ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Gene Expression Systems and Recombinant Protein Purification.

Kishwar Hayat Khan*.

Medical Biotechnology Division, School of Biosciences and Technology, VIT University, Vellore-632014, Tamil Nadu, India.

ABSTRACT

There is no substance as important as deoxyribonucleic acid as it carries within its own structure the hereditary in formations that determines the structure of proteins, the important molecule of life. Deoxyribonucleic acid transcribe messenger ribonucleic acid which finally translates protein the essential macromolecules. Through gene expression the information contained in gene can be utilized for the synthesis of functional gene product which is often a protein or an enzyme required for the cellular function of live organisms. This article discusses about the various strategies required for the expression of gene to get gene product. It also highlighted few expression systems required to express gene. The author also explained the strategies regarding the purification of expressed proteins. In this context protein affinity tags were explored and highlighted. This review suumarized the role of His-tag, GST-tag, MBP-tag, IMPACT, TAP-Tagging towards protein purification. The aim of writing this review article is to attract those researchers working in the field of gene expression and protein purification to find new molecules for better therapy.

Keywords: Proteins, Vectors, Expression system, Recombinant protein purification, Affinity protein purification tags.

Corresponding author

ISSN: 0975-8585

INTRODUCTION

One of the most important scientific legacies that our generations will handover to the successors is the detailed and elaborated information about biological sequences and structure now being determined and achieved. Deoxyribonucleic acid (DNA), the genetic material, is a long unbranched polymer having the four deoxynucleoside monophosphate namely deoxyadenosine monophosphate (dAMP), deoxythymidine monophosphate (dTMP), deoxyguanosine monophosphate (dGMP) and deoxycytidine monophosphate (dCMP). Nucleic acid is a biological molecule essential for life as it encodes for protein. Proteins are the major functional molecules of life whose properties are so useful that we employ them as therapeutic agents, catalysts and materials. Many diseases stem from mutation in proteins that cause them to loose function. In some cases, catalytic activity may be impaired, and so metabolic pathways may be altered. The example is phenylketonaria. In other cases , structural properties may be impaired, leading to a loss of physical function as seen in the case of muscular dystrophy. The ultimate goal of protein science is to be able to predict the structure and activity of a protein *de novo* and how it will bind to ligand. When this is achieved, we will be able to design and synthesize novel catalyst, materials, and drugs that will eliminate disease and minimize ill health[1].

Genetic engineering has revolutionized protein science. First recombinant DNA technology has enabled the production of large quantities of proteins that were previously either unknown or valuable only in small quantities. Second it has allowed rapid sequencing of proteins from their DNA. Third it has allowed the facile modification of protein structure by mutation of their genes. This has lead to protein engineering. Protein engineering is of fundamental importance in analyzing structure function relationships in proteins and producing novel proteins for biomedical and biotechnological use.

A gene is defined as a discrete unit of genetic information that can be utilized for the synthesis of polypeptide. It includes the coding sequence, the promoter, terminator and introns[2]. In protein production there are two aspects which require optimization that is the biology of the system and the production process itself. In order to express a gene in an efficient manner, the gene has to be inserted into an appropriate vector containing a suitable promoter. This can then be transferred into a suitable host using the modern technology.

Now a days a variety of vectors as well as expression systems are available to express gene of interest. This article is totally based on literature survey. This article has been compiled by the author by using modern and traditional methods of literature survey. The search engines used in the survey were Science direct, PubMed, Google etc. In addition to this the author also took the help of books and journals for making this manuscript. The literatures cited in this article are not limited to a particular region but is considered more or less from the entire globe. Recent research has been focused by this author. Literatures for the past ten years were paid more attention. Moreover the valuable literatures were also efficiently highlighted in this manuscript.

In his previous articles the author highlighted a number of vectors that can be used in gene manipulation[3]. In addition to this the author in his another article has described a number of gene transfer technologies that can manipulate gene in plants[4]. Prior to this work the author of this article has reported about the technologies related to transgenic animals[5]. He also described DNA vaccine [6][7]. In this article the author summarized the systems used in expressing gene. He also explored and presented the technologies that can be used to purify proteins..The author described the expression of genes in various systems like bacteria, yeast, insect and mammals. Much effort was given to the proteins affinity tag that makes the purification step easy and improves the purity and stability of recombinant protein. The aim of writing this article is to help the new researchers who are approaching the technology of gene expression and protein purification to prepare the products having high therapeutic values for living beings. A number of systems have been described below that can be used for the expression of gene. The main objective of this review article is to make awareness in the young researchers about the systems used to express gene and to purify the expressed protein.

Systems Utilized For Gene Expression

Expression is defined as transcription and translation of gene. The primary goal of gene cloning for biotechnological application is the expression of cloned gene in selected host organisms. For commercial

purposes, a high rate of production of the protein encoded by cloned gene is required. In order to achieve this, many specialized expression vectors have been constructed that provide genetic element for controlling transcription, translation, protein stability and secretion of the product of the cloned gene from the host cell. The molecular biological features that have been manipulated to modulate gene expression include the promoter and terminator sequences, the strength of ribosome binding site, the number of copies of cloned gene etc. The level of foreign gene expression also depends on the host organisms. Currently both prokaryotes and eukaryotes expression systems are available for gene expression.

To elevate protein expression it is important that an inducible expression system must be established so to enable the growth of large number of host cells before initiating the expression of a target protein[8]. The researchers made a report of gene expression in plant by *Agrobacterium tumefaciens* using green fluorescent reporter gene[9]. The author in this article presented a detailed study of a number of systems like bacteria, insect, yeast and mammals to express gene to obtain its product.

Bacteria as an expression system

Because of high knowledge about the genetics, biochemistry amd molecular biology, *E.coli* is the system of first choice for expression of genes to get required heterologous proteins. This is because the genetic manipulations are straight forward and easy. *E coli* is easy to culture and the growth is also inexpensive. Morever many foreign proteins are well tolerated and may be expressed at high level. Small cytosolic proteins and polypeptide less than hundred nucleotide in length are best expressed in *E.coli* as fusion proteins composed of carrier sequences linked by a standard peptide bond to the target protein. If only small quantities of the target proteins are required as in the case of screening a series of site directed mutants for enzymatic activity. There is little point in trying to optimize production. Most of the standard expression plasmids can be used successfully if the enzyme can be assayed in crude extract of *E.coli*. If the target protein is to be used in biochemical or cell biological studies, then maintaining or restoring protein function is important and ease of purification matters less. In some cases, direct expression vectors may be used to produce soluble, active proteins. In most cases, however, the expressed protein will be insoluble and must be purified from the inclusion bodies, solubilized, and refolded into an active form[10].

