



Review

Polyionic Tags as Enhancers of Protein Solubility in Recombinant Protein Expression

Vasiliki Paraskevopoulou and Franco H. Falcone *

Division of Molecular Therapeutics and Formulation, School of Pharmacy, University of Nottingham, Nottingham NG7 2RD, UK; paxvp@nottingham.ac.uk

* Correspondence: Franco.Falcone@nottingham.ac.uk; Tel.: +44-115846-6073

Received: 9 April 2018; Accepted: 21 May 2018; Published: 23 May 2018



Abstract: Since the introduction of recombinant protein expression in the second half of the 1970s, the growth of the biopharmaceutical field has been rapid and protein therapeutics has come to the foreground. Biophysical and structural characterisation of recombinant proteins is the essential prerequisite for their successful development and commercialisation as therapeutics. Despite the challenges, including low protein solubility and inclusion body formation, prokaryotic host systems and particularly *Escherichia coli*, remain the system of choice for the initial attempt of production of previously unexpressed proteins. Several different approaches have been adopted, including optimisation of growth conditions, expression in the periplasmic space of the bacterial host or co-expression of molecular chaperones, to assist correct protein folding. A very commonly employed approach is also the use of protein fusion tags that enhance protein solubility. Here, a range of experimentally tested peptide tags, which present specific advantages compared to protein fusion tags and the concluding remarks of these experiments are reviewed. Finally, a concept to design solubility-enhancing peptide tags based on a protein's pI is suggested.

Keywords: protein solubility; peptide tag; protein fusion tag; polycationic; polyanionic; recombinant protein expression

1. Introduction

Since the first successful attempt at recombinant production of the human peptide hormone Somatostatin in *Escherichia coli* in 1976 [1], protein therapeutics have come a long way. Until then, small amounts of proteins and enzymes had to be extracted and purified from large amounts of animal or plant tissues, or from biological fluids. However, the revolution that came along with the recombinant production of proteins, enabling large-scale production of biological macromolecules, allowed commercial success in the biopharmaceutical industry [2].

The biochemical characterisation of proteins is of utmost significance and a prerequisite prior to their commercialisation. This process requires sufficient amounts of protein, which can be generated with the use of recombinant technology. Protein stability, three-dimensional structure in order to identify active sites, binding affinity for ligands or determination of their interaction are some of the features that are useful in protein characterisation. In addition, from a regulatory point of view, post-translational modifications, protein structure and propensity to aggregate have to be defined in the development process of biosimilars [3].

Many different host systems have been exploited for recombinant protein production, including various prokaryotic systems, yeast, insect, plant and mammalian cells [4]. Although prokaryotic systems lack the mechanisms for complex post-translational protein modification, such as glycosylation, and expression of complex protein folds involving disulphide bond formation can be challenging in these systems [5], they have certain characteristics that frequently make them the host system of choice

for first-time protein production. In particular, *E. coli* is characterised by rapid growth at a low cost, with a cell doubling time of approximately 20 min [6]. This fast growth rate, in combination with the range of plasmids and safe, compatible strains that can be exploited, which can be flexibly tailored to the individual needs of recombinant production, makes this organism the ideal host [7].

2. Inclusion Bodies and Their Avoidance

Despite the clear advantages, recombinant expression of proteins in *E. coli* does not always guarantee success and is not obstacle-free, as there is not a single protocol that can be followed in order to avoid undesirable events. Insufficient yields, proteolytic degradation, protein misfolding, formation of inclusion bodies, as well as lack of protein or enzyme activity, are only a few of the possible unwanted outcomes [2].

Previously thought to be the result of unspecific hydrophobic interactions among intermediate, partially folded, products of protein expression [8,9], inclusion bodies are now recognised as ordered, dynamic structures, the organisation of which depends on specific interactions. In particular, Fourier transform infrared microspectroscopy has revealed the presence of residual, native-like secondary structures and intermolecular β -sheet structures in bacterial inclusion bodies, which resemble the organisation of amyloid fibrils—see Figure 1 [10–12]. It has also been suggested that inclusion body formation can be caused by self-association of correctly folded protein of low solubility or unfolded protein molecules of mature protein [13]. In any case, the formation of inclusion bodies can represent a major drawback during heterologous or homologous proteins' overexpression in bacterial host systems. This process is affected by and can be tuned with factors such as the environmental conditions (e.g., temperature, pH and ionic strength of culture medium), as well as the amino acid sequence of the protein [14] but can also be initiated by high protein concentration or inability of disulphide bonds to form in the reducing environment of the cytoplasmic space of the bacterial cells [13].

It should also be noted however, that more recently, formation of inclusion bodies has been seen as a potentially exploitable phenomenon, which has resulted in the development of strategies designed to enhance inclusion body formation, such as the addition of aggregation-prone tags or pull-down peptides [15,16]. Additionally, in some cases, the formation of inclusion bodies can facilitate protein purification. There are methods in place that allow the successful refolding of proteins from inclusion bodies, retaining their functionality [17]. However, the refolding process is not only time-consuming, but can also be problematic, yielding aggregated and/or inactive protein [18].

As Hippocrates taught, “prevention is better than cure.” Thus, various approaches aimed at preventing the formation of inclusion bodies, where undesirable, have been developed; these include optimisation of the growth conditions which lower the rate of protein expression and allow sufficient time for protein folding [19] and co-expression of molecular chaperones which mediate the correct folding of proteins [20,21].

Additionally, expression in the periplasmic space, where disulphide bonds can be formed in the presence of an oxidising environment [22], or mutations introduced in the reducing enzymes of the cytoplasmic space, which render them inactive and allow disulphide bond formation [23], have been exploited in order to overcome inclusion body formation. However, the most common approach is the fusion of the target protein with protein or peptide tags, which are known to enhance solubility [24]. One of the suggested mechanisms for the solubilisation effect is the increase of the net protein charge, which introduces repulsive electrostatic forces among protein molecules and promotes interaction with the solvent molecules [25,26].

A combination of the periplasmic expression approach with the addition of a peptide tag in the protein construct has been successfully used by our group in order to express a heterologous adhesin from *Helicobacter pylori* in *E. coli* [27].

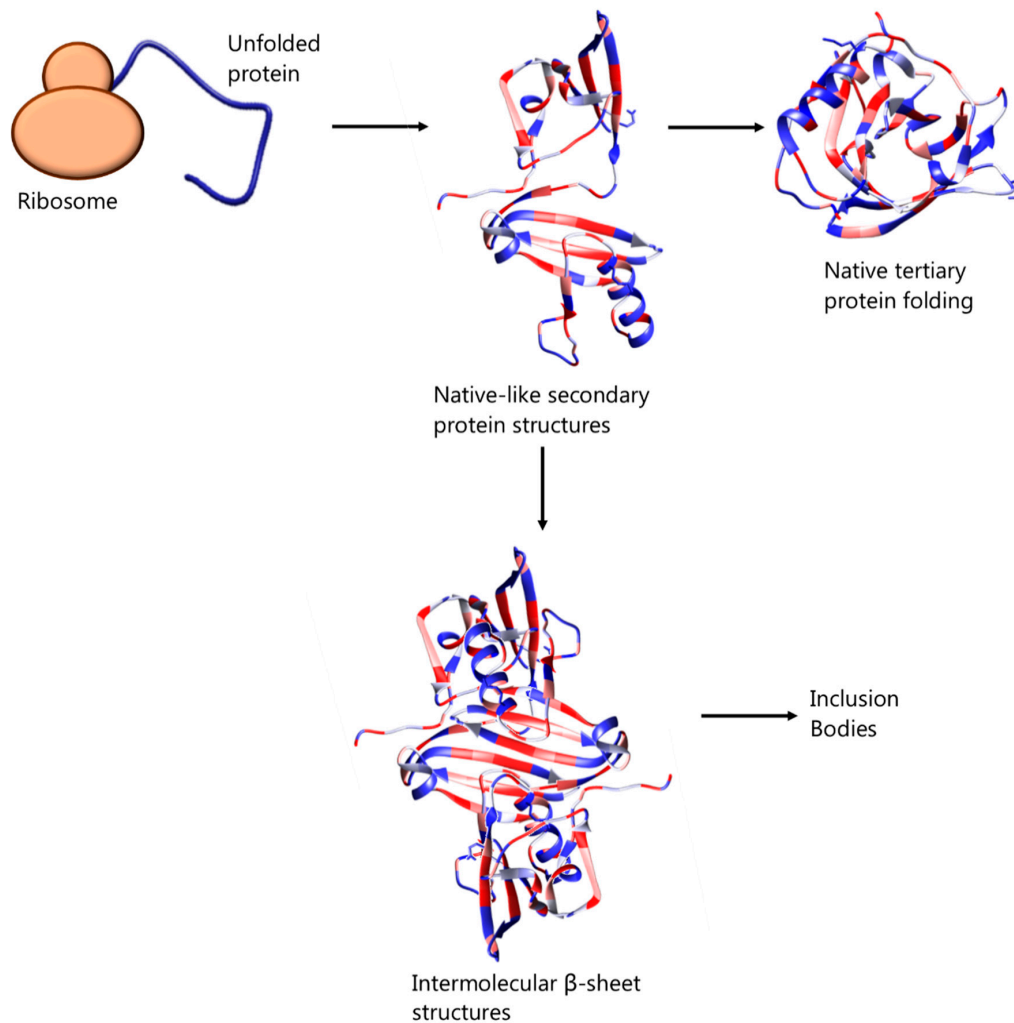


Figure 1. Suggested mechanism of native protein folding and association of intermediate, partially folded protein molecules, bearing native-like secondary structures, leading to inclusion body formation.

