HHS Public Access

Author manuscript

Curr Protoc Protein Sci. Author manuscript; available in PMC 2015 August 06.

Published in final edited form as:

Curr Protoc Protein Sci. ; 73: Unit–9.9. doi:10.1002/0471140864.ps0909s73.

Overview of Affinity Tags for Protein Purification

Michelle E. Kimple[1,*], **Allison L. Brill**[1], and **Renee L. Pasker**[1]

[1]Department of Medicine, Division of Endocrinology, Diabetes, and Metabolism, University of Wisconsin-Madison, Madison, WI 53705

EARLY AFFINITY TAGS

The first affinity tags used were large proteins utilized almost exclusively for protein expression and purification in *E. coli*. One of these tags, staphylococcal protein A, is 280 amino acids in length and, because of its relatively large size and proteolytic stability, can increase the solubility and/or expression of heterologous proteins (Sambrook et al., 1989). Protein A fusions can be purified by affinity chromatography on IgG Sepharose (K_d = 10 nM) and eluted with a low pH buffer (*UNIT 9.5*) or protease; however, the large size of the tag and/or the low pH elution can result in denaturation and loss of activity of the fusion protein, consequently altering its functional activity. Protease recovery can cleave the tagged protein itself, resulting in the same outcomes. In addition, the binding of fusion proteins to IgG complicates immunological analysis, and thus Protein A fusions are unsuitable for use as immunological reagents.

Recent advances in recovery technique have improved the ability to isolate Protein A and its associated protein complexes in its active form from IgG Sepharose. The use of Bio-Ox, a biotinylated form of the FcIII peptide, utilizes its high affinity towards IgG (K_d = 11 nM) (Strambio-de-Castillia et al., 2005). Bio-Ox competes with Protein A for binding to IgG, effectively eluting the recombinant protein. The addition of biotin increases the solubility of FcIII without affecting the binding affinity of the peptide itself (Strambio-de-Castillia et al., 2005), and thus is an ideal elution method for purifying protein A fusions in their native forms. Commercial systems are available from GE Healthcare Life Sciences (Table 9.9.1).

LacZ, also known as β-galactosidase or β-gal, is a proteolytically stable 1024–amino acid protein which can be used as an affinity tag to increase expression of fusion proteins in *E. coli* (Sambrook et al., 1989). LacZ fusion proteins can be purified by substrate affinity chromatography on immobilized *p*-amino-phenyl-β-D-thio-galactosidase (APTG) and eluted with a high pH borate buffer. Detection is then achieved using a colorimetric enzymatic assay. LacZ is relatively immunologically neutral and so it is a useful fusion partner for the generation of antibodies to a protein of interest. Because of its extremely large size and tendency to form homotetramers in solution, the LacZ affinity tag can alter the functional activity of the purified fusion protein. In addition, LacZ fusion proteins are often insoluble. This insolubility has advantages and disadvantages: protein targeting to inclusion bodies can

*Corresponding author: Michelle E Kimple, PhD, Department of Medicine, Division of Endocrinology, University of Wisconsin-Madison, 4148 UW Medical Foundation Centennial Building, 1685 Highland Ave., Madison, WI 53705, Phone: 608-616-0138, Fax: 608-263-9983, mkimple@medicine.wisc.edu.

allow the expression of gene products that are toxic to *E. coli*, but because LacZ must be correctly folded to bind APTG, LacZ fusion proteins must be refolded prior to purification and/or detection (*UNIT 6.5*).

COMMONLY USED AFFINITY TAGS

Although protein A and lacZ are still used as affinity tags today (see Table 9.9.1 for commercially available systems), many other affinity tags have recently been developed that improve upon these original protein fusion partners.

Polyhistidine

The polyhistidine affinity tag, also known as the His-tag or His_6, usually consists of six consecutive histidine residues, but can vary in length from two to ten histidine residues. Polyhistidine was first used to purify recombinant galactose dehydrogenase by immobilized metal affinity chromatography (IMAC; *UNIT 9.4*) in 1991 (Lilius et al., 1991). Previously, IMAC had been used to purify proteins with naturally occurring histidine- or tryptophan-containing sequences, but these sequences were not as generally applicable to protein expression and purification as would be desired for an affinity-tagging system (Smith et al., 1988; Ljungquist et al., 1989).

Polyhistidine is so ubiquitous an affinity tag that most companies providing expression vectors or protein expression and purification reagents offer products related to this tag (see Table 9.9.1 for examples). Histidine readily forms coordination bonds with immobilized transition metal ions. Immobilized Co^{2+}, Cu^{2+}, Ni^{2+}, Zn^{2+}, Ca^{2+}, and Fe^{3+} can all be used to purify polyhistidine fusion proteins, but Ni^{2+} is the most commonly used. If purification by Ni^{2+} is unsatisfactory, empirical determination of the most effective transition metal ion for purification of a specific polyhistidine fusion protein can be performed. There are several companies that offer IMAC resin. The matrix most widely used for IMAC is Ni(II)-nitrilotriacetic acid (Ni-NTA), available from Qiagen (Waugh, 2005). Other resins used for the immobilization of transition metal ions include iminodiacetic acid agarose (chelating Sepharose, GE Healthcare) and carboxymethylaspartate agarose (Talon resin, Clontech). Commercially available IMAC resins are unaffected by protease or nuclease activity, and are appropriate for purification of fusion proteins from crude cell lysates. Most of these resins can be regenerated and reused indefinitely. Although polyhistidine does not usually cause a protein to be targeted to *E. coli* inclusion bodies, IMAC is amenable to denaturing agents (i.e., 8 M urea, 6 M guanidine·HCl, ionic and nonionic detergents, and low concentrations of reducing agents) for the purification of insoluble or membrane-bound proteins. High concentrations of reducing agents such as dithiothreitol (DTT) can reduce the immobilized metal ion and should be avoided. The relatively small size and charge of the polyhistidine tag rarely affects protein function, and elution by imidazole gradient is relatively mild, preserving the immunogenicity of polyhistidine fusion proteins.

While purification of a highly-expressed polyhistidine fusion protein can lead to relatively pure protein (>80%) in one chromatographic step, purification from insect and mammalian cells, which contain a higher percentage of His residues in their proteins than *E. coli*, can lead to significant background binding to immobilized metal ions. This may be

circumvented by using stringent wash conditions (e.g., 5 to 10 mM imidazole), although a stringent wash may cause premature elution of the protein of interest. The location of the tag (N-terminal, C-terminal, or internal) can also have an effect on IMAC. If a change in tag location does not increase the effectiveness of IMAC, a denaturing purification can be attempted.

Primary antibodies have also been developed for the detection of polyhistidine fusion proteins *in vitro*. Again, because of the predominance of histidine residues in mammalian and insect systems, anti-polyhistidine antibodies are notoriously promiscuous. Ni^{2+} resin can also be used to precipitate a polyhistidine-tagged protein for the detection of protein-protein interactions.

Glutathione *S*-Transferase

The pGEX *E. coli* expression vectors, which encode for N-terminal glutathione *S*-transferase (GST) molecules followed by protease cleavage sites, were first designed and used to express and purify antigens of the parasite *Taenia ovis* in 1988 (Smith and Johnson, 1988; Smith, 2000). Currently, pGEX vectors are available from GE Healthcare in all three reading frames and with three different protease cleavage sites (e.g., thrombin, factor Xa, and PreScission). GST fusion proteins can be purified by affinity chromatography (*UNIT 6.6*) on commercially available glutathione (γ-glutamylcysteinylglycine) Sepharose (K_d = 0.6 nM), which is affected by γ-glutamyl transpeptidase activity in crude cell lysates. Therefore, glutathione resin has a finite lifetime and can only be regenerated and reused between four and twenty times. Glutathione affinity chromatography is amenable to low concentrations of denaturing agents (2 to 3 M urea or guanidine hydrochloride), reducing agents (<10 mM 2-mercaptoethanol or dithiothreitol), and nonionic detergents (2% v/v Tween 20), depending on the nature of the fusion protein. GST fusion proteins are incubated with glutathione sepharose in order to facilitate croslinking between the two. The elution with 10 mM glutathione is relatively mild, often preserving protein function and antigenicity. A 70 kDa *E. coli* heat-shock-induced chaperonin often copurifies with eluted GST fusion proteins (Thain et al., 1996). This contaminant can be removed by treatment of cell lysates with 5 mM $MgCl_2$ and 5 mM ATP prior to purification. Furthermore, GST can be cleaved from its fusion protein while still bound to glutathione agarose, providing a convenient method for separating the 26 kDa GST from the protein of interest.

GST fusion proteins are often expressed at high levels in *E. coli* (typical yields ~10 mg/liter), which may result in accumulation of aggregated protein in inclusion bodies. Purification from inclusion bodies (*UNIT 6.5*) has both advantages and disadvantages. Some advantages are that protein targeting to *E. coli* inclusion bodies allows the high-level expression of toxic genes and the separation of inclusion bodies serves as a significant purification step from whole-cell lysate. Unfortunately, since glutathione affinity chromatography depends on the proper three-dimensional fold of GST, insoluble fusion proteins must be refolded and buffer exchanged before purification; however, some insoluble proteins may not refold correctly or into a soluble form. Another potential disadvantage of the GST tag is that the large 26 kDa size of the tag and its dimerization in solution may affect the properties of the fusion protein.

GST fusion proteins can be detected by a colorimetric assay with the GST substrate 1-chloro-2,4-dinitrobenzene (CDNB) or with anti-GST antibodies (e.g., GE Healthcare, Sigma, BD Biosciences). Precipitations of GST fusion proteins with glutathione-coupled beads are commonly used for the detection of protein-protein or protein-DNA interactions.

