Gene Expression Systems and Recombinant Protein production & Purification.

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 plasmids 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.

☒ Bacteria as an expression system

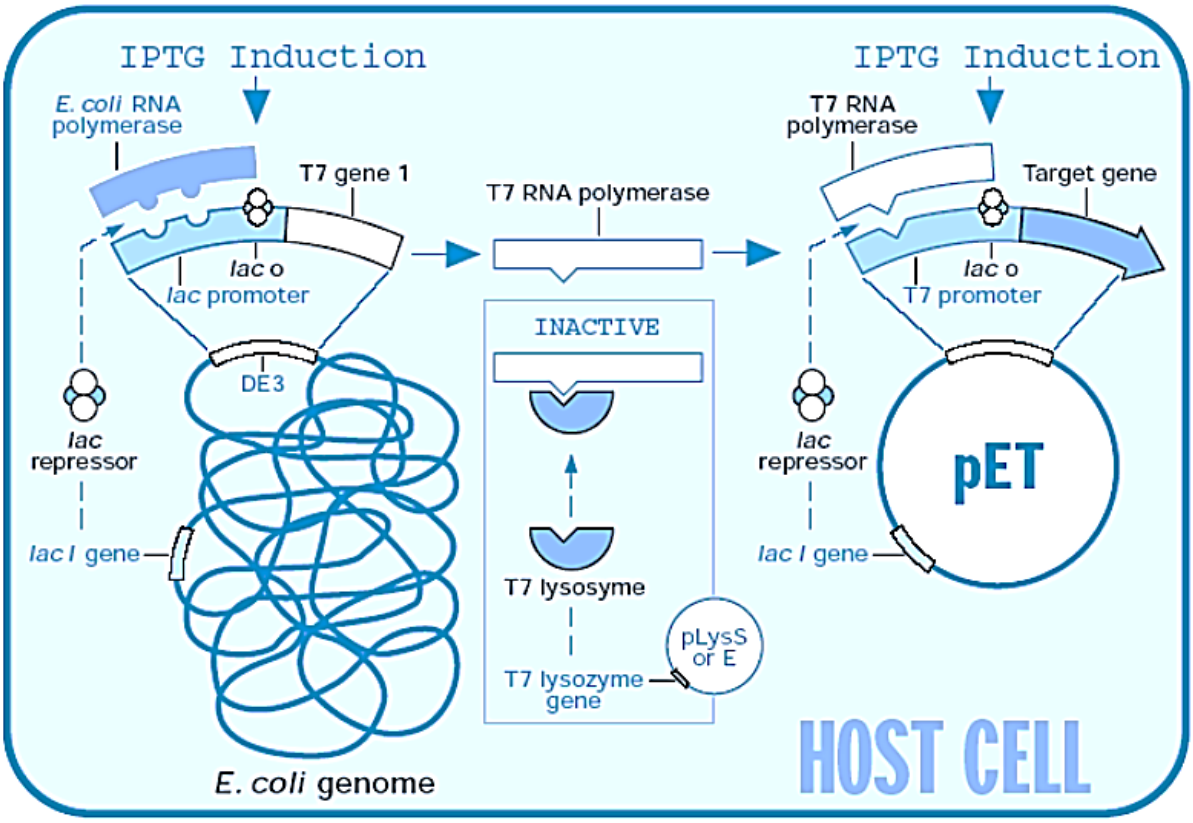
Because of high knowledge about the genetics, biochemistry and molecular biology, *E.coli* is the system of first choice for expression of genes to get required heterologous proteins. This is because the genetic manipulations are straight forward and easy. *E coli* is easy to culture and the growth is also inexpensive. Moreover 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. 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.

Escherichia coli strain to express recobinant protein

- *E. coli* **BL21** are competent cells for high-level expression of a variety of recombinant proteins. They are an ideal host for optimal expression of proteins from vectors utilizing *E. coli* promotors (this strain lacks a T7 RNA polymerase).
- **The BL21 (DE3)** are competent cells that has the T7 RNA polymerase gene under the control of the lacUV5 promoter. This arrangement is on a phage genome, called DE3. DE3 is inserted into the chromosome of BL21 to make BL21 (DE3). T7 RNA polymerase expression is induced by addition of IPTG (1 mM) to the culture. This induction allows production of T7 RNA Polymerase, which then directs the expression of the target gene located downstream of the T7 promotor in the expression vector.
- **BL21 (DE3) pLysS** is a derivative of **BL21 (DE3)**.pLysS is a plasmid that

contains the T7 lysozyme gene (LysS). The T7 lysozyme binds to T7 RNA polymerase causing inhibition until induction by the addition of IPTG. When IPTG is added, the amount of T7 RNA polymerase increases and overcomes the inhibition by LysS.



☒ Expression of recombinant proteins in animal cell culture systems

Technical advances facilitating genetic manipulation of animal cells now allow routine production of therapeutic proteins in such systems. The major advantage of these systems is their ability to carry out post-translational modification of the protein product. As a result, many biopharmaceuticals that are naturally glycosylated are now produced in animal cell lines. **Chinese hamster ovary (CHO) cells** and **Baby hamster kidney (BHK)** cells have become particularly popular in this regard.

Although their ability to carry out post-translational modifications renders their use desirable/ essential for producing many biopharmaceuticals, animal cell-based systems do suffer from a number of disadvantages. When compared with *E. coli*, animal cells display a very complex nutritional requirement, grow more slowly and are far more susceptible to physical damage. In industrial terms, this translates into increased production costs.

☒ *Yeast*

Yeast cells (particularly *Saccharomyces cerevisiae*) display a number of characteristics that make them attractive in this regard. These characteristics include: their molecular biology has been studied in detail, facilitating their genetic manipulation; most are GRAS-listed organisms ('generally regarded as safe'), and they have a long history of industrial application (e.g. in brewing and baking); they grow relatively quickly in relatively inexpensive media, and their tough outer wall protects them from physical damage; suitable industrial-

scale fermentation equipment/technology is already available; they possess the ability to carry out post-translational modifications of proteins.

☒ Transgenic animals

The production of heterologous proteins in transgenic animals has gained much attention in the recent past. The generation of transgenic animals is most often undertaken by directly microinjecting exogenous DNA into an egg cell. In some instances, this DNA will be stably integrated into the genetic complement of the cell. After fertilization, the ova may be implanted into a surrogate mother. Each cell of the resultant transgenic animal will harbour a copy of the transferred DNA.

As this includes the animal's germ cells, the novel genetic information introduced can be passed on from one generation to the next.

A transgenic animal harbouring a gene coding for a pharmaceutically useful protein could become a live bioreactor-producing the protein of interest on an ongoing basis. In order to render such a system practically useful, the recombinant protein must be easily removable from the animal, in a manner which would not be injurious to the animal (or the protein). A simple way of achieving this is to target protein production to the mammary gland. Harvesting of the protein thus simply requires the animal to be milked. Mammary-specific expression can be achieved by fusing the gene of interest with the promoter containing regulatory sequence of a gene coding for a milk-specific protein. Regulatory sequences of the whey acid protein (WAP), β-casein and α- and β-lactoglobulin genes have all been used to date to promote production of various pharmaceutical proteins in the milk of transgenic animals.

