

---

## Abstract

The proteins produced in the body control and mediate the metabolic processes and help in its routine functioning. Any kind of impairment in protein production, such as production of mutated protein, or misfolded protein, leads to disruption of the pathway controlled by that protein. This may manifest in the form of the disease. However, these diseases can be treated, by supplying the protein from outside or exogenously. The supply of active exogenous protein requires its production on large scale to fulfill the growing demand. The process is complex, requiring higher protein expression, purification, and processing. Each product needs unique settings or standardizations for large-scale production and purification. As only large-scale production can fulfill the growing demand, thus it needs to be cost-effective. The tools of genetic engineering are utilized to produce the proteins of human origin in bacteria, fungi, insect, or mammalian host. Usage of recombinant DNA technology for large-scale production of proteins requires ample amount of time, labor, and resources, but it also offers many opportunities for economic growth. After reading this chapter, readers would be able to understand the basics about production of recombinant proteins in various hosts along with the advantages and limitations of each host system and properties and production of some of the important pharmaceutical compounds and growth factors.

---

## 4.1 Introduction

The proteins produced in the body control and mediate the metabolic processes and help in its routine functioning. Any kind of impairment in protein production, such as production of mutated protein or misfolded protein, leads to disruption of the pathway controlled by that protein. This may manifest in the form of the disease. However,

these diseases can be treated, by supplying the protein from outside or exogenously. The supply of active exogenous protein requires its production on large scale to fulfill the growing demand. The process is complex, requiring higher protein expression, purification, and processing. Each product needs unique settings or standardizations for large-scale production and purification. As only large-scale production can fulfill the growing

demand, thus it needs to be cost-effective. The tools of genetic engineering are utilized to produce the proteins of human origin in bacteria, fungi, insect, or mammalian host (Fig. 4.1). Usage of recombinant DNA technology for large-scale production of proteins requires ample amount of time, labor, and resources, but it also offers many opportunities for economic growth. After reading this chapter, readers would be able to understand the basics about production of recombinant proteins in various hosts along with the advantages and limitations of each host system and properties and production of some of the important pharmaceutical compounds and growth factors.

## 4.2 Expression of Foreign Gene

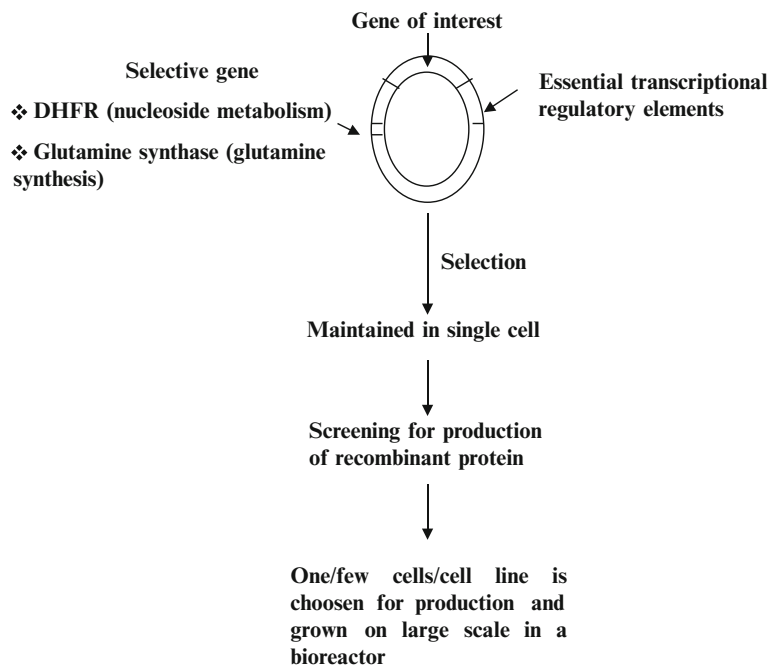
In all living cells, the expression of gene occurs where genetic information contained in DNA is passed on to RNA in the process of transcription and from RNA to protein in the process of translation. Thus, the body synthesizes RNA and then proteins according to the instructions from the DNA. DNA-dependent RNA polymerase or RNA polymerase carries the transcription. Eukaryotes

