Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Research review paper

Challenges and opportunities in the purification of recombinant tagged proteins

Ana Sofia Pina a,b, Christopher R. Lowe c, Ana Cecilia A. Roque a,*

a RQUMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal
b IBET—Instituto de Biologia Experimental Tecnológica, Oeiras, Portugal

Contents lists available at ScienceDirect

Article history:
Received 28 June 2013
Received in revised form 4 December 2013
Accepted 8 December 2013
Available online 12 December 2013

Keywords:
Affinity tags
Fusion proteins
Recombinant proteins
Affinity ligands

ABSTRACT

The purification of recombinant proteins by affinity chromatography is one of the most efficient strategies due to the high recovery yields and purity achieved. However, this is dependent on the availability of specific affinity adsorbents for each particular target protein. The diversity of proteins to be purified augments the complexity and number of specific affinity adsorbents needed, and therefore generic platforms for the purification of recombinant proteins are appealing strategies. This justifies why genetically encoded affinity tags became so popular for recombinant protein purification, as these systems only require specific ligands for the capture of the fusion protein through a pre-defined affinity tag tail. There is a wide range of available affinity pairs “tag-ligand” combining biological or structural affinity ligands with the respective binding tags. This review gives a general overview of the well-established “tag-ligand” systems available for fusion protein purification and also explores current unconventional strategies under development.

© 2013 Elsevier Inc. All rights reserved.

Contents

1. Introduction ........................................................................................................................................ 367
2. Purification of recombinant fusion proteins by “tag-ligand” strategies ...................................... 367
  2.1. Biological ligands as binding partners of affinity tags ............................................................... 367
    2.1.1. Immunoglobulin-based ligands and respective tags .............................................................. 367
    2.1.2. Streptavidin-based ligands and respective tags ...................................................................... 371
    2.1.3. Other examples of protein/peptide affinity systems ............................................................ 371
    2.1.4. Carbohydrates as biological ligands for tagged proteins ..................................................... 371
  2.2. Structural ligands as binding partners of affinity tags ................................................................. 371
    2.2.1. Metal chelators as affinity ligands for His-tagged proteins .................................................. 372
    2.2.2. Ion-exchange adsorbents for tagged protein purification .................................................... 373
    2.2.3. Covalent affinity chromatography ....................................................................................... 373
  3. Tags with versatile and unconventional properties and respective purification approaches ......... 373
    3.1. Solubility enhancers ................................................................................................................... 373
    3.2. Versatile fusion partners ........................................................................................................... 373
    3.3. Affinity tags for the purification of membrane proteins ........................................................... 374
    3.4. Unconventional fusion partners ............................................................................................... 374
      3.4.1. Hydrophobic tags ............................................................................................................... 374
      3.4.2. Oil bodies ............................................................................................................................ 374
      3.4.3. Elastin-like polypeptides (ELPs) ....................................................................................... 375
      3.4.4. Hydrophobin ....................................................................................................................... 375
      3.4.5. γ-zein (ZERA) ................................................................................................................... 375
  4. Alternative schemes for tag removal ............................................................................................... 375
  5. Concluding remarks and future trends ......................................................................................... 376

References ................................................................................................................................................. 377

* Corresponding author. Tel.: +351 212948385; fax: +351 212948550.
E-mail address: cecilia.roque@fct.unl.pt (A.C.A. Roque).

0734-9750/– see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.biotechadv.2013.12.001

© 2013 Elsevier Inc. All rights reserved.
1. Introduction

The wealth of products and methodologies for recombinant protein production and purification has increased enormously in recent years. This has contributed to the growth in the use of recombinant proteins for academic research and therapeutic and diagnostic applications as well as in industrial settings (Demain and Vaishnav, 2009; Palomares et al., 2004). The production and purification of recombinant proteins are intimately linked. The choice of host for protein production affects not only the amplification and isolation of the protein, but also the way in which the product can be subsequently purified. The advances in genetic engineering have increased the availability of large amounts of recombinant proteins produced in host cells – bacterial, mammalian, insect and yeast – and where Escherichia coli still represents the most widely used platform (Demain and Vaishnav, 2009).

Chromatography is a well-established platform for protein purification, as it is considered economically feasible and yields high recoveries at high purities with very few process steps (Carta and Jungbauer, 2010; Milne, 2011; Walsh, 2003; Walter and Gottschalk, 2010). In affinity chromatography, selectivity towards a specific target protein is introduced through the chemical functionalization of the solid support with desired affinity ligands, which can be divided into three main categories: biological, structural and synthetic (Fig. 1(A)) (Roque and Lowe, 2007). Synthetic affinity ligands have been developed in an attempt to overcome disadvantages of natural and structural ligands, by combining the best of two worlds: Molecular recognition features associated with high resistance to chemical and biological degradation and high scalability as well as low production costs and low toxicity (Clonis et al., 2000; Lowe, 2001; Lowe et al., 2001). These have been tailor-made for the purification of specific biomolecules as antibodies (Haigh et al., 2009; Qian et al., 2012; Roque et al., 2005) although they are not regarded as universal purification adsorbents for fusion proteins, and therefore will not be widely discussed in this review.

The diversity of proteins and their biochemical properties makes the development of universal purification and capturing strategies difficult. Most proteins of interest lack a suitable, specific and robust affinity ligand for capture on a solid matrix. Genetically encoded affinity tags are a viable and common option for the purification of recombinant proteins and also represent important tools for structural and functional proteomics initiatives. This approach requires the existence and availability of specific ligands for the capture of the fusion protein through an encoded affinity tag tail (Fig. 2), which can be denominated as affinity “tag–ligand” pairs. Currently available affinity “tag–ligand pairs” fall within one of these categories: protein–protein, protein–small chemical ligands, peptide–protein or peptide–metal chelating ligands.

Affinity tags display different size ranges from a single amino acid to entire proteins, and can be genetically fused to the N- or C-terminal of the target biomolecule (Arnau et al., 2006; Hedhammar et al., 2005; Waugh, 2005; Young et al., 2012). Apart from facilitating the purification process, affinity tags can also enhance protein solubility and stability, increase expression levels (Hu et al., 2001; Walls and Loughran, 2011) and allow labelling for cellular localization and imaging studies (Malhotra, 2009). An overview of the main advantages and disadvantages associated with different affinity tags has been already thoroughly discussed in the literature (Arnau et al., 2006; Hearn and Acosta, 2001; Hedhammar et al., 2005; Hu et al., 2001; Malhotra, 2009; Terpe, 2003; Walls and Loughran, 2011; Waugh, 2005; Young et al., 2012). In general, shorter affinity tags (peptides) are more attractive as they are less likely to interfere with the expression, structure and function of the target protein, and their removal can be exempt, decreasing thus the overall costs of the purification process (Hearn and Acosta, 2001; Hedhammar et al., 2005; Terpe, 2003).

This review is a comprehensive overview of the well-implemented “tag–ligand” pairs used for the purification of recombinant fusion proteins with a detailed discussion of the lesser known tags with unconventional properties currently under development.

2. Purification of recombinant fusion proteins by “tag–ligand” strategies

2.1. Biological ligands as binding partners of affinity tags

Biological ligands comprise biomolecules obtained from natural sources and from in vitro selection techniques. These are usually associated with high selectivity and affinity for the target, but also with high costs of production and purification, poor stabilization under SIP (sterilization-in-place) and CIP (cleaning-in-place) conditions, and potential leakage and end-product contamination (Clonis et al., 2000; Lowe, 2001; Roque and Lowe, 2006). Examples of common biological affinity ligands include immunoglobulins against a target protein (antigen), bacterial immunoglobulin-binding domains such as Staphylococcal proteins A, G and L (Björck and Kronvall, 1984; Duhamel et al., 1979; Füglistaller, 1989; Lindmark et al., 1983; Nilson et al., 1993), and natural lectins targeting glycoproteins (Vreeland, 1976). Novel biological affinity ligands can be obtained through in vitro selection techniques such as phage, ribosome or yeast display (Smith and Petrenko, 1997), with phage display being very popular for these purposes.

A general overview of the biological ligands and respective tags employed in the purification of tagged recombinant proteins is given in Table 1 and Fig. 1(B). Biological ligands include peptides, proteins and carbohydrates.

2.1.1. Immunoglobulin-based ligands and respective tags

Immunoglobulin-based adsorbents for affinity chromatography are usually very selective for the target proteins but the costs associated tend to be high. Also, as the interaction between ligand and tag is often strong, elution traditionally involves drastic conditions, typically extremes of pH.