3. Protein Fusion Tags

Although their mechanism of action has not been fully elucidated, protein fusion tags are very commonly used in order to enhance the solubility of recombinantly expressed proteins in *E. coli* [28]. One of the very first protein fusion tags used was a heterologous protein, glutathione S-transferase (GST), from the trematode *Schistosoma japonicum*, which has been employed for the production and purification of numerous proteins of mammalian origin in *E. coli* [29].

Another protein, thioredoxin (Trx), is thermally stable and can be overexpressed, retaining its solubility even at high concentrations. It has been employed in order to enhance the solubility and facilitate the expression of many mammalian cytokines and growth factors, previously contained in inclusion bodies. Attempts to explain how the fusion protein is resistant to forming inclusion bodies suggest that the highly soluble thioredoxin does not aggregate and allows time for correct folding of the fusion protein, among other reasons [30].

Maltose-binding protein (MBP), 42.5 kDa, a homologous *E. coli* protein used as a solubility-enhancing tag [31], is significantly larger than GST (26 kDa) and Trx (11.7 kDa)—see Table 1; however, out of the three, MBP demonstrated the biggest solubilisation effect, as well as chaperone-like behaviour [32].

Table 1. Protein fusion tags for solubility enhancement during recombinant protein production.

Name	Full Name	Size (kDa)	Reference
GST	Glutathione-S-transferase	26	Smith et al., 1988 [29]
MBP	Maltose-binding protein	42.5	Maina et al., 1988 [31]
UB	Ubiquitin	~9	Butt et al., 1989 [33]
Trx	Thioredoxin	11.7	LaVallie et al., 1993 [30]
Z-tag/ZZ-tag	IgG-binding domain from protein A	15.5/31	Samuelsson et al., 1994 [34]
GB1	Immunoglobulin-binding domain of protein G	6.2	Huth et al., 1997 [35]
DsbA	Disulphide isomerase I	21.1	Collins-Racie et al., 1998 [36]
DsbA ^{mut}			Zhang et al., 1998 [37]
NusA	N-utilization substance A	55	Davis et al., 1999 [38]
IF2 domain I (or <i>InfB</i> (1-471))	Initiation factor 2		Sorensen et al., 2003 [39]
CaBP	Calcium binding protein		Reddi et al., 2002 [40]
SUMO	Small ubiquitin-related modifier	31	Malakhov et al., 2004 [41]
FTN-H	Ferritin heavy-chain		Ahn et al., 2005 [42]
Skp	Seventeen kilodalton protein	17	Chatterjee et al., 2006 [43]
T7PK	T7 protein kinase	4.5	Chatterjee et al., 2006 [43]
Ecotin	<i>E. coli</i> trypsin inhibitor	16	Malik et al., 2006 [44]
RpoA	RNA Polymerase α -subunit	39.5	Ahn et al., 2007 [45]
PotD	Spermidine/putrescine-binding periplasmic protein	39.8	Han et al., 2007 [46]
Crr	Glucose-specific phosphotransferase (PTS) enzyme IIA component	20	Han et al., 2007 [46]
Tsf	Elongation factor Ts	30.6	Han et al., 2007 [47]
SlyD	Aggregation-resistant protein	22.2	Han et al., 2007 [48]
msyB	Acidic protein	14	Su et al., 2007 [49]
RpoS	RNA polymerase sigma factor	39	Park et al., 2008 [50]
yjgD		15	Zou et al., 2008 [51]
rpoD	σ 70 factor of RNA polymerase	20	Zou et al., 2008 [51]
HaloTag7	Inactive derivative of DhaA, a bacterial haloalkane dehalogenase	34	Ohana et al., 2009 [52]
sfGFP	Superfolder green fluorescent protein		Wu et al., 2009 [53]
Mocr	Monomeric bacteriophage T7 0.3	16.7	DelProposto et al., 2009 [54]
SNUT	Solubility enhancing Ubiquitous Tag	19	Caswell et al., 2010 [55]
EspA	<i>E. coli</i> secreted protein A	25	Cheng et al., 2010 [56]
ArsC	Arsenate reductase	16	Song et al., 2011 [57]
BLA	AmpC-type β -lactamase		Tokunaga et al., 2010 [58]
<i>InfB</i> 1-21	Entity of <i>InfB</i> (1-471) responsible for increased expression	28	Hansted et al., 2011 [59]
Fh8	<i>Fasciola hepatica</i> antigen	8	Costa et al., 2014 [28]
SmbP	Small metal-binding protein	9.9	Vargas-Cortez et al., 2016 [60]
Ffu	β -fructofuranosidase truncations	17.7–29.5	Cheng et al., 2017 [61]
TDX	Tetracopeptide domain-containing thioredoxin	35	Xiao et al., 2018 [62]
HE-MBP(Pyr)	Truncated maltotriose-binding protein with modified histidine tag		Han et al., 2018 [63]

Also, derived from yeast, a small ubiquitin-related modifier, or SUMO, has been found to have an even better solubilising effect than MBP [64]. Although its mechanism of enhancing solubility is currently unclear, it is speculated that it might act as a chaperone, similarly to Ubiquitin [65]. Alternatively, it might function as a nucleation point for the correct folding of the fusion protein [64].

In contrast to the above, it has also been reported that the introduction of MBP and Trx in the C-terminus of the mammalian proteinase procathepsin D did not prevent inclusion body formation but facilitated the recovery of soluble, yet not active, protein following refolding [66]. A few hypotheses regarding the solubility-enhancing mechanisms of protein fusion tags are described in [65]. These include the conformation of the fusion proteins into micelle-like structures, the attraction of chaperones or an intrinsic chaperone-like activity in the fusion proteins and the presence of electrostatic repulsive forces due to the protein's net charge.

The solubilising effect of the protein tags seems to rely on the tags' correct folding. However, due to their large size, their three-dimensional conformation can potentially interfere with the structure and most significantly with the activity of the expressed protein [67]. Thus, proteolytic removal of these tags after expression and purification of the fusion protein is common practice; however, the target protein's solubility after tag removal cannot be predicted and the tag removal process might exert negative effects on the quality of the protein, such as product heterogeneity due to proteolytic cleavage at multiple sites, precipitation or poor recovery [24].

Although in most cases the removal of a big protein fusion tag is desired, there have been cases where the presence of the MBP has not proven an obstacle for the resolution of a crystal structure due to conformational heterogeneity attributed to the flexible linkage between the protein tag and the target protein [68]. In fact, techniques such as surface mutagenesis of MBP, in order to decrease entropy [69], or careful design of the linker between the MBP and the target protein [70], have been employed in order to facilitate crystallisation of the fusion protein.

4. Peptide Tags

As an alternative, small peptide tags have been used as solubility-enhancing tags, almost as early as protein fusion tags [71]. These peptide tags are relatively short, usually no longer than fifteen residues and comprise mostly one or two amino acids—repeated a varying number of times. They are polar and bear a positive or negative overall charge. Due to their small size and their repetitive amino acid content, they do not necessarily have an ordered three-dimensional conformation and are usually not resolved in protein crystal structures. This was the case for a hexalysine tag which was not defined in the crystal structure of the *Helicobacter pylori* adhesin BabA [72] or the ten different tags which were mostly invisible in [73]. As a result, they can exert their solubility-enhancing effect without interfering with the structure of the protein of interest or compromising its activity [67]. Additionally, an extra step for the removal of the peptide tags after production and/or purification is not necessarily required, in contrast to the case of protein fusion tags [24]. Finally, the expression of a large fusion protein tag instead of a short peptide tag is more demanding and acts as a metabolic burden on the bacterial hosts [25].

4.1. Polycationic Tags as Enhancers of Protein Solubility in Recombinant Protein Production

Since the first reference of a polylysine tag as a protein solubilising peptide tag in 1994 [71], many studies (reviewed here) have investigated the effect of different peptide tags on protein expression and solubility, without affecting the proteins' function and activity. In this original study, a formerly chemically synthesised protein of low solubility, the minibody [74], was instead expressed in *E. coli* with a 3-lysine tag incorporated in either the N- or the C-terminus. As a result, the aqueous solubility of the tagged protein was increased by a 100-fold [71].

The introduction of two positively charged lysine residues in the N-terminus of the enzyme HemA [75] or of a hexalysine tag in the C-terminus of the protein BabA [27] led to the protection of the proteins against proteolytic degradation. Possible explanations for this stabilisation effect could be

either the interference of the positively charged lysine residues preventing the binding of the proteases, or the correct folding of the tagged proteins [75]; a few proteases, such as DegP and Tsp, are known to show preference for mis- or unfolded target proteins, respectively. The hexalysine tag also seemed to strongly enhance the solubility of the recombinantly expressed BabA protein [27].