Maltose Binding Protein

pMAL *E. coli* expression vectors were designed in 1988 (di Guan et al., 1988; Maina et al., 1988). Maltose binding protein (MBP) is often used to increase the expression level and/or solubility of its fusion partner, with typical yields of 10 to 40 mg fusion protein per liter culture. pMAL vectors are available for cytoplasmic or periplasmic expression in all three reading frames, with factor Xa, enterokinase, or genenase I protease cleavage sequences (New England Biolabs). Other MBP fusion vectors include pIVEX (Roche), which can be used for coupled *in vitro* transcription/translation.

MBP fusion proteins can be purified by affinity chromatography on cross-linked amylose resin. Amylose resins are commercially available, but are affected by amylase activity in crude cell lysates and can be regenerated and reused only three to five times. Adding glucose to bacterial growth media helps suppress amylase expression (see New England Biolabs pMAL fusion system FAQ: http://www.circuit.neb.com/neb/faqs/faq_pfp.html). Amylose affinity chromatography is not amenable to denaturing or reducing agents. Low concentrations of nonionic detergents (0.2% v/v Triton-X or Tween 20) may be used depending on the nature of fusion protein.

Like GST fusion proteins, high-level expression of MBP fusion proteins in *E. coli* may result in accumulation of insoluble protein aggregates in inclusion bodies. The large size of the MBP tag (45 kDa) may also affect protein function. Unlike glutathione affinity chromatography (*UNIT 6.6*), proteolytic cleavage of the tag while bound to amylose resin is not effective and the fusion protein must be eluted by free maltose before cleavage. A 10 mM maltose elution is sufficient and mild enough to often preserve protein function and antigenicity. If proteolytic removal of the MBP tag is performed, free maltose must be removed from the cleaved MBP by an additional chromatographic step if MBP is to rebind the amylose column (NEB recommends standard DEAE chromatography.) Anti-MBP antibodies are available for the detection of MBP fusion proteins (e.g., BD Biosciences, Sigma, Zymed).

Calmodulin Binding Peptide

The calmodulin binding peptide (CBP) purification system utilizes a C-terminal fragment from muscle myosin light-chain kinase in order to purify proteins of interest from bacteria. With low levels of calcium present at physiological pH, this 26 amino acid fragment displays a fairly strong affinity ($K_d = 10^{-9}$ M) for the protein calmodulin (Simcox et al. 1995). Removal of calcium causes calmodulin to undergo a conformational change resulting in the release of its ligand. T7-based pET expression vectors have been engineered to allow attachment of the CBP affinity tag to either the C- or N-terminal of the fusion protein.

Advantages of this system include its mild binding and elution conditions, which help to increase the ability of the fusion protein to maintain its native form following purification. One passage of crude cell lysate through the calmodulin affinity resin is sufficient, and the protein of interest is subsequently eluted at a neutral pH with 2 mM EGTA (http://www.stratagene.com/manuals/214407.PDF). The 4 kDa size of the CBP tag itself is relatively small and is much less likely to affect the properties of the protein of interest, thus making this an appealing affinity tag in comparison to tags of larger sizes.

Intein-Chitin Binding Domain

The intein-chitin binding domain (intein-CBD) tag is a combination of a protein self-splicing element (intein) with a chitin-binding domain, and allows for the purification of a native recombinant protein without need for a protease. Intein-CBD *E. coli* expression vectors were designed in 1997 (Chong et al., 1997). Commercially available vectors provide for intein-CBD expression on the N-terminus, C-terminus, or both termini of a heterologous protein of interest (IMPACT system, New England Biolabs).

CBD fusion proteins are purified by affinity chromatography on chitin resin. Commercially available resins can be regenerated at least five times. Nonspecific binding to chitin resin can be reduced by a stringent wash including high salt or detergent (2 M NaCl, 0.5% Triton X). Chitin affinity chromatography is not amenable to denaturing reagents such as urea or guanidine hydrochloride. Low concentrations of reducing agents (<1 mM dithiothreitol or <5 mM 2-mercaptoethanol) may be used during purification, but higher concentrations will prematurely activate the intein self-cleavage reaction. Low concentrations of nonionic detergents (i.e., 0.2% Triton-X or Tween 20) may also be used depending on the nature of the fusion protein.

The intein self-cleavage reaction is induced by overnight incubation with 50 mM DTT at 4°C. 2-mercaptoethanol, cysteine, or hydroxylamine may also be used but are less effective (Chong et al., 1997), with the latter two forming permanent complexes with the cleaved protein. The production of proteins possessing an N-terminal cysteine and/or C-terminal thioester can be useful for protein labeling, ligation, or cyclization (Chong et al., 1997). Typical yields from intein-CBD expression and purification are ~0.5 to 5 mg cleaved protein per liter bacterial culture.

Streptavidin/Biotin-based tags

In vivo biotinylation of heterologous proteins containing a biotinylation signal peptide (BCCP) was first reported in 1990 (Cronan, 1990). The enzyme biotin ligase is responsible for attaching biotin to a BCCP-containing protein of interest via an amide linkage between the coenzyme's carboxyl group and a lysine residue on the protein. The PinPoint Xa vectors available from Promega provide all three reading frames and a factor Xa cleavage site. Typical yields from the PinPoint expression and purification system are in the range of 1 to 5 mg fusion protein per liter bacterial culture.

Biotinylated proteins can be purified by affinity chromatography on avidin resin, although purification by traditional, or tetrameric, avidin resin has an affinity constant of $K_d = 10^{-15}$

M and thus requires a strong denaturing elution (e.g., heat, urea, or guanidine hydrochloride). Commercially available monomeric avidin resins have an affinity constant of $K_d = 10^{-7}$ M (i.e., SoftLink avidin resin, Promega) and can be regenerated at least ten times and allow elution of biotinylated fusion proteins with a mild 10 mM biotin buffer. Denaturing reagents are not compatible with avidin affinity chromatography. Low concentrations of reducing agents or nonionic detergents may be used depending on the nature of the fusion protein.

There is little nonspecific binding to avidin in *E. coli* systems. Native BCCP does not bind to avidin; however, mammalian systems have at least four biotinylated protein species that can co-elute with the biotinylated protein of interest (Hollingshead et al., 1997). Biotinylated proteins can be detected by alkaline phosphatase (AP)- or horseradish peroxidase (HRP)-coupled streptavidin, primary antibodies against biotin, or labeling with fluorescent or radioactive biotin. Biotin precipitation assays with streptavidin-coupled beads (*UNIT 9.7*) can be used for the detection of protein-protein or protein-DNA interactions. In addition, the high-affinity interaction of biotin and streptavidin ($K_d \approx 10^{-15}$ M) makes the biotin tag useful for the immobilization of fusion proteins on streptavidin-coated surfaces, such as surface plasmon resonance chips (Sensor Chip SA, Biacore).

In addition to the *in vivo* biotinylated BCCP tag, other purification systems utilize the high-affinity interaction between biotin and streptavidin or avidin. Streptavidin-binding peptide (SBP) is 38 residues in length and can be used to immobilize fusion proteins on a streptavidin matrix. In addition, the Strep-Tag and Strep II-Tag (Sigma) are eight- to nine-residue peptides that bind a mutant form of streptavidin. Strep-Tag was identified as an 8 amino acid peptide that reversibly binds to the same pocket as D-biotin. Over the years, both the peptide (Strep-Tag and Strep II-Tag) and streptavidin core (Strep-Tactin) have been engineered for optimal binding affinity. A particular advantage to using Strep II-Tag is that it does not interfere with protein folding or secretion, making it well-suited for studying functional proteins. Many commercially available expression vectors can fuse Strep II-tag to the C- or N-terminal of a protein of interest for expression in bacterial, mammalian, and insect cells. The Strep II-Tag is resistant to cellular proteases, but a protease recognition site can be included in the cloning vector. It can be released by washing with desthiobiotin, a D-biotin derivative. The column can be used an additional 3–5 times after washing with HABA (2-[4′-hydroxy-benzeneazo] benzoic acid), which removes desthiobiotin from Strep-Tactin.

Because the affinity of biotin for streptavidin is higher than that of any of the streptavidin-binding peptides described above, streptavidin-bound SBP-tagged or Strep-tagged fusion proteins can be efficiently eluted by free biotin. Streptavidin itself can also be used as an affinity tag for purification or immobilization of fusion proteins on a biotinylated matrix.

His-Patch ThioFusion

Thioredoxin has the ability to accumulate to ~40% of the total cellular protein in *E. coli*. In addition, the thioredoxin moiety can in some cases confer solubility to formerly insoluble heterologous proteins expressed in *E. coli*. Unfortunately, the phenylarsine oxide matrix used to purify thioredoxin fusion proteins does not provide high-yield, high-purity product (Terpe, 2003). To address this problem, Life Technologies developed a mutant thioredoxin

molecule that has a cluster of histidines, such that when the thioredoxin molecule is properly folded, the histidines form a patch amenable to IMAC (His-patch ThioFusion expression system; see Polyhistidine). In addition, the tag is cleavable by Factor Xa, allowing the fusion protein to be separated from the His-Patch ThioFusion molecule, which is >100 residues in length and readily dimerizes in solution. While the cost of this expression system (>$900.00) might discourage its use as a first-line affinity tag, the His-Patch ThioFusion tag might be useful when other attempts at producing large amounts of soluble protein have failed.