The first affinity “tag–ligand” pair was based on the intrinsic selectivity and affinity between the bacterial immunoglobulin-binding domain Staphylococcal protein A (SpA) and the Fc region of mammalian IgG (Nilsson and Abrahmsén, 1990). As an example, alkaline phosphatase was fused to SpA and its purification was performed in a single step using an IgC absorbent, with elution at acidic pH (Nilsson et al., 1985). It is known that SpA presents five homologous domains (E, D, A, B and C) and that IgG binds preferentially to the B domain (Nilsson and Abrahmsen, 1990). This domain was mutated to improve its resistance towards tag removal by chemical methods and denominated as the Z domain (Nilsson et al., 1987). The bacterial immunoglobulin-binding domain staphylococcal protein G (SpG) has also been studied as a fusion partner due to its bifunctional behaviour – SpG is composed of different domains (A, B, C and D) with affinity for both IgG and human serum albumin (HSA) (Akerström et al., 1987; Nygren et al., 1988) – allowing the purification of SpG tagged proteins through HSA and IgG affinity chromatography (Akerström et al., 1987; Hedhammar et al., 2005; Nilsson et al., 1997; Nygren et al., 1988).

Other affinity tags recognizing immunoglobulin-based adsorbents include the peptide epitopes FLAG, c-myc, T7, hemagglutinin antigen (HA) and Softags (Hedhammar et al., 2005; Young et al., 2012). The c-myc is a product of a proto-oncogene whose epitope presents high affinity for the monoclonal antibody 9E10 (Evans et al., 1985). This affinity pair has been mostly used as a tool for the detection of recombinant proteins through immunoblotting assays rather than for purification processes (Evans et al., 1985; Kipriyanov et al., 1996; Terpe, 2003). The same is observed for the affinity tags T7 and HA (Walls and Loughran, 2011; Young et al., 2012). The T7-tag is a leader peptide of phage T7 with eleven amino acids with affinity for the anti-T7 monoclonal antibody (Jarvik and Telmer, 1998; Studier and Moffatt, 1986), and the HA tag is a peptide epitope of the influenza virus hemagglutinin (Wilson et al., 1984) recognized by the monoclonal antibody 12 CA5 (Field et al., 1988; Foreman and Davis, 1994).

The FLAG tag technology is quite popular for purification purposes, and has been successfully employed on the purification of several
recombinant proteins such as immunoglobulins, cytokines and gene-regulatory proteins. The FLAG tag is a hydrophilic peptide with the sequence DYKDDDDK possessing high affinity for the monoclonal antibodies M1 and M2 (Einhauer and Jungbauer, 2001). As the binding between the FLAG tag and M1 adsorbents is calcium dependent, the elution of bound tagged protein can be carried out under mild conditions in the presence of metal chelators such as ethylenediaminetetraacetic acid (EDTA) and ethylene glycol tetraacetic acid (EGTA), at the expense of risking calcium co-elution. On the contrary, elution from M2 resins usually requires a decrease in pH, which in some cases can be detrimental for the protein of interest (Allen et al., 2003; Prickett et al., 1989). An attractive feature of FLAG tag technology is the possibility of simultaneous purification and removal as enterokinase recognizes the DDDDK sequence (Einhauer and Jungbauer, 2001; Hopp et al., 1988). Lately, two additional FLAG sequences were added in tandem to the original FLAG tag, the triple FLAG tag (3 × FLAG), which showed an increased detection limit of the target protein, favoured binding to the M2 antibody, and the possibility to recover the protein by competitive elution with 3 × FLAG peptide (Hernan et al., 2000). The 3 × FLAG technology has been employed in the purification of a bioactive metabolite (Ueda et al., 2011) and full-length human huntingtin (Li et al., 2006).

The epitope tags referred to as Softags are recognized by polyol-responsive monoclonal antibodies (PR-mAb). Proteins tagged with Softags can be eluted under mild conditions employing low molecular weight polyols as competing agents (Burgess and Thompson, 2002; Duellman et al., 2004). This gentle elution protocol, when compared to other immunoaffinity chromatography methods, contributes to the maintenance of the biological activity and structural integrity of the eluted target protein (Thompson et al., 2009). There are three Softags. Softag 1 is a thirteen amino acid sequence near the C-terminal of the

Fig. 1. Examples of (A) affinity ligands and (B) peptide and protein affinity tags with their respective biological ligands employed on the purification of fusion proteins based on affinity chromatography. (A) The common affinity ligands can be (i) a biological ligand (staphylococcal protein A domain, PDB: 1DEE), (ii) a structural ligand (metal chelate such as iminodiacetic acid cheated to Ni2+) and (iii) a synthetic biomimetic ligand (ligand A3C1 specific for immunoglobulins (Haigh et al., 2009). The solid support is representing agarose beads (○). (B) The (i) peptide tag is the Strep-tag, an eight amino acid sequence, with the affinity for streptavidin protein (PDB: 1RST), whilst the (ii) example of a protein used as an affinity tag is related with the staphylococcal protein G and the respective biological ligand, immunoglobulin G (PDB: 1FCC).
β′ subunit of *E. coli* RNA polymerase which can be recognized by the PR-mAb NT73 (Thompson et al., 2003a,2003b). This epitope tag is not conserved in all bacterial species. Other PR-mAb's such as 8RB13 were used for the purification of *Shewanella oneidensis* core RNA polymerase with high yield and purity (Probasco et al., 2007). Softag 2 is a heptapeptide repeat found on the C-terminal of eukaryotic RNA polymerase II, which is a highly conserved domain from all almost species and presents strong reactivity for PR-mAb8WG16 (Edwards et al., 1990; Thompson et al., 2009); Softag 3 is an epitope tag with an eight amino-acid sequence located near the N-terminal of human transcription factor IIB, and interacts with PR-mAb IIB8, an antibody mapped by phage display and site-directed mutagenesis (Duellman et al., 2004).

Fig. 2. Overview of a recombinant fusion protein purification process through the use of affinity tags fused to the target protein by conventional methods such as affinity chromatography and alternative methods based on inverse transition cycling (ITC). Both processes comprise several stages; (i) Fragment DNA construction of the fusion protein, where the fragment DNA which encodes the affinity tag is fused to the nucleotide sequence of the target protein; (ii). The fusion protein is expressed in the selected host and (iii) purified by different methods. The most conventional method involves the use of affinity chromatography, where protein capture is performed through the molecular recognition between the tag and the ligand. An alternative method for the purification of fusion proteins is based on ITC, which exploits the reversible soluble–insoluble phase transition behaviour of the affinity tags and the desired fusion protein yield can be achieved with n ITC rounds. Afterwards, in both methods, the tag can be removed leading to a pure target protein.
## Table 1
Overview of biological ligands employed as binding partners of affinity tags. The "Target protein" is a representative case and the reference listed is for the example target protein.

| Biological ligand | Tag | Tag size | Tag sequence | Elution condition | Yield | Purity | Representative Target Protein | Application | Reference |
|-------------------|-----|----------|--------------|-------------------|-------|--------|--------------------------------|-------------|-----------|
| **Peptides and Proteins** | | | | | | | | | |
| Glutathione GST | 26 kDa | – | | 20 mM reduced glutathione | 100 | 80 | DNA topoisomerase type II | Purification and solubility | Singh et al. (2011) |
| Calmodulin CBP | 26 aa | KRRWKRFNIAVAASNFKKIISSGAL | | 2 mM EDTA | n.a. | High | Nicotinamide nucleotide transhydrogenase | Purification | Egorov et al. (2004) |
| S protein (5’-fragment RNAse A) | 15 aa | KETAAKFGERHMD5 | Proteolytic cleavage | Fair | 90 | High | Recombinant human interleukin-20 | Purification | Li and He (2006) |
| **Streptavidin** | | | | | | | | | |
| Strept-tag | 9 aa | SAWKHPQFG | 1 mM iminobiotin | n.a. | High | High | Bovine heart fatty acid-binding protein | Detection and purification | Schmidt and Skerra (1993) |
| Nano tag | 9–15 aa | DVEAWLGR | 2 mM β-mercaptoethanol | 90 | 71 | High | High | Maltose-binding protein | Detection and purification | Keefe et al. (2001) |
| **Strep-Tactin (modified streptavidin)** | | | | | | | | | |
| SBP | 38 aa | MDEKTTGWRRGHVVEGLAGELEQRLEHHPQGQREP | 2 mM biotin | High | High | High | Maltose-binding protein | Detection and purification | Keefe et al. (2001) |
| Strep tag II | 8 aa | WSHHPQFEK | 2.5 mM D-desthiobiotin | n.a. | 99 | High | Human tissue transglutaminase | Detection and purification | Schmidt and Skerra (2007) |
| **NeutraAvidin** | | | | | | | | | |
| AviD-tag | 6 aa | DRATPY | 500 μM biotin | High | High | High | Green fluorescent protein (CFP) | Purification | Gaj et al. (2007) |
| IgG | Protein A (SpA) | 14–31 kDa | – | Glycine-HCl pH 3 | 95 | High | β-galactosidase | Purification | Nilsson et al. (1985) |
| | Z domain | 7 kDa | – | 0.3 M acetic acid, pH 3.1 | High | High | Klenow fragment of DNA polymerase I | DNA polymerase | Nilsson et al. (1994) |
| IgG/HSA | Protein G (SpG) | 28 kDa | – | 85°C 10 minute incubation | n.a. | 90 | DNA polymerase | DNA polymerase | Graslund et al. (1997) |
| Monoclonal antibody M1, M2 | FLAG | 8 aa | DYKDDDDK | 150 mM glycine-HCl pH 3.5 | 60 | 95 | Metal response element binding transcription factor-1 (MTERF-1) | Detection and purification | Huyck et al. (2012) |
| Monoclonal antibody 9E10 | c-myc | 10 aa | EKQISEEDL | Western blot (detection) | – | – | Tobacco etch virus (TEV) protease | Purification | Geisbrecht et al. (2006) |
| Monoclonal antibody T7 | T7 tag | 11 aa | MASMTGGQOMG | 0.1 M citric acid, pH 2.2, 5 mM glycerophosphate, 5 mM KF | High | n.a | SR protein | Detection and purification | Cazalla et al. (2005) |
| Monoclonal antibody 12 CA5 | HA tag | 9 aa | YPYDVDPYA | 0.85 M KCl | n.a. | High | Transcription factor IID | Detection and purification | Carey et al. (2010) |
| Polyspecific responsive monoclonal antibodies | Softag 1 | 13 aa | SLAEILNAGLCC | 0.7 M NaCl and 30% propylene glycol | n.a. | High | GFP | Detection and purification | Thompson et al. (2003a, 2003b) |
| | Softag 2 | 14 aa | PISPSYPSTPSYS | 0.5 M ammonium sulphate and 30% ethylene glycol | n.a. | High | RNA polymerase II | Detection and purification | Thompson et al. (1990) |
| | Softag 3 | 8 aa | TKDSPRVG | 75 M ammonium sulphate and 40% propylene glycol | n.a. | n.a | GFP | Detection and purification | Duellman et al. (2004) |
| **Carbohydrates** | | | | | | | | | |
| **Cross-linked amylose** | | | | | | | | | |
| MBP | 42 kDa | – | | | | | Tyrosine hydroxylase | Purification and solubility | Higgins et al. (2012) |
| Cellulose binding protein | Cellulose | 100 kDa | – | Pure water | 80 | High | Red fluorescent protein | Purification | Wan et al. (2011) |
| Chitin binding domain | Chitin | 5 kDa | – | Thiol induced self-cleavage | n.a. | 99 | Phosphite Dehydrogenase | Purification | Guan et al. (2013) |
| **Starch** | | | | | | | | | |
| Starch binding domain | 133 aa | – | | 10 mM glycine-NaOH pH 11 | 86 | 90 | Enhanced GFP | Purification | Lin et al. (2009) |