A slightly different solubilising tag, comprising glycine as well as lysine residues, proved to improve the solubility of the hydrophobic virus protein “u” (Vpu) from HIV-1, allowing its HPLC purification and 2D-NMR analysis in solution [76].

Other positively charged peptide tags were analysed for their effect on protein solubility, consisting of the basic amino acids arginine or histidine [25,67,77–82]. A comparison between arginine and lysine tags, from one to five residues long, in the N- or C-terminus of the poorly soluble bovine pancreatic trypsin inhibitor, revealed that the higher charges of the longer peptides had a bigger solubilising effect. Also, the position of the tag seemed to have an effect in this case, as the tags introduced to the C-terminus enhanced solubility more than the tags in the N-terminus of the protein. Finally, the arginine tags were more effective than the lysine tags of the same size in improving solubility, potentially due to the more hydrophilic character of arginine. The enhancement of the solubility was attributed to the repulsive electrostatic interactions between similarly charged tags, which prevent aggregation and allow sufficient time for correct folding, rather than their function as folding nuclei, which might have required a certain position in the expression construct [67].

Similar findings were obtained when positively charged arginine or lysine tags, comprising ten residues, were introduced into the enzyme *Candida antarctica* lipase B (CalB). The presence of these tags resulted in the transfer of the majority of the expressed protein from the insoluble to the soluble protein fraction in the cells, without affecting the protein expression yield overall [25].

As far as arginine tags are concerned, the introduction of a polyarginine tag in the C-terminus of the protein β -urogastrone, which led to the increase of the isoelectric point of the protein, has been used in order to facilitate the purification of the protein by cation exchange chromatography, which requires solubility in an aqueous system [77]. Recently, a C-terminal peptide tag rich in arginine was also exploited for the improved expression and enhanced solubility of the poorly soluble Tobacco Etch Virus (TEV) protease [78].

Histidine is the least basic amino acid out of the three, based on the pKa values of their side chains; 6.04 compared to 10.54 and 12.48 for lysine and arginine, respectively. Since the affinity of histidine-rich proteins for metal-ion resins was observed [83,84], hexahistidine tag has been established as one of the most popular affinity purification tags. The small size, the N- or C-terminal position that prevents interference with the function of the protein [85] and the highly selective interaction of histidine residues with nickel-NTA [86] are a few of the reasons that render the histidine tag so widely used in protein purification. However, when its effect on protein solubility was tested, it was found to be negative, resulting in lower protein solubility. In particular, the negative impact on protein solubility was stronger when the tag was found in the C-terminus, rather than the N-terminus, both in recombinant protein production in *E. coli* [79] and in a cell-free expression system [80].

Also, the effect of the length of a polyhistidine tag on protein expression was investigated and it was found that the longer decahistidine tag led to decreased expression yield of the protein aquaporin Z compared to the hexahistidine tag, without affecting the solubilisation of the protein by detergents [81]. Finally, the hexahistidine-tagged proteins were compared against proteins fused with other commonly used solubilising tags, such as GST and MBP and their relative solubility was found to be lower, as expected based on previous findings [82].

Chaperonins, a class of molecular chaperones that enhance protein folding in an ATP-dependent manner [87], have been found to interact with their substrate proteins based on the structural and biochemical properties of the latter. Hence, they can be classified based on their hydrophobic or polar interactions with protein substrates [88]. The molecular chaperonin CpkB from *Thermococcus kodakarensis* belongs to the second class, as it has been found that the negatively charged C-terminus of the enzyme facilitates its target protein recognition of positively charged proteins. The addition of a positively

charged tag to the target protein (see Table 2) can lead to enhanced specificity of the negatively charged chaperonin for the target protein, mediated by attractive electrostatic interactions; this results in protection of the protein against thermal denaturation and appropriate folding [89].

It has also been reported in the literature that the activity of the chaperone Hsp90 to prevent aggregation and enhance correct protein folding entirely depends on two acidic regions, bearing negative charge; upon deletion of this charge, the anti-aggregation activity is compromised [90].

4.2. Solubilising Peptide Tags in Solid-Phase Peptide Synthesis (SPPS) and Native Chemical Ligation

The solubility enhancement effect of a polycationic tag has also been investigated in the fields of SPPS and native chemical ligation. A polycationic tag, rich in but not entirely consisting of arginine, introduced in both the N- and C-termini of poorly soluble peptides synthesised by Boc or Fmoc SPPS, rendered them soluble in water and allowed their purification in an aqueous environment [91–93]. The same effect and improved purification was observed after the addition of a pentyllysine tag in the C-terminus of the poorly soluble A-chain of insulin glargine [94]. In all of the above examples, the solubilising tag was removed post purification.

In the case of native chemical ligation, which is a chemoselective reaction between an unprotected peptide with a C-terminal thioester modification and an also unprotected peptide with an N-terminal cysteine for the generation of native proteins [95], polyarginine tags, carrying positive charges, have been used. They have been shown to enhance the solubility of the peptide components of the membrane opioid receptor-like 1 [96] and the human immunodeficiency virus type 1 protease enzyme [97].

4.3. Polyanionic Tags as Enhancers of Protein Solubility in Recombinant Protein Production

Although so far positively charged polycationic tags which mostly enhance protein solubility have been reviewed, the opposite phenomenon has also been observed; polyanionic amino acid tags have been shown to enhance protein solubility too [98,99]. The addition of a negatively charged 9-aspartic acid tag led to increased solubility and expression of *Gaussia* luciferase in the soluble protein fraction [98]. Also, the presence of a polyaspartate tag resulted in increased protein expression and extracellular secretion of the periplasmic enzyme Asparaginase isozyme II [99].

Additions of single negatively charged residues, as well as longer sequences carrying negative net charge, were considered for their contribution to protein solubility of proteins prone to aggregation [100]. As with polycationic tags, the repulsive electrostatic interactions caused by the negative charge of the peptide tag seemed to enhance solubility and facilitate correct protein folding, by delaying protein aggregation, irrespectively of the size and structural conformation of the peptide tag [100]. Also, compared to commonly used solubility-enhancing fusion tags, such as MBP and Trx, peptide tags with high acidic content were found to enhance protein solubility to a greater extent [51].

4.4. Polycationic versus Polyanionic Tags

A few studies have compared homogeneous 5-amino acid long peptide tags comprising ten different amino acids with distinct biophysical properties (basic, acidic, polar and hydrophobic), side by side [73,101,102]. The overall conclusion of these studies was that the pentyllysine tag had the biggest solubilisation effect, although the majority of the tags seemed to enhance solubility to a greater or lesser extent, excluding proline and isoleucine tags. It was revealed that although the positively charged peptide tags consisting of lysine or arginine led to increased solubility of bovine pancreatic trypsin inhibitor at two different pH values, 4.7 and 7.7, the acidic tags consisting of aspartic or glutamic acid, only improved solubility at pH 7.7, where their side chains were in their ionised state [73,101]. Also, from all of the aforementioned peptide tags, only the protein bearing a pentyllysine tag was brought to high concentrations without reaching supersaturation and remained stable and aggregate-free at both pH values for up to two days [102].

A comprehensive list of polar, charged or neutral, peptide tags, which have been assessed and found to have a solubilisation effect on proteins that they are fused to, is presented in Table 2.

Following up on the observation that the majority of the signal peptides in the N-terminus of bacterial periplasmic proteins comprise basic amino acid residues, it was found that the presence of positive charges in the N-terminus is not essential for protein secretion [103]. To the contrary, when peptide tags with different biophysical properties were tested for their effect on protein expression and secretion of CalB, it was found that positively charged polylysine tags even hindered expression, while the negatively charged tags enhanced protein expression and secretion [104].

It has also been found that the presence of two arginine residues, bearing a positive charge, in the signal sequence of the mature protein alkaline phosphatase, restricted the secretion from the cytoplasmic to the periplasmic space; this could be either due to alteration of protein conformation or blockage of the secretion machinery when the positively charged N-terminus approaches the negatively charged phospholipids of the membrane [105]. Similarly, the presence of negative charge in the N-terminus leads to protein accumulation in the cytoplasm and delays protein secretion [103]. However, it has also been reported that this outcome can be reversed with the addition of a positively charged residue in the hydrophobic signal peptide [106].

Table 2. Polyionic or polar peptide tags assessed for their solubility enhancement effect during recombinant protein expression.