Tandem Affinity Purification

Tandem affinity purification (TAP) is a dual-affinity purification method based on the fusion of two affinity tags to a protein of interest. TAP not only allows purification of a tagged protein, but also allows for isolation of protein complexes interacting with the protein of interest. The TAP tag itself consists of two different affinity tags, each of which is designed to isolate the protein of interest along with its interacting proteins in two successive steps (Li, 2011). The major advantage to this method is that non-specific background is reduced due to the two purification steps.

TAP was originally done in yeast using a 21-kDa tag composed of ProtA (Protein A of Staphylococcus aureus) and CBP (calmodulin-binding peptide) separated by a TEV protease cleavage site. This tag can be applied on either the C- or N-terminal. Although 20–30% of a protein of interest can be obtained using this method, TAP does not work well in higher eukaryotes. Additionally, such a large tag (21 kDa) can disturb protein function and CBP can interfere with calcium signaling. Although the original ProtA/CBP tag is still the most widely used, these disadvantages have lead to the creation of over 30 TAP variants.

Epitope Tags

Relatively short epitope tags such as FLAG, hemagglutinin (HA), c-myc, T7, and Glu-Glu, among others (Table 9.9.1), are used for the detection of fusion proteins *in vitro* and in cell culture. Their short, linear recognition motifs rarely affect the properties of the heterologous protein of interest and are usually very specific for their respective primary antibodies. One exception is the anti-myc antibody, which is somewhat promiscuous; however, specificity can be increased by using an enzyme-linked secondary antibody to detect a conjugated anti-myc primary antibody, instead of using an HRP- or AP-anti-myc conjugate alone.

Epitope-tagged proteins can be purified using immobilized primary antibodies to the epitope tag although antibody affinity chromatography often involves a low or high pH elution which can irreversibly affect the properties of the fusion protein, and employs resin which has limited reusability. Epitope tags are therefore not the first choice when the main goals are high-level expression and fusion protein purification.

Reporter Tags

Enzymes such as β-galactosidase (β-gal), alkaline phosphatase (AP), chloramphenicol acetyl transferase (CAT), and horseradish peroxidase (HRP) serve as convenient reporters of protein expression and protein-protein interaction. These reporter tags can sometimes be

used for protein purification, but are frequently used as reporter tags for detecting protein expression.

β-Galactosidase—β-gal interacts with its substrate, 5-bromo-4-chloro-3-indoyl-β-d-galactopyranoside (X-gal), to yield a blue precipitate. The advantages and disadvantages of lacZ as a protein purification tag have been discussed above (see Early Affinity Tags).

Alkaline Phosphatase—AP interacts with 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium (BCIP/NBT) to yield a purple precipitate. AP is commonly fused to primary antibodies for immunoblotting but can also be fused to any number of nonimmune proteins for far-western blotting (i.e., detection of a membrane-transferred protein by interaction with another protein; *UNIT 19.7*). Bacterial expression vectors are available to overexpress and purify AP-tagged proteins-of-interest (Gene Therapy Systems). Since AP is normally targeted to the *E. coli* periplasmic space, crude AP fusion protein lysates can be purified by periplasmic extraction (*UNIT 5.2*). AP fusion proteins can be further purified by immunoaffinity chromatography.

Chloramphenicol Acetyl Transferase—The ease and sensitivity of the assay for CAT activity made CAT one of the first reporter genes used for studies of mammalian gene expression (Podbielski et al., 1992). The CAT gene is not found in eukaryotes, and therefore eukaryotic cells contain no background CAT activity, which can be detected using the radioisotope-labeled or fluorescently-labeled substrates acetyl-CoA or chloramphenicol. CAT fusion proteins can be purified by affinity chromatography on immobilized chloramphenicol (i.e., chloramphenicol caproate agarose, Sigma) and eluted by free chloramphenicol.

Horseradish Peroxidase—Horseradish Peroxidase (HRP) is a 44 kDa protein that catalyzes the oxidation of numerous small molecule substrates, yielding a colored or luminescent product that is detectable by spectrophotometric methods. HRP can be coupled to primary or secondary antibodies or even other protein tags, and is commonly used for Western blotting, ELISA assays, and immunohistochemistry.

Some of the most common substrates for HRP that result a colored product are 3,3-diaminobenzidine (DAB), 3,3,5,5-tetramethylbenzidine (TMB), and 2.2′-azino-di[3-ethylbenzthiazoline sulfonate (ABTS). HRP-catalyzed reactions can be stopped by addition of acid when the desired degree of color change is reached. Colorimetric substrates are often used in ELISA and immunohistochemistry protocols.

Enhanced chemiluminescent (ECL) substrates are the most commonly-used substrates for detection of proteins by Western blotting. When an HRP-tagged primary or secondary antibody is used, antibody bound to the membrane can be detected by the light that is produced, either by film or CCD cameras. Utilizing a secondary antibody increases specificity and decreases background signal. A commercially-available plasmid for tagging a protein-of-interest with HRP for direct detection is not available, to our knowledge.

DEVELOPING AFFINITY TAG TECHNOLOGIES

Some newer affinity tags have yet to be rigorously evaluated and consequently cannot be recommended above the tried-and-true affinity tags described above (see Commonly Used Affinity Tags). Still, many of these novel affinity tags contain unique properties that may make them quite useful in cases of troublesome purifications.

Small Ubiquitin-Like Modifier

Small ubiquitin-like modifier (SUMO) family proteins are ~100 amino acids in length and are highly conserved in eukaryotes. They are essential for normal cell functions such as nuclear transport, signal transduction, and protein stabilization. SUMO proteins function by forming a covalent bond with a target protein, typically at a lysine residue, which is later cleaved by SUMO protease. SUMO protease recognizes the sequence x-Gly-Gly-x, and therefore doesn't cleave the target protein. This aspect of SUMO proteases make SUMO an attractive affinity tag because no recombinant linker region needs to be constructed and the native N-terminus of the target protein is maintained. Other advantages of SUMO are that it increases protein expression and solubility. Several commercial kits are available that utilize both yeast and human SUMO proteins and proteases (Life Technologies and LifeSensors). Since SUMO is not expressed in prokaryotes, most kits use *E. coli* to express the protein and typically include a polyhistidine tag. After SUMO is cleaved, the protein can be isolated using polyhistidine antibodies, Ni-NTA or NI-IMAC or chromatography. One kit, SUMOstar, allows for the expression and purification of SUMO-tagged proteins in insect and mammalian cells.

HaloTag®

The HaloTag® (Promega) is a 33-kDa protein based on a modified haloalkane dehalogenase found in bacteria. In bacteria, native dehalogenase acts through a nucleophilic displacement reaction to remove halides from aliphatic hydrocarbons (Los et al., 2008). Asp106 is responsible for forming the enzyme-substrate complex while His272 is responsible for hydrolysis of the intermediate. A point mutation, His272Phe, impairs hydrolysis leading to a covalent bond between the enzyme and substrate. The commercially available HaloTag® has the His272Phe point mutation along with other amino acid substitutions that lead to increased ligand binding.

Although other hydrolytic enzymes could be modified, haloalkane dehalogenase has several advantages: it is a small protein that does not demand the use of cofactors or post-translational modifications for its enzymatic reaction. Also, ligands of haloalkane dehalogenase may be designed to diffuse through cell membranes, be nontoxic, and avoid cross-reactivity with other cellular proteins. (Los et al. 2008). For protein purification, an expression vector is transfected into a cell line in order to create a fusion protein. Synthetic HaloTag® ligand is then added to the cells to label the protein. The synthetic ligand can have one of many functional groups attached, including various epitope tags (e.g., GST). This versatility allows HaloTag® to be used not only for *in vitro* protein labeling and purification but also *in vivo* labeling. Cells are then lysed and the fusion protein captured on a purification matrix. The mutated hydryolase HaloTag® creates a covalent attachment to

the HaloLink® resin via an immobilized chloroalkane. This covalent bond allows for harsher wash conditions on the matrix. After washing, the protein of interest is cleaved from HaloTag with the use of TEV protease. HisLink® resin is utilized to remove TEV Protease from the protein, and the result is a highly pure, tag-free protein.

Profinity eXact

The Profinity eXact™ system (Bio-Rad) is based on the serine protease subtilisin from *Bacillus amyloliquefaciens*. Subtilisin, a member of the subtilase protein family, is somewhat promiscuous in its substrate choices. The action of subtilisin is to bind to residues P1-P4 of its substrate, resulting in the interpolation of the substrate's backbone between β-strands 100–104 and 125–129 of subtilisin (Biao et al. 2004). Engineering of subtilisin has resulted in a mutant subtilisin protein with reconstructed P1 and P4 binding pockets in order to increase sequence selectivity. Biosynthesis of subtilisin is dependent on a 75-amino acid N-terminal prodomain (Ruan et al., 2004). It is this sequence that has been modified to act as a tag. Using an expression vector, recombinant protein is created that contains the subtilisin prodomain tag on the N-terminus of the protein of interest. The recombinant protein is then applied to a column with immobilized subtilisin protease. Cleavage at the C-terminus of the tag by mature subtilisin protease is initiated by fluoride buffers and allows for the release of the purified target protein. The result is a protein with its native N-terminal that is tag-free.

Major advantages of the Profinity eXact™ system include that it does not require the use of any proteases. It allows for a quick 30 minute on-column purification and cleavage that can be performed in one step, enabling the purified proteins to maintain their native structure. Additionally, the column can be regenerated by removing the bound prodomain at pH = 2.1 (Biao et al., 2004). The subtilisin prodomain tag is known to work well with soluble heterotrimeric G-protein α-subunits, which are often problematic when attempting to purify or over-express them. A hydrophobic portion of the tag itself increases the solubility of the fusion G protein, and it has been shown that no structure or activation dynamics are altered during the purification process using the Profinity eXact™ system (Abdulaev et al., 2005).