1. Yield—Recovery of pure protein; 2. Purity—The purity of eluted proteins was evaluated by SDS-PAGE electrophoresis; GST—Glutathione-S-transferase; CBP—calmodulin binding peptide; SPB—streptavidin-binding peptide; IgG—immunoglobulin G; HSA—human serum albumin; HA—hemaglutinin antigen; MPB—maltose binding protein.
2.1.2. Streptavidin-based ligands and respective tags

Another important group of biological affinity ligands includes streptavidin. Here, the tag-ligand system is based on the selective natural binding pair avidin–bixin with a dissociation constant of 10^{-15} M (Bayer and Wilchek, 1990). The first reported tag with affinity for streptavidin was Strep-tag, a nine amino acid peptide selected from a random peptide library (Schmidt and Skerra, 1993), generally fused to the C-terminal of the target protein. According to structural studies, this affinity tag binds to the same pocket as bixin (Schmidt et al., 1996). The main advantages of Strep-tag are related to the resistance to proteolysis in vivo, the lack of interference with expression in E. coli and the mild elution conditions employed which include competitive elution with a bixin or analogue compound such as iminobixin (Schmidt and Skerra, 1993). However, a limitation of the Strep-tag system is the restriction to C-terminal fusion, and an improved version was developed and termed Strep-tag II with an equilibrium dissociation constant of 37 × 10^{-9} M (Schmidt et al., 1996). Another engineered streptavidin chromatographic support (Strep-Tactin) was developed for improved binding capacity towards Strep-tag II. This new version of streptavidin was generated from random mutagenesis on the amino acids 44–47 located at the flexible loop region near the binding site (Koonndorfer and Skerra, 2002). Although the affinity pair Strep-tag II–Strep-Tactin has been extensively used, affinity tags with higher affinity for streptavidin matrices were still investigated: The streptavidin-binding peptide (SPB) tag comprising a 38 amino acid sequence with an equilibrium dissociation constant of 2.5 × 10^{-18} M was selected from a peptide library screened against immobilized streptavidin (Keefe et al., 2001); A Nano-tag combines the small size of Strep-tag II and presents a nanomolar dissociation constant of 4 × 10^{-9} M for binding to streptavidin (Lamla and Erdmann, 2004). The latter is a 15 amino acid peptide selected from a synthetic library based on the heart fatty-acid binding protein by ribosome display (Lamla and Erdmann, 2003). The AviD-tag is another affinity tag based on the Strep-tag that displays affinity for neutAvidin, a neutral form of avidin (Gaj et al., 2007). The AviD-tag is composed by a 6-amino acid cyclic peptide selected through phage display with a dissociation constant of 12 × 10^{-6} M for both neutAvidin and avidin (Gaj et al., 2007).

2.1.3. Other examples of protein/peptide affinity systems

Calmodulin is a small (17 kDa) and stable peptide employed as an affinity ligand on the purification of calmodulin binding peptide (CBP) fusion proteins. CBP is a 26 amino acid sequence derived from the carboxy-terminal of rabbit skeletal muscle myosin light chain kinase, displaying a dissociation constant of 10^{-9} M for calmodulin (Melkko and Neri, 2003; Stofkoh-Hahn et al., 1992; Vaillancourt et al., 2000). CBP has been fused at the N-terminal of recombinant proteins that were further produced at high levels of expression. The size and stability of calmodulin ligands allow the production of calmodulin affinity resins at affordable prices. In addition, as the interaction between calmodulin and CBP is calcium dependent, protein elution can be carried out under mild conditions with calcium chelating agents such as EDTA and EGTA (Melkko and Neri, 2003; Stofkoh-Hahn et al., 1992; Vaillancourt et al., 2000).

Another example includes the S peptide tag and S-protein derived from the enzymatic cleavage of ribonuclease A by the protease subtilisin (Karpeisky et al., 1994). The S-peptide, a fifteen amino acid sequence, is used as an affinity tag recognized by the S-protein immobilized on the resin. The major advantages are related to the small size of the affinity tag and the nanomolar dissociation affinity constant of the affinity pair (Kim and Raines, 1993).

Glutathione-S-transferase (GST) is a monomeric protein (26 kDa) originating from Schistosoma japonicum which belongs to a family of enzymes that catalyse the reaction between a nucleophile, reduced glutathione, and electrophilic compounds (Boyer, 1989; Smith and Jonhson, 1988). Glutathione acts as the ligand and the complementary GST tag acts both as a purification anchoring point and as a solubility/stability enhancer, which justifies the popularity of this affinity pair system (Frangioni and Neel, 1993; Smith, 2000). In addition, GST resins are cost effective and the elution conditions are mild as reduced glutathione can be employed as a competitive agent (Smith and Jonhson, 1988). However, if reducing conditions are not guaranteed, the fusion protein can undergo oxidative aggregation due to the existence of four cysteine residues exposed at the surface of GST tag (Kaplan et al., 1997). Furthermore, fusion proteins with higher molecular weights than 100 kDa can lead to partially or completely insoluble proteins (Frangioni and Neel, 1993).

2.1.4. Carbohydrates as biological ligands for tagged proteins

Biological ligands based on carbohydrates include cross-linked amyllose, cellulose, and chitin, among others. The greatest advantages of these affinity ligands are that carbohydrates display high affinity for their respective tags, the affinity matrices are inexpensive and are prepared in a simple manner, which contributes to the ease of process scale-up. In addition, polysaccharides represent extremely interesting adsorbents for affinity based purification as they tend to be non-toxic, inert and stable (Kurek et al., 2009; Luojing et al., 1991; Tomme et al., 1998).

The maltose-binding protein (MBP) is a 42 kDa periplasmic protein engaged in the transport of maltose and maltodextrins across the bacterial cytoplasmic membrane (Kellermann and Ferenci, 1982; Nikaido, 1994). Fusion proteins containing the MBP tag fused on the N-terminus can be purified in one-single step using a cross-linked amyllose affinity resin (Guana et al., 1988). The MBP tag presents high affinity towards amyllose, a maltose analogue, with a similar dissociation constant (10^{-3} M) which is often employed as the competitive elution agent (Kellermann and Ferenci, 1982; Miller et al., 1983). This presents one of the greatest advantages of this affinity pair because the elution of the target fusion protein can be carried out under mild conditions with 10 mM of maltose (Guana et al., 1988). In addition, the MBP tag is an extremely efficient solubilizing agent when compared to GST and thioredoxin (Trx) (Karapet and Waugh, 1999), and does not contain cysteine residues and thus does not interfere with disulfide bonds of the target protein (Guana et al., 1988; Waugh, 2005).

