FEBS Letters 588 (2014) 247–252

FEBS Letters

journal homepage: www.FEBSLetters.org

Review

Targeted expression, purification, and cleavage of fusion proteins from inclusion bodies in *Escherichia coli*

Peter M. Hwang [a,b,*], Jonathan S. Pan [a], Brian D. Sykes [a]

[a] Department of Biochemistry, University of Alberta, Edmonton, AB, Canada
[b] Division of General Internal Medicine, Department of Medicine, University of Alberta, Edmonton, AB, Canada

ARTICLE INFO

Article history:
Available online 27 September 2013

Edited by Wilhelm Just

Keywords:
Fusion protein
Inclusion body
Chemical cleavage

ABSTRACT

Today, proteins are typically overexpressed using solubility-enhancing fusion tags that allow for affinity chromatographic purification and subsequent removal by site-specific protease cleavage. In this review, we present an alternative approach to protein production using fusion partners specifically designed to accumulate in insoluble inclusion bodies. The strategy is appropriate for the mass production of short peptides, intrinsically disordered proteins, and proteins that can be efficiently refolded in vitro.

There are many fusion protein systems now available for insoluble expression: TrpLE, ketosteroid isomerase, PurF, and PagP, for example. The ideal fusion partner is effective at directing a wide variety of target proteins into inclusion bodies, accumulates in large quantities in a highly pure form, and is readily solubilized and purified in commonly used denaturants. Fusion partner removal under denaturing conditions is biochemically challenging, requiring harsh conditions (e.g., cyanogen bromide in 70% formic acid) that can result in unwanted protein modifications. Recent advances in metal ion-catalyzed peptide bond cleavage allow for more mild conditions, and some methods involving nickel or palladium will likely soon appear in more biological applications.

1. Introduction

Heterologous protein overexpression in *Escherichia coli* often leads to a significant proportion of the target protein accumulating in dense insoluble aggregates known as inclusion bodies. For most applications, it is desirable to minimize this phenomenon in order to maximize the yield of soluble, well-folded, functional protein. However, in some cases, the insoluble fraction can be collected and reconstituted to yield large quantities of biologically active protein. This approach has been used to produce a number of clinically important hormones and peptides since the late 1970s [1]. Despite its historical and commercial significance, the methods associated with insoluble protein production are less well developed than the corresponding methods for soluble constructs.

Producing protein from insoluble aggregates is convenient and effective in the case of peptides that do not fold into a stable domain. Moreover, a significant proportion of genomes, particularly more complex eukaryotic ones, encode for long (>50 amino acids) protein sequences that are unstructured or only become structured upon target binding [2]. Disordered sequences are highly exposed and therefore available for protein–protein interactions and post-translational modifications like proteolytic processing and phosphorylation [3], making them very important to protein regulation and signaling networks. However, the high degree of exposure that makes these sequences biologically interesting also renders them vulnerable to proteolysis in the setting of heterologous expression in *E. coli*. Hence, targeted expression to inclusion bodies would be a useful strategy for protecting protein from proteolytic degradation.

The applicability of inclusion body-targeted protein expression is further extended by the development of efficient in vitro refolding protocols for over a thousand different protein systems (REFOLD database available on-line at refold.med.monash.edu.au [4]). Reconstitution protocols generally involve solubilizing inclusion bodies in a strong denaturant and then removing the denaturant through dialysis, rapid dilution, or chromatographic separation [5]. Refolding is a separate topic that will not be covered here.

This review will focus on the targeted expression of proteins into inclusion bodies using fusion partners, solubilization and purification under denaturing conditions, and cleavage methods to remove the fusion partner under denaturing conditions.

2. Fusion partners for targeting proteins to inclusion bodies

A number of proteins have been proposed as fusion partners for targeted expression into inclusion bodies. One may ask why a fusion partner is necessary when inclusion body formation is so

* Corresponding author at: Department of Biochemistry, University of Alberta, Medical Sciences Building Rm 4-19, Edmonton, AB T6G 2H7, Canada.
E-mail address: pmhwang@gmail.com (P.M. Hwang).

0014-5793/$36.00
http://dx.doi.org/10.1016/j.febslet.2013.09.028

common in heterologous protein overexpression. It is true that a fusion partner is not needed when a protein overexpresses robustly into inclusion bodies. However, while many difficult-to-express proteins appear to partition primarily into the insoluble fraction, this is not necessarily the fate of the majority of newly expressed protein. It is useful to think of proteins as existing in equilibrium between three states: properly folded, partially folded intermediates, and aggregated (see Fig. 1). All three states are in dynamic equilibrium and subject to the action of chaperones and proteases [6], with the partially folded state being the most prone to proteolytic degradation. Proteins that are the most difficult to express are those that exist primarily as partially folded (or misfolded) intermediates and are rapidly degraded. Expression can be improved by shifting the equilibrium towards the folded state by lowering the temperature, reducing the level of expression, co-expressing chaperones, or introducing solubilizing or stabilizing mutations. On the other hand, the protein of interest can also be protected from degradation by targeting it to inclusion bodies using an insoluble fusion partner.