PDZ domain-based tags

Post-synaptic density protein (**P**SD95), Drosophila disc large tumor suppressor (**D**lg1), and zonula occludens-1 protein (**z**o-1) were the first three proteins discovered to share the PDZ domain, a small (~10 kDa) structural domain common across species. PDZ domains bind to the C-termini of their protein targets, and are often found in signal transduction pathways where they scaffold different signaling components together. The N-terminal PDZ domain (PDZ1) of the Drosophila protein InaD has been suggested as a potential protein tagging system, as it binds the C-terminus of its target, NorpA, via an intermolecular disulfide bond. Proteins with a NorpA tag could be eluted from a PDZ-coupled resin (or vice versa) via reduction of the disulfide bond (Kimple et al., 2001; Kimple and Sondek 2002).

The PDZ/C-terminal peptide interaction has recently been harnessed for other applications. A human papilloma virus (HPV) E6 protein dipstick test is being developed that utilizes the binding affinity between E6 PDZ and its C-terminal targets (Arbor Vita). Like the InaD/

NorpA interaction, technique could also potentially be utilized to tag and detect heterologous proteins.

AFFINITY TAGS AND PROTEIN PURIFICATION

Figures 9.9.1, 9.9.2, and 9.9.3 give general schemes for selecting an affinity tag based on three different desired protein purification outcomes: a large quantity of highly pure, untagged protein (Fig. 9.9.1), a medium quantity of highly pure tagged or untagged protein (Fig. 9.9.2), and a small amount of moderate-to-highly pure tagged protein (Fig. 9.9.3). These flow charts should be used as an initial guide but more difficult protein purification problems may require the use of a system different than those recommended.

LITERATURE CITED

Abdulaev NG, Zhang C, Dinh A, Ngo T, Bryan PN, Brabazon DM, Marino JP, Ridge KD. Bacterial expression and one-step purification of an isotope-labeled heterotrimeric G-protein alpha-subunit. J Biomol NMR. 2005; 32:31–40. [PubMed: 16041481]

Berlot, CH. Expression and functional analysis of G-protein alpha subunits in mammalian cells. In: Manning, DR., editor. G-proteins: Techniques of Analysis. CRC Press; New York: 1999. p. 37-57.

Bornhorst JA, Falke JJ. Purification of proteins using polyhistidine affinity tags. Methods Enzymol. 2000; 326:245–254. [PubMed: 11036646]

Chong S, Mersha FB, Comb DG, Scott ME, Landry D, Vence LM, Perler FB, Benner J, Kucera RB, Hirvonen CA, Pelletier JJ, Paulus H, Xu MQ. Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. Gene. 1997; 192:271–281. [PubMed: 9224900]

Crespo P, Schuebel KE, Ostrom AA, Gutkind JS, Bustelo XR. Phosphotyrosine-dependent activation of Rac-1 GDP/GTP exchange by the vav proto-oncogene product. Nature. 1997; 385:169–172. [PubMed: 8990121]

Cronan JE Jr. Biotination of proteins in vivo. A post-translational modification to label, purify, and study proteins. J Biol Chem. 1900; 265:10327–10333. [PubMed: 2113052]

di Guan C, Li P, Riggs PD, Inouye H. Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein. Gene. 1988; 67:21–30. [PubMed: 2843437]

Forde GM. Preparation, analysis and use of an affinity adsorbent for the purification of GST fusion protein. Methods Mol Biol. 2008; 421:125–36. [PubMed: 18826052]

Fritze CE, Anderson TR. Epitope tagging: General method for tracking recombinant proteins. Methods Enzymol. 2000; 327:3–16. [PubMed: 11044970]

Gerdes HH, Kaether C. Green fluorescent protein: Applications in cell biology. FEBS Lett. 1996; 389:44–47. [PubMed: 8682203]

Goldstein DJ, Toyama R, Dhar R, Schlegel R. The BPV-1 E5 oncoprotein expressed in Schizosaccharomyces pombe exhibits normal biochemical properties and binds to the endogenous 16-kDa component of the vacuolar proton-ATPase. Virology. 1992; 190:889–893. [PubMed: 1387753]

Hollingshead M, Sandersen J, Vaux DJ. Anti-biotin antibodies offer superior organelle-specific labeling of mitochondria over avidin or streptavadin. J Histochem Cytochem. 1997; 45:1053–1058. [PubMed: 9267466]

Jones C, Patel A, Griffin S, Martin J, Young P, O'Donnell K, Silverman C, Porter T, Chaiken I. Current trends in molecular recognition and bioseparation. J Chromatogr. 1995; 707:3–22.

Kaldalu N, Lepik D, Kristjuhan A, Ustav M. Monitoring and purification of proteins using bovine papillomavirus E2 epitope tags. Biotechniques. 2000; 28:456–462. [PubMed: 10723557]

Karp M, Oker-Blom C. A streptavidin-luciferase fusion protein: Comparisons and applications. Biomol Eng. 1999; 16:101–104. [PubMed: 10796991]

Kimple ME, Siderovski DP, Sondek J. Functional relevance of the disulfide-linked complex of the N-terminal PDZ domain of InaD with NorpA. EMBO J. 2001; 20:4414–4422. [PubMed: 11500369]

Kimple ME, Sondek J. Affinity tag for protein purification and detection based on the disulfide-linked complex of InaD and NorpA. Biotechniques. 2002; 33:578–588. [PubMed: 12238768]

Kolodziej PA, Young RA. Epitope tagging and protein surveillance. Methods Enzymol. 1991; 194:508–519. [PubMed: 1706460]

Kuliopulos A, Walsh CT. Production, purification, and cleavage of tandem repeats of recombinant peptides. J Am Chem Soc. 1994; 116:4599–4607.

Kwatra MM, Schreurs J, Schwinn DA, Innis MA, Caron MG, Lefkowitz RJ. Immunoaffinity purification of epitope-tagged human beta 2-adrenergic receptor to homogeneity. Protein Expr Purif. 1995; 6:717–721. [PubMed: 8746622]

Lazzaroni JC, Atlan D, Portalier RC. Excretion of alkaline phosphatase by Escherichia coli K-12 pho constitutive mutants transformed with plasmids carrying the alkaline phosphatase structural gene. J Bacteriol. 1985; 164:1376–1380. [PubMed: 2999085]

Lederman L. Protein Isolation and Purification. Biotechniques. 2009; 46:87–9.

Lilius G, Persson M, Bulow L, Mosbach K. Metal affinity precipitation of proteins carrying genetically attached polyhistidine affinity tails. Eur J Biochem. 1991; 198:499–504. [PubMed: 1904025]

Ljungquist C, Breitholtz A, Brink-Nilsson H, Moks T, Uhlen M, Nilsson B. Immobilization and affinity purification of recombinant proteins using histidine peptide fusions. Eur J Biochem. 1989; 186:563–569. [PubMed: 2514094]

Los GV, Encell LP, McDougall MG, Hartzell DD, Karassina N, Zimprich C, Wood MG, Learish R, Ohana RF, Urh M, Simpson D, Mendez J, Zimmerman K, Otto P, Vidugiris G, Zhu J, Darzins A, Klaubert DH, Bulleit RF, Wood KV. HaloTag: a novel protein labeling technology for cell imaging and protein analysis. ACS Chem Biol. 2008; 3:373–82. [PubMed: 18533659]

Maina CV, Riggs PD, Grandea AG III, Slatko BE, Moran LS, Tagliamonte JA, McReynolds LA, Guan CD. An Escherichia coli vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein. Gene. 1988; 74:365–373. [PubMed: 3073105]

Makrides SC. Strategies for achieving high-level expression of genes in Escherichia coli. Microbiol Rev. 1996; 60:512–538. [PubMed: 8840785]

McLean PJ, Kawamata H, Hyman BT. Alpha-synuclein-enhanced green fluorescent protein fusion proteins form proteasome sensitive inclusions in primary neurons. Neuroscience. 2001; 104:901–912. [PubMed: 11440819]

Morandi C, Perego M, Mazza PG. Expression of human dihydrofolate reductase cDNA and its induction by chloramphenicol in Bacillussubtilis. Gene. 1984; 30:69–77. [PubMed: 6096225]

Nelson RW, Jarvik JW, Taillon BE, Tubbs KA. BIA/MS of epitope-tagged peptides directly from *E. coli* lysate: Multiplex detection and protein identification at low-femtomole to subfemtomole levels. Anal Chem. 1999; 71:2858–2865. [PubMed: 10424172]

Nilsson J, Bosnes M, Larsen F, Nygren PA, Uhlen M, Lundeberg J. Heat-mediated activation of affinity-immobilized Taq DNA polymerase. Biotechniques. 1997a; 22:744–751. [PubMed: 9105627]

Nilsson J, Stahl S, Lundeberg J, Uhlen M, Nygren PA. Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins. Protein Expr Purif. 1997b; 11:1–16. [PubMed: 9325133]

Podbielski A, Peterson JA, Cleary P. Surface protein-CAT reporter fusions demonstrate differential gene expression in the vir regulon of Streptococcus pyogenes. Mol Microbiol. 1992; 6:2253–2265. [PubMed: 1406266]

Ruan B, Fisher KE, Alexander PA, Doroshko V, Bryan PN. Engineering subtilisin into a fluoride-triggered processing protease useful for one-step protein purification. Biochemistry. 2004; 43(46): 14539–46. [PubMed: 15544324]

Rubinfeld B, Munemitsu S, Clark R, Conroy L, Watt K, Crosier WJ, McCormick F, Polakis P. Molecular cloning of a GTPase activating protein specific for the Krev-1 protein p21rap1. Cell. 1991; 65:1033–1042. [PubMed: 1904317]

Sambrook, J.; Fritsch, EF. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, N.Y.: 1989.