One of the earliest successes in this regard entailed the production of human tPA in the milk of transgenic mice. The tPA gene was fused to the upstream regulatory sequence of the mouse WAP, the most abundant protein found in mouse milk.

More practical from a production point of view, was the subsequent production of tPA in the milk of transgenic goats, again using the murine WAP gene regulatory sequence to drive expression.

☒ Transgenic plants

The production of pharmaceutical proteins using transgenic plants has also gained some attention over the last decade. The introduction of foreign genes into plant species can be undertaken by a number of means, of which *Agrobacterium*-based vector-mediated gene transfer is most commonly employed.

Agrobacterium tumefaciens and *Agrobacterium rhizogenes* are soil-based plant pathogens. Upon injection, a portion of *Agrobacterium* Ti plasmid is translocated to the plant cell and is integrated into the plant cell genome. Using such approaches, a whole range of therapeutic proteins have been expressed in plant tissue. Depending upon the specific promoters used, expression can be achieved uniformly throughout the whole plant or can be limited to, for example, expression in plant seeds.

Table Some proteins of potential/actual therapeutic interest that have been expressed (at laboratory level) in transgenic plants

Protein	Expressed in	Production levels achieved
EPO	Tobacco	0.003% of total soluble plant protein
HSA	Potato	0.02% of soluble leaf protein
Glucocerebrosidase	Tobacco	0.1% of leaf weight
IFN-α	Rice	Not listed
IFN-β	Tobacco	0.000 02% of fresh weight
GM-CSF	Tobacco	250 ng ml^{-1} extract
Hirudin	Canola	1.0% of seed weight
Hepatitis B surface antigen	Tobacco	0.007% of soluble leaf protein
Antibodies/antibody fragment	Tobacco	Various

☒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 possess the ability to produce large amount of proteins.

Baculoviruses belong to a large group of circular double stranded DNA viruses. This virus infects only invertebrates, usually insects. The virus has its genome which is 90-180 kbp. 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. 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. 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. In this system multiple genes can be expressed from a single virus. The disadvantage is that the cells grow slowly and the media is expensive. 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.

WHICH PLASMID SHOULD BE CHOSEN?

The most common expression plasmids in use today are the result of multiple combinations of replicons, promoters, selection markers, multiple cloning sites, and fusion protein/fusion protein removal strategies.

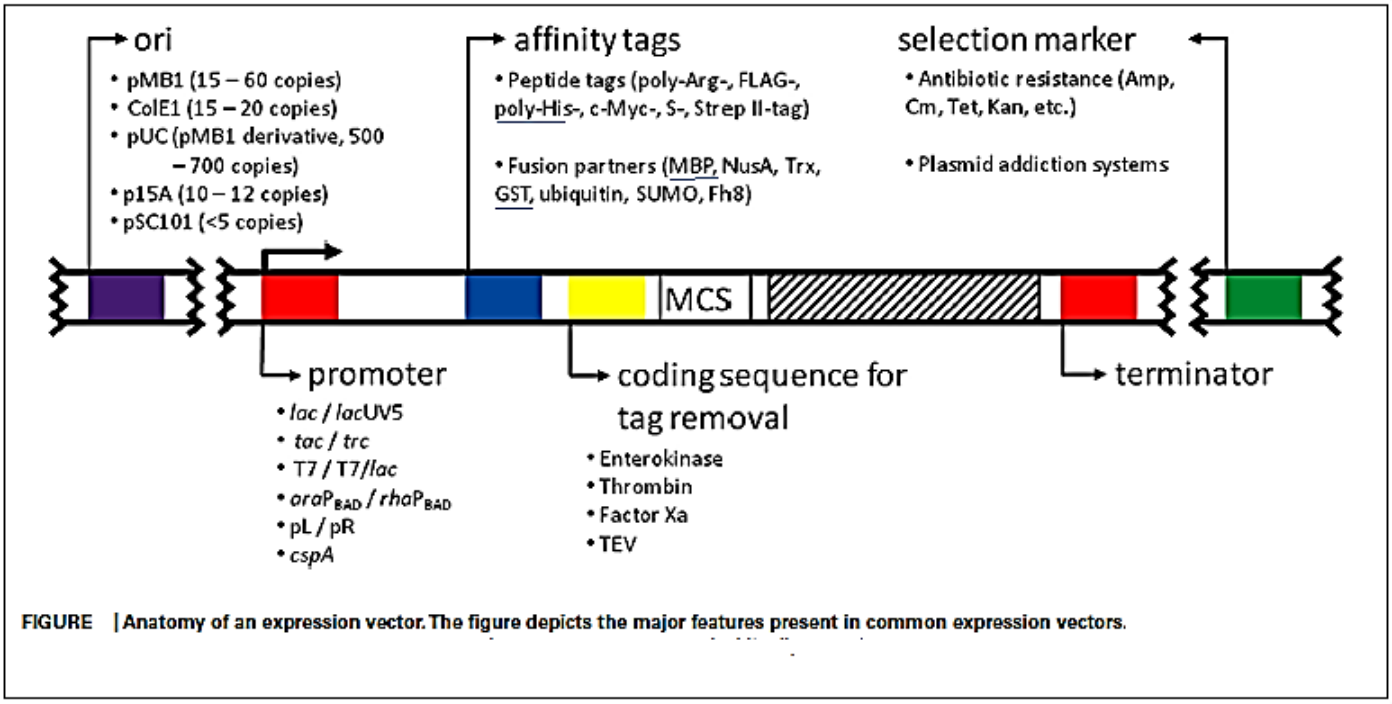


FIGURE | **Anatomy of an expression vector. The figure depicts the major features present in common expression vectors.**

Protein purification

- Affinity chromatography (Specific ligand recognition).
- Ion exchange chromatography (Charge)
- Hydrophobic interaction chromatography (Hydrophobicity).
- Size exclusion chromatography, gel filtration (Size)

Affinity chromatography

The protein of interest is purified by its ability to bind to aspecific ligand that is immobilized on chromotographic beads materials (matrix). Crude cell lysate are incubated with such a matrix under conditions to ensure specific binding of the

protein to the immobilized ligand. Other proteins that don't bind to immobilized ligand are washed out.