Just as certain proteins are more effective as solubilizing fusion partners, the same can be said of insoluble fusion partners. The ideal insoluble fusion partner accumulates efficiently in inclusion bodies even when fused to very solubilizing constructs. The TANGO algorithm has been used to predict protein sequences that are most aggregation-prone, with the most important factors being hydrophobicity and β-sheet propensity, but other factors, like alternative structural propensities, electrostatics, and H-bonding play a role as well [7].

It is difficult to assess which insoluble fusion partners are most effective at targeting to inclusion bodies, because there are few head-to-head comparisons. However, we will review the more commonly used constructs, those that have been demonstrated to work with multiple target peptides or proteins.

A very early example of a fusion partner that strongly favors inclusion body formation is the TrpΔLE 1413 sequence [8]. The *trp* operon contains a leader sequence, TrpL, which encodes a short 14-amino acid peptide sequence, MKAIFVLKGWWRTS, followed by a stop codon. A large deletion encompassing part of TrpL and the first protein of the operon, TrpE, resulted in an N-terminal 17-amino acid leader sequence, MKAIFVLKGSLDRDPEF, that produced a fusion polypeptide with the remaining carboxy-terminal fragment of TrpE. When this sequence is fused to small peptides, including somatostatin, insulin, and TGF-α [9], expression in *E. coli* is greatly enhanced. When the target protein sequence begins with methionine, the TrpΔLE peptide can be cleaved off using cyanogen bromide. This strategy has been used to express numerous peptides as well as even small membrane proteins [10].

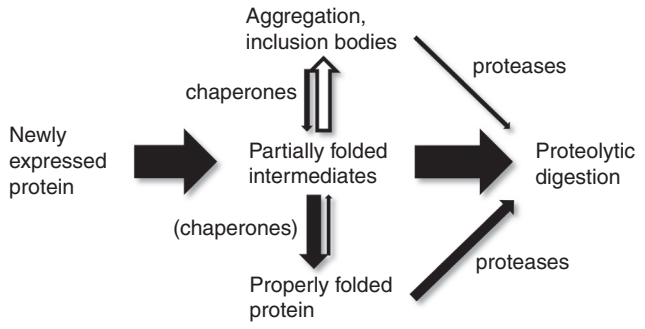

Fig. 1. Proteins exist in a dynamic equilibrium between soluble properly folded forms, partially folded intermediates, and insoluble aggregates. Difficult-to-express proteins accumulate as partially folded intermediates, the form most susceptible to proteolytic degradation. Usually, some portion of overexpressed protein will appear in insoluble aggregates, but shifting the equilibrium strongly towards inclusion body formation (clear arrow), for instance, through the use of a fusion partner, will protect the protein from proteolysis and enhance protein production.

Another commonly used insoluble fusion partner is ketosteroid isomerase (KSI) [11], a 14 kDa soluble protein that is very hydrophobic, with a strong tendency to accumulate in inclusion bodies. A KSI fusion system for expressing tandem repeats of peptides is commercially available as a pET31b (Novagen®) expression vector. The KSI protein and target peptide sequences are expressed with intervening methionine residues, so that cyanogen bromide cleavage separates out KSI and the individual peptides. The mixture of complete and partial cleavage products can then be purified using HPLC.

The strategy of targeted expression in inclusion bodies has also been used in the production of antimicrobial peptides [12], which would be useful to protect the expression host from the bactericidal effect of the peptides in theory (although it seems that soluble fusion proteins have worked as well). Besides the KSI system, multiple antimicrobial peptides have been expressed in fusion constructs utilizing a truncated *E. coli* PurF fragment [13], *Pseudomonas aeruginosa* PaP3.30 protein [14], and the histone fold domain of the human transcription factor TAF12 [15].

Another recently developed fusion partner was engineered from the N-terminal autoprotease N^{PRO} from classical swine fever virus (CSFV), termed EDDIE [16]. The native N-terminal autoprotease cleaves itself from the CSFV polyprotein to generate the N-terminus of the capsid protein. The CSFV polyprotein can be replaced by a target protein, so that autocleavage by N^{PRO} releases the target protein with its own native N-terminus. Overexpressed N^{PRO} accumulates mostly in inclusion bodies, but it is also present in the soluble fraction. The degree of soluble vs. insoluble expression in fusion constructs depended on the nature of the target protein. Fusion protein was harvested from inclusion bodies, solubilized in guanidine, and then refolded by rapid dilution, leading to autoproteolysis and N^{PRO} release. In vitro proteolytic cleavage was limited by misfolding and precipitation during the refolding process. The EDDIE protein incorporated a total of 11 point mutations to increase the solubility of the N^{PRO} protein. While these decreased the proportion of protein targeted to inclusion bodies, the mutations improved the solubilization of inclusion bodies and subsequent refolding of the protease and autoproteolytic cleavage step. This work demonstrates that if the inclusion body-targeting fusion partner can be refolded into a soluble form, it can then be cleaved off under non-denaturing conditions using enzymatic methods rather than chemical cleavage (see below).