Sano T, Vajda S, Cantor CR. Genetic engineering of streptavidin, a versatile affinity tag. J Chromatogr B Biomed Appl. 1998; 715:85–91.

Simcox TG, Zheng CF, Simcox ME, Vaillancourt P. Strategies Mol Biol. 1995; 8:40–43.

Skerra A, Schmidt TG. Use of the Strep-Tag and streptavidin for detection and purification of recombinant proteins. Methods Enzymol. 2000; 326:271–304. [PubMed: 11036648]

Smith DB. Generating fusions to glutathione S-transferase for protein studies. Methods Enzymol. 2000; 326:254–270. [PubMed: 11036647]

Smith DB, Johnson KS. Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. Gene. 1988; 67:31–40. [PubMed: 3047011]

Smith MC, Furman TC, Ingolia TD, Pidgeon C. Chelating peptide-immobilized metal ion affinity chromatography. A new concept in affinity chromatography for recombinant proteins. J Biol Chem. 1988; 263:7211–7215. [PubMed: 3284883]

Stevens RC. Design of high-throughput methods of protein production for structural biology. Structure Fold Des. 2000; 8:R177–R185. [PubMed: 10986469]

Strambio-de-Castillia C, Tetenbaum-Novatt J, Imai BS, Chait BT, Rout MP. A method for the rapid and efficient elution of native affinity-purified protein A tagged complexes. J Proteome Res. 2005; 4:2250–6. [PubMed: 16335973]

Tai TN, Havelka WA, Kaplan S. A broad-host-range vector system for cloning and translational lacZ fusion analysis. Plasmid. 1988; 19:175–188. [PubMed: 2852814]

Terpe K. Overview of tag protein fusions: From molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol. 2003; 60:523–533. [PubMed: 12536251]

Thain A, Gaston K, Jenkins O, Clarkie AR. A method for the separation of GST fusion proteins from co-purifying GroEL. Trends Genet. 1996; 12:209–210. [PubMed: 8928224]

Wang LF, Yu M, White JR, Eaton BT. BTag: A novel six-residue epitope tag for surveillance and purification of recombinant proteins. Gene. 1996; 169:53–58. [PubMed: 8635749]

Waugh DS. Making the most of affinity tags. Trends Biotechnol. 2005; 23:316–20. [PubMed: 15922084]

Li Y. The tandem affinity purification technology: an overview. Biotechnol Lett. 2011; 33:1487–99. [PubMed: 21424840]

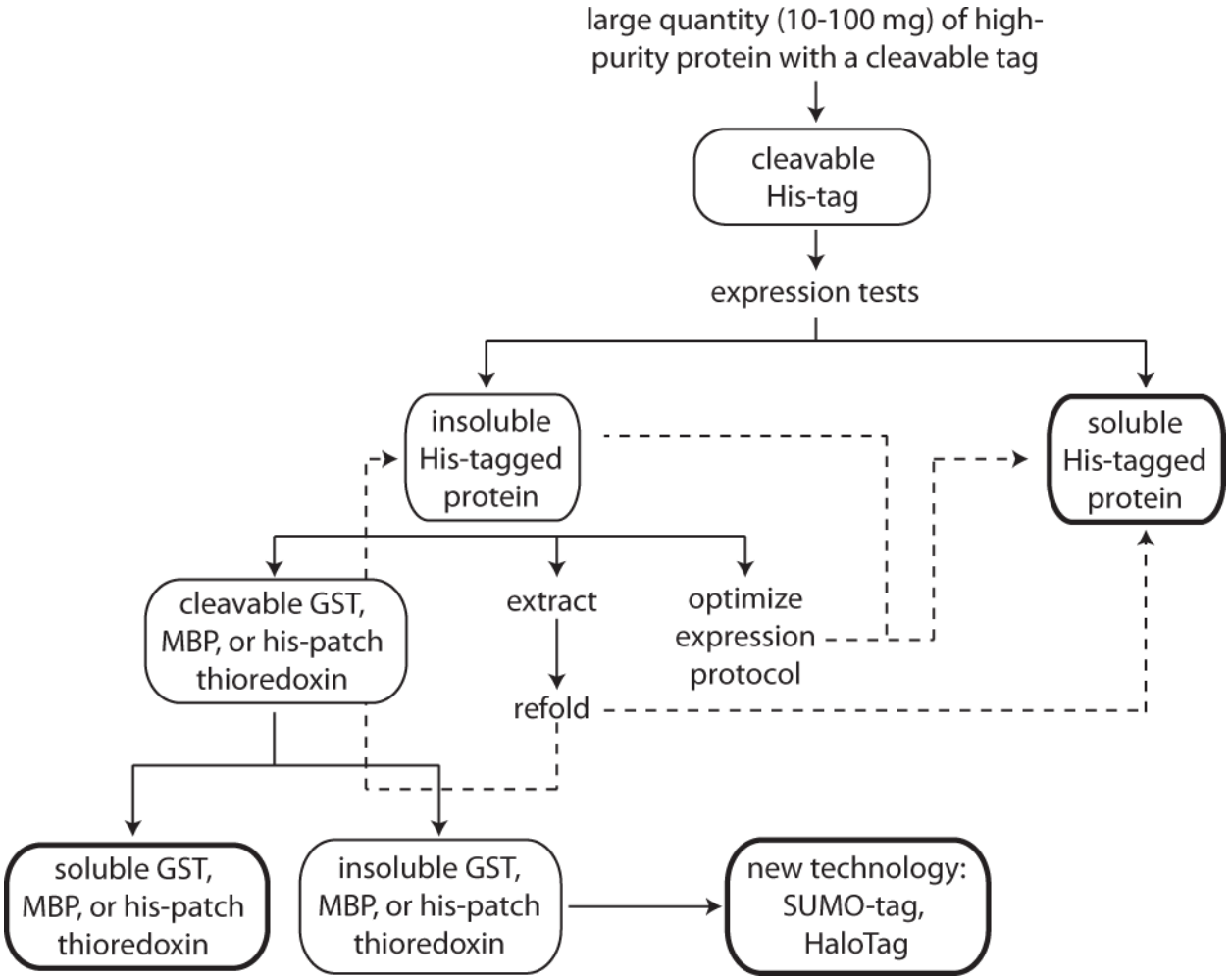

Figure 9.9.1.

Flow chart describing the general scheme for choosing an affinity tag for protein purification if a larger amount (10–100 mg) of highly pure untagged protein is required. Experimental methods requiring this high quantity of pure, native protein include, but are not limited to, protein crystallography, NMR spectroscopy, isothermal titration calorimetry, and surface plasmon resonance. Solid lines indicate forward progression in purification optimization, dashed lines indicate steps backwards, and thick boxes indicate the final desired product, a protein of high purity and quality.

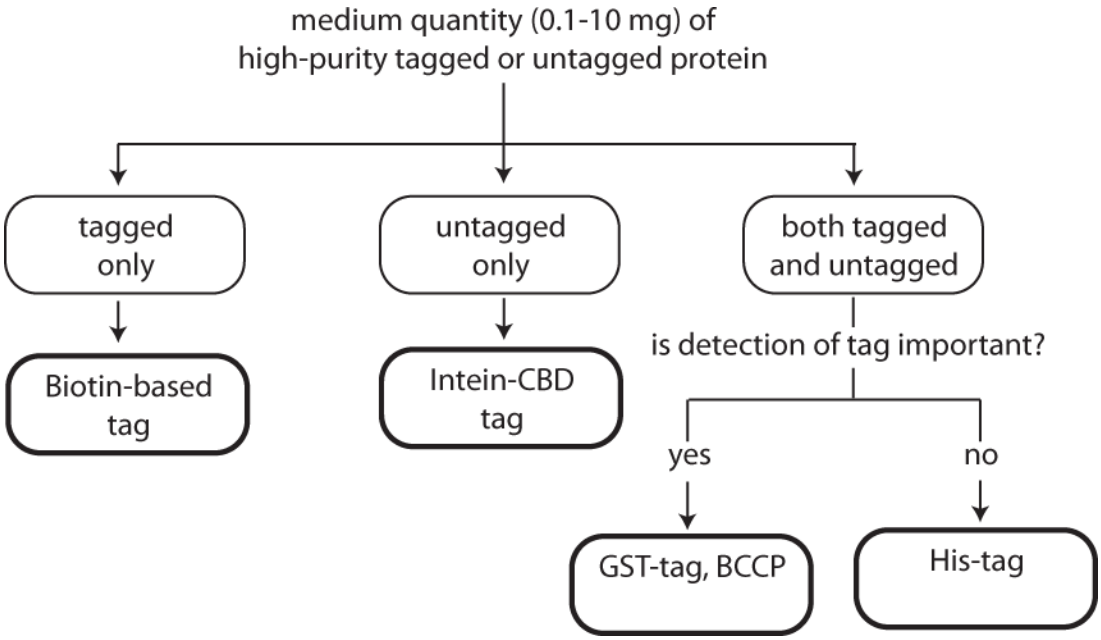

Figure 9.9.2.