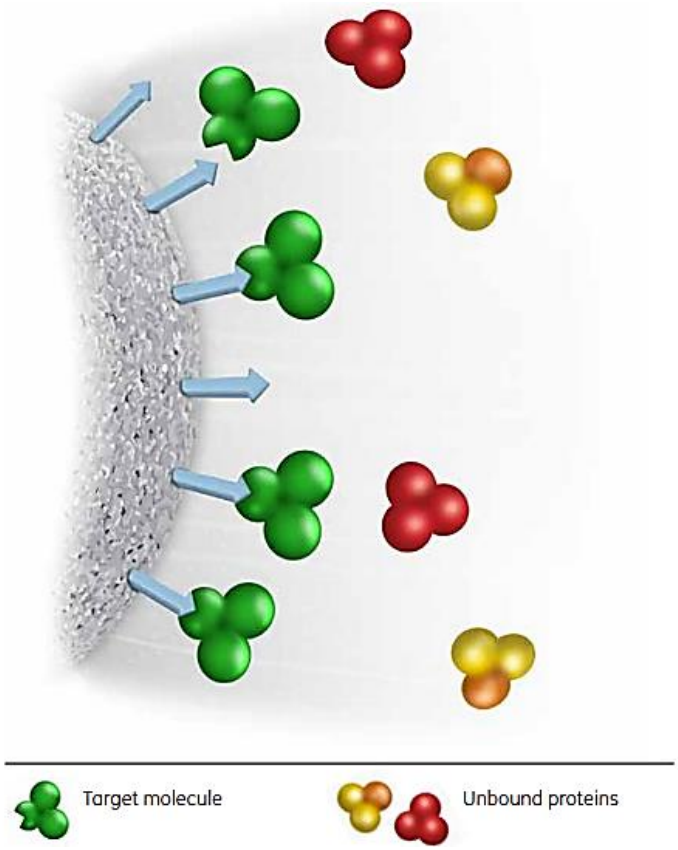


Figure: Affinity Chromotography

Fused partner	Affinity resin	Eluent
β-galactosidase	APTG	sodium borate (pH 10)
ProteinA	IgG	0.5 M acetic acid
Glutathione-S-transferase	glutathione	5 mM GSH
Maltose-binding protein	crosslinked amylose	10 mM maltose
Chloroamphenicol Acetyltransferase	Chloroamphenicol-Sepharose	chloramphenicol
Carbonic anhydrase	sulfonamide-affinity resin	Tris-sulfate
Cellulose-binding protein	cellulose	water
Poly(histidine)	immobolized Ni	imidazole
FLAG	anti-FLAG	EDTA
Poly(arginine)	S-sepharose	NaCl gradient
Poly(cysteine)	thiopropyl-sepharose	mercaptoethanol or DTT
Poly(phenylanaline)	phenyl-sepharose	ethylene glycol

Production of proteins fused to glutathione S-tranferase

Glutathione S-transferase (GST) is a 26 kDa protein derived from *Schistosoma japonicum*. GST recombinantly expressed as fusion to the N-terminus of target proteins. It was used in the development of the **pGEX** vectors.

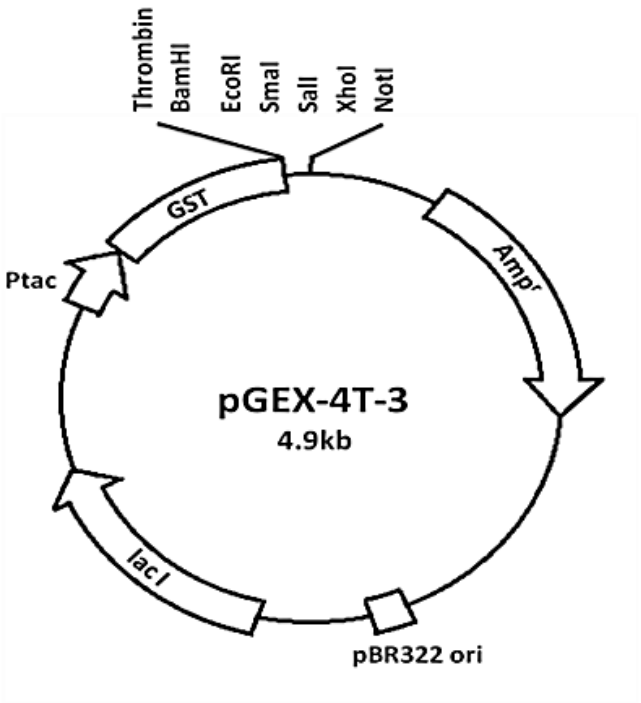
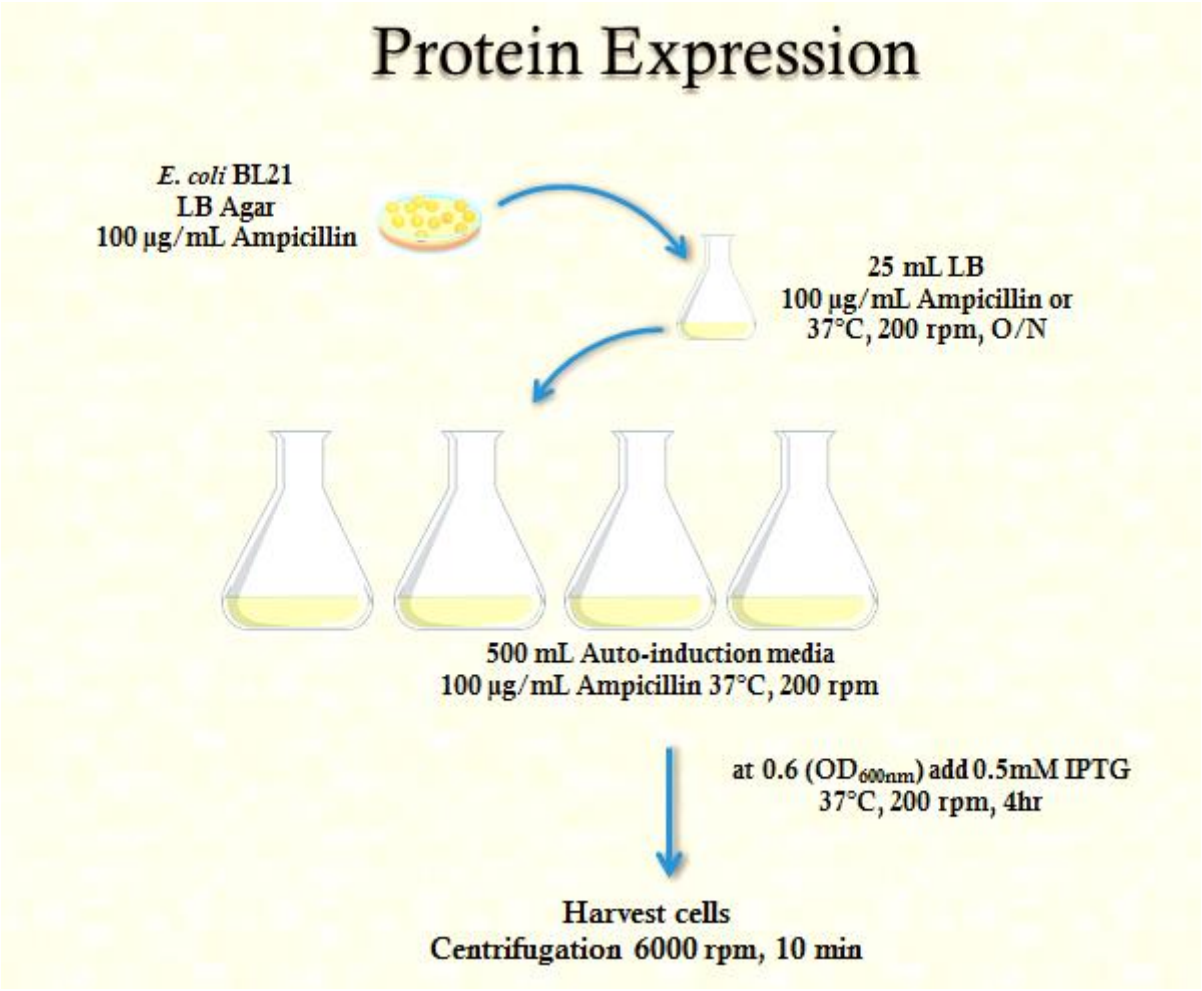