We have recently developed a new PagP-based fusion protein system for inclusion body-targeted expression [17]. PagP is a Gram-negative bacterial outer membrane protein. All integral outer membrane proteins have a β-barrel architecture with a hydrophobic exterior facing the membrane bilayer and a hydrophilic interior cavity. They are targeted to the outer membrane by a cleavable signal sequence, but when expressed in a construct lacking the signal sequence, they invariably accumulate in cytoplasmic inclusion bodies [18]. Inclusion body formation is favored because of the high β-sheet content of the protein and the inability of the protein to fold outside of the membrane environment. As a fusion partner, PagP was more effective than KSI at maintaining a tandem construct of a troponin I fragment in inclusion bodies [17]. We have found PagP effective in inclusion body-directed expression of a wide variety of peptides (unpublished results). Another theoretical advantage of using a β-barrel membrane protein is that they are surprisingly not very hydrophobic, making the inclusion bodies easier to solubilize.

3. Solubilization and purification of inclusion bodies

Inclusion bodies can be harvested from lysed cells using moderate-speed centrifugation. After centrifugation, the inclusion bodies can be washed with mild detergents or low concentrations of

denaturant [5]. Although some protocols suggest addition of a strong denaturant like guanidine hydrochloride to intact cells so as to simultaneously lyse cells and dissolve inclusion bodies, it is preferable to isolate the inclusion bodies separately, since this can be an important purification step in itself. For instance, washed inclusion bodies of PagP protein were sufficiently pure to allow crystallization for X-ray diffraction studies without any additional purification steps [19].

Solubilization is a key step in purifying proteins from inclusion bodies. The denaturant or detergent employed for solubilization is usually also used for subsequent purification or cleavage reactions. Depending on the target protein and fusion partner, not all denaturants are effective at solubilizing inclusion bodies. To our knowledge, 70% formic acid is the strongest denaturing solvent, sometimes even being used to solubilize alpha helical membrane proteins. The main issue with formic acid is that it can formylate serine and threonine residues. 70% trifluoroacetic acid has been suggested as an alternative to 70% formic acid, but it is not as strong a denaturant as formic acid or 6 M guanidine hydrochloride [20].

Guanidine hydrochloride is a commonly used denaturant and appears to be the safest in terms of minimizing protein modification side reactions. It is also compatible with nickel affinity chromatography, more recently termed "immobilized metal affinity chromatography" (IMAC), since other metal ions like cobalt and copper can be used as well. IMAC is the most widely used affinity purification method for denatured proteins. The affinity tag for IMAC is simple and easy to clone into any protein construct: a minimum 6-polyhistidine sequence, $-His_6-$. The sequence binds with high affinity to the metal ion-bound matrix and can be eluted with increasing concentrations of imidazole, acidic pH, or a chelating agent like EDTA. Ionic interactions are minimized through the use of high salt buffer (500 mM NaCl). Many of the impurities retained in nickel affinity chromatography adhere to the protein of interest through hydrophobic interactions, including other proteins as well as bacterial endotoxin. These can be removed by a washing step with detergent or organic solvents [21]. Thus, IMAC is usually more effective under denaturing conditions than under native conditions, and it is often the only chromatographic step required to obtain protein of high purity from washed inclusion bodies. It is worthwhile to note that IMAC is compatible with Gdn-HCl, urea, and various detergents like perfluorooctanoic acid [22], an anionic detergent that can also be used to solubilize inclusion bodies.

Compared to guanidine, urea has the advantages of low cost and compatibility with SDS–PAGE (guanidine precipitates with SDS). However, over time concentrated urea solutions accumulate isocyanate ions that can carbamylate protein amino groups. Another disadvantage of urea is that it is not as strong a denaturant as guanidine. Some very stable proteins can resist denaturation even in 8 M urea.

Recently, a Japanese group found that the ribosomal protein, L2, adsorbs strongly to silica particles and used it as a 273-amino acid fusion partner affinity tag called Si-tag [23]. The protein contains many disordered segments that are highly positively charged, and silica adsorption was thought to occur through electrostatic interactions. Adsorption also occurred in 8 M urea, allowing purification under denaturing conditions [24]. Elution was effectively carried out with 2 M $MgCl_2$.

In similar work, the "Z-tag" was engineered from the 58-amino acid B domain of staphylococcal protein A to contain a large positively charged patch on its surface [25]. "Z-tagged" proteins can be purified using standard cation exchange chromatography in 8 M urea [26]. Cation exchange chromatography is very effective in purifying proteins expressed in *E. coli* because few endogenous *E. coli* proteins adsorb to an anionic matrix.

It is worthwhile to note that the most commonly used chromatographic methods for protein purification, gel filtration chromatography and ion exchange chromatography, can be effectively carried out in 8 M urea.

The extremely high affinity (10^{-15} M) of biotin for avidin/streptavidin can also be exploited for affinity purification even in 8 M urea. Biotinylated protein can be produced in *E. coli* by the addition of an AviTag sequence, GLNDIFEAQKIEWHE, where the underlined K residue is biotinylated [27,28]. Biotinylation can be carried out in vivo in *E. coli* with BirA biotin ligase co-expression. One group affinity purified biotinylated protein using streptavidin beads under denaturing conditions and then "eluted" the target protein by an on-column TEV protease digestion [29].

4. Non-enzymatic chemical cleavage methods in denaturing conditions

Although some applications may not require it, fusion partner removal is often necessary. Most soluble fusion tags are connected to the target protein by a protease sensitive linker. The advantage of proteases like thrombin or TEV protease is the high degree of efficiency and specificity of cleavage. However, this may not be a viable option when the fusion partner is soluble only in denaturing conditions. There are a number of non-enzymatic chemical cleavage methods that can be performed under denaturing conditions, but these typically require harsh conditions that can result in unwanted protein modifications (see Table 1). Many of these chemical methods were developed in the 1960–1970s.