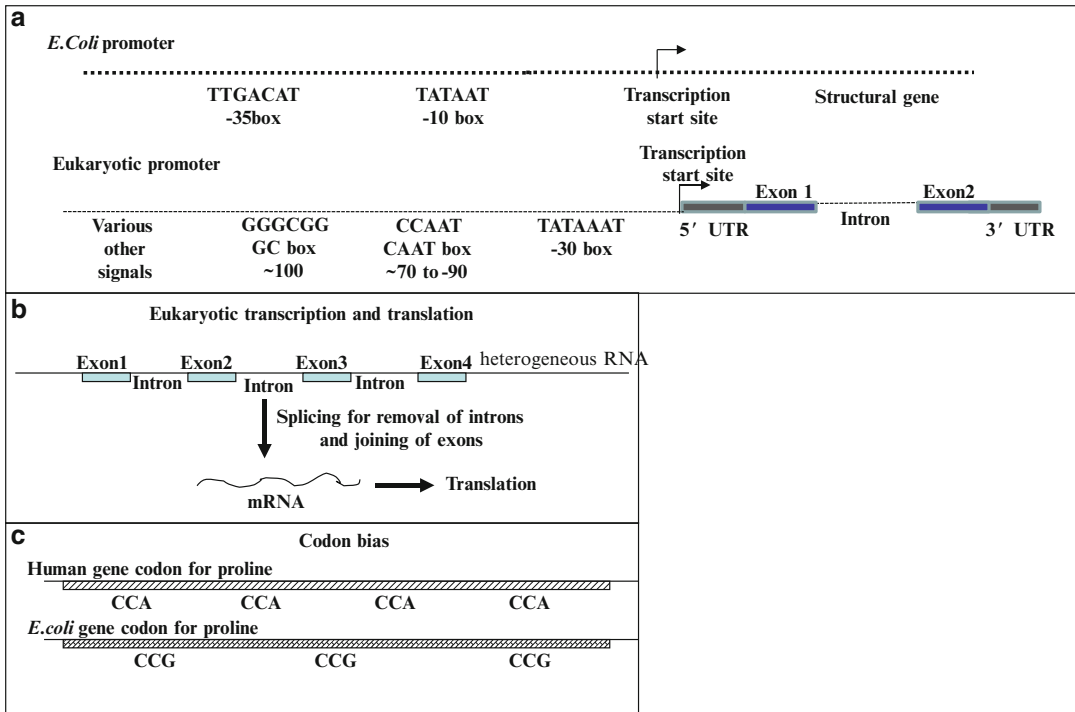
have three RNA polymerases: polymerase I transcribes ribosomal RNA genes (18S, 5.8S and 28SrRNA), polymerase II transcribes all protein-coding genes (mRNA and small RNA), and polymerase III transcribes the genes for 5SrRNA and tRNA. In prokaryotic cells (*E. coli*), RNA polymerase consists of five subunits—two identical  $\alpha$  subunits and one subunit each of  $\beta$ ,  $\beta'$ , and  $\sigma$  subunit. The  $\sigma$  subunit dissociates after polymerization ensues. Thus, the term “holoenzyme” is used for complete enzyme and “core enzyme” is used without  $\sigma$  subunit. The process of transcription is initiated by binding of RNA polymerase to DNA molecule at a very specific site called “promoters.” The promoters are critical for the start of the transcription. The promoter sites are nearly 40 bp long and are mostly located before the first base, which is copied into RNA (Fig. 4.2a). This first base is called as start point of transcription and is denoted by +1.

### 4.2.1 Promoters

Promoters for RNAPol II allow differential expression of genes and determine the rate at which the genes are transcribed. There are some

**Fig. 4.1** The gene of interest is cloned in suitable vector. For expression of the cloned gene, the gene is with all the essential regulatory elements required for transcription of the gene. The gene is attached to the selective gene, which helps in the selection of the clones with gene of interest. The cells are screened for the synthesis of desirable product and then processed for large-scale production in the bioreactor with optimum condition for high yields





**Fig. 4.2** The figure shows the problems encountered by *E. coli* due to sequence of foreign gene, when it is cloned and expressed in it. (a) It shows the promoter for *E. coli* and eukaryotes. As there are differences in the promoter of *E. coli* and eukaryotes, thus eukaryotic promoter might not work in *E. coli*, and the gene is placed under the control of *E. coli* promoter. (b) In the eukaryotes the splicing machinery removes introns from the target gene. Whereas the *E. coli* does not have any such system, therefore, the

intronless version of gene is used in *E. coli*. (c) shows codon bias in bacterial and human system. As one amino acid is encoded by more than one codon, for example, proline, where preferential codons in humans are CCA whereas *E. coli* prefers CCG, if the gene containing the preferential codons for humans is used in *E. coli*, then it might result in inefficient translation. Thus, these problems need to be taken care of while using *E. coli* as host

promoters that cause the inserted genes to be expressed all the time; in all parts of the system, they are known as “constitutive” promoters. Others allow expression only at certain stages/certain tissue/organ of individuals and at certain time points. Gene expression is under temporal and spatial regulation.

