag™) sequence, which was discovered through a systematic search for *E. coli* proteins that have the highest potential for solubility when overexpressed (11). Many proteins that are normally produced in an insoluble form in *E. coli* tend to become more soluble when fused with the 109 aa N-terminal thioredoxin (Trx•Tag™) sequence. The Trx•Tag expressed from pET-32 vectors not only enhances the solubility of many target proteins, but appears to catalyze the formation of disulfide bonds in the cytoplasm of *trxB* mutants (12).

Schistosomal glutathione-S-transferase (GST) is commonly used as an N-terminal fusion partner when expressing proteins in *E. coli*. Although not specifically designed for this purpose, the 220 aa GST•Tag™ sequence has been reported to enhance the solubility of its fusion partners. The pET-41 and -42 series of vectors encode the GST•Tag sequence driven by the powerful T7*lac* promoter. Note that these vectors carry kanamycin resistance so are not recommended for use with *trxB* mutant hosts.

An alternative strategy to obtain active, soluble proteins is to use vectors that enable export into the periplasm, which is a more favorable environment for folding and disulfide bond formation. For this purpose vectors carrying signal peptides are used. DsbA and DsbC are periplasmic enzymes that catalyze the formation and isomerization of disulfide bonds, respectively. The 208 aa DsbA•Tag™ [pET-39b(+)] and 236 aa DsbC•Tag™ [pET-40b(+)] vectors enable fusion of target polypeptides to these enzymes, which include their N-terminal secretion signals. If the fusion protein is exported to the periplasm, the Dsb partner can assist in proper disulfide bond formation. Other pET vectors that carry signal sequences without the additional DsbA or DsbC coding regions are also available.

Some purification strategies optimize production of insoluble inclusion bodies in the cytoplasm. Inclusion bodies are extracted and solubilized; then the target protein is refolded *in vitro*, (e.g., with Novagen's Protein Refolding Kit). This procedure usually produces the highest yields of initial protein mass and protects against proteolytic degradation in the host cell. However, the efficiency of refolding into active protein varies significantly with the individual protein and can be quite low. The pET-31b(+) vector is specifically designed for the generation of insoluble fusion proteins and provides a powerful method for the production of small proteins and peptides.

Fusion Tags for Different Needs

If a fusion sequence is tolerated by the application you are using, it is useful to produce fusion proteins carrying the S•Tag™, T7•Tag®, GST•Tag, His•Tag®, or HSV•Tag® peptides for easy detection on Western blots. These peptides are small in size (except for the GST•Tag sequence) and the detection reagents for them are extremely specific and sensitive. The His•Tag, GST•Tag, S•Tag, and T7•Tag sequences can also be used for affinity purification using the corresponding resins and buffer kits.

Fusion proteins can be accurately quantified in crude extracts or purified form using S•Tag and GST•Tag Assay Kits. The FRETWorks™ S•Tag Assay Kit is based on a novel substate that enables fluorescent detection of less than 1 fmol of fusion protein in a homogeneous format. (See Chapter 7, Protein & Gene Analysis.)

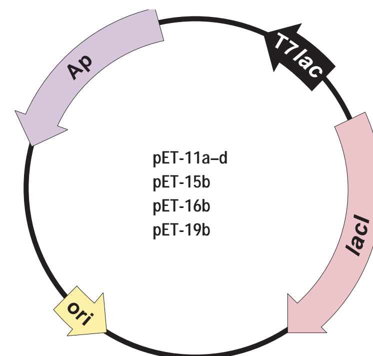
The His•Tag sequence is very useful as a fusion partner for purification of proteins in general. It is especially useful for those proteins initially expressed as inclusion bodies, because affinity purification can be accomplished under totally denaturing conditions that solubilize the protein.

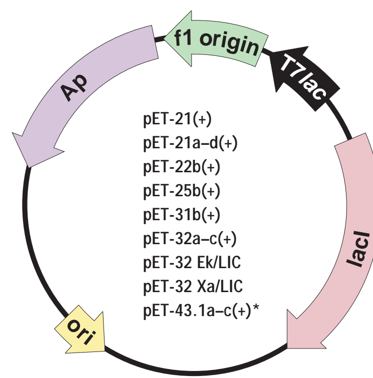
The CBD•Tag™ sequences are also generally useful for low cost affinity purification. They are also uniquely suited to refolding protocols [especially pET-34b(+) and 35b(+), which contain the CBD_{clos}•Tag sequence]. Because only properly refolded CBDs bind to the cellulose matrix, the CBIND™ affinity purification step can remove improperly folded molecules from the preparation. While any of the tags can be used to immobilize target proteins, the CBD•Tag sequences are ideally suited for this purpose due to the inherent low non-specific binding and biocompatibility of the cellulose matrix.

The Nus•Tag, Trx•Tag and GST•Tag sequences have been reported to enhance the solubility of their fusion partners. The Nus•Tag and Trx•Tag vectors are compatible with Origami, Origami B, and Rosetta-gami host strains, which facilitate disulfide bond formation in the cytoplasm.