Most chemical cleavage reactions require acidic or basic conditions. Under the harsh acidic conditions used for amino acid analysis (e.g., 6 M HCl, 110 °C), Asn and Gln are quantitatively converted to Asp and Glu [30]. Trp and Cys are partially destroyed at a slower rate [31]. However, these reactions are not significant under the milder acidic conditions used in the methods described below. A study of Asn deamidation found a half-life of about 40 h in 0.1 N HCl at 60 °C, which is on par with deamidation rates at neutral pH [32]. At increased temperatures and acidic conditions, Asp-X can form a cyclic succinimide, which may proceed to peptide bond cleavage, more likely to occur when X = Pro (see below) [33].

Under alkaline conditions, deamidation of Asn is greatly accelerated, especially when it is followed by a Gly [32]. The deamidation occurs through a cyclic succinimide intermediate that can then proceed to peptide bond cleavage or conversion to Asp or iso-Asp. Cys is converted to dehydroalanine and Arg to ornithine at 0.1 M NaOH at 55 °C [34]. At 80 °C and higher concentrations of NaOH, Ser and Thr are also destroyed.

Some of the chemicals used for peptide bond cleavage can also promote amino acid oxidation. Cys and Met are most susceptible, with oxidation occurring spontaneously at room temperature, so addition of oxidizing agents or increased temperatures can lead to complete oxidation of these residues. Tyr and Trp, and to a lesser extent, His, are also prone to oxidation under harsher conditions [35]. Amino acid oxidation can occur under acidic, neutral, or alkaline conditions, although Cys oxidation in particular is facilitated by alkaline pH (Cys pKa is about 8.3) [36].

Perhaps the most commonly used chemical cleavage method is cyanogen bromide (CNBr) cleavage [37], likely due to the nearly quantitative reaction that does not require elevated temperatures (see Table 1). CNBr cleaves the peptide bond C-terminal to the Met residue, leaving behind a C-terminal homoserine in place of methionine and a new N-terminus. Near-complete cleavage is possible, as long as the Met is not followed by Ser or Thr [38,39]. The cleavage reaction is typically carried out in 70% formic acid or 6 M Gdn-HCl with 0.1 M HCl. The reaction is quite specific under acidic conditions. More alkaline conditions cause increased side reactions

Table 1
Chemical peptide bond cleavage methods. For references and more details, please refer to text.

Sequence cleaved	Sample conditions used	Observed side reactions
–M–↓–X– X = not S, T	100 mM CNBr in 0.1 M HCl + 6 M Gdn-HCl or 100 mM CNBr in 70% formic acid	M to homoserine; S, T formylation
–D–↓–P–	0.1 M HCl + 6 M Gdn-HCl or 70% formic acid, >40 °C	–D–X– succinimide formation, cleavage; S, T formylation; C, M oxidation
–N–↓–G–	1 M NH_2OH + 6 M Gdn-HCl, pH 9, 45 °C	N deamidation; N hydroxamate formation; C, M oxidation
–↓–C–	20 mM CNTB in 6 M Gdn-Hcl, pH 9, 1 M glycine or 1 M NH_4OH, 37 °C	C to dehydroalanine; K carbamylation; N deamidation; C, M oxidation
–W–↓–	10 mM BNPS-skatole, 80% acetic acid	C, M, W, Y oxidation, Y halogenation

Table 2
Metal ion-catalyzed peptide bond cleavage methods. For references and more details, please refer to text.

Sequence cleaved	Sample conditions used
–M–↓–	5 mM *cis*-$[Pt(en)(H_2O)_2]^{2+}$, pH 2.5, 40 °C
–D/E–K/R–↓–S/T–H–	2.5 mM Cu^{2+} or Ni^{2+}, 50 mM Tris–HCl, pH 9, 62 °C
–↓–S/T–X–H–Z– X = H, K, R Z = I, K, L, R, W	2 mM Ni^{2+}, 10 mM Tris–HCl, pH 8.5, 37 °C
–↓–Z–H/M– Z = any amino acid	5 mM Pd^{2+}, pH 7 at 40 °C or pH <3 at 20 °C for Z = P; pH <3 at 60 °C for Z = everything else
–C–H–↓–	1 mM Pd^{2+}, 100 mM HBF_4, $HClO_4$, or CF_3COOH, pH 2, 40 °C

with amino groups and methionine oxidation [40], which prevents cleavage. Methionine oxidation occurs spontaneously in proteins, but a pre-cleavage reduction step with β-mercaptoethanol or potassium iodide can increase cleavage efficiency with CNBr [17]. Tryptophan oxidation was minimized through the use of fresh CnBr [41]. Besides the toxicity of cyanogen bromide, the biggest disadvantage of cyanogen bromide cleavage is that the protein of interest cannot contain internal methionine residues.

Cysteine residues can also be cyanylated using 2-nitro-5-thiocyanatobenzoic acid (NTCB), and the modified cysteine residue then undergoes cleavage of its N-terminal peptide bond at pH 9 [42]. Recent re-examination of this reaction found that cleavage efficiency could be improved by using a high concentration of ammonia or glycine. Major side products of the reaction are carbamylated Lys residues as well as dehydroalanine resulting from β-elimination of thiocyanate from cyanocysteine [43].