The cellulose and chitin binding proteins present high affinity for the polysaccharides, cellulose, and chitin respectively, which act as affinity ligands for purification purposes. Cellulose binding proteins can be found in a wide range of carbohydrates, displaying dissociation constants in the low micromolar range for cellulose supports, and varying in size from small (33–36 amino acids) to large domains with more than 180 amino acids (Tomme et al., 1998). For the particular application of fusion protein expression and purification, smaller domains are preferred. Cellulosic supports are commercially available in a wide range of geometries including fibres, beads, membranes and hydrogels, with a diverse range of porosities and functionalities. Chitin binding domains present high affinity for chitin, the second most abundant polymer after cellulose, and are part of the enzymes involved in the bio-degradation of chitin and chitosan (Kurek et al., 2009). Starch binding domains have also been explored as affinity tags. These domains were fused to the N-terminal of β-galactosidase which was then purified by using starch granules (Luojing et al., 1991).

2.2. Structural ligands as binding partners of affinity tags

Structural ligands are produced at affordable prices with relatively facile chemistries, and possess high resistance to harsh conditions. These ligands comprise ion-exchange (Gräslund et al., 2000; Hedhammar et al., 2004, 2006), hydrophobic (Cummins and O’Connor, 2011), metal chelate (Charlton and Zachariou, 2007), thiophilic (Porath, 1987; Roque and Lowe, 2006, 2007), boronate (Borlido et al., 2012; Bouriotis et al., 1981; Liu, 2006; Liu and Scouten, 1994, 2000), and mixed-mode affinity ligands (Brockier et al., 2008; Burton et al., 1997; Mant and Hodges, 2008; Yon, 1981; Zhao et al.,...
Multimodal affinity chromatography relies on multiple forms of interactions and the respective mixed-mode affinity ligands usually combine a hydrophobic moiety with ionic and hydrophilic groups (Zhao et al., 2009). However, their limited selectivity and low affinity for specific targets usually necessitates the design of multi-step purification schemes.

Structural affinity ligands, mainly cation and anion-exchange, metal chelate and hydrophobic ligands are extremely popular as binding partners of affinity tags and are widely employed for the affinity purification of fusion proteins (Table 2).

2.2.1. Metal chelators as affinity ligands for His-tagged proteins

The most implemented structural ligands are metal chelates complementary to the His-tag (6 × His). Purification of the fusion protein is performed by immobilized metal affinity chromatography (IMAC) which has been comprehensively reviewed in the literature (Block et al., 2009; Charlton and Zachariou, 2007; Cheung et al., 2012; Gaberc-Porekar and Menart, 2001; Porath, 1992). The concept of IMAC was developed by Porath and co-workers in 1975, through the exploitation of the affinity between proteins and heavy metal ions (Zn2+, Cu2+, Ni2+), in particularly zinc and copper metal which strongly absorb polypeptides containing histidine and cysteine residues in aqueous solutions (Porath et al., 1975). Based on this principle, the same authors coupled iminodiacetic acid (IDA) on to agarose to bind metal ions and further chelate polypeptides through histidine and cysteine residues (Porath and Olín, 1983). Later, nitrilotriacetic acid (NTA), another highly stable quadridentate chelating adsorbent, was found to bind strongly to Ni2+ and Cu2+ and to further chelate adjacent histidine residues (Hochuli et al., 1988). NTA technology was then reported for the production and purification of a recombinant protein. The hexapeptide 6 × His was genetically fused to the N-terminal of mouse dihydrofolate reductase protein, produced in E. coli and then purified by IMAC using a Ni2+-NTA adsorbent. Subsequently, protein elution was attempted by pH gradient (pH 8–5) in the presence of denaturing agents (e.g. guanidine-HCl), and the tag was further removed by carboxypeptidase A cleavage (Hochuli et al., 1988). However, the major drawback of this pioneer purification study was the employment of denaturing conditions, which triggered the development of the procedure by conducting binding under native conditions and elution with low concentrations of a competitor imidazole or EDTA (Janknecht et al., 1991; Westra et al., 2001).

The compatibility of His-tag with denaturing agents can be a benefit for recombinant proteins produced as inclusion bodies, as they require steps of solubilization with chaotropic agents prior to binding to the column (Hochuli et al., 1988). The main disadvantage of IMAC adsorbents is the metal leaching that contributes to end-product contamination and additional steps of purification with increased costs. In addition, IMAC adsorbents can bind to native histidine patches in proteins or to other amino acids (glutamic and aspartic acid, tyrosine, cysteine, arginine, and methionine) resulting in low selectivity. Other disadvantages are related to evidence that the imidazoyl side chain groups of histidine can interfere with protein expression and folding and also mask the biological activity of the protein (Hearn and Acosta, 2001). Nonetheless, the universe of applications for the IMAC/His-tag system is widely used for protein purification and is extended to matrix-assisted refolding (Dashivets et al., 2009), detection and immobilization, protein microarrays (Kato et al., 2005; Wegner et al., 2003; Wilson and Nock, 2002), and to the development of other affinity tag technologies.

The HAT tag is one of these examples and comprises a polyhistidine peptide sequence that can be fused on either ends of the target protein. The purification can be carried out under the same conditions as for Histag; however, the adsorbent is based on carboxymethylaspartate (CMA) complexed with Co2+ with a very high capacity (Chaga et al., 1999). Novel metal chelating agents based on 1,4,7-triazacyclononane (TACN) complexed with Ni2+ were developed in an attempt to overcome the limitations of conventional IMAC matrices such as leaching.
and lack of stability and selectivity (Mooney et al., 2011). The respective affinity tags correspond to random heptapeptide sequences displaying multiple histidine, tryptophan, and/or tyrosine residues selected from a phage display library (Mooney et al., 2011). The selected peptide tag (HHHNSWDD) was fused to the N-terminal of green fluorescent protein (GFP), produced in E. coli and then purified under physiological conditions with high salt concentrations. Recovery was executed with imidazole due to the presence of three histidine residues (Mooney et al., 2011).

### 2.2.2. Ion-exchange adsorbents for tagged protein purification

Ion-exchange chromatographic resins present affinity for several tags comprising of charged amino acids through the establishment of electrostatic interactions with opposite charges. The polyarginine-tag (polyArg) is such an example and consists of six arginine residues with high retention on cationic exchange resins. A great advantage of using cation-exchange chromatography for the purification of recombinant proteins produced in E. coli is that most host proteins are neutral or acidic, which can contribute to the reduction of non-specific interactions at physiological and alkaline pH values (Gräslund et al., 2000; Hearn and Acosta, 2001). The first reported example was the production and two-step purification of human urogastrone fused to the polyArg tail. The tagged protein eluted under a NaCl linear gradient at alkaline pH, with 44% yield and purity higher than 95% (Sassenfeld and Brewer, 1984). The polyArg-tag has also been successfully employed for on-column refolding studies of tagged proteins produced as inclusion bodies (Kweon et al., 2004; Stempfer et al., 1996). However, the polyArg-tag/cation-exchange affinity affinity pair shows low selectivity and the tag promotes aggregation of the protein due to charge–charge interactions (Hedhammar et al., 2005). Poly-anionic affinity tails from one to eight glutamic residues were used for the anion-exchange chromatography based purification of functional human growth hormone (Dalboge et al., 1987) and virus-like particles (Stubenrauch et al., 2000) produced in E. coli. Negatively charged affinity tags based on aspartate residues also yielded stable fusion proteins with unaltered activity (Zhang et al., 2001).

The Z domain, a variant of the B domain of SpA, has been also engineered to create a highly charged Z domain -- Zbasic -- through the introduction of charged residues at specific positions, which allowed the interaction with negatively charged matrices under physiological conditions (Gräslund et al., 2000; Hedhammar et al., 2006). The Z domain proved to be an excellent protein scaffold for engineering as it is soluble, does not display disulfide bonds and presents reversible folding after exposure to chaotropic agents (Hedhammar et al., 2006). Due to its excellent properties, the Zbasic was explored for matrix assisted purification of proteins solubilised with chaotropic agents after being produced as inclusion bodies. The Z domain was also on the basis of Zaacidic, engineered to be negatively charged and therefore to be purified by anion-exchange chromatography (Hedhammar et al., 2004).

### 2.2.3. Covalent affinity chromatography

The vast majority of the affinity “tag-ligand” pairs exploit the reversible binding interaction between a tag and its respective ligand. Recently, a novel pair based on the covalent linkage between a tag and a synthetic ligand has been developed. This novel system is composed by the Halo-tag and a chloroalkane linker. The Halo-tag was rationally designed based on a bacterial haloalkane dehalogenase (DaH) derived from a Rhodococcus spp. to bind covalently to synthetic ligands with a chloroalkane linker. Halotag7 appeared as an improved version of Halotag created to enhance protein expression, solubility and structural compatibility but with the same ability to bind covalently and irreversible to a chloroalkane linker (Ohana et al., 2008). The covalent bond formed between the Halo-tags and the linker is highly specific, irreversible, stable and occurs rapidly under physiological conditions, which facilitates the attachment to a variety of molecules or solid surfaces containing diverse functional groups (Los and Wood, 2006; Los et al., 2008; Ohana et al., 2009). These proteins were therefore explored in protein labelling technology for cell imaging and protein analysis (Los et al., 2008), immobilization (Taniguchi and Kawakami, 2010) and purification of recombinant proteins (Chumanov et al., 2011; Ohana et al., 2011). However, the formation of covalent bonds between the tag and the ligands makes the re-use of the solid support difficult, which is particularly problematic in protein purification.