Figure : Vector Map of pGEX-4T3 expression vector

Expression conditions

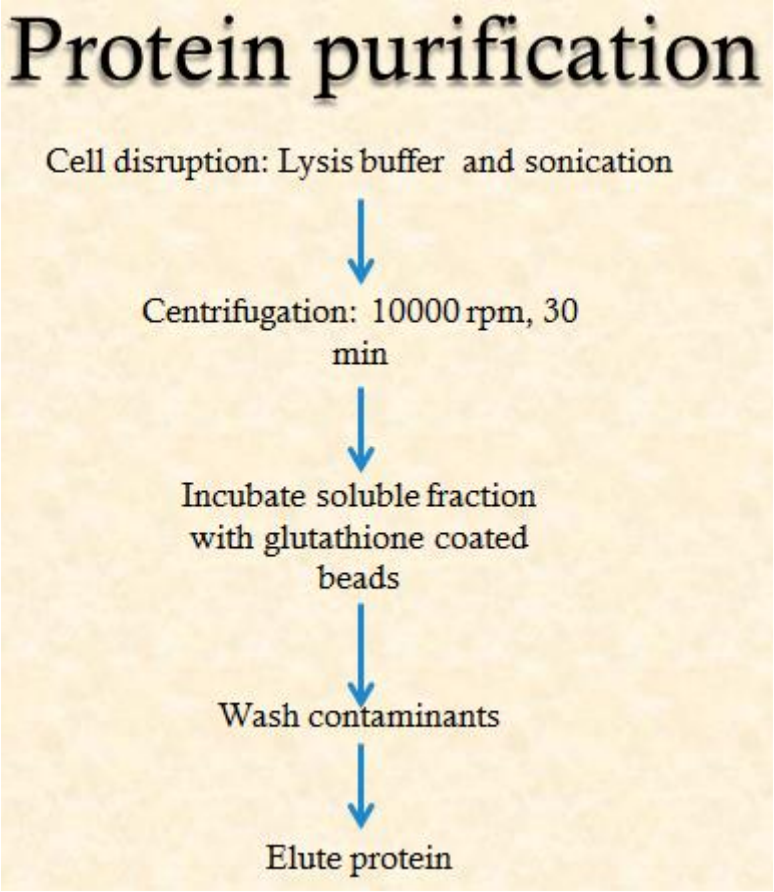
Protein expression can be induced either with the chemical inducer isopropyl-β-D-thiogalactoside (IPTG) or by manipulating the carbon sources during *E. coli* growth (auto-induction).

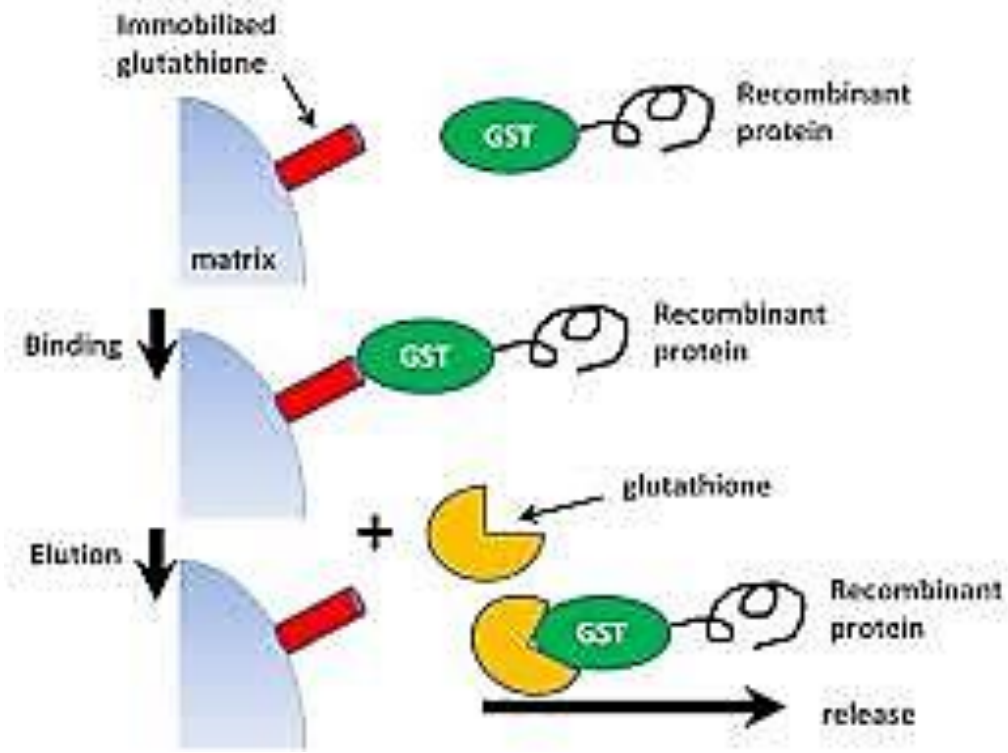


Purification of proteins fused to glutathione S-tranferase

The 26KDa GST moiety binds with high affinity to glutathione coupled to a Sepharose matrix. This binding is reversible and the protein can be eluted under mild, non-denaturing conditions by the addition of reduced glutathione to the elution buffer. A specific protease site engineered between the GST moiety and the protein of interest allows removal of the GST moiety from the target recombinant protein. The GST can be removed from the sample by re-chromatography on a glutathione

column.Thrombin, or Factor Xa protease recognition sites for cleaving the desired protein from the fusion product