The uptodate research and development in the fundamental understanding of phenomenon of transcription, translation, and protein folding in *E. coli* and the indirect discoveries and the availability of modern genetic tools are further making *E.coli* more important than ever for the expression of complex eukaryotic proteins[11]. The advantages of using *E. coli* as an expression system is that it produce large amount of protein. Moreover the growth of this bacterium is very fast as compared to mammalian cells, giving the opportunity to purify, analyze and use the expressed protein in a much shorter time period. In addition to this minimal amount of DNA is required to transform *E coli* cells and transformation experiment is also easy.

The *E. coli* T7 system is regarded as the most widely used system for high-level gene expression. In contrast to other *E. coli* expression systems using host RNA polymerases for heterologous gene expression, an appropriate T7 system yields higher protein amounts since the bacteriophage RNA polymerase exhibits enhanced processivity. For this reason, the *E. coli* T7 expression system has been recently recommended by leading structural genomics consortia as a 'what to try first' system for the expression of soluble, globular and stably folded pro- and eukaryotic proteins[12].

Recently the researchers reported the expression and purification of chimeric peptide comprising EGFR B-cell epitope and measles virus fusion protein T-cell epitope in *E. coli* [13]. Recombinant expression, purification and characterization of the native glutamate racemase from *Lactobacillus plantarum* NC8 was also described[14]. High-level expression, purification, and characterization of *Staphylococcus aureus dihydroorotase* (PyrC) as a cleavable His-SUMO fusion was explained[15]. High expression of HPV16L2N120E7E6 fusion protein in *E. coli* and its inhibitory effect on tumor growth in mice was also studied by the researchers[16]. Moreover the researchers also reported on the production of pentameric cholera toxin B subunit in *E.coli*[17]. The researchers studied the over-production of soluble protein complex and validating protein-protein interaction through a new bacterial co-expression system[18]. A novel T7 RNA polymerase-dependent expression system was described for high-level protein production in *Rhodobacter capsulatus* [19].

Yeast as an expression systems

Yeast belongs to both microorganism and eukaryote and so have the advantages over the other expression system. The main reason for selecting yeast as an expression system is that unlike *E. coli*, yeast provides advanced protein folding pathways for heterologous proteins. In addition to this when yeast signal sequences are utilized then yeast can easily secrete correctly folded and processed proteins. The proteins can be released into the culture media. Moreover simple growth media is the requirement for the growth and multiplication of yeast.

The production of protein through recombinant DNA route utilizes *Saccharomyces cerevisiae* as the most favoured microbial eukaryotes. A number of other yeast like *Hansela polymorpha*, *Schizosaccharomyces pombe*, *Kluveromyces lactis. Pichia pastoris* and *Yarrouvia lipolytica* are also used for the same purpose. Unlike bacteria the proteins expressed in yeast are subjected to post translational modification. In addition to this, there is usually a higher degree of authenticity with respect to three dimensional confirmation and the immunogenic properties of the protein. Thus, in a situation where the biological properties of the protein are critical, yeast may provide a better product than prokaryotic hosts. A variety of strong constitutive promoters have been utilized to carry on target gene expression in yeast. The examples in this context are the promoters for the genes encoding phosphoglycerate kinase (PGK), glyceraldehyde-3-phosphate dehydrogenase (GPD) and alcohol dehydrogenase (ADH1). All of these can be used to produce target proteins[20].

Pichia pastoris is the most frequently used yeast system for heterologous protein production today[21]. A yeast-based rapid prototype platform for gene control elements in mammalian cells was reported[22]. Fibronectin C-terminal heparin-binding domain could be expressed and purified successfully in *Pichia pastoris*[23]. Systematic analysis of a dipeptide library for inhibitor development using human dipeptidyl peptidase IV produced by a *Saccharomyces cerevisiae* expression system were performed[24]. Purified recombinant Japanese encephalitis virus (JEV) envelope protein expressed in the *P. pastoris* expression system holds great promise for use in the development of a subunit vaccine against JEV[25].

Expression of gene in insect cells

Baculovirus infect insects and does not appear to infect mammalian cells. Therefore any system based on such viruses has the immediate attraction of low risk of human infection because of their colligative nature. Baculovirus expression systems are considered important as they posses the ability to produce large amount of proteins

Baculoviruses belong to a large group of circular double stranded DNA viruses. This virus infect only invertebrates, usually insects[26]. The virus has its genome which is 90-180 kbp[27]. The cell lysis takes place after three to five days of initial infection. The nuclear polyhedrosis viruses produce occlusion bodies in the nucleus of infected cells. These occlusion bodies consist primarily of protein. High level of polyhedron gene is transcribed in the late transfection process[2]. To express the target gene the polyhedron promoter can be easily utilized. The *Autographa californica* nuclear polyhedrosis virus (AcNPV) which is an example of baculovirus has become a famous tool to make recombinant protein particularly in insect cells[28] . The advantage of protein production in baculovirus infected insect cells is that it produces very high levels of protein relative to other eukaryotic expression systems[29][30]. In this system multiple genes can be expressed from a single virus. The disadavantge is that the cells grow slowly and the media is inexpensive. Moreover the construction and purification of recombinant baculovirus vectors for the expression of target genes in insect cells can take four to six weeks which is a long time. In addition to this chance of contamination of culture is also there. The main demerit of this system is that the expression of the target protein is controlled by a very late viral promoter and peaks when the cells are dying due to the infection from virus[31].

Insect cells have been reported as a production platform of complex virus-like particles[32]. High-level expression of neutrophil gelatinase-associated lipocalin lipocalin2 by baculovirus expression was reported[33]. The researchers engineered the baculovirus genome in order to produce galactosylated antibodies in lepidopteran cells[34]. Insect cell-based expression and characterization of a single-chain variable antibody fragment directed against blood coagulation factor VIII were also studied[35]. Moreover expression of recombinant human IFNa-2b/IgG4 Fc fusion protein in a baculovirus insect cell system were also explained[36].