Name	pI	Size (kDa)	Reference
Polycationic tags			
(Arg) ₁₋₅	10.00–12.62	0.174–0.799	Kato et al., 2007 [67]
(Arg) ₅	12.62	0.799	Islam et al., 2015 [73]
(His) ₅	7.66	0.704	Islam et al., 2012 [101]
			Khan et al., 2015 [102]
(Arg) ₁₀	12.95	1.6	Jung H-J et al., 2011 [25]
			Johnson et al., 2007 [97]
(Arg-Gly-Gly) ₃ -Gly	12.30	0.886	Englebretsen et al., 1999 [92]
Poly(Arg)			Smith et al., 1984 [77]
(Gly-Arg) ₃ -(Arg) ₃	12.91	1.5	Kalpana et al., 2018 [78]
(Gly-Arg) ₄	12.48	0.871	Englebretsen et al., 1996 [91]
Gly-(Arg) ₅	12.62	0.856	Sato et al., 2005 [96]
Gly(Arg-Gly-Gly) ₃	12.30	0.886	
Gly(Lys-Gly) ₆	10.70	1.2	Choma et al., 1998 [93]
(Gly) ₂ -(Arg) ₂ -Gly-Arg	12.30	0.658	
Gly-Lys-Gly-(Lys) ₂	10.28	0.517	Gao et al., 2017
(Gly) ₂ (Lys) ₄	10.47	0.645	Park et al., 2003 [76]
(Lys) ₁₋₅	8.88–10.61	0.146–0.659	Kato et al., 2007 [67]
(Lys) ₂	10.00	0.274	Wang et al., 1999 [75]
(Lys) ₃	10.28	0.402	Bianchi et al., 1994 [71]
			Islam et al., 2015 [73]
(Lys) ₅	10.61	0.659	Hossain et al., 2009 [94]
			Islam et al., 2012 [101]
			Khan et al., 2015 [102]
(Lys) ₆	10.70	0.787	Hage et al., 2015 [27]
(Lys) ₁₀	10.94	1.3	Englebretsen et al., 1999 [92]
Polyanionic tags			
(Asp) ₅	3.34	0.593	Kim et al., 2015 [99]
			Kim et al., 2014 [104]
(Asp) ₅			Islam et al., 2015 [73]
(Glu) ₅	3.34	0.664	Islam et al., 2012 [101]
			Khan et al., 2015 [102]
[Gly-(Asp) ₃] ₃	3.16	1.2	Rathnayaka et al., 2011 [98]
Negative peptide extensions (>–6)			Zhang et al., 2004 [100]
Polar tags			
(Asn) ₅	5.50	0.588	Islam et al., 2015 [73]
(Gln) ₅	5.50	0.659	Islam et al., 2012 [101]
(Ser) ₅	5.50	0.453	Khan et al., 2015 [102]

It is worth mentioning that homogeneous polyionic tags, both positively and negatively charged, have been exploited in different applications in research. These include matrix-assisted refolding from

inclusion bodies and protein purification, both mediated by the reversible immobilisation of tagged protein on an ion-exchange resin [107]. An example is the use of a polyanionic peptide tag consisting of varying number of glutamic acid residues for the purification of the polyoma coat protein VP1 with anion-exchange chromatography [108].

This immobilisation feature can also facilitate the functionalisation of flat surfaces, by immobilising protein molecules on the surface in a specific and consistent orientation. Last but not least, the generation of chimeric bifunctional proteins, through the electrostatic attraction of two proteins with oppositely charged tags has been described; due to poor stability, however, the introduction of cysteine residues has also been studied for the formation of more stable, covalent disulphide bonds [107]. In particular, polyionic peptides have been exploited for the heterodimerisation of α -glucosidase fused with a 10-arginine tag and a modified Fab fragment fused with a 10-glutamic acid tag, both enhanced with a cysteine residue. The chimeric product retained both the enzymatic activity and antigen-binding capacity [109].

4.5. Polyionic Tags Displaying the Opposite Effect

It has also been reported in the literature that the presence of a hexalysine tag has led to recombinant protein production in inclusion bodies, due to the intramolecular attractive electrostatic interactions between the positively charged polylysine tag and the negatively charged protein at the intracellular pH 7.0 [26].

In addition, polypeptides comprising either lysine or glutamic acid residues have been exploited for the reversible precipitation of a range of proteins in low ionic strength solutions, which were then redissolved at physiological ionic strength (150 mM NaCl). Negatively charged proteins were precipitated by mixing with polylysine peptides and positively charged proteins were precipitated by mixing with polyglutamic acid peptides. The cause of precipitation is the intermolecular attractive electrostatic interactions between the proteins and the free peptides [110].

5. Supercharging of Proteins

Based on all of the aforementioned, it is also important to consider the effect of the overall protein charge on solubility, not found localised in one of the two termini but spread across the protein sequence and surface. So, although it is known that proteins are least soluble at their isoelectric points where they do not bear any net positive or negative charge, it was desired to prove that protein charges prevent aggregation [111]. The mutation of positively charged arginine residues of eukaryotic proteins expressed in *E. coli* to negatively charged aspartate residues resulted in enhanced protein solubility [37]. Increased solubility was also observed following mutation of residues of green fluorescent protein exposed to the solvent to positively or negatively charged residues, leading to highly charged protein molecules [111].

This process, called supercharging, prevented both thermally and chemically induced protein aggregation [111]. The same effect of enhanced solubility and stability was also observed after supercharging a human enteropeptidase; it was speculated that a small increase in the protein's net charge by supercharging the protein resulted in more significant increase in protein solubility than the solubility enhancement conveyed by peptide tags [112]. Nonetheless, the point mutations involved in the supercharging of protein surfaces could result in loss of protein activity and/or alter its biochemical properties [73].

6. Discussion

As explained above, there are several factors that can render a protein insoluble or lead to the formation of inclusion bodies during recombinant protein production; not only high protein concentrations can result in aggregation but also large proteins are more prone to it. Of course, the composition of a protein in its primary amino acid sequence is crucial for its propensity to aggregate, as long hydrophobic regions will make the protein less soluble [113].

On the other hand, protein net average charge, as well as hydrophilicity, are known to be related to protein solubility [114]. The average charge of a protein at a certain pH value depends on the pI, which can be calculated from the pKa values of the side chains of the residues that are ionised. Due to protein folding, the experimental pKa values, hence the pI of the protein, can be slightly different from the calculated value. As mentioned earlier, proteins are the least soluble at their pIs and their solubility increases at differing pH values [115].

Thus, the introduction of net charge by the addition of even a single amino acid residue can enhance solubility by introducing repulsive electrostatic interactions between protein molecules that allow sufficient time for the correct folding of proteins, or by disrupting hydrophobic interactions between or within the same protein molecule [25]. Such peptide tags amplify the solubility properties of the amino acids regardless of their position in the N- or C-terminus of the protein, which offers the advantage of flexibility [66]. Another advantage of peptide tags over protein fusion tags is that in order to exert their solubilisation effect, an ordered secondary structure is not required, as is the case for the protein fusion tags, in the cases when they potentially function as a folding nucleus [101]. The above rationale has been summarised in Figure 2.

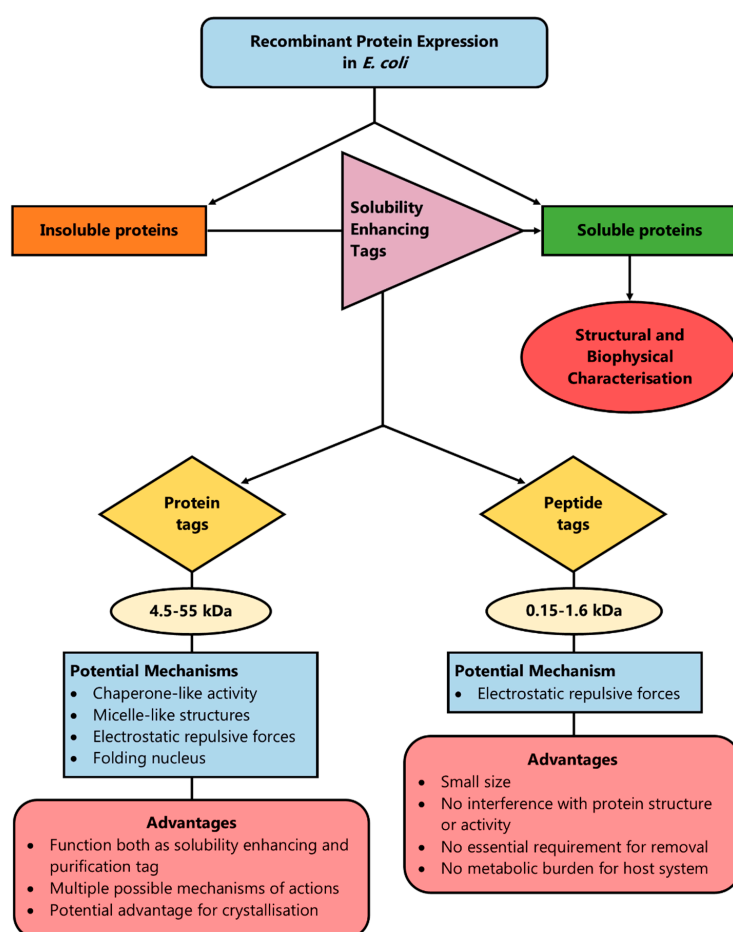


Figure 2. A method to enhance protein solubility during recombinant protein production is the introduction of solubility-enhancing tags (protein or peptide) in the recombinant plasmid. By having a few potential mechanisms of action, protein tags can cover a wider range of proteins in order to enhance solubility and most of them act simultaneously as solubility and purification tags. However, peptide tags are more versatile and smaller in size, which means their removal is not always essential, they do not pose a burden on the host system's metabolism and they do not affect the target protein's structure or function.