Flow chart describing the general scheme for selecting an affinity tag for protein purification if a mid-range amount (0.1–10 mg) of high purity tagged and/or untagged protein is needed. An example of an experimental method requiring this quantity of tagged protein is surface plasmon resonance (SPR), where a biotin moiety could be sued to attach a protein ligand to a streptavidin-coated SPR chip. Other possible applications are in GST and Ni^{2+} precipitation assays, and mass spectrometry. Certain variants of the intien-CBD domain are useful for methods requiring a completely native protein. Finally, the His-tag and GST-tag systems are useful because both tagged and untagged proteins can be easily prepared from the same cellular lysate, and there exist many secondary reagents specifically for use with these two systems. Thick lines indicate the desired result has been reached.

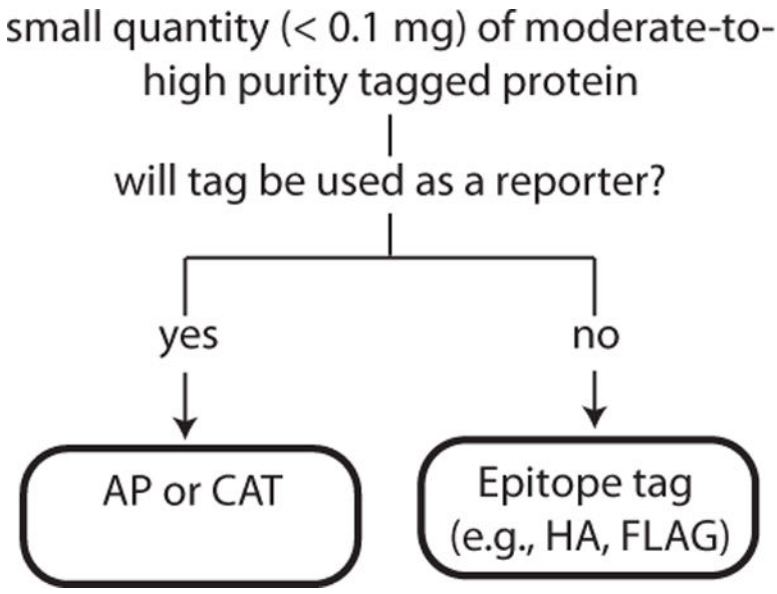

Figure 9.9.3.

Flow chart describing the general scheme for selecting an affinity tag for purification of a small amount (<0.1 mg) of moderate-to-high purity tagged protein. Such a protein might be used in co-immunoprecipitation, Western blotting, far-Western blotting, ELISA, or gel overlay assays. Since only small amounts of protein are required, it is not usually necessary to optimize protein expression, leaving the choice of affinity tag to be based on the final use for the fusion tag—i.e., as a reporter or for general detection. Examples of epitope tags include the HA-tag, FLAG-tag, T7-tag, Myc-tag, and others.

Table 9.9.1

Characteristics of Protein Affinity Tags

Tag	Length (sequence)	Position	Matrix/elution	Commercial systems (supplier)	Typical use(s)	Comments	References
Albumin-binding protein (ABP)	137	N-term or C-term	Albumin/low pH or denaturation (e.g., heat, urea)	NA	Purification, increased expression and increased solubility	May increase proteolytic stability of fusion proteins to increase expression; may improve fusion solubility; high background binding to albumin in eukaryotic systems; matrix has limited reusability; large tag or low pH elution conditions may affect fusion properties	Nilsson et al. (1997a)
Alkaline Phosphatase (AP)	444	Primary amines (lysine side chain epsilon-amines and N-terminal α-amines)	mAb/low pH	EasyLink Alkaline phosphatase Conjugation Kit (AbCam); Lightning-Link™ Alkaline Phosphatase Kit (Innova); gWIZ secreted AP expression vector (Genlantis)	Detection and purification	Colorimetric detection; useful for Western bloting, far-Western blotting, southern blotting, sandwich ELISA, and subcellular localization in bacterial and mammalian expression systems; convenient purification of crude periplasmic extract from bacteria; antibody purification does not give high yields; low pH elution may irreversibly affect protein properties; matrix is of limited reusability; AP dimerization and large size may affect properties of fusion.	Lazzaroni et al. (1985)
AU1 epitope	6 (DTYRYI)	N-term or C-term	mAb/low pH	Epitope Tag Peptide, AU1 Kit (Covance)	Detection and purification	Antibody purification does not give high yields; low pH elution may irreversibly affect protein properties; matrix is of limited reusability	Goldstein et al. (1992)
AU5 epitope	6 (TDFYLK)	N-term or C-term	mAb/low pH	Epitope Tag Peptide, AU5 Kit (Covance)	Detection and purification	Antibody purification does not give high yields; low pH elution may irreversibly affect protein properties; matrix is of limited reusability	Crespo et al. (1997)
Bacteriophage T7 epitope (T7-tag)	11 (MASMTGGQQMG)	N-term or internal	mAb/low pH	T7-tag affinity purification kit (EMD Millipore)	Purification and increased expression	N-term eleven amino acids of phage T7 gene 10; may increase expression of fusion proteins; antibody purification does not give high yields; low pH elution may irreversibly affect protein properties; matrix is of limited reusability	Makrides (1996)
Bacteriophage V5 epitope (V5-tag)	14 (GKPIPNPLLGLDST)	C-term	NA	Selected pET directional TOPO, pBAD, and Gateway systems	Detection	Short, linear recognition motif; antibody specific in bacterial	McLean et al. (2001)

Tag	Length (sequence)	Position	Matrix/elution	Commercial systems (supplier)	Typical use(s)	Comments	References
				(Life Technologies); V5 tag peptide and mAb (AbCam) (Life Technologies); V5 tag peptide and mAb (AbCam)		lysates, although some cross-reactivity in mammalian lysates; often used in combination with His-tag for protein purification	
Biotin-carboxy carrier protein (BCCP)	100	N-term or C-term	Avidin or streptavidin/biotin or denaturation (e.g. heat, urea)	PinPoint system (Promega); Panorama® Antibody Arrays (Sigma-Aldrich)	Detection, purification, and immobilization	Tag is biotinylated *in vivo;* one-step purification of fusion protein; protein may be secreted for convenient purification; biotin tag can be used to immobilize fusion to streptavadin-coated surfaces, e.g. surface Plasmon resonance (SPR) chips; high background binding to avidin in eukaryotic systems; elution with biotin may be inefficient; denaturing elution requires refolding of fusion protein; Promega SoftLink avidin, part of the PinPoint system, allows elution under mild conditions	Nilsson et al. (1997b)
Bluetongue virus tag (B-tag)	6 (QYPALT)	N-term or C-term	mAb/low pH	NA	Detection and purification	VP7 region of bluetongue virus; antibody purification does not give high yields; low pH elution may irreversibly affect protein properties; matrix is of limited reusability	Wang et al. (1996)
Calmodulin binding peptide (CBP)	26	N-term or C-term	Calmodulin/EGTA or EGTA and high salt	The Affinity Protein Expression and Purification System (Agilent Technologies)	Purification and expression	Relatively short recognition motif; no endogenous *E. coli* proteins that bind calmodulin; PKA target sequence allows ^{32}P-labeling; high yield matrix is compatible with reducing agents and detergents, but of limited reusability; not useful for purification from eukaryotic cells; tag at N terminus may reduce translation efficiency	Terpe (2003)
Chloramphenicol Acetyl Transferase (CAT)	218	N-term	Chloramphenicol/chloramphenicol	gWiz™ Vectors High-Expression Vectors (Genlantis)	Detection, purification, and increased solubility	Enzymatic assay available for protein quantification; chloramphenicol purification does not give high yields, large tag may affect properties of fusion protein	Podbielski et al. (1992)
Cellulose binding domain (CBP)	27–189	N-term, C-term, or internal (domain-dependent)	Cellulose/denaturation (family I CBDs) or ethylene glycol (family II or III CBDs)	pCAL-n, pCAL-n-EK, pCAL-n-FLAG, and pCAL-c vector set (Agilent Technologies)	Purification, secretion and immobilization	Irreversible binding of some CBDs to cellulose useful for immobilization; reversible binding of CBD I, II, and III families useful for purification; purification of family I CBD fusion necessitates protein refolding; purified family II CBD fusions must be buffer exchanged	Terpe (2003)

Tag	Length (sequence)	Position	Matrix/elution	Commercial systems (supplier)	Typical use(s)	Comments	References
Chitin binding domain (CBD)	51	N-term or C-term	Chitin (when fused with intein)/ thiol-containing reducing agent (e.g., DTT or β-ME)	IMPACT system (New England Biolabs)	Purification	Typically used in conjunction with self-splicing intein tag; one-step purification of nearly 100% pure protein with low milligram yields; matrix compatible with high salt and nonionic detergents; purification must be done in absence of thiol-containing reducing agents until elution step; fusion protein may have effect on intein cleavage efficiency	Terpe (2003)
Choline-binding domain (CBD)	145	N-term	Choline/choline or CBD peptide	LYTAG Two-Phase Protein Purification (Microlytic)	Purification and immobilization	N-term CBD can immobilize fusion protein on choline-coated gold chip for SPR or microscopy studies; large tag or C-term tag placement may affect fusion proteins	Jones et al. (1995)
Dihydrofolate reductase (DHFR)	227	N-term or C-term	Methotrexate/folate	pQE vector set (Qiagen)	Increased expression	May increase proteolytic stability of fusion proteins to increase expression; little immunogenicity in mouse and rat, therefore, useful for generation of antibodies to fusion; methotrexate purification does not give high yields; usually coupled with a short affinity tag (e.g., His-tag, Strep-tag) for purification	Morandi et al. (1984)
E2 epitope	10 (SSTSSDFRDR)	N-term, C-term or internal	mAb/low pH	E2-Tagging and Detection Kit (Abcam)	Detection and purification	Derived from bovine papilloma-virus type-1 transactivator protein E2; primary antibodies available for colorimatric and chemiluminescent detection; antibody purification does not give high yields, low pH elution may irreversibly affect protein properties; matrix is of limited reusability	Kaldalu et al. (2000)
FLAG epitope	8 (DYKDDDK)	N-term or C-term	mAb/low pH, EDTA, or FLAG peptide	FLAG system (Sigma-Aldrich)	Detection and purification	Short, linear recognition motif; moderately pure protein in one step; enterokinase cleaves after C-term Lys to completely remove tag, depending on identity of first amino acid of fusion; M1 antibody can only bind tag at N-term; antibody purification does not give high yields; low pH elution may irreversibly affect protein properties; matrix is of limited reusability	Terpe (2003)