In prokaryotes the position before start site at  $-10$  and  $-35$  can interact with  $\sigma$  subunit of holo-enzyme of RNA polymerase. In eukaryotes the sequence (analogous to  $-10$ —consensus sequence of prokaryotes), TATAAAA, is present at  $-30$  position in the promoter region. Eukaryotic promoter is shown (Fig. 4.2a) with TATA box forming the core promoter at  $-30$  position (from  $-30$  to  $-100$ ), upstream of transcription start site. CAAT box and GC box are at

approximately  $-70-90$  and  $-100$ , respectively. The location of promoter is always on the same DNA molecule which they regulate. They are referred as cis-acting elements. The spacing of various elements is more important and much is dependent on locus-specific activators, either at core promoter or at distant sites. Various other signals as enhancers are also involved which are far apart from the target gene. They exert stimulatory effects on promoter activity and can be upstream, downstream or in the middle of the gene. Promoters of housekeeping genes or genes with complex patterns of expression have CpG islands rather than TATA box. The gene to be transcribed has 5'-untranslated region (5' UTR), exons (coding region) separated by introns (non-coding region). The end has 3'-untranslated

region (3' UTR). In the cloning for gene expression, usually intronless version of the gene is used. In eukaryotes, the transcription is further enhanced by enhancers, which may be 2,000–3,000 bases away from the promoter region but are able to affect the rate of transcription.

#### 4.2.2 General Considerations for Protein Production

The simplest host for the work of recombinant DNA technology is prokaryotic bacterial system. In the early 1980s, the first recombinant FDA-approved pharmaceutical, the human insulin (Humulin-US/Humuline-EU), was obtained from genetically engineered *Escherichia coli* (*E. coli*) for treatment of diabetes.

Due to increasing demand, many strains of microbial species are being designed with increased throughput and better recovery of the therapeutic protein from large-scale culture. The recombinant proteins approved by FDA are obtained either from *Escherichia coli* or other prokaryotes; from *Saccharomyces cerevisiae* or other fungal species; from insect cells, mammalian cells, or human cells; or from transgenic plants or animals. Cloning and production of protein in a particular host system are dependent upon a property of host to clone and express the desirable size of protein-encoding genes; production of correctly modified, folded, and functional protein; high yield of the protein; and low-cost requirements. The choice of host systems requires best system, which can fulfill the requirements [17].

The advantages of producing proteins using recombinant DNA technology are:

- As human gene may be cloned and expressed, it minimizes the risk of immune reaction and the specific activity of the protein is high.
- The therapeutic protein can be produced efficiently, maintaining its cost-effectiveness.
- It minimizes the risk of transmission of unknown pathogens present in animal and human sources.

- Appropriate modification with higher specificity, increased half-life, and improved functionality.
- Allow to create critical changes for better specificity and activity.

#### 4.2.3 High Protein Expression in the Host

Expression and production of eukaryotic protein require cloning of its cDNA in an expression vector and subsequent transfer of the vector in suitable host. DNA is modified, cloned, and expressed in other host for the production of the protein; thus, optimum production conditions are required in each host. However, there are yield variations in different expression systems but high-level expression of the protein may be achieved by considering the following points:

- The recombinant gene should be with all the necessary elements for effective transcription initiation.
- Use is made of strong viral or cellular promoter/enhancer for efficient driving of transcription. The usage of viral T7 promoter in bacteria can result in higher yield of recombinant proteins. Transgene may be expressed under the control of either polyhedrin promoter of baculovirus, E1 promoter of adenovirus, or p7.5 promoter of *Vaccinia virus*. These are suitable for wide range of cell types. Mammalian cells may use promoter and enhancer of SV40 or the long-terminal-repeat promoter and enhancer of the *Rous sarcoma virus* or early promoter of the human cytomegalovirus (refer to Chap. 2 for promoters and expression vectors).
- Polyadenylation signals are helpful in eukaryotic genes. These terminators are required for defined 3' end to the mRNA which extends by addition of poly A tail ultimately increasing the stability of RNA and facilitating its export to the cytoplasm.
- Removal of 3' and 5' UTRs (untranslated repeats) may influence gene expression. They may interfere with initiation of translation and

their secondary structure prevents efficient translation.

