Fusion tags for enhancing the expression of recombinant proteins: A review

Authors: Nandini B Nataraj, Raja Sudhakaran

VIT, Vellore Campus, Vellore - 632 014 Tamilnadu, India.

Abstract

E. coli has always been the first organism of choice for the researchers for the production of recombinant proteins. It offers rapid and economical production of recombinant proteins and is known to be a well-established expression platform. However, the major problems associated with the *E. coli* expression system is recovering maximum properly folder proteins. To overcome this, typically some of the fusion tags are used for the expression of proteins. This tags not only helps in increasing the yield of fusion proteins produced. But some of them allow affinity based chromatographic purification. Subsequent removal of these fusion tags using specific proteases.

Aim and Objective: This article gives an overview of the fusion tag technology used for overproduction of recombinant proteins.

Key words: *E. coli: Escherichia coli; TRX: Thioredoxin; SUMO: Small Ubiquitin like modifier protein; GST: Glutathione S transferase; CPD: Cysteine Protease Domain; CBD: Chitin Binding Domain*

Corresponding Author:

Raja Sudhakaran

Sudhakaran.r@vit.ac.in

Phone: +919444070739

Fax: 044 3993 255

Introduction:

E. coli was the first host used for overexpression of recombinant proteins. Heterologous protein expression of recombinant proteins often results in aggregation of protein molecules into inclusion bodies. Inclusion bodies are dense insoluble aggregates which has to solubilized externally using some chaotropic agents and for proper folding of proteins, providing oxidized conditions is a major requirement. There are several factors which influences the formation of inclusion bodies such as induction temperature, high inducer concentrations, strong promoters and protein characteristic.

Recombinant products produced from *E. coli* is found in all major industry sectors such as enzyme industry, Agriculture, pharmaceutical/therapeutic use, functional analysis and structure determination[1]. The major limitation in the *E. coli* system is the lack of sophisticated machinery for post translational modification of expressed proteins. This results in reduced solubility and more of inclusion bodies[2]. Almost 75% of the proteins produced in *E. coli* are in soluble form, whereas 25% of the proteins are in active soluble form. This could be due to the lack of proper disulphide bond formation, absence of chaperons and codon bias[3]. Solubility tags helps to attain maximal solubility by providing oxidized environment in the cytoplasm. There are several fusion proteins which are used for production of various proteins, the choice of fusion proteins depends on these aspects; i) For improving the solubility, ii) To help in the purification, iii) If the proteins to be expressed is a peptide/toxic to cells, iv) Allows HTS.

Expression proteins in its insoluble form is convenient and effective in case of peptides as it is difficult for them to fold into its native structure. Fusion tag technology is another effective tool to enhance the expression of heterologous proteins. These fusion tags not only enhance the yield they also stabilize the protein. Earlier fusion tags were categorized into soluble type

and affinity tags. But now we can categorize them into solubility tags, hydrophobic tags, affinity tags and self-cleaving tags[4].

A major change is to design the protease cleavage site for efficient removal of fusion tags and protein without leaving a scar sequence[5]. Some of the example of solubility enhancing tags are SUMO (Small Ubiquitin like modifiers), MBP (maltose binding proteins), TRX (Thioredoxin), NUSA (N-Utilization substance A), GST (Glutathione S transferase). These are several hypotheses which are suggested to support/understand the mechanism behind the possibility of increasing the solubility by these tags. Some of them are i) These tags might attract chaperons, thereby leading to proper folding, ii) Might processes intrinsic chaperons like activity, intern presenting aggregation and promoting proper folding, iii) Net charge of the fusion protein; acidic fusion partner might inhibit protein aggregation[6,7,8,9].