Tag	Length (sequence)	Position	Matrix/elution	Commercial systems (supplier)	Typical use(s)	Comments	References
Galactose-binding protein (GBP)	509	N-term or C-term	Galactose/galactose	NA	Purification and increased solubility	Fusion protein may be targeted to periplasm for convenient purification of crude periplasmic extracts; may increase solubility of fusion proteins; large tag may affect properties of fusion protein	Jones et al. (1995)
Green fluorescent protein (GFP)	220	N-term or C-term	NA	phCMV Fusion Stable Reporter Vectors (Genlantis); CT-GFP Fusion TOPO Cloning kit (Life Technologies); gWIZ GFP mammalian expression vector (Genlantis)	Detection	Intrinsic fluorescence can permit native detection without antibody; anti-GFP antibodies specific; expression in prokaryotic and eukaryotic expression systems, as well as whole organisms (e.g., worms, plants, mice); useful to monitor gene expression, protein folding, and targeting, as well as protein-protein interactions, by FRET; some GFP fusions nonspecifically targeted to nucleus; very large tag or GFP dimerization may affect properties of fusion	Gerdes and Kaether (1996)
Glu-Glu (EE-tag)	6 (EYMPME or EFMPME)	N-term, or C-term, or internal	mAb/low pH or 30°C incubation	Epitope Tag Peptide, GLU-GLU Kit (Covance)	Detection and purification	Short, linear recognition motif; available antibody recognizes only EYMPME motif; antibody purification does not give high yields, low pH or 30°C elution may irreversibly affect protein properties; matrix is of limited reusability	Rubinfeld et al. (1991)
Glutathione S-transferase (GST)	211	N-term or C-term	Glutathione/reduced glutathione	pGEX System (GE Healthcare); GST•Bind™ System (EMD Millipore)	Detection, purification, increased expression, increased solubility, and immobilization	Very common purification tag; one-step purification of relatively pure protein; detection antibodies specific; kits for GST fusion proteins common (e.g., SPR, enzymatic assays); matrix relatively reusable (four to twenty times); GST is highly antigenic; purification under native conditions only; some GST fusions insoluble; GST dimerization and/or glutathione elution may affect fusion protein properties	Smith (2000)
Human influenza hemagglutinin (HA)	31	N-term, C-term or internal	mAb/low pH or HA peptide	HA-tag vector set (Clontech); μMACS and multiMACS HA Isolation Kit (Miltenyi Biotec)	Detection and purification	Anti-HA antibodies specific; useful in mammalian expression systems; antibody purification does not give high yields; low pH elution may irreversibly affect protein properties; matrix is of limited reusability	Tai et al. (1988)

Tag	Length (sequence)	Position	Matrix/elution	Commercial systems (supplier)	Typical use(s)	Comments	References
HaloTag®	312	N-term or C-term	HaloLink™ Resin/HaloTag® Buffer and TEV Protease	HaloTag® Protein Purification System (Promega)	Purification, increased solubility and expression	A different ligand must be purchased for each different experiment; stringent wash conditions may affect properties of fusion; works quickly and in living cells	Los et al. (2008)
Histidine affinity tag (HAT)	19 (KDHLIHNVHKEFHAHAHNK)	N-term or C-term	Divalent metal (Ni^{2+}, Co^{2+}, Cu^{2+}, or Zn^{2+})/imidazole or low pH	HAT protein expression and purification system (Clontech)	Purification	Natural protein sequence that was found to chelate divalent metals; short recognition motif; contains six histidines interspersed among other amino acids; semipure protein in one step; matrix may be regenerated and reused almost infinitely; does not bind metal affinity resin as tightly as His-tag; tag or elution conditions may affect fusion properties	Terpe (2003)
Horseradish Peroxidase (HRP)	400	Primary amines (lysine side chain epsilon-amines and N-terminal α-amines)	NA	EZ-Link Plus Activated Peroxidase and Kit (Pierce)	Detection	Chemiluminescent detection; useful for Western blotting, sandwich ELISA, and immunohistochemistry; not recommended as a protein purification tag.	Imagawa et al. (1982)
HSV epitope	11 (QPELAPED)	C-term	mAb/low pH	HSV•Tag® Kit (EMD Millipore)	Purification	C-term tag placement only; antibody purification does not give high yields; low pH elution may irreversibly affect protein properties; matrix is of limited reusability	Fritze and Anderson (2000)
Ketosteroid isomerase (KSI)	125	N-term	NA	pET-31b(+) vector (EMD Millipore)	Increased expression	Insoluble protein is targeted to inclusion bodies; usually coupled with a short affinity tag (His-tag, Strep-tag) for purification of toxic proteins, but refolding necessary	Kuliopulos and Walsh (1994)
KT3 epitope	11 (KPPTPPPEPET)	N-term or C-term	mAb/low pH	NA	Detection and purification	Short, linear recognition motif; antibody purification does not give high yields; low pH elution may irreversibly affect protein properties; matrix is of limited reusability	Kwatra et al. (1995)
LacZ	1024	N-term or C-term	APTG/high pH borate buffer	LacZ & Beta-Gal Vectors and Assay Kit (Clontech); gWIZ β-galactosidase mammalian expression vector (Genlantis)	Detection, purification, and increased expression	Also known as β-galactosidase or β-gal; enzymatic assay available for protein quantification; may increase proteolytic stability of fusion proteins to increase expression; fusion proteins may be insoluble; extremely large tag which forms tetramers in solution,	Tai et al. (1988)

Tag	Length (sequence)	Position	Matrix/elution	Commercial systems (supplier)	Typical use(s)	Comments	References
						potentially affecting properties of fusion protein potentially affecting properties of fusion protein	
Luciferase	551	N-term	NA	gWIZ luciferase mammalian expression vector (Genlantis); Luciferase assay system (Promega)	Detection	Luminescent; can serve as a reporter immediately upon translation; useful for studies involving in situ hybridization, RNA processing, RNA transfection or coupled in vitro transcription/translation, protein folding, and imaging; can be labeled with ^{35}S; no more than five codons can be removed from the N- or C-term to maintain enzymatic activity; very large tag may affect properties of fusion	Karp and Oker-Blom (1999)
Maltose-binding protein (MBP)	396	N-term or C-term	Cross-linked amylose/maltose	Vector Fusion-Aid MBP Kit (Vector Laboratories); pMAL system (New England Biolabs)	Detection, purification, increased expression, and increased solubility	One-step purification of relatively pure protein (>70%); matrix compatible with nonionizing detergents and high salt, but not reducing agents; can increase expression of eukaryotic proteins in bacteria; anti-MBP antibodies specific; tag at N-term can decrease translation efficiency; very large tag size may affect fusion protein properties	Nilsson et al. (1997b), Terpe (2003)
Myc epitope	11 (CEQKLISEEDL)	N-term, C-term or internal	mAb/low pH	pDual Expression System (Agilent Technologies); pET Express & Purify Kits (Clontech)	Detection and purification	Short, linear recognition motif; anti-myc antibody somewhat promiscuous; antibody purification does not give high yields; low pH elution may irreversibly affect protein properties; matrix is of limited reusability	Kolodziej and Young (1991)
NusA	495	N-term or C-term	NA	pET NusA Fusion System (EMD Millipore)	Increased expression and solubility	Anti-transcription termination factor; increases solubility and expression of fusion proteins; must be used in conjunction with another affinity tag for protein purification; large tag may affect properties of fusion protein	Terpe (2003)
PDZ domain	80–90	N-term, C-term, or internal	Immobilized C-terminal PDZ ligand sequence/reducing agent (InaD PDZ1)	NA	Detection, purification, and immobilization	Short, linear recognition motif; PDZ-ligand interaction can be very specific and high-affinity; reducing elution may disrupt properties of fusion protein	Kimple and Sondek (2002)
PDZ ligand	5–7 (varies)	C-term	Immobilized PDZ domain/reducing agent (NorpA C-terminus)	NA	Detection, purification, and immobilization	Short, linear recognition motif; PDZ-ligand interaction can be very specific and high-affinity; reducing elution may disrupt properties of	Kimple and Sondek (2002)