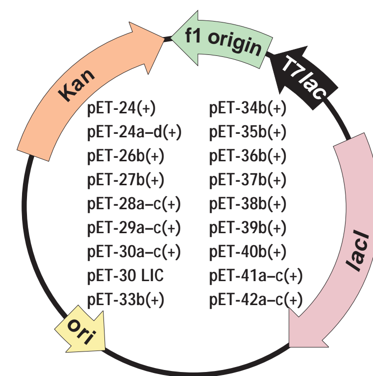
The various fusion tags available and their corresponding pET vectors are listed in the following table. A number of pET vectors encode several of the fusion tags in tandem as amino-terminal fusion partners. In addition, many vectors enable expression of fusion proteins carrying a peptide tag on each end by allowing in-frame read-through of the target gene sequence. Using vectors with protease cleavage sites (thrombin, Factor Xa, enterokinase) between the amino terminal tag and the target sequence enables optional removal of one or more tags following purification. Vectors that represent a

continued on next page





* Ap gene in opposite orientation in pET-43.1 series



pET vector backbones: T7lac promoter vectors

Protein Expression

Prokaryotic Expression

pET System Tutorial *continued*

good selection for cellular localization and affinity tag configurations are pET-30 Ek/LIC, pET-32 Ek/LIC, pET-41 Ek/LIC and pET-43.1 Ek/LIC. A single preparation of insert can be used with all of the ready-to-use Ek/LIC vectors to allow convenient construction of several target gene configurations at once.

How to Order

The pET Vectors are available configured into Expression Systems containing one or more vectors, plus glycerol stocks and competent cells of popular host strains for cloning and expression. Individual plasmids and cell strains are available separately.

1. Moffatt, B. A., and Studier, F. W. (1986) *J. Mol. Biol.* **189**, 113–130.
2. Rosenberg, A. H., Lade, B. N., Chui, D., Lin, S., Dunn, J. J., and Studier, F. W. (1987) *Gene* **56**, 125–135.
3. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) *Methods Enzymol.* **185**, 60–89.
4. Studier, F. W. (1991) *J. Mol. Biol.* **219**, 37–44.
5. Phillips, T. A., Van Bogelen, R. A., and Neidhardt, F. C. (1984) *J. Bacteriol.* **159**, 283–287.
6. Leahy et al. (1992) *Science* **258**, 987–991.
7. Derman, A. I., Prinz, W. A., Belin, D., and Beckwith, J. (1993) *Science* **262**, 1744–1747.
8. Dubendorff, J. W. and Studier, F. W. (1991) *J. Mol. Biol.* **219**, 45–59.
9. Novy, R., Drott, D. Yaeger, K., and Mierendorf, R. (2001) ***inNovations*** **12**, 1-3.
10. Mierendorf, R., Yaeger, K., and Novy, R. (1994) ***inNovations*** **1**, 1–3.
11. Harrison, R. (2000) ***inNovations*** **11**, 4–7
12. Stewart et al. (1998) *EMBO J.* **17**, 5543–5550.

Table 2. Fusion tags available for pET constructs

Tag	N/C Terminal or Internal	Size (aa)	Basis for Detection and/or Purification	Applications	pET Vector Series
T7•Tag®	N, I	11 or 260	monoclonal antibody	Western blot, immunoprecipitation, purification	3, 9, 11, 17 17x, 21, 23, 24, 28, 33
S•Tag™	N, I	15	S-protein (104 aa) affinity	Western blot, quantitative assay, purification	29, 30, 32, 34–37 39, 40, 41, 42, 43.1
His•Tag®	N, C, I	6, 8, or 10	metal chelation chromatography (native or denaturing)	His•Bind® resin purification	14-16, 19–43.1
HSV•Tag®	C	11	monoclonal antibody	Western blot, immunofluorescence	25, 27, 43.1
pelB/ompT	N	20/22	potential periplasmic localization	protein export/folding	12, 20, 22 25, 26, 27
KSI	N	125	highly expressed hydrophobic domain	small protein/peptide production/purification, insoluble protein	31
Trx•Tag™	N	109	thioredoxin	soluble protein, cytoplasmic disulfide bond formation in *trxB⁻* hosts	32
PKA site	N	5	protein kinase A recognition site	*in vitro* phosphorylation	33
CBD_{clos}•Tag	N	156	polyclonal antibody, cellulose binding domain	Western blot, purification, noncovalent immobilization	34, 35
CBD_{cenA}•Tag	N	114	polyclonal antibody, cellulose binding domain, periplasm/media	protein export, Western blot, purification, non-covalent immobilization	36, 37
CBD_{cex}•Tag	C	107	polyclonal antibody, cellulose binding domain, periplasm/media	protein export, Western blot, purification, non-covalent immobilization	38
Dsb•Tag™	N	208 (DsbA) 236 (DsbC)	potential periplasmic localization	soluble protein, periplasmic disulfide bond formation, isomerization	39, 40
GST•Tag™	N	220	glutathione affinity monoclonal antibody enzymatic activity	purification, Western blot, quantitative assay	41, 42
Nus•Tag™	N	495	NusA	soluble protein, cytoplasmic disulfide bond formation in *trxB⁻* hosts	43.1