SUMO (Small ubiquitin like modifier protein) is a 11kDa protein is known to modulate protein structure and function by covalent modification of target protein in eukaryotes. SUMO proteases cleave the C-termini of SUMO tag (-GG). This cleavage is based on the tertiary structure of the SUMO protein. The advantages of SUMO fusion are enhanced expression solubility and utilization of specific proteases. SUMO tag can be converted into an affinity tag by addition of HIS tag at the N-terminus of SUMO tag this will aid in purification of fusion protein using Ni-NTA column[10,11,12].

GST (Glutathione S transferase) is a 26 kDa protein found in parasitic helminth, *Schistosoma japonicum*[13]. GST fusion tag is not only known for enhancing the solubility but also used as affinity tag for purification.it has high binding affinity to glutathione moiety coupled to Sepharose matrix. This binding is reversible and can be eluted under mild non-denaturing conditions by addition of reduced glutathione[14]. A specific protease cleavage site has to be designed for cleavage of GST from PDI.

MBP is a large (43 kDa) periplasmic protein which is highly soluble in *E. coli*. it has a native affinity towards maltose as it plays an important role in translocation of maltose. Structure of MBP is the key factor for enhancing the solubility of the fusion proteins[15] intrinsic chaperon activity of MPB promotes proper folding of the target protein. MBP binds to immobilized amylase resin for easy purifications[16].

NusA-N utilization pathway is 55kDa protein and used as a fusion partner to enhance the solubility of the target protein. NusA is a transcription termination/anti termination protein that promotes/prevents RNA polymerase forming anti-termination complex. NusA increases the solubility of protein by showing down translation at the transcription time, thus providing more time for proper folding of protein[17,18,19]. Thrombin or enterokinase cleavage sites are usually introduced between NusA tag and target proteins.

TRX is a 12 kDa protein used as fusion partner for increasing the solubility of fusion protein. TRX is used to avoid the inclusion bodies. It is employed as fusion tag considering that its intrinsic oxido-reductase activity is responsible for the reduction of disulfide bonds through thio-disulfide exchange[20,21,22]. TRX can be fused to either N/C terminus of target protein with a proper protease cleavage site. Addition of His tag at the N-/C-terminus of TRX makes it the right candidate for affinity-based purification. Truncating this tag is another strategy to reduce the molecular weight of the fusion tag which might have an impact on fusion protein yield. Nandini et.al.,[23] have successfully truncated TRX protein into peptide tags of 10 – 15 amino acids. This peptide was fused to proinsulin gene and expressed in *E. coli* as inclusion bodies. They achieved an yield of 156 mg/g of IB's.

Fh8 is a small antigen (8 kDa) secreted by the parasite *F. hepatica* during the early onset of infection. When recombinantly produced in *E. coli* for the development of diagnosis method. It was observed that this protein is highly soluble and prosses unusual thermal stability[24,25].

Fh8 has high homology with calcium binding protein and calmodulin like proteins. Tertiary structure revealed that 11 amino acids in the N-terminus is involved in the calcium binding. This was named as H sequence/H tag. Studies were performed to compare the stability and solubility of both Fh8 and H tag[26,27,28]. It was observed that H tag showed lesser solubility of fusion protein compared to Fh8 tag. Due to its native Ca-binding nature, it can be clubbed with HIC for the purification. There is also experimental proof available for HIC-His based purification of Fh8 fusion protein allowing a dual protein purification strategy which was used sequentially to obtain active and pure protein.

Next set of fusion tags are the ones which can induce inclusion body formation. There are a number of proteins which are used as fusion partners for targeted expression of proteins as inclusion bodies. It is a well-known fact that, if proteins are over produced in *E. coli* they form aggregates. There is a dynamic equilibrium between the three states of protein expression: properly folder, partially folded and aggregated. Balancing these three states depends on the availability of chaperons[29]. Some of the well-known fusion tags which are used for producing the proteins as aggregates are TrpΔLE, KSI, Pag P, NPro.