### 3. Tags with versatile and unconventional properties and respective purification approaches

The affinity tags described so far were developed to be the anchoring point for purification by affinity chromatography where the existence of a complementary affinity ligand, immobilized in a separation matrix, is required. However, tags alone can be auto-sufficient for enhanced protein solubility and purification. This section will describe tags with versatile and unconventional properties that can be combined with alternative purification protocols.

#### 3.1. Solubility enhancers

The production of recombinant proteins in bacterial hosts can lead to the formation of protein aggregates with unfolded or partially folded structures (Demain and Vaishnav, 2009; Lange and Rudolph, 2008), termed inclusion bodies (IBs) (García-Fruitós et al., 2012). When these proteins are fused to solubility tags, solubility is improved during protein expression, although the mechanism is not fully understood (Walls and Loughran, 2011; Waugh, 2005; Young et al., 2012). Thioredoxin A (TrxA) (Katti et al., 1990), small ubiquitin-related modifier (SUMO) (Li and Hochstrasser, 1999; Marblestone et al., 2006), and N-terminus inactivation (NusA) (Gusarov and Nudler, 2001; Liu and Hanna, 1995) are examples of widely employed tags with the purpose of increasing the solubility of the fusion partner rather than acting as a handle for further purification. As a consequence, a purification tag must be also added to the fusion partner to facilitate purification through affinity chromatography. The TrxA is a small (12 kDa) oxidoreductase protein with intrinsic thermal stability, solubility and robust folding properties that when fused to a target protein contributes to high yields of soluble protein (LaVallie et al., 1993, 2000). The SUMO protein (11 kDa) is involved in post-translational modification in eukaryotic cells, enhances solubility and stability of the binding partner when fused to the N-terminal (Marblestone et al., 2006), and allows simultaneous cleavage as it is recognized by the SUMO protease (Saccharomyces cerevisiae Ulp1) at the conserved dipeptide glycine–glycine (Li and Hochstrasser, 1999). However, when used in eukaryotic hosts, the affinity tag can be cleaved due to the presence of SUMO proteases in vivo (LaVallie et al., 2000; Malakhov et al., 2004). The NusA is a 55 kDa transcription elongation and anti-termination factor, also known to increase the solubility of fusion proteins, however due to their large size can contribute to a high metabolic burn for the host cell (Harrison, 2000).

#### 3.2. Versatile fusion partners

In order to combine several functionalities, a multipurpose peptide tag with the sequence HYDHYD which comprises a bi-repetition of the tripeptide constituted by a histidine, tyrosine and aspartate residues was developed (Becker et al., 2008). This peptide can be fused to the N- or C-terminal of the target protein enabling the purification by a combination of several techniques including IMAC, IEX, aqueous two phase system (ATPS) and hydrophobic interaction chromatography (HIC). This versatile and multifunctional affinity tag was employed for the purification of GFP, lactate dehydrogenase and human haemoglobin (Becker et al., 2008).
Novel dual-tagging methods have been also developed and the concept of tandem affinity chromatography (TAP) has emerged as a generic platform for complex protein purification combined with mass spectrometry, for the isolation, identification and characterization of protein partners in multi-protein complexes (Günzl and Schimanski, 2009; Puig et al., 2001).

The original C-terminal TAP tag was developed by Puig et al by using S. cerevisiae hosts and consists of two IgG binding domains of SpA and a CBP, separated by a tobacco etch virus (TEV) protease cleavage site. The N-terminal version was also developed, being an inverse version of C-terminal (Puig et al., 2001). The main advantage related to this strategy is the highly efficient recovery of target protein (Günzl and Schimanski, 2009; Puig et al., 2001; Xu et al., 2010). Nowadays, the TAP strategy is successfully employed in protein–protein interaction studies in prokaryotic and eukaryotic cells (Ahmed et al., 2011; Günzl and Schimanski, 2009; Ma et al., 2012; Müller et al., 1998; Xu et al., 2010).

A wide variety of affinity TAP tags emerged recently by using different combinations or repetitions of single affinity tags, the most commonly used being SpA, CBP, His tag, HA, FLAG, SBP and Strep-tag II (Li, 2010). Examples of combinations of affinity tags comprise a tandem histidine and FLAG epitope tag for the purification of a Candida albicans septin complex, a complex of cytoskeletal proteins (Kaneko et al., 2004). In addition, combinations of S-tag and Strep-tag II separated by HRV 3C protease cleavage were used for the isolation of protein complexes such as immune cell proteins, adhesion and degranulation promoting adapters (ADAP) (Lehmann et al., 2009). Other examples enclose combinations of streptavidin-binding peptide (SBP) and calmodulin-binding peptide as TAP tag to isolate and identify proteins directly after purification (Schimanski et al., 2009; Puig et al., 2001). The main advantage related to this strategy is the highly efficient recovery of target protein and a decrease of recovery yield on purification stage are observed (Lienqueo et al., 2008). It has been reported that as the length of the tag increases, a negative effect on cellular growth, stability of the plasmids, secretion efficiency of the target protein and a decrease of recovery yield on purification stage are observed (Lienqueo et al., 2008). According to Collén et al., during protein expression, fully exposed hydrophobic tags can reduce target protein solubility, compromising the secretion pathway and leading to the formation of insoluble protein aggregates (Collén et al., 2001) and the chance that the target protein will be membrane-associated increases (Febxy and Bülow, 2004). Additionally, it was found that tags with higher tryptophan content in the peptide (n = 6) are more prone to proteolytic cleavage when produced in E. coli and fungi hosts (Collén et al., 2001). Therefore, different combinations comprising tryrosine and proline (YP)n have been exploited, as tryrosine presents a less hydrophobic character than tryptophan. Febxy et al have developed a peptide tag with three tyrosine and proline residues fused to GFP which increased protein expression and protein recovery yield by ATPS based purification. The main advantages of using hydrophobic tags with APTS systems rather than with affinity chromatography based methods rely on the combination of rapid separation and high capacity with inexpensive costs (Febxy et al., 2004).

3.4.1. Hydrophobic tags

Examples of hydrophobic tags for recombinant protein purification include tags based on tryptophan and tyrosine, exploited for aqueous two phase separation (ATPS) (Febxy and Bülow, 2004), and tags based on repetitions of the dipeptide tryptophan–proline (WP)n (Lienqueo et al., 2008). It has been reported that as the length of the tag increases, a negative effect on cellular growth, stability of the plasmids, secretion efficiency of the target protein and a decrease of recovery yield on purification stage are observed (Lienqueo et al., 2008). According to Collén et al., during protein expression, fully exposed hydrophobic tags can reduce target protein solubility, compromising the secretion pathway and leading to the formation of insoluble protein aggregates (Collén et al., 2001) and the chance that the target protein will be membrane-associated increases (Febxy and Bülow, 2004). Additionally, it was found that tags with higher tryptophan content in the peptide (n = 6) are more prone to proteolytic cleavage when produced in E. coli and fungi hosts (Collén et al., 2001). Therefore, different combinations comprising tryrosine and proline (YP)n have been exploited, as tryrosine presents a less hydrophobic character than tryptophan. Febxy et al have developed a peptide tag with three tyrosine and proline residues fused to GFP which increased protein expression and protein recovery yield by ATPS based purification. The main advantages of using hydrophobic tags with APTS systems rather than with affinity chromatography based methods rely on the combination of rapid separation and high capacity with inexpensive costs (Febxy et al., 2004).

3.4.2. Oil bodies

An alternative affinity system to enable purification of tagged recombinant proteins relies on the interaction between oil bodies (OBs) and oleosin fusion proteins. OBs are spherical structures that contain a triacylglycerol (TAG) matrix surrounded by a phospholipid monolayer and several proteins, mainly oleosins, caleosins and stereoleosins (Bhatia et al., 2010; McLean et al., 2012; Peng et al., 2004; Roberts et al., 2008). Oleosin, the main structural protein of the oil bodies’ membrane, acts as an affinity tag. Oleosin fusion proteins have been produced in plants, which also presents several advantages over bacterial hosts such as simple handling, cost-effectiveness and easy scalability, as well as effective polypeptide folding (Bhatia et al., 2010). However the main disadvantages of this host are related to regulatory demands and production time (Wilken and Nikolov, 2012). Oleosin fusion proteins are targeted to naturally occurring OBs in the seed, and purified by several washing/centrifugation steps, as the OBs along with oleosin fusion protein are easily separated by floating (Bhatia et al., 2010; Roberts et al., 2008). Examples of proteins purified by this technology include recombinant human insulin, revealing a high protein recovery (Nykiforuk et al., 2006). Apart from naturally occurring oil bodies,
3.4.3. Elastin-like polypeptides (ELPs)

Elastin-like polypeptides (ELPs) are composed by repetitions of the pentapeptide VPCGX (where X is any amino acid, with the exception of proline) and undergo a reversible phase transition under a certain temperature, which is designated as lower critical solution temperature (LCST), also known as inverse transition temperature (Tt). Below transition temperature (Tt), ELPs are highly water soluble whereas at temperatures higher than Tt, they aggregate and become insoluble (Floss et al., 2010; Hassouneh et al., 2012; Meyer and Chilkoti, 1999). Due to their unique properties, ELPs have been used as tags, and therefore successfully employed on purification of recombinant fusion proteins (Banki et al., 2005; Meyer and Chilkoti, 1999). The purification of ELP fusion proteins can be carried out through a simple method termed inverse transition cycling (ITC), which allows the purification of fusion proteins by cycling the crude lysate through the insoluble and soluble phases (Hassouneh et al., 2012). Factors such as higher ionic strength, polymer concentration and length can decrease LCST as well as significant hydrophobic surface areas on the target protein. On the other hand, a higher charged surface area can increase LCST (Hassouneh et al., 2001).