The peptide bond of Asp-Pro has been found to be susceptible to cleavage under acidic conditions and increased temperature [44]. The major issue with this method is the low efficiency in some cases, likely due to sequence-dependent factors that have yet to be determined. Some groups have reported essentially complete cleavage for their construct [33]. However, in almost all cases, high temperatures (>40 °C) and long incubation times are needed to achieve reasonable cleavage rates, increasing the incidence of unwanted side reactions: Asp succinimide formation followed by cleavage at non-Asp-Pro sites and formylation of Ser/Thr if formic acid is used. It is important to determine the mildest conditions of acidity and temperature compatible with an acceptable cleavage yield.

Another method for chemical cleavage is hydroxylamine cleavage at an Asn–Gly bond [45]. The reaction can be carried out in guanidine- or urea-containing buffer and is favored by high concentrations of hydroxylamine (>1 M), high pH (>8.5), and high temperature. Increasingly harsh conditions improve efficiency but also promote competing side reactions. Along with the peptide bond cleavage, the Asn residue can be converted to a hydroxamic acid derivative [46]. As mentioned earlier, the alkaline conditions cause Asn residues to form cyclic succinimide structures that are subsequently hydrolyzed to Asp and iso-Asp [32]. Bond cleavage can also occur at Asn-X sites, where X is another amino acid besides Gly, albeit at lower rates [32,45].

There are a number of oxidizing reagents that can cleave at oxidation-susceptible residues. One of the more mild agents is BNPS-skatole [2-(2-nitrophenyl)-3-methyl-3-bromoindolenine], which cleaves Trp–X bonds through oxidative halogenation of the indole ring of Trp using under acidic conditions (80% acetic acid) [47,48]. This is still a fairly harsh agent that leads to complete oxidation of Cys and Met residues as well.

5. Metal ion-catalyzed cleavage methods

More recently, various metal ions and chelates have been used to catalyze peptide bond cleavage under more mild conditions (see Table 2). Although there is a large body of chemical literature in this area, these methods have yet to find widespread biological application. One potential problem with metal ion-catalyzed peptide cleavage is the tendency of metal ions to catalyze oxidation of certain amino acids (His, Arg, Lys, Pro, Met, and Cys) [49]. Oxidative cleavage can be encouraged by the addition of hydrogen peroxide or ascorbate [50], but these should probably be avoided due to the potential for collateral damage to the protein of interest.

We will cover examples of metal ion-catalyzed cleavage where stringent sequence specificity would be useful for biological applications. Since these methods are new and relatively untested compared to the much older chemical cleavage techniques, there is far less data concerning side reactions. However, most of the techniques require increased temperatures, and one would predict that amino acid oxidation, particularly Cys and Met, would be an important consideration.

A proposed alternative to CNBr cleavage (at the peptide bond C-terminal to Met residues) is platinum-catalyzed cleavage using *cis*-$[Pt(en)(H_2O)_2]^{2+}$ at pH 2.5 and 40 °C [51].

The sequence Asp-Lys-Thr-His was found to be susceptible to copper(II) ion-catalyzed peptide bond cleavage between Lys and Thr under basic conditions [52]. Further investigation of this phenomenon showed that the mechanism was hydrolytic rather than oxidative [53]. The cleavage efficiency could be increased by changing the peptide sequence to Asp-Lys-Ser-His, allowing

for near-complete cleavage at pH 8.0 and 62 °C. The cleavage could also be carried out at a slower rate using nickel(II) ions instead.

The Bal group has done the most comprehensive work to date in optimizing amino acid sequences for Ni(II)-catalyzed cleavage. They focused on the sequence Ser/Thr-X-His-Y, where cleavage occurs N-terminal to the Ser/Thr [54]. Combinatorial library screening of the four hundred possible amino acid combinations for positions X and Y determined that a positively charged residue is favored at position X and a hydrophobic or positively charged residue at position Y, and the optimal sequence was Ser-Arg-His-Trp. This combinatorial approach could be an important means of optimizing sequences for any cleavage method. This sequence was used as part of a linker (SRHWAP) between a protease inhibitor protein, SPI2, and a poly-His tag, and near-complete nickel-catalyzed cleavage was demonstrated under mildly alkaline conditions (pH 8.2), both in solution and bound to a nickel affinity column [55]. Recent work by our lab has confirmed the general utility of this sequence as a cleavable linker for fusion protein constructs. However, one proviso is that metal ion binding sequences can lead to irreversible aggregation in the setting of inclusion body-targeted expression (results to be published).

Palladium(II) binds to His, Cys, and Met and has been found to catalyze cleavages near these residues. One study found that for X–Z-Met and X–Z-His sequences, the X–Z bond was susceptible to Pd^{2+}-catalyzed cleavage under acidic (pH <3.0) conditions (Cys-containing peptides were not examined) [56]. X-Pro-Met and X-Pro-His sequences were found to be particularly susceptible to Pd^{2+}-catalyzed cleavage even under neutral pH conditions. The susceptibility of the Pro peptide bond to acid-catalyzed cleavage has already been noted. In this regard, it is interesting to note that the sequence Gly-Asp-Pro-His was found to be particularly susceptible to acid-catalyzed cleavage (pH <6.0) in human MUC2 mucin [57].