A method, based on highly conserved amino acid sequences in a range of soluble proteins, has been described for the design of novel solubility-enhancing peptide tags [116]. It is suggested that the solubility of a protein can be theoretically calculated and controlled. However, it is acknowledged that there always needs to be compatibility between the protein of interest and the solubility controlling peptide tag [116].

7. Conclusions

From all the above, it becomes obvious that the choice of the most appropriate solubility-enhancing tag depends on the individual protein and requires careful design; generalisation should be avoided [64]. Peptide tags have overall benefits compared to the protein fusion tags, due to their small size and versatility [73]. It is speculated that the introduction of a peptide tag bearing similar charge as the protein of interest at a certain pH value in either of the protein's termini will enhance solubility due to inter- and intramolecular repulsive interactions. Peptide tags of the opposite charge to the protein of interest should be avoided, as they could lead to protein precipitation instead.

Acknowledgments: V.P. and F.H.F.'s research is funded by the EPSRC (Grant EP/L01646X), AstraZeneca R&D and the University of Nottingham, through the Centre for Doctoral Training in Advanced Therapeutics and Nanomedicines.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Itakura, K.; Hirose, T.; Crea, R.; Riggs, A.D.; Heyneker, H.L.; Bolivar, F.; Boyer, H.W. Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. *Science* **1977**, *198*, 1056–1063. [[CrossRef](#)] [[PubMed](#)]
2. Rosano, G.L.; Ceccarelli, E.A. Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Front. Microbiol.* **2014**, *5*, 172. [[CrossRef](#)] [[PubMed](#)]
3. Berkowitz, S.A.; Engen, J.R.; Mazzeo, J.R.; Jones, G.B. Analytical tools for characterizing biopharmaceuticals and the implications for biosimilars. *Nat. Rev. Drug Discov.* **2012**, *11*, 527–540. [[CrossRef](#)] [[PubMed](#)]
4. Demain, A.L.; Vaishnav, P. Production of recombinant proteins by microbes and higher organisms. *Biotechnol. Adv.* **2009**, *27*, 297–306. [[CrossRef](#)] [[PubMed](#)]
5. Grangeasse, C.; Stülke, J.; Mijakovic, I. Regulatory potential of post-translational modifications in bacteria. *Front. Microbiol.* **2015**, *6*, 500. [[CrossRef](#)] [[PubMed](#)]
6. Sezonov, G.; Joseleau-Petit, D.; D'Ari, R. *Escherichia coli* Physiology in Luria-Bertani Broth. *J. Bacteriol.* **2007**, *189*, 8746–8749. [[CrossRef](#)] [[PubMed](#)]
7. Sorensen, H.P.; Mortensen, K.K. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J. Biotechnol.* **2005**, *115*, 113–128. [[CrossRef](#)] [[PubMed](#)]
8. Mitraki, A.; Haase-Pettingell, C.; King, J. Mechanisms of Inclusion Body Formation. In *Protein Refolding*; ACS Publications: Washington, DC, USA, 1991; pp. 35–49.
9. Ami, D.; Natalello, A.; Gatti-Lafranconi, P.; Lotti, M.; Doglia, S.M. Kinetics of inclusion body formation studied in intact cells by FT-IR spectroscopy. *FEBS Lett.* **2005**, *579*, 3433–3436. [[CrossRef](#)] [[PubMed](#)]
10. Carrio, M.; Gonzalez-Montalban, N.; Vera, A.; Villaverde, A.; Ventura, S. Amyloid-like properties of bacterial inclusion bodies. *J. Mol. Biol.* **2005**, *347*, 1025–1037. [[CrossRef](#)] [[PubMed](#)]
11. Ami, D.; Natalello, A.; Taylor, G.; Tonon, G.; Doglia, S.M. Structural analysis of protein inclusion bodies by Fourier transform infrared microspectroscopy. *Biochim. Biophys. Acta Proteins Proteom.* **2006**, *1764*, 793–799. [[CrossRef](#)] [[PubMed](#)]
12. Wang, L.; Maji, S.K.; Sawaya, M.R.; Eisenberg, D.; Riek, R. Bacterial Inclusion Bodies Contain Amyloid-Like Structure. *PLoS Biol.* **2008**, *6*, e195. [[CrossRef](#)] [[PubMed](#)]
13. Fink, A.L. Protein aggregation: Folding aggregates, inclusion bodies and amyloid. *Fold. Des.* **1998**, *3*, R9–R23. [[CrossRef](#)]
14. Strandberg, L.; Enfors, S.O. Factors influencing inclusion body formation in the production of a fused protein in *Escherichia coli*. *Appl. Environ. Microbiol.* **1991**, *57*, 1669–1674. [[PubMed](#)]

15. Rinas, U.; Garcia-Fruitos, E.; Corchero, J.L.; Vazquez, E.; Seras-Franzoso, J.; Villaverde, A. Bacterial Inclusion Bodies: Discovering Their Better Half. *Trends Biochem. Sci.* **2017**, *42*, 726–737. [[CrossRef](#)] [[PubMed](#)]
16. Garcia-Fruitos, E.; Vazquez, E.; Diez-Gil, C.; Corchero, J.L.; Seras-Franzoso, J.; Ratera, I.; Veciana, J.; Villaverde, A. Bacterial inclusion bodies: Making gold from waste. *Trends Biotechnol.* **2012**, *30*, 65–70. [[CrossRef](#)] [[PubMed](#)]
17. Middelberg, A.P.J. Preparative protein refolding. *Trends Biotechnol.* **2002**, *20*, 437–443. [[CrossRef](#)]
18. Cabrita, L.D.; Bottomley, S.P. Protein expression and refolding—A practical guide to getting the most out of inclusion bodies. *Biotechnol. Annu. Rev.* **2004**, *10*, 31–50. [[CrossRef](#)] [[PubMed](#)]
19. Georgiou, G.; Valax, P.; Ostermeier, M.; Horowitz, P.M. Folding and aggregation of TEM beta-lactamase: Analogies with the formation of inclusion bodies in *Escherichia coli*. *Protein Sci.* **1994**, *3*, 1953–1960. [[CrossRef](#)] [[PubMed](#)]
20. Caspers, P.; Stieger, M.; Burn, P. Overproduction of bacterial chaperones improves the solubility of recombinant protein tyrosine kinases in *Escherichia coli*. *Cell. Mol. Biol.* **1994**, *40*, 635–644. [[PubMed](#)]
21. Cole, P.A. Chaperone-assisted protein expression. *Structure* **1996**, *4*, 239–242. [[CrossRef](#)]
22. Makrides, S.C. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev.* **1996**, *60*, 512–538. [[PubMed](#)]
23. Derman, A.I.; Prinz, W.A.; Belin, D.; Beckwith, J. Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*. *Science* **1993**, *262*, 1744–1747. [[CrossRef](#)] [[PubMed](#)]
24. Esposito, D.; Chatterjee, D.K. Enhancement of soluble protein expression through the use of fusion tags. *Curr. Opin. Biotechnol.* **2006**, *17*, 353–358. [[CrossRef](#)] [[PubMed](#)]
25. Jung, H.-J.; Kim, S.-K.; Min, W.-K.; Lee, S.-S.; Park, K.; Park, Y.-C.; Seo, J.-H. Polycationic amino acid tags enhance soluble expression of *Candida antarctica* lipase B in recombinant *Escherichia coli*. *Bioprocess Biosyst. Eng.* **2011**, *34*, 833–839. [[CrossRef](#)] [[PubMed](#)]
26. Kim, S.-G.; Min, W.-K.; Rho, Y.-T.; Seo, J.-H. Electrostatic interaction-induced inclusion body formation of glucagon-like peptide-1 fused with ubiquitin and cationic tag. *Protein Expr. Purif.* **2012**, *84*, 38–46. [[CrossRef](#)] [[PubMed](#)]
27. Hage, N.; Renshaw, J.G.; Winkler, G.S.; Gellert, P.; Stolnik, S.; Falcone, F.H. Improved expression and purification of the *Helicobacter pylori* adhesin BabA through the incorporation of a hexa-lysine tag. *Protein Expr. Purif.* **2015**, *106*, 25–30. [[CrossRef](#)] [[PubMed](#)]
28. Costa, S.J.; Almeida, A.; Castro, A.; Domingues, L.; Besir, H. The novel Fh8 and H fusion partners for soluble protein expression in *Escherichia coli*: A comparison with the traditional gene fusion technology. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 6779–6791. [[CrossRef](#)] [[PubMed](#)]
29. Smith, D.B.; Johnson, K.S. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene* **1988**, *67*, 31–40. [[CrossRef](#)]
30. LaVallie, E.R.; DiBlasio, E.A.; Kovacic, S.; Grant, K.L.; Schendel, P.F.; McCoy, J.M. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Biotechnology* **1993**, *11*, 187–193. [[CrossRef](#)] [[PubMed](#)]
31. Maina, C.V.; Riggs, P.D.; Grandea, A.G.; Slatko, B.E.; Moran, L.S.; Tagliamonte, J.A.; McReynolds, L.A.; di Guan, C. An *Escherichia coli* vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein. *Gene* **1988**, *74*, 365–373. [[CrossRef](#)]
32. Kapust, R.B.; Waugh, D.S. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **1999**, *8*, 1668–1674. [[CrossRef](#)] [[PubMed](#)]
33. Butt, T.R.; Jonnalagadda, S.; Monia, B.P.; Sternberg, E.J.; Marsh, J.A.; Stadel, J.M.; Ecker, D.J.; Croke, S.T. Ubiquitin fusion augments the yield of cloned gene products in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 2540–2544. [[CrossRef](#)] [[PubMed](#)]
34. Samuelsson, E.; Moks, T.; Uhlen, M.; Nilsson, B. Enhanced in vitro Refolding of Insulin-like Growth Factor I Using a Solubilizing Fusion Partner. *Biochemistry* **1994**, *33*, 4207–4211. [[CrossRef](#)] [[PubMed](#)]
35. Huth, J.R.; Bewley, C.A.; Jackson, B.M.; Hinnebusch, A.G.; Clore, G.M.; Gronenborn, A.M. Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR. *Protein Sci.* **1997**, *6*, 2359–2364. [[CrossRef](#)] [[PubMed](#)]
36. Collins-Racie, L.A.; McColgan, J.M.; Grant, K.L.; DiBlasio-Smith, E.A.; McCoy, J.M.; LaVallie, E.R. Production of recombinant bovine enterokinase catalytic subunit in *Escherichia coli* using the novel secretory fusion partner DsbA. *Biotechnology* **1995**, *13*, 982–987. [[CrossRef](#)] [[PubMed](#)]