ISSN: 0975-8585

Mammalian expression system

Mammalian cell expression has become the dominant recombinant protein production system for clinical applications because of its capacity for post-translational modification and human protein-like molecular structure assembly[37]. In mammalian expression systems the protein, or complex, can be expressed in its native cell type, under physiological conditions, with numerous molecular systems working together for efficient production and quality control[38].

As compared to bacteria and other microbes, mammalian expression system is a better system as far as recombinant human protein is concerned. Growth media are more expensive and complex. Mammalian cells are generally less robust as compared to microbes when large scale fermentation is involved. Difficulties are also there for processing of the products. Despite these difficulties, a variety of vectors are now present to express protein in mammalian cells. Vectors possess selectable markers and also utilizes promoters that enable expression of the cloned gene sequence[2]. The commonly used promoters are based on cytomegalovirus (CMV) or simian virus (SV40). The proteins obtained in mammalian system have the best structural and functional features that are usually most close to their cognate native form. It can satisfy the applications, needs or utility. The author did the extensive literature survey and finally summarized the proteins in table 1.

Viral- based vectors for gene expression in mammalian cells are listed in the table 2.
Transfection in mammalian cell has been illustrated in figure 1 and 2.

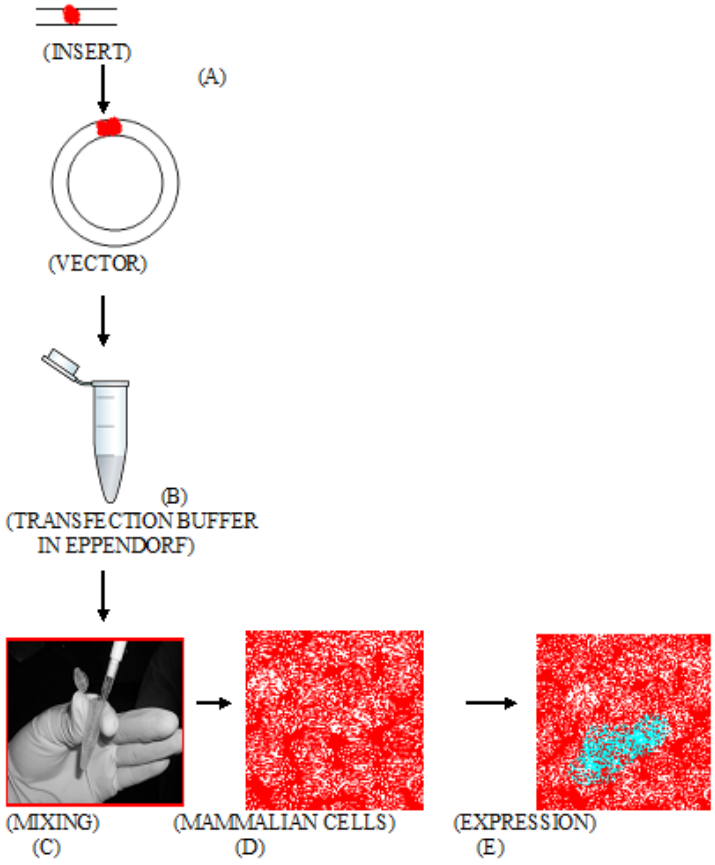

Figure 1: Diagramatic representation of transfection by calcium chloride co precipitation method: (A) DNA of interest is inserted into plasmid. (B) Vector containing the insert is introduced into eppendorf containing the transfection buffer. (C) Calcium chloride is released slowly to form fine precipitate. (D) Fine precipitate is then allowed to enter into mammalian cells (E) Expression of protein.

ISSN: 0975-8585

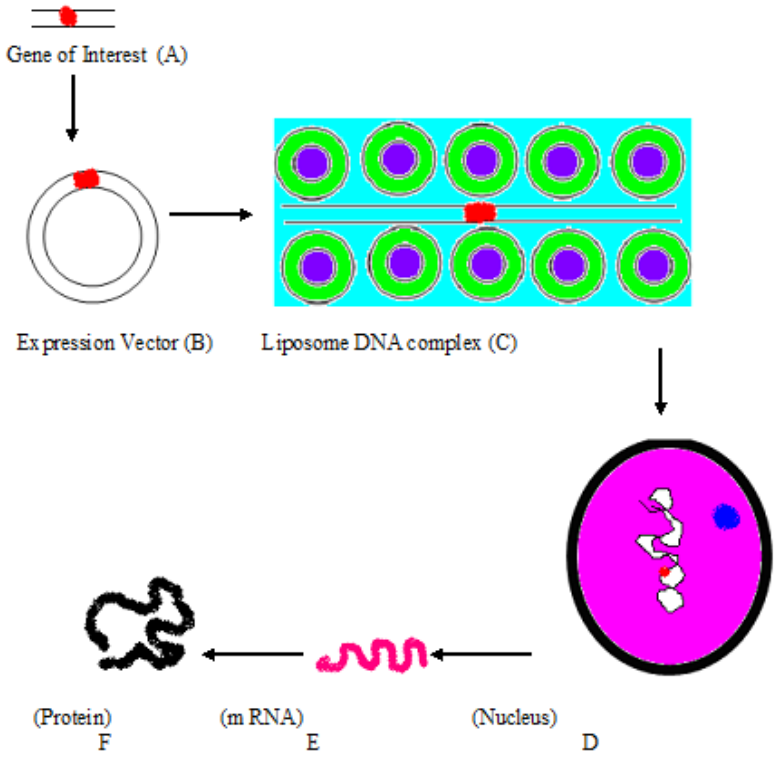

Figure 2: Diagrammatic representation introduction of DNA into cells using cationic liposomes: A) Gene of required interest B) Gene is ligated into an appropriate vector C) The vector DNA containing the insert gets surrounded by cationic liposomes forming liposome DNA complex. D) Cationic liposomes are able to interact with the negatively charged cell membrane resulting in the delivery of DNA across the membrane. The DNA escape endosome mediated degradation and gets integrated into the DNA present in the nucleus. E) mRNA gets transcribed F) Protein is then formed.