TrpΔLE 1413 sequence is known to strongly favors the inclusion body formation. TrpL is the leader sequence in the TRP operon, it is a short sequence comprising of 14 amino acids. TrpE resulted in an N-terminal 17 amino acid leader sequence. A fusion of Trp L and TrpE leader sequences led the derivation of TrpLE fusion sequence. This sequence was fused with several small proteins/peptides/enzymes and expressed in *E. coli*and was found to greatly enhance the expression level. "MKAIFVLKSGLDRDPEF" was the fusion sequence and was found to be highly hydrophobic[30,31].

Ketosteroid Isomerase (KSI), another widely used fusion partner which induces aggregation formation of the fusion protein.it is 14 kDa soluble protein but highly hydrophobic with strong

tendency to accumulate in inclusion bodies. pET31b (Novagen) with expression vector has been commercialized with KSI fusion tag. There is a methionine amino acid between fusion tag and PDI and the tag is separated by the addition of cyanogen bromide which cleaves after methionine. The major drawback which is seen in this system is the presence of both cleaves and uncleaved proteins. This mixture is purified using HPLC, but there will be a drop in the yield observed due to the presence of uncleaved proteins[32].

Another recently developed fusion tag is Pag-P based protein for inclusion body formation. It is a gram negative bacterial outer membrane and integral with a hydrophobic exterior facing and hydrophilic interior core. Inclusion body formation is favored due to high β-sheet content in the protein and the insolubility of the protein to fold[33]. This β-barrel nomenclature protein is surprisingly not very hydrophobic, thus making the solubilization of fusion protein aggregates earlier. Several proteases are being used for the removal of tags, but still there is a necessity to use these tags due to inefficient processing of proteases and the presence of heterologous mixture of proteins.

Tag removal has always been a topic of discussion due to these disadvantages. Usually, fusion partners are separated from the fusion proteins by enzymatic cleavage or chemical cleavage. In enzymatic cleavage, site specific proteases identify the target region and cleave the bonds between the amino acids. There are usually endo proteases and the verity of such proteases are being studied by researchers. Some of the well-known proteases are serine proteases such as Enterokinase (DDDDK/), factor Xa (IEGR/), Thrombin (LVPR/GS) and some viral proteases such as TEV (ENLYFQ/G) and Prescission (LEVLFQ/GP). These endo-proteases cleaves at the c-terminal end of the fusion proteins leaving the N-terminus of the protein intact. However, apart from enterokinase all other enzymes leave a scar sequence which becomes a part of protein of interest. If the protein to be expressed is a therapeutic[34,35].

i. Protein and the addition amino acid present at the N-terminus would pose the regulatory hurdles[36,37,38,39].

ii. Exoproteases – they are often in combination with endoprotease for the removal of C-terminal fusion tags. Widely used exoprotease are carboxypeptidases and aminopeptidases.

Removal of fusion tag is usually accomplished in two purification steps, initial affinity purification (using his tag located at the N-terminus of fusion tag) to separate fusion protein from other host cell proteins. Followed by protease digestion step for separation of tag and protein. Second affinity purification step to recover the protein. This will cause a loss in protein yield and recovery. There are several reasons for the reduction in protease cleavage efficiently and they are i) Steric hindrance, ii) Presence of unfavorable residues around the cleavage site, iii) Inefficient processing of protease, iv) Protease concentration required, v) Low protein yield, vi) High cost of protease.

Hence fusion tag removal adds another layer of complexity and expense to the process. This can be overcome by employing self-cleaving tags as fusion partners. These self-cleaving tags can be combined with some affinity tags for efficient protein separation. Some of the well studies self-cleaving tags are Inteins, Npro, sortase, CPD and FrpC.