37. Zhang, Y.; Olsen, D.R.; Nguyen, K.B.; Olson, P.S.; Rhodes, E.T.; Mascarenhas, D. Expression of Eukaryotic Proteins in Soluble Form in *Escherichia coli*. *Protein Expr. Purif.* **1998**, *12*, 159–165. [[CrossRef](#)] [[PubMed](#)]
38. Davis, G.D.; Elisee, C.; Newham, D.M.; Harrison, R.G. New fusion protein systems designed to give soluble expression in *Escherichia coli*. *Biotechnol. Bioeng.* **1999**, *65*, 382–388. [[CrossRef](#)]
39. Sorensen, H.P.; Sperling-Petersen, H.U.; Mortensen, K.K. A favorable solubility partner for the recombinant expression of streptavidin. *Protein Expr. Purif.* **2003**, *32*, 252–259. [[CrossRef](#)] [[PubMed](#)]
40. Reddi, H.; Bhattacharya, A.; Kumar, V. The calcium-binding protein of *Entamoeba histolytica* as a fusion partner for expression of peptides in *Escherichia coli*. *Biotechnol. Appl. Biochem.* **2002**, *36*, 213–218. [[CrossRef](#)] [[PubMed](#)]
41. Malakhov, M.P.; Mattern, M.R.; Malakhova, O.A.; Drinker, M.; Weeks, S.D.; Butt, T.R. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J. Struct. Funct. Genom.* **2004**, *5*, 75–86. [[CrossRef](#)] [[PubMed](#)]
42. Ahn, J.-Y.; Choi, H.; Kim, Y.-H.; Han, K.-Y.; Park, J.-S.; Han, S.-S.; Lee, J. Heterologous gene expression using self-assembled supra-molecules with high affinity for HSP70 chaperone. *Nucleic Acids Res.* **2005**, *33*, 3751–3762. [[CrossRef](#)] [[PubMed](#)]
43. Chatterjee, D.K.; Esposito, D. Enhanced soluble protein expression using two new fusion tags. *Protein Expr. Purif.* **2006**, *46*, 122–129. [[CrossRef](#)] [[PubMed](#)]
44. Malik, A.; Rudolph, R.; Söhling, B. A novel fusion protein system for the production of native human pepsinogen in the bacterial periplasm. *Protein Expr. Purif.* **2006**, *47*, 662–671. [[CrossRef](#)] [[PubMed](#)]
45. Ahn, K.-Y.; Song, J.-A.; Han, K.-Y.; Park, J.-S.; Seo, H.-S.; Lee, J. Heterologous protein expression using a novel stress-responsive protein of *E. coli* RpoA as fusion expression partner. *Enzyme Microb. Technol.* **2007**, *41*, 859–866. [[CrossRef](#)]
46. Han, K.-Y.; Seo, H.-S.; Song, J.-A.; Ahn, K.-Y.; Park, J.-S.; Lee, J. Transport proteins PotD and Crr of *Escherichia coli*, novel fusion partners for heterologous protein expression. *Biochim. Biophys. Acta* **2007**, *1774*, 1536–1543. [[CrossRef](#)] [[PubMed](#)]
47. Han, K.-Y.; Song, J.-A.; Ahn, K.-Y.; Park, J.-S.; Seo, H.-S.; Lee, J. Enhanced solubility of heterologous proteins by fusion expression using stress-induced *Escherichia coli* protein, Tsf. *FEMS Microbiol. Lett.* **2007**, *274*, 132–138. [[CrossRef](#)] [[PubMed](#)]
48. Han, K.-Y.; Song, J.-A.; Ahn, K.-Y.; Park, J.-S.; Seo, H.-S.; Lee, J. Solubilization of aggregation-prone heterologous proteins by covalent fusion of stress-responsive *Escherichia coli* protein, SlyD. *Protein Eng. Des. Sel.* **2007**, *20*, 543–549. [[CrossRef](#)] [[PubMed](#)]
49. Su, Y.; Zou, Z.; Feng, S.; Zhou, P.; Cao, L. The acidity of protein fusion partners predominantly determines the efficacy to improve the solubility of the target proteins expressed in *Escherichia coli*. *J. Biotechnol.* **2007**, *129*, 373–382. [[CrossRef](#)] [[PubMed](#)]
50. Park, J.-S.; Han, K.-Y.; Lee, J.-H.; Song, J.-A.; Ahn, K.-Y.; Seo, H.-S.; Sim, S.-J.J.; Kim, S.-W.; Lee, J. Solubility enhancement of aggregation-prone heterologous proteins by fusion expression using stress-responsive *Escherichia coli* protein, RpoS. *BMC Biotechnol.* **2008**, *8*, 15. [[CrossRef](#)] [[PubMed](#)]
51. Zou, Z.; Cao, L.; Zhou, P.; Su, Y.; Sun, Y.; Li, W. Hyper-acidic protein fusion partners improve solubility and assist correct folding of recombinant proteins expressed in *Escherichia coli*. *J. Biotechnol.* **2008**, *135*, 333–339. [[CrossRef](#)] [[PubMed](#)]
52. Ohana, R.F.; Encell, L.P.; Zhao, K.; Simpson, D.; Slater, M.R.; Urh, M.; Wood, K. V HaloTag7: A genetically engineered tag that enhances bacterial expression of soluble proteins and improves protein purification. *Protein Expr. Purif.* **2009**, *68*, 110–120. [[CrossRef](#)] [[PubMed](#)]
53. Wu, X.; Wu, D.; Lu, Z.; Chen, W.; Hu, X.; Ding, Y. A Novel Method for High-Level Production of TEV Protease by Superfolder GFP Tag. *J. Biomed. Biotechnol.* **2009**, *2009*. [[CrossRef](#)] [[PubMed](#)]
54. DelProposto, J.; Majmudar, C.Y.; Smith, J.L.; Brown, W.C. Mocr: A novel fusion tag for enhancing solubility that is compatible with structural biology applications. *Protein Expr. Purif.* **2009**, *63*, 40–49. [[CrossRef](#)] [[PubMed](#)]
55. Caswell, J.; Snoddy, P.; McMeel, D.; Buick, R.J.; Scott, C.J. Production of recombinant proteins in *Escherichia coli* using an N-terminal tag derived from sortase. *Protein Expr. Purif.* **2010**, *70*, 143–150. [[CrossRef](#)] [[PubMed](#)]