Table 1: Proteins produced by mammalian expression systems and their importance

S.No	Application, needs and utility of proteins produced by mammalian expression system.	Protein	References
1.	Therapeutic application	Factor VII, Indoleamine 2,3-dioxygenase (IDO). Tissue plasminogen activator Human erythropoietin Human GH Human a1-antitrypsin	[39] [40] [41] [42] [43] [44][45]
2.	Diagnostic application	Monoclonal antibody	[46]
3.	Pathology studies	Recombinant pigment epithelium-derived factor	[47]
4.	Cell line development	Mcl-1	[48]
5.	Drug target discovery	Human IgMs	[49]
6.	Transgene expression	ceramide transfer protein (CERT)	[50]
7.	Prophylactic (vaccine) development	HeLa based cell substrate	[51]
8.	Immunogen for antibodies development	HBsAg antigen	[52]
9..	Protein-protein interaction. experiments	Glutathione-S-transferase	[53]
10.	Protein engineering	Recombinant antibodies B72.3 Fv	[54] [55]
11.	Mutagenesis studies	Proteins with sugars of the form $Man_5GlcNAc_2$	[56]

Table2: Vectors based on virus used for gene transfer in mammalian cell lines

S.No	Vectors	S.No	Vectors
1.	Retroviruses	2.	Poliovirus
3.	Epstein-barr virus	4.	Adenovirus
5.	Herpes simplex virus	6.	Simian virus 40
7.	Sindbis virus	8.	Baculovirus
9.	Lentiviruses	10.	Semliki forest virus
11.	Adeno-associated virus	12.	Vaccinia virus

Enhanced expression of secretable influenza virus neuraminidase in suspension mammalian cells was studied[57]. Rapid, simple and high yield production of recombinant proteins in mammalian cells using a versatile episomal system was reported[58]. The researchers reported about the different matrix attachment regions flanking a transgene, effectively enhance gene expression in stably transfected Chinese hamster ovary cells[59]. High level expression, purification and activation of human dipeptidyl peptidase I from mammalian cells was also reported[60]. High levels of human recombinant cyclooxygenase-1 expression in mammalian cells using a novel gene amplification method was also studied[61].

Protein Purification

The development of techniques and methods for the separation and purification of biological macromolecules such as proteins has been an important prerequisite for many of the advancement made in bioscience and biotechnology over the past five decades. Improvement in materials, utilization of computerized instruments, and an increase use of *invivo* tagging have made proteins separations more predictable and controllable. In recent years the success of genomics has provided the impetus to explore and understand the molecular events that happens within the cell. The genome provides the information to manufacture a protein, but it provides non of the information on the activity or function of that protein once it has been synthesized in the cell.

The main aim of protein analysis and purification is to elucidate its structure and function. The purification of protein is a multistep process. It utilizes biochemical and biophysical characteristics of the target protein, such as its source, relative concentration, solubility, charge and hydrophobicity. The ideal purification strives to obtain the maximum recovery of the desired protein, with minimal loss of activity, combined with the maximum removal of other non-target proteins. The methods applied for protein purification must be mild to protect and preserve the native conformation of the molecule and its bioactivity. One should aim the following as listed in the Table 3 while designing the purification protocol for protein.

Table 3: Factors to be considered during protein purification

S,No	Key considerations during protein purification
1.	Purification step should be simple. The number of steps involved should be kept as minimium as possible.
2.	Process should be cheap. Expensive technique shouls be avoid or replace with cheap techniques. Reagent used must be cheap.
3.	Recovery must be high
4.	Assays are developed to monitor the purification.
5.	End product must be highly purified.
6.	Process must be convenient with regard to time.
7.	Reproducibility within the lab, in other labs and also when either scaled up or down.
8.	A notice will be kept on yields and activity throughout
9.	Reliable techniques and apparatus must be used
10.	Delay and slow equipment must be avoided

There is an increasing requirement for biological scientists of all disciplines (at all levels) to be able to device protocols to enrich a low abundance protein for subsequent analysis. The two important parameters utilized to develop most purification protocols are the physical properties and the chemical structure of the protein. Charge density, isoelectric points (PI), PH stability are some of the properties of proteins that can be

exploited during purification. A number of techniques are now available that are capable of resolving protein on the basis of differences in net charge. These include gel electrophoresis and ion exchange chromatography.

Various affinity procedure to purify protein is listed in the table 4.

Table 4: Affinity procedure for purifying protein

S.No	Affinity procedure
1.	Affinity chromatography
2.	Covalent chromatography
3.	Dye legand affinity chromatography
4.	Immobilized metal (ion) affinity chromatography (IMAC)
5.	Immunoaffinity chromatography
6.	Lectin affinity chromatography
7.	IMAC for purifying recombinant protein
8.	Affinity partitioning (Precipitation)

Protein stability is a crucial issue in biotechnology. The most potentially useful protein is of little value if cannot be produced, isolated, and stably stored. It is often difficult to purify recombinant proteins. The purification will require multiple time-consuming chromatographic steps to get an acceptable level of purity. Now a day to make protein purification easy the researchers were taking the help of protein tags. These tags are protein sequences. These sequences exhibit high affinity binding properties for particular molecules. The main function of the tag is that it allows the target protein to bind to a solid support, usually in the form of a column matrix, to which very few (if any) other proteins are able to bind. A number of steps are envolved for the purification of tagged proteins from host cells. Initially the host cell is allowed to lyse. Then the tagged proteins are allowed to bind to an affinity column. Untagged proteins are then removed by washing. Then the tagged protein was eluted out. A number of tag proteins are now available to make the purification of recombinant protein a convenient procedure.

The Glutathione S-Transferase-tag

This is also an important tag used in protein purification. Glutathione comprises of glutamic acid, glycine and cysteine. The enzyme Glutathione-S-transferace (GST) binds to glutathione with greater affinity. The gene responsible for encoding this protein is fused in the correct reading frame, to the target gene and a fusion protein is produced from an expression vector. The host cells making the fusion protein are then observed. The cells producing the protein are lysed. The soluble proteins are then applied to a column containing glutathione (eg glutathione-agarose). The specific interaction between GST and glutathione take place. It helps in the binding of the fusion protein to the column. Most of the proteins remain unbound to the column. The bonded protein can then be eluted by washing with a high concentration of glutathione to compete for the interaction with the column. The GST portion then can be removed from the fusion protein[2].