- Kozak's consensus 5'-CCRCCAUGG-3' with purine at -3 position and guanine at +4 position affects transgene expression.
- Inclusion of one intron sequence, which is located between the promoter and cDNA coding sequence, gives better yield; thus, most expression vectors include at least one intron sequence. Intron presence is of importance when transgene needs to be expressed in mammalian cells.
- For eliminating the slow rates of translation due to codon biasing, the gene may be converted to a high-expressing gene by changing the t-RNA codons to the most abundant ones.
- The integration site has a major effect on the rate of the transcription of the recombinant gene (position effect).
- Some selective genes like dihydrofolate reductase (DHFR) or glutamine synthetase (GS) gene are incorporated. The genes are involved in nucleotide biosynthesis and glutamine synthetase, respectively; the selection occurs when the appropriate metabolite is missing preventing the growth of non-transformed cells.
- For increasing the yield of the recombinant protein, the protein is often fused with endogenous protein sequence (see Chap. 5 for selectable markers, reporters).
- To aid purification, the protein-encoding DNA contains coding DNA for specific protein or peptide that can be a target for affinity chromatography.
- For affinity purification, two systems are readily employed, (1) glutathione S-transferase (GST)-glutathione affinity and (2) polyhistidine-nickel ion affinity. GST has high affinity for glutathione and protein with side chains of histidine has high affinity for nickel ions.
- Stable transfection gives high yield of protein. After recombinant DNA is transfected into animal cells, it either can be integrated (stable expression) into the host genome or maintained in episomal form (transient expression). In stable expression systems, the foreign gene is passed on to the next descendants, and the expression is maintained generation after generation.

### 4.3 Microbial System for Production of Therapeutic Protein

Though various host systems are available for production of recombinant proteins, microbial hosts offer several advantages over other systems, as production is fast, cheap, and economic:

1. The molecular biology and physiology is well characterized and documented.
2. Easy to maintain and manipulate.
3. Utilizes inexpensive nutrition sources.
4. Rapid growth and biomass accumulation to achieve high cell densities.
5. Scale-up is easy and convenient.
6. Their expression machinery can be with variety of strong inducible promoters.

#### Inducible Promoters

The inducible promoter may require an inducer, or the depletion or addition of a specific nutrient, or pH change or changes in physicochemical factors in order to initiate the process of gene expression. The inducible systems suffer from the disadvantage that chemical inducers may be expensive and toxic and would require elimination during downstream processing when the product is intended for human usage. Thus, usage of thermoregulated systems has been used for production of recombinant pharmaceutical proteins as expression is dependent upon strong heat-regulated promoter minimizing the risks of any addition of chemical agent.

As with the cellular structure of bacteria, it can rapidly adapt to culturing conditions with very short replication time (20 min). The media requirement of bacterial cell is simple and consists of simple carbon and nitrogen source. Thus, the overall inputs in bacteria are 90% lower than for mammalian cells. Some of the

approved bacteria-derived (by either the European Union or FDA, USA) therapeutics include hormones (human insulin and insulin analogs, calcitonin, human growth hormone, glucagons, parathyroid hormone, somatropin, and insulin-like growth factor 1), interferons (alfa-1, alfa-2a, alfa-2b, and gamma-1b), interleukins 11 and 2, light and heavy chains raised against vascular endothelial growth factor-A, tumor necrosis factor alpha, cholera B subunit protein, granulocyte colony-stimulating factor, and plasminogen activator [3].

The microbe of first choice for production of recombinant protein is enterobacterium *E. coli*. The system offers quick and easy modifications, ease of growing in manageable environmental conditions and short life cycle. The bacterial cell can tolerate and adapt to changes in the environment rapidly, thus scale-up is easier. However, the system suffers from some of the disadvantages [11,16].