56. Cheng, Y.; Gu, J.; Wang, H.; Yu, S.; Liu, Y.; Ning, Y.; Zou, Q.; Yu, X.; Mao, X. EspA is a novel fusion partner for expression of foreign proteins in *Escherichia coli*. *J. Biotechnol.* **2010**, *150*, 380–388. [[CrossRef](#)] [[PubMed](#)]
57. Song, J.-A.; Lee, D.-S.; Park, J.-S.; Han, K.-Y.; Lee, J. A novel *Escherichia coli* solubility enhancer protein for fusion expression of aggregation-prone heterologous proteins. *Enzyme Microb. Technol.* **2011**, *49*, 124–130. [[CrossRef](#)] [[PubMed](#)]
58. Tokunaga, H.; Saito, S.; Sakai, K.; Yamaguchi, R.; Katsuyama, I.; Arakawa, T.; Onozaki, K.; Arakawa, T.; Tokunaga, M. Halophilic beta-lactamase as a new solubility- and folding-enhancing tag protein: Production of native human interleukin 1alpha and human neutrophil alpha-defensin. *Appl. Microbiol. Biotechnol.* **2010**, *86*, 649–658. [[CrossRef](#)] [[PubMed](#)]
59. Hansted, J.G.; Pietikainen, L.; Hog, F.; Sperling-Petersen, H.U.; Mortensen, K.K. Expressivity tag: A novel tool for increased expression in *Escherichia coli*. *J. Biotechnol.* **2011**, *155*, 275–283. [[CrossRef](#)] [[PubMed](#)]
60. Vargas-Cortez, T.; Morones-Ramirez, J.R.; Balderas-Renteria, I.; Zarate, X. Expression and purification of recombinant proteins in *Escherichia coli* tagged with a small metal-binding protein from *Nitrosomonas europaea*. *Protein Expr. Purif.* **2016**, *118*, 49–54. [[CrossRef](#)] [[PubMed](#)]
61. Cheng, C.; Wu, S.; Cui, L.; Wu, Y.; Jiang, T.; He, B. A novel Ffu fusion system for secretory expression of heterologous proteins in *Escherichia coli*. *Microb. Cell Fact.* **2017**, *16*, 231. [[CrossRef](#)] [[PubMed](#)]
62. Xiao, W.; Jiang, L.; Wang, W.; Wang, R.; Fan, J. Evaluation of rice tetratricopeptide domain-containing thioredoxin as a novel solubility-enhancing fusion tag in *Escherichia coli*. *J. Biosci. Bioeng.* **2018**, *125*, 160–167. [[CrossRef](#)] [[PubMed](#)]
63. Han, Y.; Guo, W.; Su, B.; Guo, Y.; Wang, J.; Chu, B.; Yang, G. High-level expression of soluble recombinant proteins in *Escherichia coli* using an HE-maltotriose-binding protein fusion tag. *Protein Expr. Purif.* **2018**, *142*, 25–31. [[CrossRef](#)] [[PubMed](#)]
64. Marblestone, J.G.; Edavettal, S.C.; Lim, Y.; Lim, P.; Zuo, X.; Butt, T.R. Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. *Protein Sci.* **2006**, *15*, 182–189. [[CrossRef](#)] [[PubMed](#)]
65. Costa, S.; Almeida, A.; Castro, A.; Domingues, L. Fusion tags for protein solubility, purification and immunogenicity in *Escherichia coli*: The novel Fh8 system. *Front. Microbiol.* **2014**, *5*, 63. [[CrossRef](#)] [[PubMed](#)]
66. Sachdev, D.; Chirgwin, J.M. Solubility of Proteins Isolated from Inclusion Bodies Is Enhanced by Fusion to Maltose-Binding Protein or Thioredoxin. *Protein Expr. Purif.* **1998**, *12*, 122–132. [[CrossRef](#)] [[PubMed](#)]
67. Kato, A.; Maki, K.; Ebina, T.; Kuwajima, K.; Soda, K.; Kuroda, Y. Mutational analysis of protein solubility enhancement using short peptide tags. *Biopolymers* **2007**, *85*, 12–18. [[CrossRef](#)] [[PubMed](#)]
68. Smyth, D.R.; Mrozkiewicz, M.K.; McGrath, W.J.; Listwan, P.; Kobe, B. Crystal structures of fusion proteins with large-affinity tags. *Protein Sci.* **2003**, *12*, 1313–1322. [[CrossRef](#)] [[PubMed](#)]
69. Moon, A.F.; Mueller, G.A.; Zhong, X.; Pedersen, L.C. A synergistic approach to protein crystallization: Combination of a fixed-arm carrier with surface entropy reduction. *Protein Sci.* **2010**, *19*, 901–913. [[CrossRef](#)] [[PubMed](#)]
70. Jin, T.; Chuenchor, W.; Jiang, J.; Cheng, J.; Li, Y.; Fang, K.; Huang, M.; Smith, P.; Xiao, T.S. Design of an expression system to enhance MBP-mediated crystallization. *Sci. Rep.* **2017**, *7*, 40991. [[CrossRef](#)] [[PubMed](#)]
71. Bianchi, E.; Venturini, S.; Pessi, A.; Tramontano, A.; Sollazzo, M. High level expression and rational mutagenesis of a designed protein, the minibody. From an insoluble to a soluble molecule. *J. Mol. Biol.* **1994**, *236*, 649–659. [[CrossRef](#)] [[PubMed](#)]
72. Hage, N.; Howard, T.; Phillips, C.; Brassington, C.; Overman, R.; Debreczeni, J.; Gellert, P.; Stolnik, S.; Winkler, G.S.; Falcone, F.H. Structural basis of Lewis(b) antigen binding by the *Helicobacter pylori* adhesin BabA. *Sci. Adv.* **2015**, *1*, e1500315. [[CrossRef](#)] [[PubMed](#)]
73. Islam, M.M.; Nakamura, S.; Noguchi, K.; Yohda, M.; Kidokoro, S.; Kuroda, Y. Analysis and Control of Protein Crystallization Using Short Peptide Tags That Change Solubility without Affecting Structure, Thermal Stability, and Function. *Cryst. Growth Des.* **2015**, *15*, 2703–2711. [[CrossRef](#)]
74. Pessi, A.; Bianchi, E.; Cramer, A.; Venturini, S.; Tramontano, A.; Sollazzo, M. A designed metal-binding protein with a novel fold. *Nature* **1993**, *362*, 367. [[CrossRef](#)] [[PubMed](#)]
75. Wang, L.; Wilson, S.; Elliott, T. A mutant HemA protein with positive charge close to the N terminus is stabilized against heme-regulated proteolysis in *Salmonella typhimurium*. *J. Bacteriol.* **1999**, *181*, 6033–6041. [[PubMed](#)]

76. Park, S.H.; Mrse, A.A.; Nevzorov, A.A.; Mesleh, M.F.; Oblatt-Montal, M.; Montal, M.; Opella, S.J. Three-dimensional structure of the channel-forming trans-membrane domain of virus protein “u” (Vpu) from HIV-1. *J. Mol. Biol.* **2003**, *333*, 409–424. [[CrossRef](#)] [[PubMed](#)]
77. Smith, J.C.; Derbyshire, R.B.; Cook, E.; Dunthorne, L.; Viney, J.; Brewer, S.J.; Sassenfeld, H.M.; Bell, L.D. Chemical synthesis and cloning of a poly(arginine)-coding gene fragment designed to aid polypeptide purification. *Gene* **1984**, *32*, 321–327. [[CrossRef](#)]
78. Nautiyal, K.; Kuroda, Y. A SEP tag enhances the expression, solubility and yield of recombinant TEV protease without altering its activity. *New Biotechnol.* **2018**, *42*, 77–84. [[CrossRef](#)] [[PubMed](#)]
79. Woestenenk, E.A.; Hammarstrom, M.; van den Berg, S.; Hard, T.; Berglund, H. His tag effect on solubility of human proteins produced in *Escherichia coli*: A comparison between four expression vectors. *J. Struct. Funct. Genom.* **2004**, *5*, 217–229. [[CrossRef](#)]
80. Busso, D.; Kim, R.; Kim, S.-H. Expression of soluble recombinant proteins in a cell-free system using a 96-well format. *J. Biochem. Biophys. Methods* **2003**, *55*, 233–240. [[CrossRef](#)]
81. Mohanty, A.K.; Wiener, M.C. Membrane protein expression and production: Effects of polyhistidine tag length and position. *Protein Expr. Purif.* **2004**, *33*, 311–325. [[CrossRef](#)] [[PubMed](#)]
82. Hammarström, M.; Hellgren, N.; van den Berg, S.; Berglund, H.; Härd, T. Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. *Protein Sci.* **2002**, *11*, 313–321. [[CrossRef](#)] [[PubMed](#)]
83. Porath, J.; Carlsson, J.; Olsson, I.; Belfrage, G. Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* **1975**, *258*, 598–599. [[CrossRef](#)] [[PubMed](#)]
84. Nair, P.S.; Robinson, W.E. Purification and characterization of a histidine-rich glycoprotein that binds cadmium from the blood plasma of the bivalve *Mytilus edulis*. *Arch. Biochem. Biophys.* **1999**, *366*, 8–14. [[CrossRef](#)] [[PubMed](#)]
85. Bornhorst, J.A.; Falke, J.J. [16] Purification of Proteins Using Polyhistidine Affinity Tags. *Methods Enzymol.* **2000**, *326*, 245–254. [[PubMed](#)]
86. Crowe, J.; Dobeli, H.; Gentz, R.; Hochuli, E.; Stuber, D.; Henco, K. 6xHis-Ni-NTA chromatography as a superior technique in recombinant protein expression/purification. *Methods Mol. Biol.* **1994**, *31*, 371–387. [[CrossRef](#)] [[PubMed](#)]
87. Rüdiger, S.; Buchberger, A.; Bukau, B. Interaction of Hsp70 chaperones with substrates. *Nat. Struct. Biol.* **1997**, *4*, 342. [[CrossRef](#)] [[PubMed](#)]
88. Hirtreiter, A.M.; Calloni, G.; Forner, F.; Scheibe, B.; Puype, M.; Vandekerckhove, J.; Mann, M.; Hartl, F.U.; Hayer-Hartl, M. Differential substrate specificity of group I and group II chaperonins in the archaeon *Methanosarcina mazei*. *Mol. Microbiol.* **2009**, *74*, 1152–1168. [[CrossRef](#)] [[PubMed](#)]
89. Gao, L.; Hidese, R.; Fujiwara, S. Function of a thermophilic archaeal chaperonin is enhanced by electrostatic interactions with its targets. *J. Biosci. Bioeng.* **2017**, *124*, 283–288. [[CrossRef](#)] [[PubMed](#)]
90. Wayne, N.; Bolon, D.N. Charge-rich regions modulate the anti-aggregation activity of Hsp90. *J. Mol. Biol.* **2010**, *401*, 931–939. [[CrossRef](#)] [[PubMed](#)]
91. Englebretsen, D.R.; Alewood, P.F. Boc SPPS of two hydrophobic peptides using a “solubilising tail” strategy: Dodecaalanine and chemotactic protein 1042-55. *Tetrahedron Lett.* **1996**, *37*, 8431–8434. [[CrossRef](#)]
92. Englebretsen, D.R.; Robillard, G.T. An N-terminal method for peptide solubilisation. *Tetrahedron* **1999**, *55*, 6623–6634. [[CrossRef](#)]
93. Choma, C.T.; Robillard, G.T.; Englebretsen, D.R. Synthesis of hydrophobic peptides: An Fmoc “solubilising tail” method. *Tetrahedron Lett.* **1998**, *39*, 2417–2420. [[CrossRef](#)]
94. Hossain, M.A.; Belgi, A.; Lin, F.; Zhang, S.; Shabanpoor, F.; Chan, L.; Belyea, C.; Truong, H.-T.; Blair, A.R.; Andrikopoulos, S.; et al. Use of a temporary “solubilizing” peptide tag for the Fmoc solid-phase synthesis of human insulin glargine via use of regioselective disulfide bond formation. *Bioconjug. Chem.* **2009**, *20*, 1390–1396. [[CrossRef](#)] [[PubMed](#)]
95. Dawson, P.E.; Muir, T.W.; Clark-Lewis, I.; Kent, S.B. Synthesis of proteins by native chemical ligation. *Science* **1994**, *266*, 776–779. [[CrossRef](#)] [[PubMed](#)]
96. Sato, T.; Saito, Y.; Aimoto, S. Synthesis of the C-terminal region of opioid receptor like 1 in an SDS micelle by the native chemical ligation: Effect of thiol additive and SDS concentration on ligation efficiency. *J. Pept. Sci.* **2005**, *11*, 410–416. [[CrossRef](#)] [[PubMed](#)]