Table 5: Recognition and cleavage site of protease

S.No	Protease	Recognition and cleavage site	References
1.	PreScission	LeuGluValLeuPheGln↓GlyPro	[62]
2.	Factor Xa	IleGluGlyArg↓	[63]
3.	TEV	GluAsnLeuTyrPheGln↓Gly	[64]
4.	Enterokinase	AspAspAspAspLys↓	[65]
5.	Thrombin	LeuValProArg↓GlySer	[66]

To make this goal possible, an appropriate expression vector is first selected. DNA segment that has the capability of coding the amino acid sequence of a specific protease cleavage site is to be introduced between the GST and the target gene placed in an expression vector. The protein is then expressed. The expressed purified protein is treated with protease. This causes the formation of two polypeptides. One peptide will be free target protein and other will be GST itself. To separate the target protein from GST, it is again applied to the glutathione column. The GST will again bind to the column, but the target protein will not. The column flow through can be eluted and will contain the purified target protein. Several specific protease

have been used to cleave target fusion protein obtained by using purification tags. A number of these proteases with recognition and cleavage site have been listed in the Table 5.

Reports were made regarding the *E.coli* expression system and baculovirus-insect cell expression system that were used to produce the kinase, followed by purification using His-tag[67]. Researchers also described the method to purify soluble LubX protein using GST-tag and *E. coli* overexpression systems[68]. Studies were also made regarding Human PNAS-4 (hPNAS-4) which is a novel pro-apoptotic protein in mammalian cells. The hPNAS-4 gene was first cloned into the pGEX-6p-1 vector with GST tag. The recombinant hPNAS-4 was then purified[69].

The Maltose Binding Protein tag

The bacteria *E.coli* posses a gene known as *malE* gene. This gene encodes maltose binding protein (MBP). The expression vector uses this gene for protein purification purpose. In an expression vector the target gene is inserted downstream from the *malE* gene of *E. coli*. This results in the production of an MBP fusion protein[70]. One step purification of fusion protein is done using the affinity of MBP for cross linked amylase[71]. The target protein get bound to the column of amylase and eluted with maltose. The MBP-target fusion is cleaved with a protease. It is again applied to the amylase column. The protein of interest will not get attached to the column and thus get separated from MBP.

The MBP tag was used in concerned with expression, purification and characterization of non-specific Serratia nuclease in *E. coli*[72]. Expression of proteins in *E. coli* as fusions with maltose-binding protein to rescue non-expressed targets in a high-throughput protein-expression and purification was also reported[73]. Elevated solubility of integrin beta A domain by using maltose-binding protein as a fusion tag was described[74]. Mitochondrial fraction of apoptotic cells contains membrane protein called as p18Bax. Increased expression and purification of p18 form of Bax as an MBP-fusion protein was also reported[75].

The Histidine-tag

Affinity tags are highly efficient tools that can be used for protein purification. These tags permit the purification of virtually any type of protein without having any prior knowledge of its biochemical properties. These affinity tags have wide applications in several areas of research. The simplest among all the protein tag used for protein purification is the Histidine tag (His tag). This tag comprises of six histidine residues. For every protein it is the DNA that codes for it. The DNA for the histidine residue is cloned into the target gene. The cloning is performed in such a way that the produced protein contains at some point in its polypeptide sequence, six consecutive histidine residues[76]. During the time of cloning the His tag is placed either at the extreme amino or extreme carboxyl-terminal end of the protein so that it is less likely to impair the function of protein. If the central region of the protein is already recognized to be non-essential, then the tag can be placed in the middle of the protein[77]. Certain metal ions are available with which histidine can bind non-covalently and with high affinity. Nickel is the metal ions that are bound to a resin matrix and are used to capture protein containing his tag[78]. To achieve the above aim the most commonly used resins have nitrocellulose acid (NTA) covalently attached to them. The NTA has four coordination sites that bind very tightly a single nickel ion. At least six histidine residues are needed to provide the necessary binding affinity to firmly adhere the tagged protein to the column. The other protein will not bind to the column and were eluted out leaving behind the tagged protein attached to the column. The tagged proteins were finally eluted by changing the concentration of buffer used in elution process[2].

The recombinant L -arabinose isomerase was purified to homogeneity by one-step His-tag affinity chromatography was described[79]. Expression and purification of human PYY(3-36) in *E. coli* using a His-tagged small ubiquitin-like modifier fusion was also reported[80]. His-tag truncated butyrylcholinesterase as a useful construct for *in vitro* characterization of wild-type and variant butyrylcholinesterases was described[81]. The researchers also reported about histidine-tag-directed chromophores for tracer analyses in the analytical ultracentrifuge[82]. His-tag was also used to study the expression and purification of recombinant human coagulation factor VII[39].

Tandem-affinity purification tag

The reason behind the popularity of affinity tags is that considerable purification can be performed in just a single chromatography step. To achieve the desired purity a single step is usually not sufficient. Dual affinity tags have been developed in the recent years in which two different affinity tags are expressed in tandem. This method was originally developed in yeast and is called tandem affinity purification (TAP) [83]. TAP tagging systems, developed by the research group of Bertrand Seraphin and others, are a means of isolating physiologically relevant protein and protein-nucleic acid complexes[84]. This tag facilitates fast purification of complexes from a relatively small number of cells. The purification takes place without prior information of the complex composition, activity or function[85][86]. The purification procedure consists of two steps. It is highly specific and can isolate contaminant-free protein complexes. At the 3^{1}-end of a target gene, the DNA Tap -tag is cloned so that little disruption is made to its ability to be transcribed, and the fusion protein should be produced at the same level as the wild-type target protein. The Tap-tag encodes a calmodulin binding peptide and protein A from *Staphylococcus aureus*. These two are separated from each other by a TEV protease cleavage site[87]. Tagged protein containing cells were lysed. It is then applied to a column containing IgG, which binds with greater affinity to protein A. The fusion protein and its associated proteins are removed from the column using TEV protease and then applied directly to a calmodulin bead column, in the presence of Ca^{2+} and eluted using chelating agent like EDTA.

The researchers reported about *in vivo* investigation of protein-protein interactions for helicases using TAP[88]. The TAP method is an efficient system for protein complex purification and protein interaction identification[89]. Moreover an enhanced strategy for TAP-tagging of *Schizosaccharomyces pombe* genes was sttudied[90]. Reports were also made regarding the targeted TAP of PSD-95 recovers core postsynaptic complexes and schizophrenia susceptibility proteins[91]. The studies were also made in concwerned with a modified version of TAP tagging to identify proteins interacting with HIV-1 Rev in human[92]. The author listed various in the Table 6.

Table 6: Tag used for purification purpose.