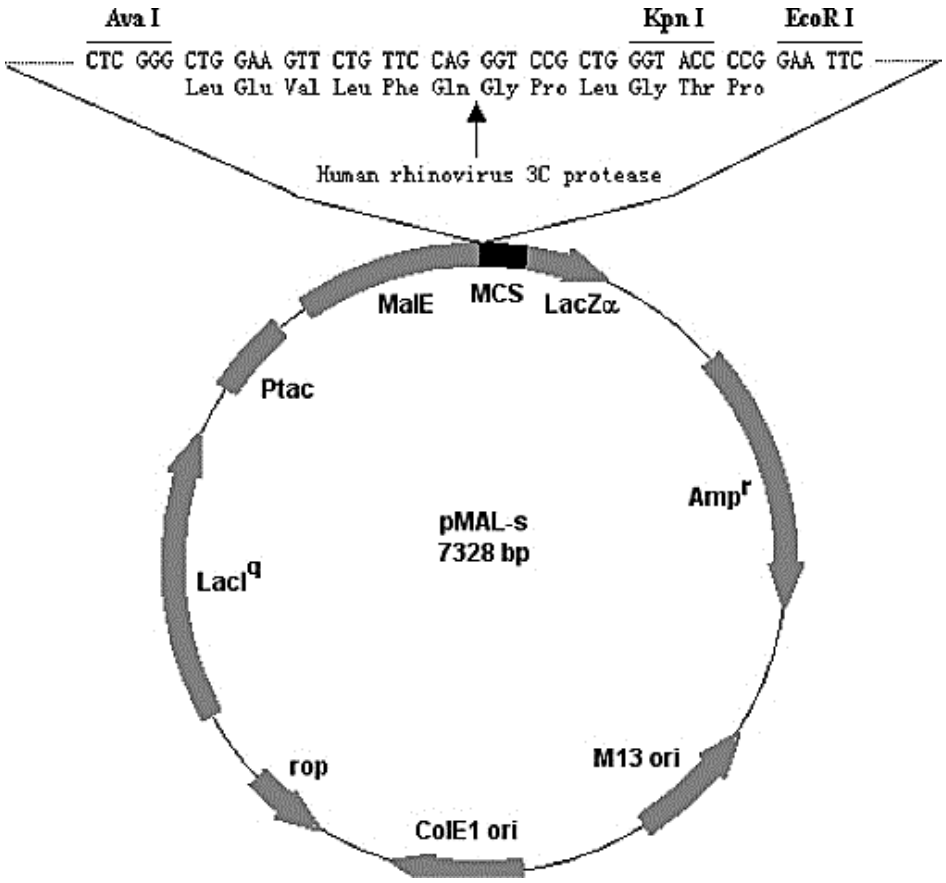




Maltose-Binding Protein (MBP) for purification of recombinant proteins:

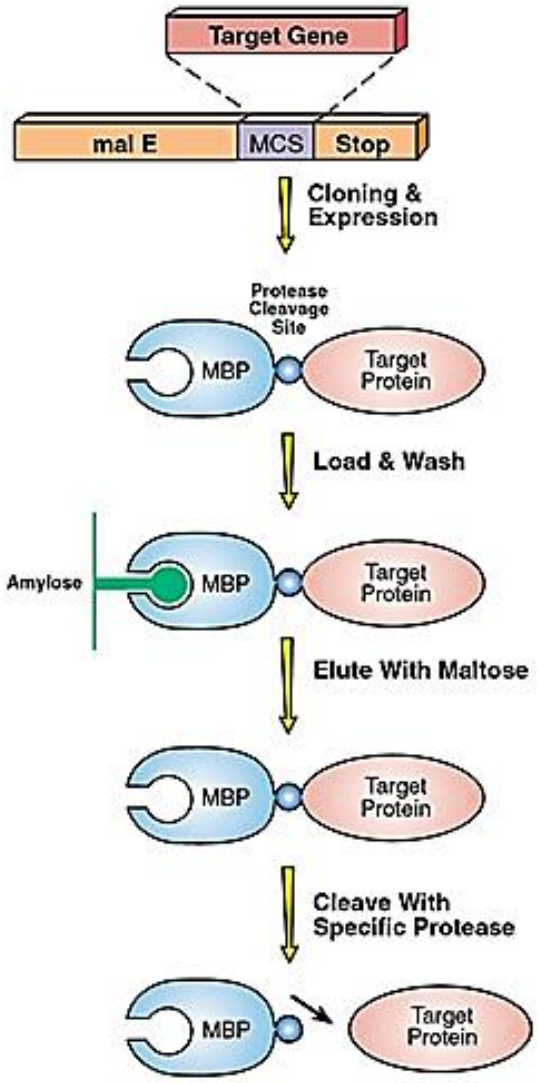
A cloned gene is inserted into a pMAL vector down-stream from the *malE*gene, which encodes maltose-binding protein (MBP). This results in the expression of an MBP-fusion protein. The technique uses the P_{tac} promoter and the translation initiation signals of MBP to express large amounts of the fusion protein.

The *malE* gene can be expressed in *E. coli* producing MBP fusion protein (MBP fused with protein of interest containing factor Xa cleavage sequence between MBP and protein of interest.



This MBP-fusion protein is purified using amylose column, MBP has affinity for the amylose ligand and finally fusion protein can be eluted using maltose gradient. Finally MBP tag can be cleaved from fusion protein using factor Xa protease (as there is factor Xa cleavage site between MBP and protein of interest). Generally factor Xa protease used for the cleavage is in minute amount and removal is not required (or an immobilized factor Xa protease may be used which may be removed by simple centrifugation after completion of cleavage reaction). Cleaved MBP tag

can be separated from the protein of interest by loading it again to amylose column. This time cleaved MBP tag will bind to column but protein of interest will go in unbound fraction.

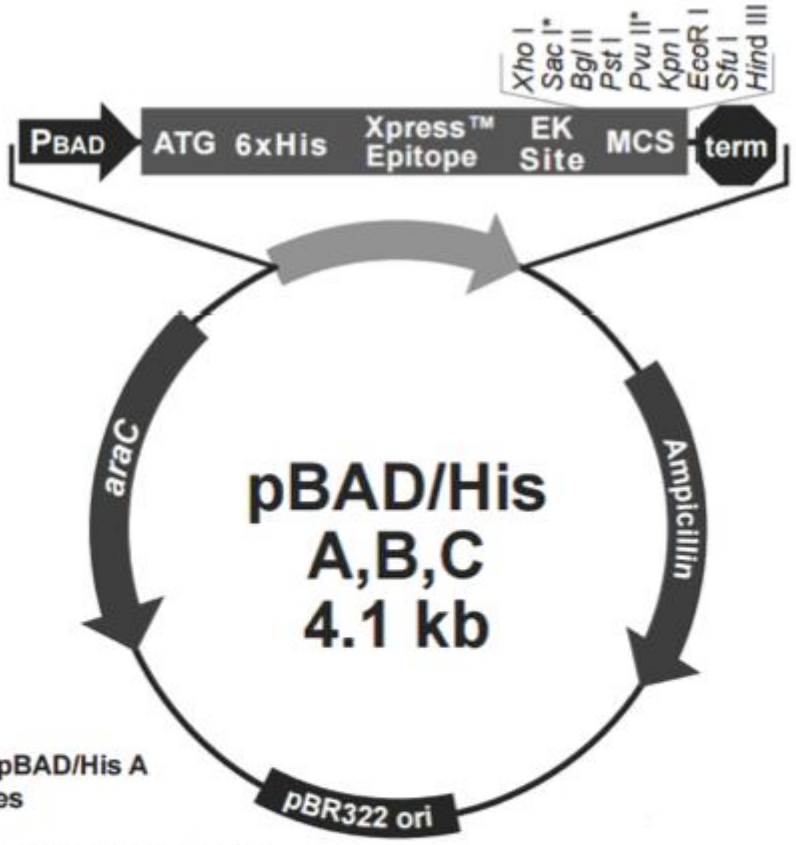


Purification of His-tagged Proteins

A his-tag, or polyhistidine tag, is a string of histidine residues at either the N or C terminus of a recombinant protein. There can be from four to ten residues in a string, although commonly there are six histidine residues — a hexahistidine tag. Some recombinant proteins are engineered to have two hexahistidine tags.