97. Johnson, E.C.B.; Malito, E.; Shen, Y.; Rich, D.; Tang, W.-J.; Kent, S.B.H. Modular Total Chemical Synthesis of a Human Immunodeficiency Virus Type 1 Protease. *J. Am. Chem. Soc.* **2007**, *129*, 11480–11490. [[CrossRef](#)] [[PubMed](#)]
98. Rathnayaka, T.; Tawa, M.; Nakamura, T.; Sohya, S.; Kuwajima, K.; Yohda, M.; Kuroda, Y. Solubilization and folding of a fully active recombinant Gaussia luciferase with native disulfide bonds by using a SEP-Tag. *Biochim. Biophys. Acta* **2011**, *1814*, 1775–1778. [[CrossRef](#)] [[PubMed](#)]
99. Kim, S.-K.; Min, W.-K.; Park, Y.-C.; Seo, J.-H. Application of repeated aspartate tags to improving extracellular production of *Escherichia coli* L-asparaginase isozyme II. *Enzyme Microb. Technol.* **2015**, *79–80*, 49–54. [[CrossRef](#)] [[PubMed](#)]
100. Zhang, Y.-B.; Howitt, J.; McCorkle, S.; Lawrence, P.; Springer, K.; Freimuth, P. Protein aggregation during overexpression limited by peptide extensions with large net negative charge. *Protein Expr. Purif.* **2004**, *36*, 207–216. [[CrossRef](#)] [[PubMed](#)]
101. Islam, M.M.; Khan, M.A.; Kuroda, Y. Analysis of amino acid contributions to protein solubility using short peptide tags fused to a simplified BPTI variant. *Biochim. Biophys. Acta Proteins Proteom.* **2012**, *1824*, 1144–1150. [[CrossRef](#)] [[PubMed](#)]
102. Khan, M.A.; Islam, M.M.; Kuroda, Y. Analysis of protein aggregation kinetics using short amino acid peptide tags. *Biochim. Biophys. Acta* **2013**, *1834*, 2107–2115. [[CrossRef](#)] [[PubMed](#)]
103. Vlasuk, G.P.; Inouye, S.; Ito, H.; Itakura, K.; Inouye, M. Effects of the complete removal of basic amino acid residues from the signal peptide on secretion of lipoprotein in *Escherichia coli*. *J. Biol. Chem.* **1983**, *258*, 7141–7148. [[PubMed](#)]
104. Kim, S.-K.; Park, Y.-C.; Lee, H.H.; Jeon, S.T.; Min, W.-K.; Seo, J.-H. Simple amino acid tags improve both expression and secretion of *Candida antarctica* lipase B in recombinant *Escherichia coli*. *Biotechnol. Bioeng.* **2015**, *112*, 346–355. [[CrossRef](#)] [[PubMed](#)]
105. Li, P.; Beckwith, J.; Inouye, H. Alteration of the amino terminus of the mature sequence of a periplasmic protein can severely affect protein export in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 7685–7689. [[CrossRef](#)] [[PubMed](#)]
106. Sung, C.Y.; Gennity, J.M.; Pollitt, N.S.; Inouye, M. A positive residue in the hydrophobic core of the *Escherichia coli* lipoprotein signal peptide suppresses the secretion defect caused by an acidic amino terminus. *J. Biol. Chem.* **1992**, *267*, 997–1000. [[PubMed](#)]
107. Lilie, H.; Richter, S.; Bergelt, S.; Frost, S.; Gehle, F. Polyionic and cysteine-containing fusion peptides as versatile protein tags. *Biol. Chem.* **2013**, *394*, 995–1004. [[CrossRef](#)] [[PubMed](#)]
108. Stubenrauch, K.; Bachmann, A.; Rudolph, R.; Lilie, H. Purification of a viral coat protein by an engineered polyionic sequence. *J. Chromatogr. B Biomed. Sci. Appl.* **2000**, *737*, 77–84. [[CrossRef](#)]
109. Richter, S.A.; Stubenrauch, K.; Lilie, H.; Rudolph, R. Polyionic fusion peptides function as specific dimerization motifs. *Protein Eng. Des. Sel.* **2001**, *14*, 775–783. [[CrossRef](#)]
110. Kurinamaru, T.; Maruyama, T.; Izaki, S.; Handa, K.; Kimoto, T.; Shiraki, K. Protein-poly(amino acid) complex precipitation for high-concentration protein formulation. *J. Pharm. Sci.* **2014**, *103*, 2248–2254. [[CrossRef](#)] [[PubMed](#)]
111. Lawrence, M.S.; Phillips, K.J.; Liu, D.R. Supercharging Proteins Can Impart Unusual Resilience. *J. Am. Chem. Soc.* **2007**, *129*, 10110–10112. [[CrossRef](#)] [[PubMed](#)]
112. Simeonov, P.; Berger-Hoffmann, R.; Hoffmann, R.; Strater, N.; Zuchner, T. Surface supercharged human enteropeptidase light chain shows improved solubility and refolding yield. *Protein Eng. Des. Sel.* **2011**, *24*, 261–268. [[CrossRef](#)] [[PubMed](#)]
113. Schein, C.H. Production of Soluble Recombinant Proteins in Bacteria. *Nat. Biotechnol.* **1989**, *7*, 1141. [[CrossRef](#)]
114. Wilkinson, D.L.; Harrison, R.G. Predicting the solubility of recombinant proteins in *Escherichia coli*. *Biotechnology* **1991**, *9*, 443–448. [[CrossRef](#)] [[PubMed](#)]

115. Shaw, K.L.; Grimsley, G.R.; Yakovlev, G.I.; Makarov, A.A.; Pace, C.N. The effect of net charge on the solubility, activity, and stability of ribonuclease Sa. *Protein Sci.* **2001**, *10*, 1206–1215. [[CrossRef](#)] [[PubMed](#)]
116. Hirose, S.; Kawamura, Y.; Mori, M.; Yokota, K.; Noguchi, T.; Goshima, N. Development and evaluation of data-driven designed tags (DDTs) for controlling protein solubility. *New Biotechnol.* **2011**, *28*, 225–231. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).