The ITC method (Fig. 2) is an alternative strategy for the purification of recombinant proteins without the use of chromatographic resins (Meyer and Chilkoti, 1999). This simple strategy comprises cycles of “hot spin” and “cold spin”, differing mainly in the temperature at which the centrifugation step is performed. The “hot spin” is related to the centrifugation step performed after the aggregation of ELP fusion proteins from cell lysate. Aggregation can be triggered by an increase of temperature above Tt or an increase of salt concentration to decrease Tt; the increase of salt concentration is usually preferred due to protein denaturation. This step allows the separation of soluble host protein contaminants and recovery of the pellet containing the insoluble ELP fusion protein which is further used on the following step termed “cold spin”. This step consists of the transition of the ELP fusion proteins to a soluble form by dissolving the pellet at low temperature and ionic strength buffer so that the insoluble contaminants can be further removed by centrifugation. The supernatant will contain the purified ELP fusion protein (Hassouneh et al., 2001, 2012). The number of ITC rounds depends on the properties of the target protein and final purity desired. This sample methodology has been already extended for the purification of several recombinant proteins (Table 3).

3.4.4. Hydrophobin

Hydrophobins are small surface proteins (7–12 kDa) enrolled in numerous functions on filamentous fungi life cycle. The unique properties of these small proteins are related to their amphiphilic nature and self-assembly capacity (Mustalhti et al., 2013). There are two classes (I and II) of hydrophobins differing on the solubility of the formed aggregates and the amino acid composition. The hydrophobin classes I and II do not present as much identity in terms of amino acid composition but present the same pattern of eight conserved cysteine residues (Linder, 2009). In terms of aggregates solubility, class I forms more stable aggregates than class II and therefore, in the first case, the aggregates are only able to dissociate under the presence of strong conditions (e.g. formic acid and trifluoroacetic acid) whilst the aggregates formed by class II hydrophobins are much easier to dissolve with aqueous diluted organic solvents (Linder et al., 2005; Mustalhti et al., 2013).

Hydrophobins class II HFBII from *Trichoderma reesei* have been the most exploited hydrophobin. This hydrophobin fusion technology can simultaneously improve the yield of recombinant protein production by protein body formation in filamentous fungi, and allow subsequent protein recovery by simple methods as ATPS. The protein bodies were found to enhance the accumulation of soluble protein, and to stabilize the storage of large amounts of recombinant protein. The formation of protein bodies also presents other advantages such as protection of fusion proteins from proteolysis degradation and possible toxic effects resultant from the accumulation of the target protein. The hydrophobin fusion technology has been reported in literature as a strategy for GFP purification by ATPS with selective recovery up to 91% (Joensu et al., 2010; Mustalhti et al., 2013).

3.4.5. γ-zein (ZERA)

The γ-zein is a 27 kDa maize storage protein composed of a signal peptide, a N-terminal proline-rich sequence (PPPVLH)n, a Pro-X motif and a hydrophobic cysteine rich domain (Llompart et al., 2010). The N-terminal of γ-zein that comprises the repeat domain (PPPVLH)n and Pro-X motif is also referred to as ZERA, presented to be sufficient to self-assemble and then induce the formation of protein bodies, which are natural endoplasmic reticulum (ER) derived organelles in the maize seed and which provide great stability to accumulated proteins (Khan et al., 2012; Llompart et al., 2010). ZERA has been exploited as a fusion partner for the production of fusion proteins as well as for protein body inducer in order to facilitate further steps of purification (Torrent et al., 2009). This strategy was extended not only to plant based hosts but also in eukaryotic hosts (mammalian and insect cells or filamentous fungi). By taking advantage of the PB properties such as high density, the protein can be easily recovered by density. Examples of this include human growth hormone (Llompart et al., 2010) and enhanced cyan fluorescent protein (Torrent et al., 2009).

4. Alternative schemes for tag removal

For most applications, affinity tags are innocuous and do not interfere with the structure and biological function of the target protein. However, for therapeutic proteins there is a demand for tag cleavage to avoid immunological responses and to guarantee authenticity of the protein structure. The removal of the affinity tag can be performed by enzymatic cleavage or harsh chemical treatments and has been extensively reviewed (Arnau et al., 2006; Hearn and Acosta, 2001; Terpe, 2003; Waugh, 2011). Chemical cleavage methods mostly used involve cyanogen bromide or hydroxylamine treatments, which are an inexpensive option but with low specificity and stringent conditions detrimental to the protein. Therefore, enzymatic methods are usually preferred as they operate under milder conditions and present higher selectivity as endoproteases recognize specific amino acid sequences or motifs. The main endoproteases used for tag removal comprise enterokinase (Choi et al., 2001), tobacco etch virus (TEV) (Dougherty et al., 1988) Factor Xa (Jenny et al., 2003), thrombin (Chang, 1985; Jenny et al., 2003) and SUMO protease (Li and Hochstrasser, 1999). An attractive feature of enterokinase and Factor Xa is that after cleavage there is no additional amino acid residues on the protein structure, contributing to an intact end product (Choi et al., 2001; Hearn and Acosta, 2001; Jenny et al., 2003). The main limitations in the use of proteases are related to the tag incompatibility under certain cleavage conditions (e.g. buffers, temperature), thus influencing its operation, and accessibility to the cleavage site (Waugh, 2011). The high costs associated with proteases limit the scalability of the purification process (Fong et al., 2010). Strategies on the reusability of these proteases without
efficiency loss have been studied, as recently reported for the immobilization of enterokinase onto magnetic supports (Santana et al., 2012).

On the same basis, alternative strategies have been exploited to overcome the traditional tag cleavage methods. Self-cleaving tags have emerged as an alternative strategy for tag removal as they possess inducible proteolytic activity, thus representing an extremely interesting class of affinity tags (Fong et al., 2010; Li, 2011). When combining with a purification tag, the recombinant proteins can be purified with the successive tag removal in a single step. Moreover, the use of self-cleaving tags can simplify the entire process at reduced costs, considering that the use of additional steps is not required for the separation of the target protein from the protease and the affinity tag (Wang et al., 2010).

The intein-based system is a well-established example where purified fusion protein cleavage can be induced by thiolcs such as β-mercaptotoeanol or 1,4-dithiothreitol (DTT), or by a shift in pH or temperature (Fong et al., 2010; Li, 2011). The pH-induced inteins are usually fused on C-terminal and are preferred in terms of costs and minor interference with the target protein over the N-or C-terminal thiol-induced intein. This is related to the fact that the presence of thiol groups can disrupt disulfide bonds and therefore confine the purification of the target protein (Fong et al., 2010). The pH-induced inteins combined with ELPs have been involved on a non-chromatographic purification scheme of several proteins, among them GFP (Wu et al., 2006), β-lactamase (Shi et al., 2013) and chloramphenicol acetyl transferase (CAT) (Liu et al., 2012) with further tag removal under mild conditions with a pH shift for 6.0–6.5. Apart from ELPs, the pH-induced inteins were also fused to chitin binding domain (CBD) for standard affinity chromatography purification via intein-mediated cleavage under the same mild conditions. This strategy has been employed for different proteins such as human antibody fragments, E. coli maltose binding protein (Wu et al., 2011) and human epidermal growth factor (Esipov et al., 2008). On the other hand, examples of thiol-induced inteins have been reported in the literature as fusion partners of oleosin for purification of industrial enzymes.