Inteins are protein fragments which can excise themselves from protein precursors and can rejoin the flanking region. Inteins themselves can be divided into three major domains, endonuclease region sandwiched between N-terminal splicing domain and C-terminal splicing domain. Recently there are some mini-inteins isolated which are of size 134 to 198 amino acids and were isolated from ribonucleoside diphosphate reductase gene of Methanobacterium thermoautotrophicum. One of the most interesting mini-inteins known till date is trans splicing inteins from the catalytic subunit of a DNA pol III, which was present in Cyanobacterium

synechocystis. Most studied inteins are *Saccharomyces cerevisiae* VMA gene which encodes for vacuolar ATPase[40,41]. A comprehensive data base is available for various inteins and its sequences known as inbase[42] at NEB studied about these inteins in great detail and developed a technology known as intein based affinity purification system. Here the C terminus of target protein is fused to chitin binding domain (CBD) which serves as an affinity tag when passed through the chitin column. Intein mediated cleavage is triggered by thiol reagents such as DTT or β – mercaptoethanol. This system is known as IMPACT (Intein mediated purification with an affinity chitin binding tag)[43]. In this system, intein tag can be used either on N or C terminus, which provides a greater flexibility for researchers. Various intein tags are available ranging from Sce VMA intein which is of 51 kDa to mini intein system Ssp intein which is of 17kDa. In some of the cases intein tags have not contributed to the increase in the yield of fusion protein, but they were helpful during the purification and cleavage. To overcome these, researchers have fused TRX and SUMO tags at the N – terminus of the intein tag followed by protein of interest. This combination has relatively increased the yield as well as solubility. Presence of His tag in the SUMO / TRX also helps in purifying proteins employing IMAC purification technology[44,45]. There are two preferred methods of cleavage, one is thiol mediated, which may not be cost effective at large scale and the other one is pH/temperature induced cleavage. Above all the fusion protein needs to be expressed in soluble and correctly folded form for efficient cleavage of proteins. Cleavage efficiency is high when there are amino acids such as Asp, Glu, Arg, His or Thr and the cleavage efficiency is low in the presence of Asn, Cys or Pro. This is applicable when intein tag is present at the C terminus of the target protein[46]. Inefficient cleavage was also observed due to the steric hindrance caused by the first few amino acids of the target protein which in turn leads to less yield. Another drawback is the uncontrolled invivo cleavage of fusion protein[47,48]. A variety of proteins was expressed using intein tag and yield was observed was in the range of 0.05 mg/L to 60 mg/L[49].

Sortases are bacterial transpeptidases that recognize the sequence motif LPXTG (where X is any amino acid) in the C terminal region of bacterial cell wall, proteins to subsequently cleave after the Threonine residue[50]. Mao and team developed the self - cleavable fusion system exploiting the SrtA self - cleavage activity. Calcium acts as a co factor and triggers SrtA's activity. The fusion construct was developed by introducing His tag at the N terminus of SrtA tag with LPXTG motif which is followed by protein of interest. The protein expressed was highly soluble, which was purified using affinity chromatography (IMAC) column. Since Calcium triggers the cleavage of this protein, on resin cleavage was induced by incubating in calcium buffer. The cleavage happens between Threonine and Glycine, leaving behind glycine residue at the N – Terminus of the target protein. Mao successfully expressed and purified three different proteins: GFP (soluble, 25.5 mg/L), Cre (partially soluble, 0.8 mg/L) and p27 (soluble, 1.9 mg/L). Even though SrtA tag has its own advantages such as enhancing the solubility of fusion protein and efficient cleavage of the tag in the presence of Ca ions (cost effective inducer), there are some disadvantages associated with this. One of it is the presence of additional amino acid at the N – terminus of the target protein and the other one is cleavage during the protein expression within the cells.