The same group also found that the His-X peptide bond in the sequence Cys-His-X was susceptible to Pd^{2+}-catalyzed cleavage under acidic conditions (pH 2), so long as the anion of the acid did not coordinate Pd^{2+}: e.g., HBF_4, $HClO_4$, or CF_3COOH [58].

The above examples of sequence-specific metal ion-catalyzed cleavage show great promise, but further work is needed to fully delineate not only the optimal sequences for lower temperature cleavage, but also the full range of susceptible sequences so that unwanted cleavages can be minimized.

6. Future directions

Targeted expression of proteins to inclusion bodies using an insoluble fusion partner is an effective means of producing large amounts of very pure protein. Although the strategy is quite old, it is only recently being refined through modern recombinant DNA and combinatorial library screening techniques. The ideal fusion partner for maximizing expression and purity remains to be determined. As well, the full range of amino acid sequences for non-oxidative chemical cleavage under acid-, base-, and metal ion-catalyzed conditions needs to be further clarified. While most studies have focused solely on the effect of amino acid sequence for residues immediately adjacent to the cleaved peptide bond (−1 and +1 positions), it is conceivable that longer range local effects need to be considered as well (from the −4 to +4 positions).

Acknowledgments

PMH holds a Canadian Institutes of Health Research (CIHR) Phase I Clinician Scientist award and an Alberta Innovates–Health Solutions (AIHS) Clinical Fellowship Incentive award.

References

[1] Marston, F.A. (1986) The purification of eukaryotic polypeptides synthesized in *Escherichia coli*. Biochem. J. 240, 1–12.
[2] Dyson, H.J. and Wright, P.E. (2005) Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell Biol. 6, 197–208.
[3] Iakoucheva, L.M., Radivojac, P., Brown, C.J., O'Connor, T.R., Sikes, J.G., Obradovic, Z. and Dunker, A.K. (2004) The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res. 32, 1037–1049.
[4] Chow, M.K., Amin, A.A., Fulton, K.F., Fernando, T., Kamau, L., Batty, C., Louca, M., Ho, S., Whisstock, J.C., Bottomley, S.P. and Buckle, A.M. (2006) The REFOLD database: a tool for the optimization of protein expression and refolding. Nucleic Acids Res. 34, D207–D212.
[5] Misawa, S. and Kumagai, I. (1999) Refolding of therapeutic proteins produced in *Escherichia coli* as inclusion bodies. Biopolymers 51, 297–307.
[6] Villaverde, A. and Carrio, M.M. (2003) Protein aggregation in recombinant bacteria: biological role of inclusion bodies. Biotechnol. Lett. 25, 1385–1395.
[7] Fernandez-Escamilla, A.M., Rousseau, F., Schymkowitz, J. and Serrano, L. (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat. Biotechnol. 22, 1302–1306.
[8] Miozzari, G.F. and Yanofsky, C. (1978) Translation of the leader region of the *Escherichia coli* tryptophan operon. J. Bacteriol. 133, 1457–1466.
[9] Derynck, R., Roberts, A.B., Winkler, M.E., Chen, E.Y. and Goeddel, D.V. (1984) Human transforming growth factor-alpha: precursor structure and expression in *E. coli*. Cell 38, 287–297.
[10] Cook, G.A., Stefer, S. and Opella, S.J. (2011) Expression and purification of the membrane protein p7 from hepatitis C virus. Biopolymers 96, 32–40.
[11] Kuliopulos, A. and Walsh, C.T. (1994) Production, purification, and cleavage of tandem repeats of recombinant peptides. J. Am. Chem. Soc. 116, 4599–4607.
[12] Li, Y. (2011) Recombinant production of antimicrobial peptides in *Escherichia coli*: a review. Protein Expr. Purif. 80, 260–267.