His-tag purification uses the purification technique of immobilized metal affinity chromatography, or IMAC. In this technique, transition metal ions are immobilized on a resin matrix using a chelating agent such as iminodiacetic acid. The most common ion for his-tag purification of a recombinant protein is Ni^{2+}, though Co^{2+}, Cu^{2+}, and Zn^{2+} are also used. The his-tag has a high affinity for these metal ions and binds strongly to the IMAC column. Most other proteins in the lysate will not bind to the resin, or bind only weakly. The use of a his-tag and IMAC can often provide relatively pure recombinant protein directly from a crude lysate.

Imidazole competes with the his-tag for binding to the metal-charged resin and thus is used for elution of the protein from an IMAC column. Typically, a low concentration of imidazole is added to both binding and wash buffers to interfere with the weak binding of other proteins and to elute any proteins that weakly bind. His-tagged protein is then eluted with a higher concentration of imidazole.



Comments for pBAD/His A
4102 nucleotides

*ara*BAD promoter region: bases 4-276
Initiation ATG: bases 319-321
Polyhistidine tag: bases 331-348
Xpress™ epitope: bases 388-411
Enterokinase recognition site: bases 397-411
Multiple cloning site: bases 430-470
*rrn*B transcription termination region: bases 553-710
Ampicillin ORF: bases 989-1849
pBR322 origin: bases 1994-2667
AraC ORF: bases 4076-3198

Feature	Benefit
*ara*BAD promoter (P_{BAD})	Provides tight, dose-dependent regulation of heterologous gene expression
Optimized ribosome binding site	Increases efficiency of recombinant fusion protein expression
Initiation ATG	Provides a translational initiation site for the fusion protein
N-terminal polyhistidine tag	Forms metal-binding site for affinity purification of recombinant fusion protein on a metal-chelating resin. In addition, it allows detection of the recombinant protein with the Penta-His™ Mouse IgG1 Monoclonal Antibody
Anti-Xpress™ epitope (Asp-Leu-Tyr-Asp-Asp-Asp-Asp-Lys)	Permits detection of recombinant fusion protein by appropriate antibodies
Enterokinase cleavage site (Asp-Asp-Asp-Asp-Lys)	Allows removal of the N-terminal peptide by enterokinase for production of native protein
Multiple cloning site	Allows insertion of your gene for expression
*rrn*B transcription termination region	Strong transcription termination region
Ampicillin resistance gene (β-lactamase)	Allows selection of the plasmid in *E. coli*
pBR322 origin	Low copy replication and growth in *E. coli*
*ara*C gene	Encodes the regulatory protein for tight regulation of the P_{BAD} promoter

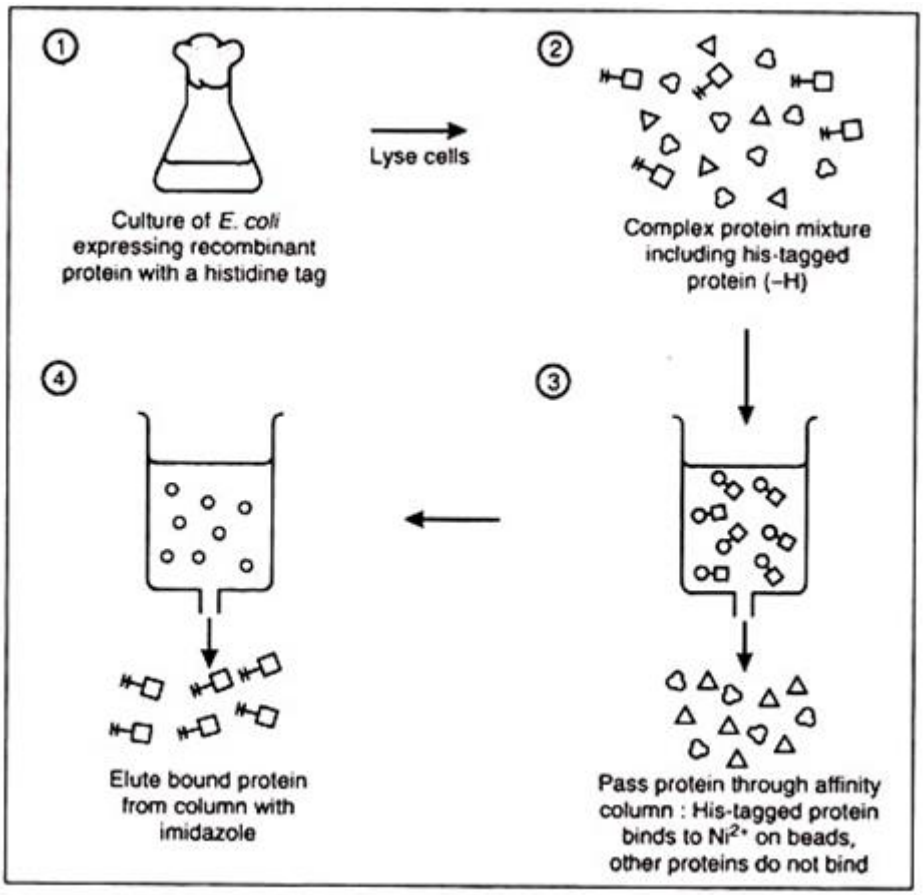
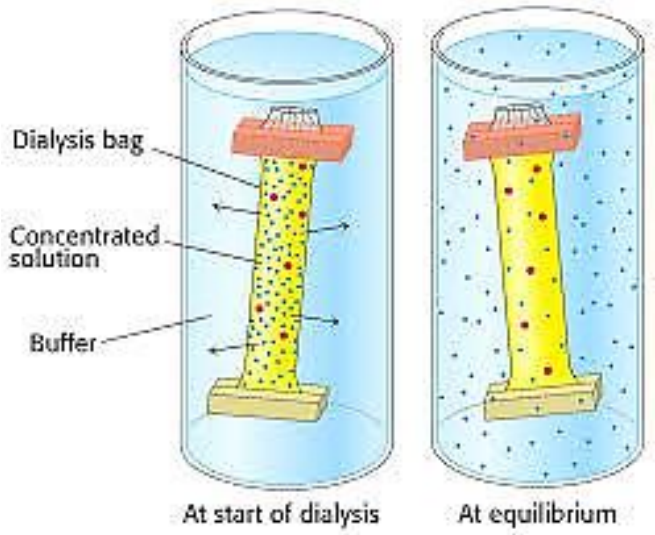


Fig. Purification of his-tagged protein on an affinity column

Dialysis:

The solution of protein is enclosed in a partially permeable membrane. Smaller solute molecules and ions pass through this, while the larger protein molecules are held back.