Npro is an autoproteolytic protein derived from pestiviruses. Its proteolytic activity is exploited in recombinant expression of fusion protein. The major short coming of this fusion tag is the molecular weight of the protein which is ~ 19kDa and this protein is highly hydrophobic. Hydrophobicity of this protein facilitates the fusion protein aggregation[51]. The advantages of using this tag are high yields of fusion protein, stable expression of toxic peptides and proteins and easier downstream processing. Precise and efficient cleavage of N pro tag at the C terminal residue (C168) can be achieved by rapid refolding the solubilized proteins using kosmotropic buffers. This N pro fusion protein tends to form insoluble aggregates during refolding, this inturn inhibits autoproteolysis. To overcome this Achmuller et.al., 2007 A, constructed a

mutant variant of N pro called EDDIE. This was designed by exchanging 11 amino acids. EDDIE demonstrated the reduced invivo cleavage, imptoved invitro refolding, solubility of fusion protein and efficient cleavage. The cleavage efficiency of Npro tag is greatly affected in the presence of proline as the first amino acid of the target protein. It was observed that N terminal proline blocks the cleavage completely[52].

Since the N pro protein is a 19 kDa protein, the overall yield of the fusion protein is greatly impacted. To overcome this[53], developed a complementation strategy. This strategy enabled the reduction of Npro tag size from 168 aa to 58 aa. The N terminal Npro sequence was expressed individually, whereas the C terminal fragment was fused to the target protein. When both the fragments were mixed, due to complementation the autoproteolytic activity was observed.

Cysteine protease domain (CPD)

MARTX toxin which is secreted by Vibrio cholerae as a large multifunctional auto processing repeats undergoes autocatalytic cleavage during translocation. Proteolysis of the CPD is induced by specific small molecule InsP6 (Inositol hexa bisphosphate). P4 – P4' residues identified from several known cleavage sites indicates that Leucine at P1 position[54] is the most preferred amino acid for cleavage. Ins P6 molecule is not present in the bacterial cells hence intact CPD fusion proteins can be purified from bacterial cell lysates. His_6 tagged at the C terminus of the CPD fusion enables purification of this fusion protein using IMAC column. Column bound InsP6 was used to trigger CPD mediated cleavage activity. Released protein was further purified by His tag binding of CPD to IMAC column; flow through would contain protein of Interest. Shen et.al., have tried to express two different proteins using CPD fusion which showed a yield of 5.9 mg/l (gp 130) and 26.9 mg/l (Bir A). The major disadvantage of

this combination is the residual amino acids and the specificity of P1' to P4' position leaves residual four amino acids in the protein of interest.

FrpC is another autoproteolytic protein produced in the gram-negative bacterium Neisseria meningitides. It is an iron regulated protein which is induced for autocatalytic activity by calcium ions. In the presence of calcium ions bond between Asp^{414} and Pro^{415} is autocatalytically cleaved[55,56]. Major drawbacks of this system are i) size of fusion protein, ii) presence of residual amino acid at the N terminus of the protein.

If it is complex / toxin proteins, then solubility enhancing tags are compatible as they help in promoting protein solubility. There are several solubility enhancing tags available and each of them have been constantly challenged. These tags have proven to promote solubility of various target proteins in *E. coli* cytoplasm and also cost-effective protein purification. Major obstacle in the fusion tag technology is the separation of the fusion tag and target protein. Enzymes involved in the removal of fusion tag add a significant cost to the process. To overcome this self - cleaving tags were identified, but one of the major hindrances using self - cleaving tags is that, it leaves a scar except Intein and N-pro tag. This depends on the P1' amino acid.

Although there are advantages and disadvantages associated with these fusion tags, hydrophobic tags which are known to increase the fusion protein yield is widely used. These tags helps in maximizing protein expression and since the recombinant proteins are expressed as inclusion bodies, they are highly pure. Recovery of these proteins can be maximized by optimizing the solubilization and refolding steps. *E. coli* has been successfully used to produce industrial scale proteins using fusion tag technology. Identifying the correct fusion partner which complements the target protein solves the protein yield issues.