Tag	Length (sequence)	Position	Matrix/elution	Commercial systems (supplier)	Typical use(s)	Comments	References
						fusion protein; tag must be at C-terminus to bind to PDZ domain	fusion protein; tag must be at C-terminus to bind to PDZ domain
Polyarginine (Arg-tag)	5–6 (usually 5; RRRRR)	C-term	Cation exchange resin/high pH salt gradient	NA	Purification and immobilization	Can immobilize targets on mica for microscopy studies; short, linear recogition motif; very pure protein in one step; charged tag may affect tertiary structure of protein and/or protein properties; limited success of tag cleavage by carboxypeptidase B	Terpe (2003)
Polyaspartate (Asp-tag)	5–16 (DDDDD)	C-term	Anion exchange resin/low-neutral pH salt gradient	NA	Purification	Short, linear recognition motif; polar tag may affect tertiary structure of protein and/or protein properties	Stevens (2000)
Polycysteine (Cys-tag)	4 (CCCC)	N-term	Thiopropyl-Sepharo sethiol-containing reducing agent (E.G., DTT, β-ME)	NA	Purification	Short, linear recognition motif; moderately pure protein in one step; purification must be performed in absence of thiol-containing reducing agents until elution step; reducing elution may disrupt properties of fusion protein	Stevens (2000)
Polyhistidine (His-tag)	2–10 (usually 6; HHHHHH)	N-term or C-term	Divalent metal (i.e., Ni^{2+}, Co^{2+}, Cu^{2+}, or Zn^{2+})/imidazole or low pH	QIAexpress system (Qiagen); Selected pET directional TOPO, pBAD, and Gateway systems (Life Technologies)	Detection, purification and immobilization	Most common purification tag; short, linear recognition motif; one-step purification of 20%–80% pure protein, depending on fusion protein expression levels; denaturing purification possible; matrix may be regenerated and reused indefinitely; can be used to immobilize fusion to Ni-NTA SPR chip, but significant dissociation complicates data analysis; tag or elution may affect protein properties; detection antibodies highly promiscuous	Bornhorst and Falke (2000)
Polyphenylalanine (Phe-tag)	11 (FFFFFFFFFFF)	N-term	Phenyl-Sepharose/ethylene glycol	NA	Purification	Short, linear recognition motif; moderately pure protein in one step; nonpolar tag or ethylene glycol elution may disrupt properties of fusion protein	Stevens (2000)
Profinity eXact	75	N-term	Immobilized subtilisin protease/ Fluoride buffer and mature subtilisin protease	Profinity eXact™ Fusion-Tag System (Bio-Rad)	Expression and purification	Quick, one-time purification and cleavage (30 min); column can be regenerated indefinitely; mild elution conditions do not affect fusion properties; has been shown to work well with heterotrimeric G proteins	Abdulaev et al. (2005), Biao et al. (2004), Ruan et al. (2004)

Tag	Length (sequence)	Position	Matrix/elution	Commercial systems (supplier)	Typical use(s)	Comments	References
Protein C	12	N-term or C-term	mAb/Ca^{2+} buffer	Protein C Epitope Tagging System (Roche)	Detection and purification	Short, linear recognition motif; anti-PC antibody binds in Ca^{2+}-dependent manner; elution by Ca^{2+} in physiological buffer conditions; antibody purification does not give high yields	Fritze and Anderson (2000)
S1-tag	9 (NANNPDWDF)	N-term or C-term	mAb/low pH	NA	Detection and purification	Hepatitis B virus S1 region; short, linear recognition motif; AP1 antibody specific; has been tested in bacterial and mammalian expression systems; relatively pure protein in one step; antibody purification does not give high yields, low pH elution may irreversibly affect protein properties, and matrix is of limited reusability	Berlot (1999)
S-tag	15 (KETAAAKFERQHMDS)	N-term, C-term or internal	S-fragment of RNase A/low pH	S•Tag™ Purification Kit (EMD Millipore)	Detection and purification	Short, linear recognition motif; RNase S assay possible for quantitative assay of expression levels; colorimetric assays used for detection without antibody; tag or low pH elution may irreversibly affect protein properties; matrix is of limited reusability	Fritze and Anderson (2000)
Streptavadin-binding peptide (SBP)	38	C-term	Streptavadin/biotin	SBP Bacterial System Expression Vectors (Sigma-Aldrich)	Purification and immobilization	Relatively short recognition motif; immobilization of protein to various media (e.g., streptavadin-coated beads, SPR chips); tag at C-term only	Terpe (2003)
Staphylococcal protein A (Protein A)	280	N-term	IgG/low pH or IgG	pEZZ 18 and pRIT2T vectors (GE Healthcare)	Purification and increased solubility	Proteolytically stable; may increase solubility of fusion; fusion proteins secreted; purification does not give high yields; large tag size and/or low pH elution may irreversibly affect protein properties; matrix is of limited reusability	Nilsson et al. (1997b)
Staphylococcal protein G (Protein G)	280	N-term or C-term	Amylose/low pH or amylose	Protein AG Agarose and Kits (Thermo Scientific)	Purification and increased solubility	Proteolytically stable; may increase solubility of fusion; fusion proteins secreted; purification does not give high yields; large tag size and/or low pH elution may irreversibly affect protein properties; matrix is of limited reusability	Nilsson et al. (1997b)
Strep-tag	8–9 (WSHPQFEK) or (AWAHPQPGG)	N-term or C-term	Strep-Tactin (modified streptavadin)/biotin or desthiobiotin	*Strep*-tag® – *Strep*-Tactin® system (IBA GmbH); Strep-tag® Expression Vectors (Neuromics)	Detection, purification and immobilization	Short, linear recognition motif; matrix regenerable; useful for purification under anaerobic conditions, eukaryotic cell surface	Skerra and Schmidt (2000)

Tag	Length (sequence)	Position	Matrix/elution	Commercial systems (supplier)	Typical use(s)	Comments	References
						display, and immobilization to streptavadin-coated surfaces (e.g., SPR chips); s	
						display, and immobilization to streptavadin-coated surfaces (e.g., SPR chips); s	
						display, and immobilization to streptavadin-coated surfaces (e.g., SPR chips); s	
						display, and immobilization to streptavadin-coated surfaces (e.g., SPR chips); s	
						display, and immobilization to streptavadin-coated surfaces (e.g., SPR chips); s	
Streptavadin	159	N-term or C-term	Biotin/biotin or denaturation (e.g., heat, urea)	NA	Detection, purification, increased expression and immobilization	May increase proteolytic stability of fusion proteins to increase expression; extremely high affinity for biotin useful for immobilization of fusion on surfaces such as SPR chips; large size or tetramer formation may disrupt properties of fusion protein; fusion protein may not be released upon addition of free biotin, necessitating denaturing elution followed by refolding; newer streptavadin mutants that have lower affinities for biotin useful for purification	Sano et al. (1998)
Small Ubiquitin-like Modifier (SUMO)	100	N-term	Ni-NTA, NI-IMAC, or chromatography/SUMO Protease	Champion™ pET SUMO Expression System (Life Technologies); SUMOstar Kits (LifeSensors); Expresso® SUMO Cloning and Expression Systems (Lucigen)	Increased solubility and expression	Allows for cellular protein-protein interaction studies; available for bacterial, yeast, insect and mammalian expression systems; desumoylases and SUMO tags contain His6 tags to make their removal efficient; may be used for protein labeling *in vitro* as well as *in vivo*	
Tandem Affinity Purification (TAP)	Variable	N-term or C-term	Dependent on TAP variants	InterPlay Mammalian TAP System (Agilent Technologies); FLAG® HA System (Sigma-Aldrich)	Purification	Two-step purification results in highly pure protein with reduced non-specific background; mild elution conditions do not affect fusion properties; protein of interest remains in native form; use of protease unnecessary for elution; over 30 TAP variants available	Li (2011)
T7 epitope	260	N-term	mAb/low pH	NA	Purification and increased expression	May increase expression of fusion proteins; insoluble protein is targeted to inclusion bodies; denaturing purification of toxic proteins necessitates refolding	Stevens (2000)
Thioredoxin (Trx)	109	N-term or C-term	Phenylarsinine oxide/thiol-containing reducing agent (e.g., DTT, β-ME)	ThioFusion system (Life Technologies)	Increased solubility	Heat stable; may increase solubility of fusion proteins; convenient purification of crude periplasmic extract from bacteria; purification must be done in absence of thiol-containing reducing agents until elution step;	Terpe (2003)

Curr Protoc Protein Sci. Author manuscript; available in PMC 2015 August 06.

Tag	Length (sequence)	Position	Matrix/elution	Commercial systems (supplier)	Typical use(s)	Comments	References
						large tag or elution conditions may affect properties of fusion protein	
TrpE	25–336	N-term or C-term	mAb/low pH	NA	Purification and increased expression	Larger constructs expressed may be targeted to inclusion bodies (allowing high-level expression of toxic genes); antibody purification does not give high yields; low pH elution may irreversibly affect protein properties; matrix is of limited reusability; large tag may affect properties of fusion protein	Stevens (2000)
Ubiquitin	76	N-term	NA	NA	Increased solubility	May increase solubility of proteins expressed in *E. coli*; not useful for expression in eukaryotic cells	Stevens (2000)
Universal	6 (HTTPHH)	N-term, C-term or internal	mAb/low pH	NA	Detection and purification	Sequence HTTPHH is translated regardless of reading frame for ease in cloning; multiple tag copies increase antibody specificity; immobilized mAb can bind multiple tag copies in SPR studies; antibody purification does not give high yields; low pH elution may irreversibly affect protein properties; matrix is of limited reusability	Nelson et al. (1999)
VSV-G	11 (YTDIEMNRLGK)	C-term	mAb/low pH	pVPack-VSV-G Vector (EMD Millipore)	Detection and purification	C-term residues of VSV-G; relatively pure protein in one step; antibody purification does not give high yields; low pH elution may irreversibly affect protein properties; matrix is of limited reusability	