SDS-PAGE analysis

SDS – sodium dodecyl sulfate

PAGE – polyacrylamide gel electrophoresis

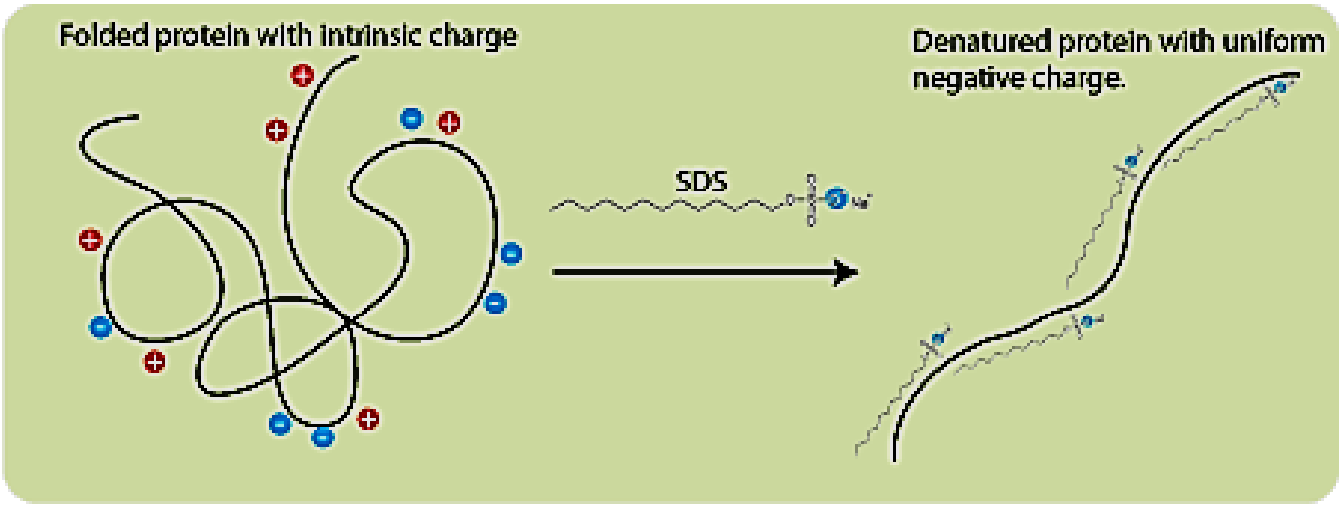
SDS-PAGE is an analytical technique to separate proteins based on their molecular weight. When proteins are separated by electrophoresis through a gel matrix, smaller proteins migrate faster due to less resistance from the gel matrix. Other influences on the rate of migration through the gel matrix include the structure and charge of the proteins.

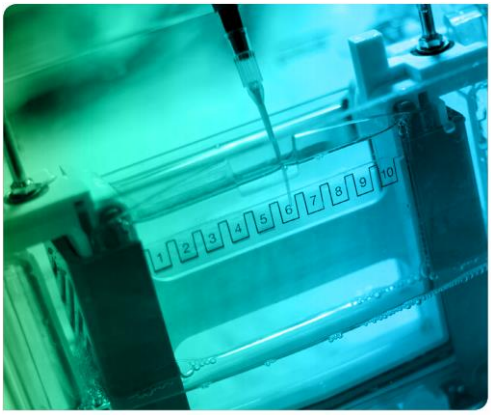
Sodium dodecyl sulfate (SDS) is negatively charged detergent used to denature and linearize the proteins and coated the proteins with negatively charged.

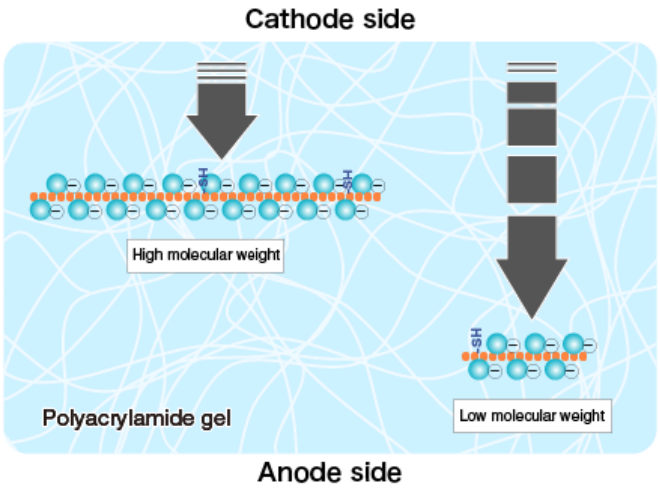
Polyacrylamide is used to form a gel, a matrix of a pores which allow the molecules migrate at different rates.

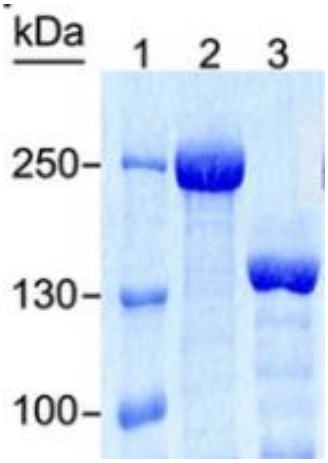
To visualize protein bands

Visualizes the band under UV light using either Coomassie Blue or Silver Stain .


Folded protein with intrinsic charge
SDS
Denatured protein with uniform negative charge.




Cathode side
High molecular weight
Polyacrylamide gel
Low molecular weight
Anode side


kDa
1
2
3
250
130
100

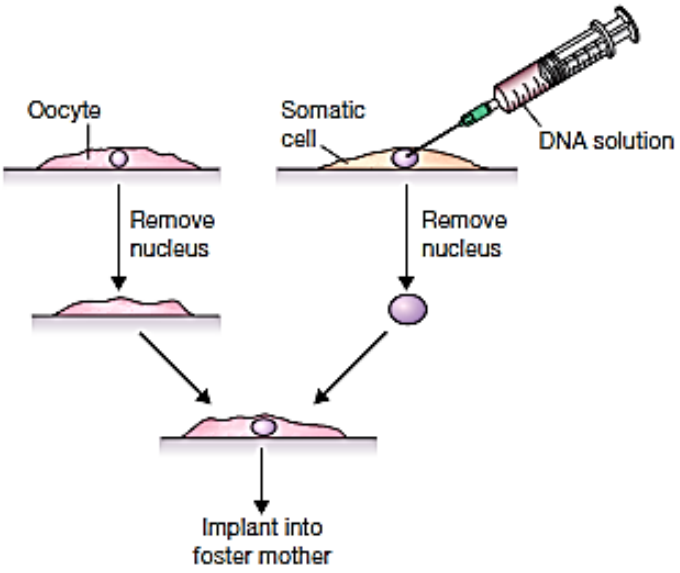
Gene therapy

Gene therapy is a novel treatment method which utilizes genes or short oligonucleotide sequences as therapeutic molecules, instead of conventional drug compounds. This technique is widely used to treat those defective genes which contribute to disease development. Gene therapy involves the introduction of one or more foreign genes into an organism to treat hereditary or acquired genetic defects. In gene therapy, DNA encoding a therapeutic protein is packaged within a "vector", which transports the DNA inside cells within the body. The disease is treated with minimal toxicity, by the expression of the inserted DNA by the cell machinery.

Gene therapy for inherited diseases

There are two basic approaches to gene therapy: germline therapy and somatic cell therapy.

- ✓ In **germline therapy**, a fertilized egg is provided with a copy of the correct version of the relevant gene and re-implanted into the mother. If successful, the gene is present and expressed in all cells of the resulting individual. Germline therapy is usually carried out by **microinjection** of a somatic cell followed by nuclear transfer into an oocyte, and theoretically could be used to treat any inherited disease.