Conclusion

Growing demand of recombinant proteins gives rise to understanding and identifying the important key factors responsible for overexpressing these. One of the major factors is the recombinant protein yield, to achieve these various parameters play a major role. One of it is the fusion tag strategy which is well exploited for increasing the protein yield. From the above information it is very evident that choosing the right fusion tag will help in enhancing the yield of recombinant proteins. In some cases, peptide tags are beneficial compared to protein tags due to the flexibility they bring in the system as they are small.

Acknowledgement: Authors acknowledge the scholars whose articles are cited and included in references of this manuscript.

Conflict of interest: None

Source of funding: Nil

References:

1. Demain AL, Vaishnav P. Production of recombinant proteins by microbes and higher organisms. Biotechnol Adv 2009; 27, 297 – 306.
2. Kamionka M. Engineering of therapeutic proteins production in Escherichia coli. Curr Pharm Biotechnol 2011; 12, 268 - 274.
3. Terpe K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol 2006; 72, 211–222.
4. Malhotra A. Tagging for protein expression. Methods Enzymol 2009; 463:239-258.
5. Waugh DS. Making the most of affinity tags. Trends Biotechnol. 2005; 23, 316-320.
6. Huang YS, Chuang DT. Mechanisms for GroEL/GroES mediated folding of a large 86-kDa fusion polypeptide invitro. J Biol Chem 1999; 274, 10405-10412.
7. Kapust RB, Waugh DS. *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.
8. Zhang YB, Howitt J, Mc corkle S, Lawrence P, Springer K, Freimuth P. Protein aggregation during over expression limited by peptide extensions with large net negative charge. Protein Expr Purif. 2004; 36, 207-216.
9. Su Y, Zou ZR, Feng SY, Zhou P, Cao LJ. 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.

10. Panavas T, Sanders C, Butt TR. SUMO fusion technology for enhanced protein production in prokaryotic and eukaryotic expression systems. Methods Mol Biol 2009; 303–317.
11. Marblestone JG, Edavettal SC, Lim Y, Lim P, Zuo X, Butt TR. Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. Protein Sci 2006; 15, 182–189.
12. Wang ZY, Li N, Wang YY, Wu YP, Mu TY, Zheng Y, et al. Ubiquitin- intein and SUMO2-intein fusion systems for enhanced protein production and purification. Protein Expr Purif 2012; 82, 174–178.
13. Smith DB, Johnson KS. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione-S-transferase. Gene 1998; 67, 31–40.
14. Ketterer B. A bird's eye view of the glutathione transferase field. Chem Biol Interact 2001; 138, 27–42.
15. Fox JD, Kapust RB, Waugh DS. Single amino acid substitutions on the surface of *Escherichia coli* maltose – binding protein can have a pro found impact on the solubility of fusion proteins. Protein Sci 2001; 10, 622–630.
16. Pryor KD, Leiting B. High-level expression of soluble protein in *Escherichia coli* using a His(6) - tag and maltose binding protein - double affinity fusion system. Protein Expr Purif 1997; 10, 309–319.
17. De Marco V, Stier G, Blandin S, DeMarco A. The solubility and stability of recombinant proteins are increased by their fusion to NusA. Biochem Biophys Res Commun 2004; 322, 766–771.
18. Dummler A, Lawrence AM, DeMarco A. Simplified screening for the detection of soluble fusion constructs expressed in *E. coli* using a modular set of vectors. Microb Cell Fact 2005; 4, 34.