Table 3

| Tag                  | ELP           | Yield (%) | Purity (%) | n ITC round | Tl (°C) | Representative Target Protein | Reference                  |
|----------------------|---------------|-----------|------------|-------------|---------|-------------------------------|---------------------------|
| Elastin-like peptides (ELP) | V5,78        | 67        | na         | 2           | 25      | Organophosphorus hydrolase (OPH) | Shimazu et al. (2003)     |
|                      | V5A,GC90     | 80        | na         | 2           | 25      | Chloramphenicol acetyltransferase (CAT) | Trabbc-Carlson et al. (2004) |
|                      | V5A,G90      | 75        | na         | 1           | 25      | Blue fluorescent protein (BFP) | Trabbc-Carlson et al. (2004) |
|                      | V5A,G90      | na        | na         | 6           | 25      | Calmodulin (CaM)              | Trabbc-Carlson et al. (2004) |
|                      | V5A,G90      | 75        | na         | 1           | 25      | Thioredoxin (Trx)             | Trabbc-Carlson et al. (2004) |
|                      | V5A,G90      | 75        | na         | 2           | 45      | Green fluorescent protein (GFP) | Chow et al. (2006)         |
|                      | V5-40        | 98        | na         | 1           | 17      | Levensucrase                   | Kang et al. (2007)         |
|                      | V5A,G95-110  | 18        | na         | na          | 25      | β-Lactamase                   | Fong et al. (2009)         |
|                      | V5-80        | 90        | 95         | 1           | 37      | β-Galactosidase               | Yu and Liu (2012)          |
|                      | (V5-F30)     | 50        | 95         | 3-4         | 42      | Trx                           | Lim et al. (2007)          |

Currently there is a wide range of affinity “tag-ligand” pairs that can be used for the production and purification of fusion proteins. Commercial purification supports possess either biological or structural ligands and are presented in Tables 4 and 5, respectively. The choice of the affinity tag should take into account the target protein, the host and production conditions, as well as the respective binding partner.

Unconventional fusion partners have emerged as promising affinity tags, which due to their unique properties of hydrophobicity and self-assembly do not require chromatographic methods for the selective protein recovery. Non-chromatographic affinity tags are attractive choices as they can easily overcome the drawbacks associated with the affinity tag and therefore present high potential as large-scale purification platforms. However, the main limitation of unconventional fusion partners is the low purity of the recovered target protein when compared with the standard affinity tags commonly employed.

The application of genetically encoded tags for the affinity purification of recombinant protein is mostly in manufacturing and research. In manufacturing tag strategies are most applicable to situations where the purity of the target protein is less strict as it is the case for industrial enzymes.

When considering therapeutic proteins, the use of affinity pairs is not the preferred choice due to the need to strictly control the presence of the tag in the final product. Moreover, the additional steps of tag removal and further purification contribute to higher associated costs and low yields.

Therefore, the use of affinity pairs has a higher impact on research rather than manufacturing as their use is often employed in a wide range of applications within the scientific community being used not only for purification processes. These applications are related to protein–protein interaction studies in cell signaling, protein immobilization based strategies for sensing and diagnostics, protein labelling and cell imaging. Genetically encoded affinity tags were initially developed to provide universal strategies for recombinant protein purification that could be “protein-independent”. However, time has shown that there is no ideal affinity pair that can be applied to all proteins—the choice of the right tag-ligand (or combinations thereof) is often dependent on the protein’s properties and final application as well as the production host.

5. Concluding remarks and future trends

Nowadays there is still a demand to develop processes for recombinant protein purification that can combine selectivity with affordable prices. Affinity chromatography is the most frequent choice for protein purification; however this technique is dependent on the adsorbents’ availability for each particular target protein. Within the range of available ligands for affinity chromatography separations, those of biological origin are still the preferred option due to the high selectivity. Structural ligands are produced at low cost but the conditions for their use need to be optimised for each case due to the lack of specificity, often resulting in purification protocols with multiple steps. Synthetic ligands raised to overcome the high costs associated with biological ligands and the lack of selectivity of structural ligands have been so far tailor-made to very specific high value proteins. As each protein presents different biochemical and physical properties, the use of affinity pairs “tag-ligand” emerged as a universal purification strategy.
Examples of commercially available affinity tags and respective purification supports based on biological ligands.

| Biological ligands | Affinity solid support and supplier | Affinity tag | Capacity (mg/ml) |
|--------------------|------------------------------------|--------------|------------------|
| Peptides and proteins B | Glutathione sepharose (GE Healthcare) | GST tag | 10 |
| | Glutathione superflow (Clontech) | | 10 |
| | Glutathione HiCap matrix (Qagen) | | 20 |
| | Pierce glutathione agarose/magnetic beads (Thermo Scientific) | | 40/5–10 |
| | Glutathione agarose and magnetic agarose beads (Sigma) | | 5–10/15 |
| | Calmodulin resin (VWR) | CBP | 3 |
| | MagneZeo™-CAM (BioWorld) | na | |
| | Calmodulin resin (G Biosciences) | | 3 |
| | S protein Agarose (Novagen-EMD Millipore) | S-tag | na |
| | Streptactin™ (GE Healthcare) | Strept-tag II | 6 |
| | Strept-Tactin superfload/magnetic beads (Qagen) | | 9/0.3 |
| | Strept-Tactin superfload (Novagen-EMD Millipore) | | 3 |
| | IgG agarose (GE Healthcare) | Protein | 2 |
| | IgG agarose (Sigma) | A/G tag | 1 |
| | Anti-FLAG M1, M2 resin (Sigma-Aldrich) | Flg-tag | 0.6 |
| | Anti-FLAG antibody agarose/magnetic beads (Clontech) | na/1.3 |
| | EZview™ red Anti-myc resin (Sigma) | c-myc tag | 0.05 |
| | Anti-c-myc agarose (Sigma-Aldrich) | | 0.2 |
| | Anti c-myc antibody agarose (Clontech) | | 0.5 |
| | Anti-MyC-tag mAb-magnetic beads (MBL) | na | |
| | Pierce c-myc-Tag IP (Thermo Scientific) | | na |
| | T7 Tag® antibody agarose (EMD Millipore) | T7-tag | 0.3 |
| | T7 tag antibody agarose (Genetex) | | 0.5 |
| | Anti-T7 tag® antibody agarose (Abcam) | | na |
| | Pierce Anti-HA agarose (Thermo Scientific) | HA tag | 150 nmol/ml |
| | Pierce Anti-HA magnetic beads (Thermo Scientific) | | 10 µg/mg |
| | EZview™ anti-HA affinity gel (Sigma) | | 0.4 |
| | Anti-HA affinity matrix (3F10) (Roche) | | 9 nmol/ml |
| | Anti-HA-tag mAb-magnetic beads (MBL) | na | |
| | Softag immune affinity resin (Neoclude) | Softag | 0.3 |
| Carbohydrates | Amlyose resin/magnetic beads (New England Biolabs) | MBP | 7–10/0.01 |
| | Cellulose (Sigma) | Cellulose binding domain | 15 |
| | Chitin agarose (New England Biolabs) | Chitin binding domain | 3 |

Acknowledgements

The authors thank the financial support from Fundação para a Ciência e a Tecnologia through grant no. PEst-C/EBB/LA0006/2011, contracts PTDC/BIO/65383/2006, EBB-BIO/102163/2008, PTDC/EBB-BIO/098961/2008, PTDC/EBB-BIO/118317/2010 and PhD fellowship SFRH/BD/48804/2008 for A.S.P.

References

Ahmed S, Daulat A, Angers S. Tandem affinity purification and identification of heterotrimERIC g protein-associated proteins. In: Luttrell LM, Ferguson SC, editors. Signal Transduction Protocols. Humana Press; 2011. p. 357–70.

Table 5

Examples of commercially available affinity tags and respective purification supports based on structural ligands.

| Structural ligands | Affinity solid support and supplier | Affinity tag | Capacity (mg/ml) |
|--------------------|------------------------------------|--------------|------------------|
| Metal chelate | Ni-sepharose high performance (GE Healthcare) | His-tag | 40 |
| | His Mag sepharose (GE Healthcare) | | 50 |
| | TALON® superfload™ (GE Healthcare) | | 20 |
| | HisPur Ni-NTA resin/magnetic beads (Thermo Scientific) | | 20/0.5 |
| | HisPur cobalt superfload agarose (Thermo Scientific) | | 20 |
| | His60 Ni resin/magnetic beads (Clontech) | | 60/n.a |
| | TALON® resin/magnetic beads (Clontech) | | 5–18/n.a |
| | Ni-NTA agarose (Invitrogen) | | 5–10 |
| | Dynabeads® His-Tag (Invitrogen) | | 0.04 |
| | NTAGarose/Magnetic Agarose Beads (Qagen) | | 50/2 |
| | Profinity™ IMAC Nickel Charged resin (BIORAD) | | 15 |
| | IMAC HyperCel™ (PALL) | na | |
| | TALON® metal affinity resin (Clontech) | HAT tag | 18 |
| Ion-exchange (IEC) | Anion or cation exchange resins (several sources) | Charged peptides | 50–120 |
| Covalent | HaloLink™ (Promega) | Halo tag | 7 |

na—data not available.
Hu S, Wang A, Wang T. Expression tags for protein production. Encyclopedia of life sciences; 2001.

Huybrechts AM, Keightley A, Laijty J. Expression and purification of full length mouse metal response element binding transcription factor-1 using Pichia pastoris. Protein Expr Purif 2012;85:86–93.

Janknecht R, Martynoff GD, Lou J, Hipkins R, Nordheim A, Stunnenberg H. Rapid and efficient purification of native histidine-tagged protein expressed by recombinant vaccinia virus. Proc Natl Acad Sci 1991;88:8972–6.

Jarvik J, Telmer C. Epitope tagging. Ann Rev Genet 1998;32:601–18.