Figure

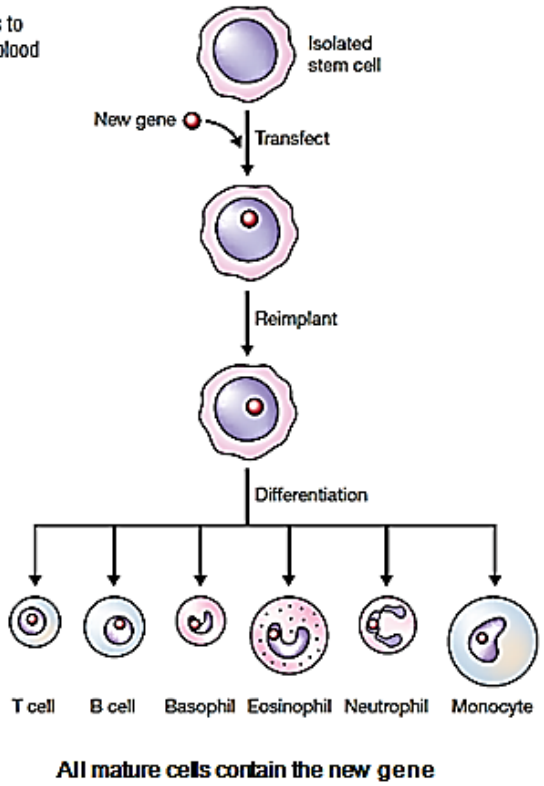
Transfer of the nucleus from a transgenic somatic cell to an oocyte.

✓ **Somatic cell therapy** involves manipulation of cells, which either can be removed from the organism, transfected, and then placed back in the body, or transfected *in situ* without removal. The technique has most promise for inherited blood diseases (e.g., hemophilia and thalassaemia), with genes being introduced into stem cells from the bone marrow, which give rise to all the specialized cell types in the blood. The strategy is to prepare a bone extract containing several billion cells, transfect these with a **retrovirus-based vector**, and then re-implant the cells. Subsequent replication and differentiation of transfectants leads to the added gene being present in all the mature blood cells. The advantage of a retrovirus is that this type of vector has an extremely high transfection frequency,

enabling a large proportion of the stem cells in a bone marrow extract to receive the new gene.

Figure

Differentiation of a transfected stem cell leads to the new gene being present in all the mature blood cells.

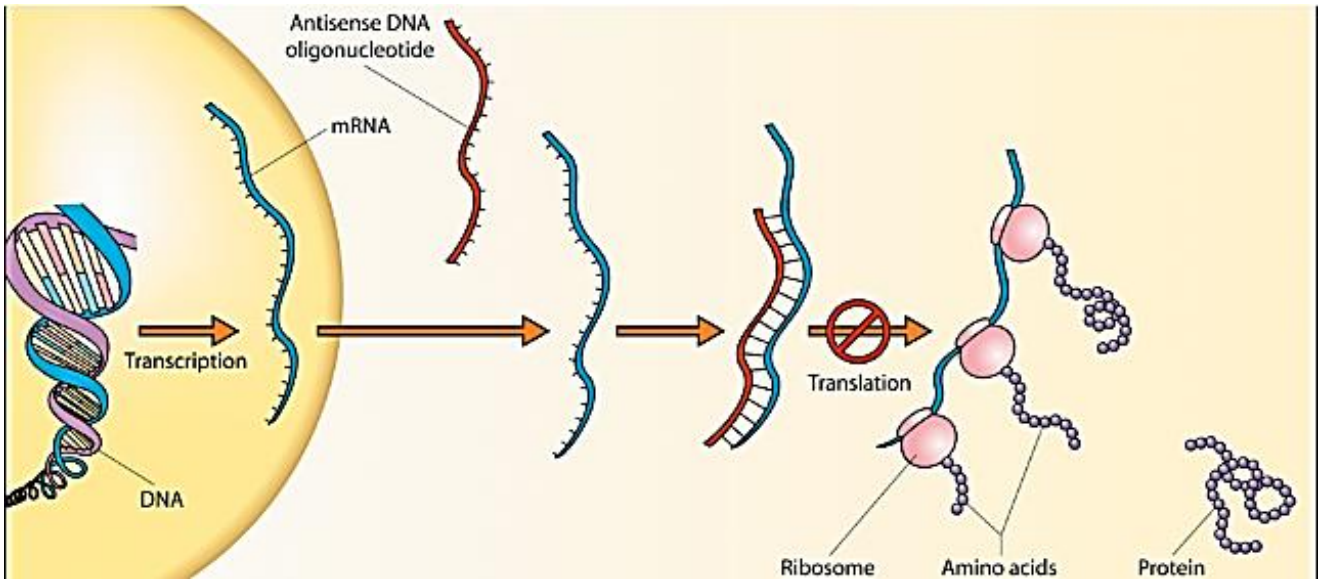


Somatic cell therapy also has potential in the treatment of lung diseases such as cystic fibrosis, as DNA cloned in adenovirus vectors or contained in liposomes is taken up by the epithelial cells in the lungs after introduction into the respiratory tract via an inhaler. However, gene expression occurs for only a few weeks, and as yet this has not been developed into an effective means of treating cystic fibrosis.

Gene therapy and cancer

Most cancers result from **activation of an oncogene that leads to tumor formation**, or **inactivation of a gene that normally suppresses formation of a tumor.** In both cases a gene therapy could be envisaged to treat the cancer. Inactivation of a tumor suppressor gene could be reversed by introduction of the correct version of the gene. Inactivation of an oncogene would, however, require a more subtle approach, as the objective would be to prevent expression of the oncogene, not to replace it with a non-defective copy. One possible way of doing this would be to introduce into a tumor a gene specifying an **antisense** version of the mRNA transcribed from the oncogene. This known as **Antisense gene therapy** is a **gene silencing** technique. The therapy is called a gene silencing technique because, instead of repairing the gene, it aims to “silence” the gene’s effect.

An antisense RNA is the reverse complement of a normal RNA, and can prevent synthesis of the protein coded by the gene it is directed against, probably by hybridizing to the mRNA producing a double-stranded RNA molecule that is rapidly degraded by cellular ribonucleases. The target is therefore inactivated.



An alternative would be to introduce a gene that selectively kills cancer cells or promotes their destruction by drugs administered in a conventional fashion. This is called **suicide gene therapy** and is looked on as an effective general approach to cancer treatment, because it does not require a detailed understanding of the genetic basis of the particular disease being treated. Many genes that code for toxic proteins are known, and there are also examples of enzymes that convert non-toxic precursors of drugs into the toxic form. Introduction of the gene for one of these toxic proteins or enzymes into a tumor should result in the death of the cancer cells, either immediately or after drug administration. It is obviously important that the introduced gene is targeted accurately at the cancer cells, so that healthy cells are not killed. This requires a very precise delivery system, such as direct inoculation into the tumor, or some other means of ensuring that the gene is expressed only in the cancer cells. One possibility is to place the gene under control of the human telomerase promoter, which is active only in cancerous tissues.