S.No	Tag	S.No	Tag
1.	Avi Tag	2.	V5tag
3.	Xpress tag	4.	Isopep tag
5.	HA-tag	6.	Nus-tag
7.	Spytag	8.	Thioredoxin-tag
9.	S-tag	10.	TC tag
11.	Ty tag	12.	Strep-tag
13.	SBP-tag	14	Green fluorescent protein-tag
15	Myc-tag	16	Calmodulin-tag
17	Softag 3	18	Softag 1
19	FLAG-tag	20	BCCP tag

Intein-mediated purification with an affinity chitin binding tag

The use of protein fusion and affinity technology has simplified the purification of recombinant proteins. Intein-mediated purification with an affinity chitin binding tag are also used by the researchers for protein purification. It is also known as IMPACT. Inteins are proteins. They are present in a number of organisms, which excise themselves from a precursor protein and in the process, ligate the flanking protein sequences[93]. IMPACT uses the protein self splicing of inteins to remove the purification tag and give pure isolated protein in single chromatographic steps. Most inteins have asparagine at their carboxyl terminal and a cystine residue at their amino terminal end. All the information required for the splicing reaction is stored within the intein itself, and if these sequences are placed in the context of target protein they still splice themselves out. An expression vector is taken to which the target gene is cloned such that a three component fusion protein is obtained, in which a target protein —intein—chitin binding domain fusion protein is produced.

E.coli is the microorganism that is used nn IMPACT system and fusion protein is made in the said microrganism. For protein purification a chitin column is made. The fusion protein is passed through this column. The protein will bind the chitin column. Dithiothreitol (DTT) can be used to cleave off the protein

from the column at 4 ^{o}C. The process is slow and so it required overnight incubation to complete. This is problematic if the target protein is not stable under these conditions. The target protein produced by this method is native except for DTT thioester moiety attached at the carboxyl terminal end. This thioester is unstable. It will hydrolyse to yield native protein.

The researchers reported Intein-mediated expression, purification, and characterization of thymosin α1-thymopentin fusion peptide in *E. coli*[94]. Intein-mediated one-step purification of *E. coli* secreted human antibody fragments was also studied[95]. By using the intein-mediated purification with a affinity chitin binding tag system, the peptide thioester M-[A(49)]-SDF-1(1-49)-MESNA was expressed[96]. Using IMPACT system, the thioester of enhanced green fluorescent protein was prepared[97].

CONCLUSION

The main conclusion is that the author made a systemic study on various expression system. The reader can know through this article about the gene to be expressed. More over a concise study has been represented regarding protein purification. A number of protein purification tags were also described by the author which can make the purification step easy. It is the requirement of the present time to discover new expression system that can express and yield protein without damaging its biological properties. The system should be designed in such a way so that it should be less time consuming and also require less effort. The entire system should be cheap and must have high productivity to meet the required demand.

ACKNOWLEDGEMENTS

Dr. K.H.Khan thanks his mother Zarina Khanam at Ward N0 33, Holding N0 135, Sahamat Hussain Road, IslamNagar, Bhikhanpur-Barahpura, Gumti No-3; Bhagalpur-812001, Bihar, India.

REFERENCES

[1] Fersht A. Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. W.H. Freeman and company, New York. Printed in the United States of America. ISBN: 0-7167-3268-8. 1999, pp1-631.
[2] Reece RJ. Analysis of Genes and Genomes. John Wiley and Sons Ltd. 2004, pp 1-400.
[3] Khan KH. Advanced Biotech Journal online., Tutorial review. 2009; 1-8. www.advancedbiotech.in/Vectors%20webonline.pdf
[4] Khan KH. Recent Research in Science and Technologies 2009; 1 (3): 116-123.
[5] Khan KH. Journal of Ecobiotechnology 2009; 1 (1): 32-40.
[6] Khan KH. GERMS 2013;3(1):26-35.
[7] Maurya S, Priya T, Khan KH. International Journal of Science Innovations and Discoveries 2013; 3 (1): 34-48.
[8] Glick BR, Pasternak JJ. 2007. Molecular Biotechnology: Principles and applications of recombinant DNA. 3rd Edition. ASM Press, Washington DC. 2007, pp 121-162.
[9] Subramaniam S, Rahman ZA. Emir. J. Food Agric 2010; 22 (2): 103-116.
[10] Sambrook J, Russell DW. Molecular cloning a laboratory manual. Volume 3. rd Internation Edition. Cold Spring harbor Laboratory press. Cold spring harbor. New York. 2001, pp 15.2.
[11] Baneyx F. Curr Opin Biotechnol 1999; 10 (5): 411-421.
[12] Katzke N, Arvani S, Bergmann R, Circolone F, Markert A, Svensson V, Jaeger KE, Heck A, Drepper T. Protein Expression and Purification 2010; 69 (2): 137-146.
[13] Wu M, Zhao L, Zhu L, Chen Z, Li H. Protein Expression and Purification 2013; 88 (1): 2013, 7-12.
[14] Böhmer N, Dautel A, Eisele T, Fischer L. Protein Expression and Purification 2013; 88 (1): 54-60.
[15] Truong L, Hevener KE, Rice AJ, Patel K, Johnson ME, Lee H. Protein Expression and Purification 2013; 88 (1): 98-106.
[16] Zhao L, Gao M, Gao J, Ren J, Zhang H, Tian HW, Tan WJ, Ruan L. Zhonghua Zhong Liu Za Zhi 2012; 34(11): 810-815.
[17] Dakterzada F, Mobarez AM, Roudkenar MH, Forouzandeh M. Avicenna J Med Biotechnol 2012; 4(2): 89-94.
[18] [18] Zeng J, Zhang L, Li Y, Wang Yi, Wang M, Duan X, He ZG. Protein Expression and Purification 2010; 69 (1): 47-53.