19. Turner P, Holst O, Karlsson EN. Optimized expression of soluble cyclo malto dextrinase of thermophilic origin in *Escherichia coli* by using a soluble fusion-tag and by tuning of inducer concentration. Protein Expr Purif 2005; 39, 54–60.
20. Stewart EJ, Aslund F, Beckwith. Disulfide bond formation in the *Escherichia coli* cytoplasm: an in vivo role reversal for the thioredoxins. EMBO J 1998; 17, 5543–5550.
21. LaVallie ER, Lu ZJ, Diblasio-Smith EA, Collins-Racie LA, Mccoy JM. Thioredoxin as a fusion partner for production of soluble recombinant proteins in *Escherichia coli*. Appl Chimeric Genes Hybrid Proteins PtA 2000; 326, 322-340.
22. Young CL, Britton ZT, Robinson AS. Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. Biotechnol J 2012; 7, 620–634.
23. Nandini BN, Sunilkumar S, Ganesh S, Sudhakaran R. Truncated Thioredoxin peptides serves as an efficient fusion tag for production of Proinsulin. Protein and peptide letters 2020; 27:5, 419 - 431.
24. Silva E, Castro A, Lopes A, Rodrigues A, Dias C, Conceicao A, et al. A recombinant antigen recognized by Fasciola hepatica infected hosts. J Parasitol 2004; 90, 746–751.
25. Fraga H, Faria TQ, Pinto F, Almeida A, Brito RMM, Damas AM. FH8- a small EF-hand protein from *Fasciola hepatica*. FEBS J 2010; 277, 5072–5085.
26. Costa SJ, Almeida A, Castro A, Domingues, 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 2013a; 97, 6779–6791.
27. Costa SJ, Coelho E, Franco L, Almeida A, Castro A, Domingues L. The Fh8 tag:a fusion partner for simple and cost-effective protein purification in *Escherichia coli*. Protein Expr Purif 2013b; 92, 163–170.

28. Costa SJ, Silva P, Almeida A, Conceição M, Domingues L, Castro A. A novel adjuvant free H fusion system for the production of recombinant immunogens in Escherichia coli: its application to a 12 kDa antigen from *Cryptosporidium parvum*. Bioengineered 2013c; 4, 1–7.
29. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence dependent and mutational effects on the aggregation of peptides and proteins. Nat. Biotechnol 2004; 22, 1302–1306.
30. Derynck R, Roberts AB, Winkler ME, Chen EY, Goeddel DV. Human transforming growth factor-alpha: precursor structure and expression in *E. coli*. Cell 1984; 38, 287–297.
31. Cook GA, Stefer S, Opella SJ. Expression and purification of the membrane protein p7 from hepatitis C virus. Biopolymers 2011; 96, 32–40.
32. Kuliopulos A, Walsh CT. Production, purification, and cleavage of tandem repeats of recombinant peptides. J Am Chem Soc 1994; 116, 4599–4607.
33. Hwang PM, Pan JS, Sykes BD. A Pag P fusion protein system for the expression of intrinsically disordered proteins in *Escherichia coli*. Protein Expr Purif 2012; 85, 148–151.
34. Ramos R, Domingues L, Gama FM. Escherichia coli expression and purification of LL37 fused to a family III carbohydrate binding module from *Clostridium thermocellum*. Protein Expr Purif 2010;71, 1–7.
35. Ramos R, Moreira S, Rodrigues A, Gama M, Domingues L. Recombinant expression and purification of the antimicrobial peptide magainin-2. Biotechnol Prog 2013; 29, 17–22.
36. Kapust RB, Tozser J, Fox JD, Anderson DE, Cherry S, Copeland TD et.al. Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. Protein Eng 2011; 14, 993–1000.

37. Parks TD, Leuther KK, Howard ED, Johnston SA, Dougherty WG. Release of proteins and peptides from fusion proteins using a recombinant plant-virus proteinase. Anal Biochem 1994; 216, 413–417.

38. Choi SI, Song HW, Moon JW, Seong BL. Recombinant enterokinase light chain with affinity tag expression from *Saccharomyces cerevisiae* and its utilities in fusion protein technology. Biotechnol Bioeng 2001; 75, 718–724.

39. Jenny RJ, Mann KG, Lundblad RL. A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. Protein Expr Purif 2003; 31, 1-11.

40. Hirata R, Ohsumk Y, Nakano A, Kawasaki H, Suzuki K, Anraku Y. Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)- translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. J Biol Chem 1990; 265(12):6726-3

41. Kane PM, Yamashiro CT, Wolczyk DF, Neff N, Goebl M, Stevens TH. Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase; Science 1990 2;250(4981):651-7.