[13] Lee, J.H., Kim, J.H., Hwang, S.W., Lee, W.J., Yoon, H.K., Lee, H.S. and Hong, S.S. (2000) High-level expression of antimicrobial peptide mediated by a fusion partner reinforcing formation of inclusion bodies. Biochem. Biophys. Res. Commun. 277, 575–580.
[14] Rao, X.C., Li, S., Hu, J.C., Jin, X.L., Hu, X.M., Huang, J.J., Chen, Z.J., Zhu, J.M. and Hu, F.Q. (2004) A novel carrier molecule for high-level expression of peptide antibiotics in *Escherichia coli*. Protein Expr. Purif. 36, 11–18.
[15] Vidovic, V., Prongidi-Fix, L., Bechinger, B. and Werten, S. (2009) Production and isotope labeling of antimicrobial peptides in *Escherichia coli* by means of a novel fusion partner that enables high-yield insoluble expression and fast purification. J. Pept. Sci. 15, 278–284.
[16] Achmuller, C., Kaar, W., Ahrer, K., Wechner, P., Hahn, R., Werther, F., Schmidinger, H., Cserjan-Puschmann, M., Clementschitsch, F., Striedner, G., Bayer, K., Jungbauer, A. and Auer, B. (2007) N(pro) fusion technology to produce proteins with authentic N termini in *E. coli*. Nat. Methods 4, 1037–1043.
[17] Hwang, P.M., Pan, J.S. and Sykes, B.D. (2012) A PagP fusion protein system for the expression of intrinsically disordered proteins in *Escherichia coli*. Protein Expr. Purif. 85, 148–151.
[18] Booth, P.J. and Curran, A.R. (1999) Membrane protein folding. Curr. Opin. Struct. Biol. 9, 115–121.
[19] Cuesta-Seijo, J.A., Neale, C., Khan, M.A., Moktar, J., Tran, C.D., Bishop, R.E., Pomes, R. and Prive, G.G. (2010) PagP crystallized from SDS/cosolvent reveals the route for phospholipid access to the hydrocarbon ruler. Structure 18, 1210–1219.
[20] Rodriguez, J.C., Wong, L. and Jennings, P.A. (2003) The solvent in CNBr cleavage reactions determines the fragmentation efficiency of ketosteroid isomerase fusion proteins used in the production of recombinant peptides. Protein Expr. Purif. 28, 224–231.
[21] Franken, K.L., Hiemstra, H.S., van Meijgaarden, K.E., Subronto, Y., den Hartigh, J., Ottenhoff, T.H. and Drijfhout, J.W. (2000) Purification of his-tagged proteins by immobilized chelate affinity chromatography: the benefits from the use of organic solvent. Protein Expr. Purif. 18, 95–99.
[22] Ramjeesingh, M., Li, C., Garami, E., Huan, L.J., Hewryk, M., Wang, Y., Galley, K. and Bear, C.E. (1997) A novel procedure for the efficient purification of the cystic fibrosis transmembrane conductance regulator (CFTR). Biochem. J. 327, 17–21.
[23] Ikeda, T., Ninomiya, K., Hirota, R. and Kuroda, A. (2010) Single-step affinity purification of recombinant proteins using the silica-binding Si-tag as a fusion partner. Protein Expr. Purif. 71, 91–95.
[24] Ikeda, T., Motomura, K., Agou, Y., Ishida, T., Hirota, R. and Kuroda, A. (2011) The silica-binding Si-tag functions as an affinity tag even under denaturing conditions. Protein Expr. Purif. 77, 173–177.
[25] Graslund, T., Lundin, G., Uhlen, M., Nygren, P.A. and Hober, S. (2000) Charge engineering of a protein domain to allow efficient ion-exchange recovery. Protein Eng. 13, 703–709.
[26] Hedhammar, M., Alm, T., Graslund, T. and Hober, S. (2006) Single-step recovery and solid-phase refolding of inclusion body proteins using a polycationic purification tag. Biotechnol. J. 1, 187–196.
[27] Beckett, D., Kovaleva, E. and Schatz, P.J. (1999) A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. Protein Sci. 8, 921–929.
[28] Cull, M.G. and Schatz, P.J. (2000) Biotinylation of proteins in vivo and in vitro using small peptide tags. Methods Enzymol. 326, 430–440.