Jenny R, Mann K, Lundblad R. A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. Protein Expr Purif 2003;31:1–11.

Joensuu J, Conley A, Lieberman M, Brandle J, Linder M, Menassa R. Hydrophobin fusions for high-level transient protein expression and purification in Nicotiana benthamiana. Plant Physiol 2010;152:622–33.

Kano A, Unoymoto T, Okada M, Monk B, Uehara Y, Nimi M. Tandem affinity purification of the Candida albicans septin protein complex. Yeast 2004;21:1025–33.

Kang H, Kim J, Chang W, Kim E, Koo Y. Heterologous expression and optimized one-step separation of levanucrase via elitain-like polypeptide tags. J Microbiol Biotechnol 2007;17:1751–7.

Kaplan W, Erhardt J, Sluss-Crenner N, Dirr H, Hülsler P, Klumpp H. Conformational stability of pGEX-expressed Schistosoma japonicum glutathione S-transferase: a detoxification enzyme and fusion-protein affinity tag. Protein Sci 1997;6:399–406.

Kapost R, Waukh D. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci 1999;8:1668–74.

Karpeisky M, Senchenko V, Dianoova M, Kanevsky V. Formation and properties of S. cerevisiae complex with S-peptide-containing fusion protein. FEBS Lett 1994;339:209–12.

Kato K, Sato H, Iwata H. Immobilization of histidine-tagged recombinant proteins onto immobilized metal affinity chromatography. Protein Folding Handbook. Wiley-VCH Verlag GmbH; 20081245

Kapust R, Waugh D. Immobilization of histidine-tagged recombinant proteins onto cation-exchange resin. Biotechnol Prog 2004;20:277–80.

Khan I, Twyman R, Arcalis E, Stoger E. Using storage organelles for the accumulation and encapsulation of recombinant proteins. Biotechnol J 2012;7:1099–1102.

Kim J, Raines R. Ribonuclease S-peptide as a carrier in fusion proteins. Protein Sci 1993;2:406–14.

Kim J, Lustig S, Madan R, Chen W. Engineering a high-affinity scaffold for non-covalent chromoaffinity protein purification via in vitro-mediated cleavage. Biotechnol Bioeng 2012;109:2829–35.

Kim J, Duan C, Zhao X, Zien J, Cheng K, Pai C. Cloning of a rung of fumal xylanase gene and purification of the recombinant enzyme via artificial oil bodies. Appl Microbiol Biotechnol 2008;79:225–33.

Kim K, Hanna M. NusA contacts nascent RNA in Escherichia coli transcription complexes. J Biol Chem 1995;270:547–58.

Kim X. Boronic acids as ligands for affinity chromatography. Chin J Chromatogr 2006;24:73–80.

Kim X, Scutten W. New ligands for boronate affinity chromatography. J Chromatogr A 1994;687:81–96.

Kim X, Scutten W. Boronate affinity chromatography. In: Bailon P, Ehrlich GK, Fung WJ, Berthold B, editors. Affinity Chromatography. Humana Press; 2000. p. 119–28.

Kloppmar B, Illop-Tous I, Marzabal P, Torrent M, Pallasí R, Bastida M, et al. Production protein from recombinant protein bodies. Process Biochem 2010;45:1816–20.

Klöppner C, Wood K, The HaloTag™ system: in Taylor DH, Hinkel J, Giolou K, editors. High Content Screening. Humana Press; 2006; 205-208.

Ko GV, Encell LP, McDougall MG, Hartzell DD, Karassina N, Zimprich C, et al. HaloTag: a novel protein labeling technology for cell imaging and protein analysis. ACS Chem Biol 2008;3:373–82.

Kow C. Combinatorial approaches to affinity chromatography. J Chromatogr Biomed Appl 2001;2:548–56.

Kow C, Lowe A, Gupta G. New developments in affinity chromatography with potential application in the production of biopharmaceuticals. J Biochem Biophys Methods 2001;49:561–74.

Knipping P, Ford C, Nikolov Z. Adsorption of a 50-gallatinosidase fusion protein on immobilized metal affinity chromatography of the starch-branching enzyme of Aspergillus glaucaroseae. Gene 1991;99:121–6.

Ku H, McLean J, Chao L, Mama-Capelli S, Paramasivam M, Haggstrom K, et al. A highly efficient multifunctional tandem affinity purification approach applicable to diverse organisms. Mol Cell Proteomics 2012;11:501–11.

Klakibov M, Mattens M, Klakibová O, Drinker M, Weeks S, Butt T. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. J Struct Funct Genomics 2004;5:75-86.

Kmalová A. Chapter 16 tagging for protein expression. In: Burgess RR, Deutscher MP, editors. Methods in Enzymology. Academic Press; 2009. p. 239–50.

Kraft C, Hodges R. Mixed-mode hydrophilic interaction/cation-exchange chromatography (HLIC/ECX) of peptides and proteins. J Sep Sci 2008;31:2754–73.

Kronberger J, Edavallet S, Lim Y, Lim P, Zuo X, Butt T. Comparison of SUMO fusion technology with traditional carrier sequence space systems: enhanced expression and solubility with SUMO. Protein Sci 2006;15:182–9.

Kremer M, Chen R, Yu D, Mah K, Teat J, Wang H, et al. Production of the therapeutic antibody trastuzumab from genetically modified plants using safEf Protein Expression Protocols. In: Bailon P, Ehrlich GK, Fung WJ, editors. Affinity Chromatography. Humana Press; 2000. p. 119–28.

Kruif K, Sier D. Calmodulin as an affinity purification tag. In: Vaillancourt P, editor. Methods in Molecular Biology—Coli gene expression protocols. Academic Press; 2003. p. 67.

Kroemer M, Keightley A, Laijty J. Expression and purification of full length mouse metal response element binding transcription factor-1 using Pichia pastoris. Protein Expr Purif 2012;85:86–93.

Kruif K, Sier D. Calmodulin as an affinity purification tag. In: Vaillancourt P, editor. Methods in Molecular Biology—Coli gene expression protocols. Academic Press; 2003. p. 67.

Kruif K, Sier D. Calmodulin as an affinity purification tag. In: Vaillancourt P, editor. Methods in Molecular Biology—Coli gene expression protocols. Academic Press; 2003. p. 67.

Kruif K, Sier D. Calmodulin as an affinity purification tag. In: Vaillancourt P, editor. Methods in Molecular Biology—Coli gene expression protocols. Academic Press; 2003. p. 67.
Nilsson B, Moks T, Jansson B, Abrahamsson L, Elmiabadi A, Holmgren E, et al. A synthetic IgC-binding domain based on staphylococcal protein A. Protein Eng 1987;1:103–9.

Nilsson J, Nilsson P, Williams Y, Pettersson L, Uhlen M, Nygren P. Competitive elution of protein A fusion proteins allows specific recovery under mild conditions. Eur J Biochem 1994;224:103–8.

Nilsson S, Rossin L, Ludvigson G, Uhlen M, Nygren P. Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins. Protein Expr Purif 1997;11:1–16.

Nygren P, Elssson A, Abrahamsson L, Uhlen M, Palmertz A. Analysis and use of the serum albumin-binding domain of streptococcal protein G. J Mol Recognit 2001;14:523–33.

Nylander J, Celsius J, Ljungdahl P, Sallstedt M, Taljedal M, Lhrlie H. Purification of a viral coat protein by an engineered polynucleic sequence. J Chromatogr B 2000;737:77–84.

Nylander J, Celsius J, Ljungdahl P, Sallstedt M, Taljedal M, Lhrlie H. Purification of a viral coat protein by an engineered polynucleic sequence. J Chromatogr B 2000;737:77–84.

Nylander J, Celsius J, Ljungdahl P, Sallstedt M, Taljedal M, Lhrlie H. Purification of a viral coat protein by an engineered polynucleic sequence. J Chromatogr B 2000;737:77–84.

Nylander J, Celsius J, Ljungdahl P, Sallstedt M, Taljedal M, Lhrlie H. Purification of a viral coat protein by an engineered polynucleic sequence. J Chromatogr B 2000;737:77–84.

Nylander J, Celsius J, Ljungdahl P, Sallstedt M, Taljedal M, Lhrlie H. Purification of a viral coat protein by an engineered polynucleic sequence. J Chromatogr B 2000;737:77–84.
Yu S, Liu Y. Expression and one-step purification of a β-galactosidase by fusion with elastin-like polypeptides. Process Biochem 2012;47:1108–14.

Zhang C, Love R, Jilka J, Glatz C. Genetic engineering strategies for purification of recombinant proteins from canola by anion exchange chromatography: an example of β-glucuronidase. Biotechnol Prog 2001;17:161–7.

Zhao G, Dong X, Sun Y. Ligands for mixed-mode protein chromatography: principles, characteristics and design. J Biotechnol 2009;144:3–11.

Zuo X, Li S, Hall J, Mattern M, Tran H, Shoo J, et al. Enhanced expression and purification of membrane proteins by SUMO fusion in Escherichia coli. J Struct Funct Genomics 2005;6:103–11.