42. Chong S, Shao Y, Paulus H, Benner J, Perler FB, Xu M Q. Protein splicing involving the Saccharomyces cerevisiae VMA intein. The steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an in vitro splicing system. J Biol Chem 1996; 271(36):22159-68.

43. Ke Xu, David B Evans, Kei Kawabata, Riadh Zeramdini, Jan Klavus, Christopher J L Murray. Household catastrophic health expenditure: a multicountry analysis. Lancet 2003; 12;362(9378):111-7.

44. Wang ZY, Li N, Wang YY, Wu YP, Mu TY, Zheng Y, et al. Ubiquitin- intein and SUMO2 intein fusion systems for enhanced protein production and purification. Protein Expr Purif; 2012; 82, 174–178.

45. Xingmei Qi, Qian Lu, Jing Ping Hu, Sidong Xiong. Spontaneous C-cleavage of a truncated intein as fusion tag to produce tag-free VP1 inclusion body nanoparticle vaccine against CVB3-induced viral myocarditis by the oral route. Microbial Cell Factories 2019; 18:66.
46. Xu S, Huang HK, Kaiser P, Latterich M, Hunter T. Phosphorylation and spindle pole body localization of the Cdc15p mitotic regulatory protein kinase in budding yeast. *Curr Biol 2000;*10 (6):329-3
47. Wood DW. Simplified protein purification using engineered self-cleaving affinity tags. J Chem Technol Biotechnol 2003; 78:103–110.
48. Cui C, Zhao W, Chen J, Wang J, Li Q. Elimination of in vivo cleavage between target protein and intein in the intein - mediated protein purification systems. Protein Expr Purif 2006; 50:74–8.
49. Yifeng Li. Self-cleaving fusion tags for recombinant protein production. Biotechnol Lett (2011) 33:869–881.
50. Mao H. A self-cleavable sortase fusion for one-step purification of free recombinant proteins. Protein Expr Purif 2004; 37:253–263.
51. Stark R, Meyers G, Ru¨menapf T, Thiel HJ. Processing of pestivirus polyprotein: cleavage site between autoprotease and nucleocapsid protein of classical swine fever virus. J Virol 1993; 67:7088–7095.
52. Achmu¨ller C, Kaar W, Ahrer K, Wechner P, Hahn R, Werther F, Schmidinger H, Cserjan-Puschmann M, Clementschitsch F, Striedner G, Bayer K, Jungbauer A, Auer B. N(pro) fusion technology to produce proteins with authentic N termini in *E. coli*. Nat Methods 2007; 4:1037–1043
53. Schindler S, Missbichler B, Walther C, Sponring M, Cserjan-Puschmann M, Auer B, Schneider R, Dürauer A. Npro fusion technology: On-column complementation to improve

efficiency in biopharmaceutical production Protein Expression and Purification 2016; 120, 42-50.

54. Prochazkova K, Satchell KJ. Structure-function analysis of inositol hexakisphosphate-induced autoprocessing of the Vibrio cholerae multifunctional autoprocessing RTX toxin. J Biol Chem 2008; 283:23656–23664.
55. Thompson SA, Wang LL, Sparling PF. Cloning and nucleotide sequence of frpC, a second gene from Neisseria meningitidis encoding a protein similar to RTX cytotoxins. Mol Microbiol 1993; 9:85–96.
56. Osicka R, Procha´zkova´ K, Sulc M, Linhartova´ I, Havlı´cek V, Sebo P. A novel ''clip-and-link'' activity of repeat in toxin (RTX) proteins from gram-negative pathogens. Covalent protein cross-linking by an Asp–Lys iso peptide bond upon calcium-dependent processing at an Asp–Pro bond. J Biol Chem 2004; 279:24944–24956.