[19] Katzke N, Arvani S, Bergmann R, Circolone F, Markert A, Svensson V, Jaeger KE, Heck A, Drepper T. Protein Expression and Purification 2010; 69 (2): 137-146.
[20] Cereghino GP, Cregg JM. Curr. Opin. Biotechnol 1999; 10 (5): 422-427.
[21] Gasser B, Prielhofer R, Marx H, Maurer M, Nocon J, Steiger M, Puxbaum V, Sauer M, Mattanovich D. Future Microbiol 2013; 8 (2):191-208.
[22] Wei KY, Chen YY, Smolke CD. Biotechnol Bioeng 2013; 110 (4):1201-1210.
[23] Chen X, Chen X, Zou Q, Wu Y, Chen Y. Sheng Wu Gong Cheng Xue Bao 2012; 28 (10):1265-1273.
[24] Hikida A, Ito K, Motoyama T, Kato R, Kawarasaki Y. Biochem Biophys Res Commun 2013; 430 (4):1217-1222.
[25] Kwon WT, Lee WS, Park PJ, Park TK, Kang H. J Microbiol Biotechnol 2012; 22(11):1580-1587.
[26] Granados RR, Federici BA. The biology of baculoviruses. CRC Press, Boca Raton, Florida.1986, pp1-200.
[27] Ayres MD, Howard SC, Kuzio J, Ferber M L, and Possee RD. Virology 1994; 202 (2): 586-605.
[28] Fraser MJ. Curr. Top. Microbiol. Immunol 1992; 158: 131-172.
[29] Possee RD. Curr. Opin. Biotechnol 1997; 8 (5): 569-572.
[30] Joshi L, Davis TR, Mattu TS, Rudd PM, Dwek RA, Shulter ML, Wood HA. Biotechnol. Prog 2000; 16 (4): 650-656.
[31] Verma R, Boleti E, George AJT. Journal of Immunological Methods 1998; 216 (1-2): 165-181.
[32] Fernandes F, Teixeira AP, Carinhas N, Carrondo MJ, Alves PM. Expert Rev Vaccines 2013; 12 (2):225-236.
[33] Rouhbakhsh M, Halabian R, Masroori N, Mohammadi Pour M, Bahmani P, Mohammadi Roush A, Jahanian-Najafabadi A, Habibi Roudkenar M. Iran J Basic Med Sci 2012; 15 (3):845-52.
[34] Juliant S, Lévêque M, Cérutti P, Ozil A, Choblet S, Violet ML, Slomianny MC, Harduin-Lepers A, Cérutti M. Methods Mol Biol 2013; 988:59-77.
[35] Kurasawa JH, Shestopal SA, Jha NK, Ovanesov MV, Lee TK, Sarafanov AG. Protein Expr Purif 2013; 88 (2):201-206.
[36] Ji ZX, Chen YN, Zhang YR, Yang YX, Wang CR, Han SY. Zhonghua Gan Zang Bing Za Zhi 2012; 20(8):617-20.
[37] Jianwei Zhu. Biotechnology Advances 2012; (30) 5: 1158-1170.
[38] Zhao Y, Bishop B, Clay JE, Lu W, Jones M, Daenke S, Siebold C, Stuart DL, Jones EY, Aricescu AR. Journal of Structural Biology 2011; 175 (2): 209-215.
[39] Halabian R, Fathabad ME, Masroori N, Roushandeh AM, Saki S, Amirizadeh N, Najafabadi AJ, Gharehbaghian A, Roudkenar MH. Blood Transfus 2009; 7 (4): 305-312.
[40] Li R, Wei F, Yu J, Li H, Ren X, Hao X. Cancer Biol Ther 2009; 8 (14): 1402-1408.
[41] Watson JD, Gilman H, Witkowski J, Zoller M. Recombinant DNA, 2nd ed., Scientific American Books, Distributed by Freeman and company, 41 Medison Avenue, New York.1992 pp1-400.
[42] Rodriguez A, Castro FO, Aguilar A, Ramos B, Del Barco DG, Lleonart R, Fuente J. De la. Biol Res 1995; 28: 141- 153.
[43] Hammer RE, Pursel VG, Rexroad JCE, Wall RJ, Bolt DJ, Ebert KM, Palmiter RD, Brinster RL. Nature 1985; 315: 680- 683.
[44] Massoud M, Bischoff R, Dalemans W, Pointu H, Attal J, Schultz H, Clesse D, Stinnakre MG, Pavirani A, Houdebine LM. C R Acad Sci Ser III Sci Vie 1990; 311: 275-280.
[45] Massoud M, Bischoff R, Dalemans W, Pointu H, Attal J, Schultz H, Clesse D, Stinnakre M G, Pavirani A, Houdebine LM. J Biotechnol 1991; 18: 193-203.
[46] Rita CA, Elisa RM, Henriques M, Azeredo J, Oliveira R. Eur J Pharm Biopharm 2010; 74 (2):127-138.
[47] Chen J, Ye L, Zhang L, Jiang WG. Int. J. Oncol 2009; 35 (1): 159-166.
[48] Majors BS, Betenbaugh MJ, Pederson NE, Chiang GG. Biotechnol Prog 2009; 25 (4):1161-1168.
[49] Tchoudakova A, Hensel F, Murillo A, Eng B, Foley M, Smith L, Schoenen F, Hildebrand A, Kelter AR, Ilag LL, Vollmers HP, Brandlein S, McIninch J, Chon J, Lee G, Cacciuttolo M, MAbs 2009; 1(2):163-171.
[50] Florin L, Pegel A, Becker E, Hausser A, Olayioye MA, Kaufmann H. J Biotechnol 2009; 141 (1-2): 84-90.
[51] Tatalick LM, Gerard CJ, Takeya R, Price DN, Thorne BA, Wyatt LM, Anklesaria P. Vaccine 2005; 23 (20): 2628-2638.
[52] Luo W, Sun X, Yi X, Zhang Y. J Biosci Bioeng 2005; 100 (4):475-477.
[53] Einarson MB, Orlinic JR. Identification of protein-protein interaction with Glutathione-S-transferase fusion protein. In: Erica Golemis. (Eds). Protein-protein interactions: A molecular cloning manual. Cold Spring Harbor laboratory Press. Cold Spring harbor, New York. Printed at USA. 2002, pp 37-43.