[29] Tagwerker, C., Flick, K., Cui, M., Guerrero, C., Dou, Y., Auer, B., Baldi, P., Huang, L. and Kaiser, P. (2006) A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with in vivocross-linking. Mol. Cell. Proteomics 5, 737–748.
[30] Darragh, A.J., Garrick, D.J., Moughan, P.J. and Hendriks, W.H. (1996) Correction for amino acid loss during acid hydrolysis of a purified protein. Anal. Biochem. 236, 199–207.
[31] Yamada, H., Moriya, H. and Tsugita, A. (1991) Development of an acid hydrolysis method with high recoveries of tryptophan and cysteine for microquantities of protein. Anal. Biochem. 198, 1–5.
[32] Tyler-Cross, R. and Schirch, V. (1991) Effects of amino acid sequence, buffers, and ionic strength on the rate and mechanism of deamidation of asparagine residues in small peptides. J. Biol. Chem. 266, 22549–22556.
[33] Long, S., Wendt, D.J., Bell, S.M., Taylor, T.W., Dewavrin, J.Y. and Vellard, M.C. (2012) A novel method for the large-scale production of PG-CNP37, a C-type natriuretic peptide analogue. J. Biotechnol. 164, 196–201.
[34] Provansal, M.M., Cuq, J.L. and Cheftel, J.C. (1975) Chemical and nutritional modifications of sunflower proteins due to alkaline processing. Formation of amino acid cross-links and isomerization of lysine residues. J. Agric. Food Chem. 23, 938–943.
[35] Fontana, A., Dalzoppo, D., Grandi, C. and Zambonin, M. (1981) Chemical cleavage of tryptophanyl and tyrosyl peptide bonds via oxidative halogenation mediated by *o*-iodosobenzoic acid. Biochemistry 20, 6997–7004.
[36] Griffiths, S.W., King, J. and Cooney, C.L. (2002) The reactivity and oxidation pathway of cysteine 232 in recombinant human alpha 1-antitrypsin. J. Biol. Chem. 277, 25486–25492.
[37] Gross, E. and Witkop, B. (1962) Non-enzymatic cleavage of peptide bonds: the methionine residues in bovine pancreatic ribonuclease. J. Biol. Chem. 237, 1856–1860.
[38] Schroeder, W.A., Shelton, J.B. and Shelton, J.R. (1969) An examination of conditions for the cleavage of polypeptide chains with cyanogen bromide: application to catalase. Arch. Biochem. Biophys. 130, 551–556.
[39] Kaiser, R. and Metzka, L. (1999) Enhancement of cyanogen bromide cleavage yields for methionyl-serine and methionyl-threonine peptide bonds. Anal. Biochem. 266, 1–8.
[40] Joppich-Kuhn, R., Corkill, J.A. and Giese, R.W. (1982) Oxidation of methionine to methionine sulfoxide as a side reaction of cyanogen bromide cleavage. Anal. Biochem. 119, 73–77.
[41] Kraft, P., Mills, J. and Dratz, E. (2001) Mass spectrometric analysis of cyanogen bromide fragments of integral membrane proteins at the picomole level: application to rhodopsin. Anal. Biochem. 292, 76–86.
[42] Jacobson, G.R., Schaffer, M.H., Stark, G.R. and Vanaman, T.C. (1973) Specific chemical cleavage in high yield at the amino peptide bonds of cysteine and cystine residues. J. Biol. Chem. 248, 6583–6591.
[43] Tang, H.Y. and Speicher, D.W. (2004) Identification of alternative products and optimization of 2-nitro-5-thiocyanatobenzoic acid cyanylation and cleavage at cysteine residues. Anal. Biochem. 334, 48–61.
[44] Piszkiewicz, D., Landon, M. and Smith, E.L. (1970) Anomalous cleavage of aspartyl–proline peptide bonds during amino acid sequence determinations. Biochem. Biophys. Res. Commun. 40, 1173–1178.
[45] Bornstein, P. and Balian, G. (1970) The specific non-enzymatic cleavage of bovine ribonuclease with hydroxylamine. J. Biol. Chem. 245, 4854–4856.
[46] Milner, S.J., Thomas, S.M., Ballard, F.J. and Francis, G.L. (1996) Optimization of the hydroxylamine cleavage of an expressed fusion protein to produce recombinant human insulin-like growth factor (IGF)-I. Biotechnol. Bioeng. 50, 265–272.
[47] Rahali, V. and Gueguen, J. (1999) Chemical cleavage of bovine beta-lactoglobulin by BNPS-skatole for preparative purposes: comparative study of hydrolytic procedures and peptide characterization. J. Protein Chem. 18, 1–12.
[48] Omenn, G.S., Fontana, A. and Anfinsen, C.B. (1970) Modification of the single tryptophan residue of staphylococcal nuclease by a new mild oxidizing agent. J. Biol. Chem. 245, 1895–1902.
[49] Stadtman, E.R. (1993) Oxidation of free amino acids and amino acid residues in proteins by radiolysis and by metal-catalyzed reactions. Annu. Rev. Biochem. 62, 797–821.
[50] Andberg, M., Jantti, J., Heilimo, S., Pihkala, P., Paananen, A., Koskinen, A.M., Soderlund, H. and Linder, M.B. (2007) Cleavage of recombinant proteins at poly-His sequences by Co(II) and Cu(II). Protein Sci. 16, 1751–1761.
[51] Milovic, N.M., Dutca, L.M. and Kostic, N.M. (2003) Transition-metal complexes as enzyme-like reagents for protein cleavage: complex cis-[Pt(en)(H_2O)(2)](2+) as a new methionine-specific protease.. Chem.-Eur. J. 9, 5097–5106.
[52] Smith, M.A., Easton, M., Everett, P., Lewis, G., Payne, M., Riveros-Moreno, V. and Allen, G. (1996) Specific cleavage of immunoglobulin G by copper ions. Int. J. Pept. Protein Res. 48, 48–55.
[53] Humphreys, D.P., King, L.M., West, S.M., Chapman, A.P., Sehdev, M., Redden, M.W., Glover, D.J., Smith, B.J. and Stephens, P.E. (2000) Improved efficiency of site-specific copper(II) ion-catalysed protein cleavage effected by mutagenesis of cleavage site. Protein Eng. 13, 201–206.
[54] Krezel, A., Kopera, E., Protas, A.M., Poznanski, J., Wyslouch-Cieszynska, A. and Bal, W. (2010) Sequence-specific Ni(II)-dependent peptide bond hydrolysis for protein engineering. Combinatorial library determination of optimal sequences. J. Am. Chem. Soc. 132, 3355–3366.
[55] Kopera, E., Belczyk-Ciesielska, A. and Bal, W. (2012) Application of Ni(II)-assisted peptide bond hydrolysis to non-enzymatic affinity tag removal. PLoS ONE 7, e36350.
[56] Milovic, N.M. and Kostic, N.M. (2003) Palladium(II) complex as a sequence-specific peptidase: hydrolytic cleavage under mild conditions of X-Pro peptide bonds in X-Pro-Met and X-Pro-His segments. J. Am. Chem. Soc. 125, 781–788.
[57] Lidell, M.E., Johansson, M.E. and Hansson, G.C. (2003) An autocatalytic cleavage in the C terminus of the human MUC2 mucin occurs at the low pH of the late secretory pathway. J. Biol. Chem. 278, 13944–13951.
[58] Zhu, L.G., Qin, L., Parac, T.N. and Kostic, N.M. (1994) Site-specific hydrolytic cleavage of cytochrome-c and of its heme undecapeptide, promoted by co-ordination complexes of palladium(II). J. Am. Chem. Soc. 116, 5218–5224.