[54] Birch J R, Racher A J. Advanced Drug Delivery Reviews 2006; 58: 671- 685.
[55] [55] King DJ, Byron OD, Mountain A, Weir N, Harvey A, Lowson AD, Proudfoot KA, Baldock D, Harding SE. Yarranton GT. Biochem. J 1993; 290 (Pt3): 723-725.
[56] [56] Davis SJ, Puklavec MJ, Ashford DA, Harlos K, Jones EY, Stuart DI., Williams AF. Protein Eng 1993; 2 (6): 229-232.
[57] Nivitchanyong T, Yongkiettrakul S, Kramyu J, Pannengpetch S, Wanasen N Journal of Virological Methods 2011; 178 (1–2): 44-51.
[58] Magistrelli G, Malinge P, Lissilaa R, Fagète S, Guilhot F, Moine V, Buatois V, Delneste Y, Kellenberger S, Gueneau F, Ravn U, Kosco-Vilbois M, Fischer N. Protein Expression and Purification 2010; 72 (2): 209-216.
[59] Wang F, Wang TY, Tang YY, Zhang JH, Yang XJ. Gene, 2012; 500 (1):. 59-62.
[60] Yang W, Xia W, Mao J, Xu D, Chen J, Feng S, Wang J, Li H, Theisen CF, Petersen JM, Thórólfsson M, Rasmussen HB, Junker F, Boel E, Su J. Protein Expression and Purification 2011; 76 (1): 59-64.
[61] Yoshimura H, Sekine S, Adachi H, Uematsu Y, Mitani A, Futaki N, Shimizu N. Protein Expression and Purification 2011; 80 (1): 41-46.
[62] Walker PA, Leong LE, Ng PW, Tan SH, Waller S, Murphy D, Porter AG. Biotechnology 1994; 12 (16): 601-605.
[63] Nagai K, Perutz MF, Poyart C. Proc. Natl. Acad. Sci. USA 1985; 82 (21): 7252-7255.
[64] Dougherty WJ, Park TD, Cary SM, Bazan JF, Fletterick RJ. Virology. 1989; 172 (1): 302-310.
[65] LaVallie ER, Rehemtulla A, Racie LA, DiBlasio EA, Ferenz C, Grant KL, McCoy JM. J.Biol.Chem 1993; 268 (31): 23311-23317.
[66] Chang JY. Eur. J. Biochem 1985; 151(2): 217-224.
[67] XU SF, XU J, LI MT. Nan Fang Yi Ke Da Xue Xue Bao 2011; 31(3):397-402.
[68] Nagai H, Kubori T. Methods Mol Biol 2013; 954:347-54.
[69] Deng H, Jiang Q, Liang S, Yan F, Hou S., Qian Z, Li J, Wen Y., Yang J, Wei Y Biotechnol Appl Biochem 2010; 55 (2): 63-72.
[70] Kellermann OK, Frenchi T. Methods Enzymol 1982; 90 (PtE): 459-463.
[71] Di Guan C, P Li, Riggs PD, Inouye H. Gene 1988; 67 (1): 21-30.
[72] Chen P, Yang H, Li H, Yang L, Li X. Sheng Wu Gong Cheng Xue Bao 2011; 27 (8):1247-1257.
[73] Hewitt SN, Choi R, Kelley A, Crowther GJ, Napuli AJ, Van Voorhis WC. Acta Crystallogr Sect F Struct Biol Cryst Commun 2011; 67(Pt 9):1006-1009.
[74] Lee NP, Tsang S, Cheng RH, Luk JM. Protein Pept Lett 2006; 13 (5): 431-435.
[75] Eliseev R, Alexandrov A, Gunter T. Protein Expr Purif 2004; 35 (2): 206-209.
[76] Hoffman A, Roeder RG. Nucleic Acids Res 1991;19 (22): 6337-6338.
[77] Zenke FT, Engles R, Vollenbroich V, Meyer J, Hollenberg CP, Breunig KD. Science 1996; 272 (5268): 1662-1665.
[78] Yip TT, Hutchens TW. Immobilized Methods Mol. Biol. 1996; 59: 197-210.
[79] Zhou X, Wu JC. World J Microbiol Biotechnol 2012; 28(5):2205-2212.
[80] Fazen CH, Kahkoska AR, Doyle RP. Protein Expr Purif 2012; 85(1):51-59.
[81] Ralph EC, Xiang L, Cashman JR, Zhang J. Protein Expr Purif 2011; 80 (1):22-27.
[82] Hellman LM, Zhao C, Melikishvili M, Tao X, Hopper JE, Whiteheart SW, Fried MG. Methods 2011; 54(1):31-38.
[83] Stamsås GA, Håvarstein LS, Straume D. Journal of Microbiological Methods 2013; 92 (1): 59-63.
[84] Blackwell C, Brown JD. Methods Mol Biol 2009; 499:133-148.
[85] Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B. Nat. Biotechnol 1999; 17: 1030-1032.
[86] Gavin AC, Bösche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, Remor M, Höfert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T, Gnau V, Bauch A, Bastuck S, Huhse B, Leutwein C, Heurtier MA, Copley RR, Edelmann A, Querfurth E, Rybin V, Drewes G, Raida M, Bouwmeester T, Bork P, Seraphin B, Kuster B, Neubauer G, Superti-Furga G. Nature 2002; 415 (6868): 141-147.
[87] Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, Wilm M, Seraphin B. Methods 2001; 24: 218-229.
[88] Jessulat M, Buist T, Alamgir M, Hooshyar M, Xu J, Aoki H, Ganoza MC, Butland G, Golshani A Methods Mol Biol 2010; 587:99-111.
[89] Xu X, Song Y, Li Y, Chang J, Zhang H, An L. Protein Expression and Purification 72, (2): 2010; 149-156
[90] Cipak L, Spirek M, Novatchkova M, Chen Z, Rumpf C, Lugmayr W, Mechtler K, Ammerer G, Csaszar E, Gregan J. Proteomics 2009; 9 (20): 4825-4828.

[91] Fernández E, Collins MO, Uren RT, Kopanitsa MV, Komiyama NH, Zografos L, Armstrong JD, Choudhary JS, Grant SG. Targeted tandem affinity purification of PSD-95 recovers core postsynaptic complexes and schizophrenia susceptibility proteins. Mol Syst Biol 2009; 5:269.
[92] Cochrane A, Murley LL, Gao M, Wong R, Clayton K, Brufatto N, Canadien V, Mamelak D, Chen T, Richards D, Zeghouf M, Greenblatt J, Burks C, Frappier L. Virology 2009; 388 (1): 103-111.
[93] Cooper AA, Stevens TH. Trends Biochem. Sci 1995; 20 (9): 351-356.
[94] Li J, Zheng L, Li P, Wang F. Protein Expr Purif 2012; 84(1):1-8.
[95] Wu WY, Miller KD, Coolbaugh M, Wood DW. Protein Expr Purif. 2011; 76(2):221-8.
[96] Baumann L, Beck-Sickinger AG. Biopolymers 2010; 94 (6):771-778.
[97] Yu HH, Nakase I, Pujals S, Hirose H, Tanaka G, Katayama S, Imanishi M, Futaki S. Biochim Biophys Acta 2010; 1798 (12):2249-2257.