1. Human or mammalian genes cloned in bacteria cannot undergo splicing due to lack of splicing machinery, thus intron-less version of the gene is cloned for optimum results (Fig. 4.2b).
2. The signals involved in transcription of genes may vary; thus, the gene of interest is usually fused with bacterial gene under the control of its promoter, and the protein is obtained as a fusion product, which can be later cleaved, purified, and used.
3. Lac promoter is one of the most popular bacterial promoters. However, for high-level expression, T7 promoter is also preferred (present in pET vectors). It can drive the target protein expression to nearly 50% of total cell protein. Gene of interest can be placed under the control of regulated promoter of phage.
4. For translation, abundance of t-RNA is related to the frequency of appearance of different codons (codon biasing). Therefore, codons rare for *E. coli* may cause amino acid misincorporation or premature termination affecting the yield of the therapeutic protein. This can be solved by either site-directed replacement of rare codons to the codons (for the same amino acid) preferred by *E. coli*. Another approach may be co-expression of rare t-RNAs in *E. coli* (strains of *E. coli*, BL21 codon plus, and Rosetta were designed for this purpose). For addition of amino acids during the process of translation, as there are more than one codon for several amino acids, thus, codon biasing occurs which is the preference of a particular codon of amino acid in a particular species (Fig. 4.2c).
5. Complexity: Eukaryotic cells have the advantage of producing fully functional and properly folded proteins. The antibodies with four subunits may be secreted by eukaryotic cells in fully functional form. On the other hand, it is very difficult to obtain multidomain protein from *E. coli*. Even if protein is obtained, the renaturation and folding in laboratory condition may either be very expensive or protein may lack its activity.
6. Lack of posttranslational modifications (PTMs) is the problem, which cannot be solved, and is mandatory for activity of many therapeutic proteins. The glycosylation is the most common modifications, and others are phosphorylation and formation of disulfide bond, which are essentially required for the full functional capability of many human proteins. PTMs play an important role in proper protein folding, processing, stability, tissue targeting, activity, immune reactivity, and half-life of the protein. Lack of these results in insoluble, unstable, or inactive product.
 

However, the N-linked glycosylation system of *Campylobacter jejuni* has been successfully transferred to *E. coli*, thus opening a possibility for the production of glycosylated protein in it. Certain mutant *E. coli* are being developed to promote disulfide bond formation (AD494, Origami, Rosetta-gami) with reduced protease activity (BLZ1).
7. Overproduction of recombinant protein in bacteria might result in the loss of solubility and deposition of many protein species as protein aggregates or inclusion bodies. Alteration in growth conditions might render the product in insoluble form. Many eukaryotic proteins are found trapped in inclusion bodies with resistance to further processing. Success has

been obtained in purification of insulin and betaferon from inclusion bodies. The retrieval of proteins using denaturing condition with subsequent refolding and renaturation might not always be easy and prove to be extremely expensive.

8. With the *E. coli* host, it is very difficult to obtain protein larger than 60 kDa in soluble form.

Due to certain limitations for production of proteins in *E. coli*, other host systems are being discussed for production of proteins.

### Glycosylation

Covalent attachment of carbohydrate group to the protein to form glycoprotein is called glycosylation. In glycoproteins, proteins constitute a major fraction. These play important roles in various physiological processes and are components of cell membranes.

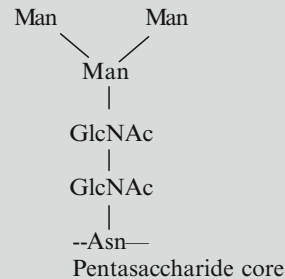
The carbohydrates commonly attached to proteins may be fucose (Fuc), galactose (Gal), N-acetylgalactosamine (GalNAc), glucose (Glc), N-acetylglucosamine (GlcNAc), mannose (Man), and sialic acid (Sia).

These sugar moieties may associate through amide nitrogen atom of side chain of asparagine (Asn) termed as N-linked glycosylation or to the oxygen atom in the side chain of serine (Ser) or threonine (Thr) termed as O-linked glycosylation.

Not all the asparagine (Asn) present in polypeptide can accept the carbohydrate moiety. The residues with the sequence Asn-X-Ser or Asn-X-Thr, where X is any amino acid except proline, are targets for glycosylation. Not only the residual sequence but other aspects of the structure of the protein and cell type determine the glycosylation site.

All the N-linked sugar residues have a common core of pentasaccharides. These pentasaccharide consists of three mannose

and two N-acetylglucosamine residues. The core may attach to different oligosaccharides to form different glycoproteins.



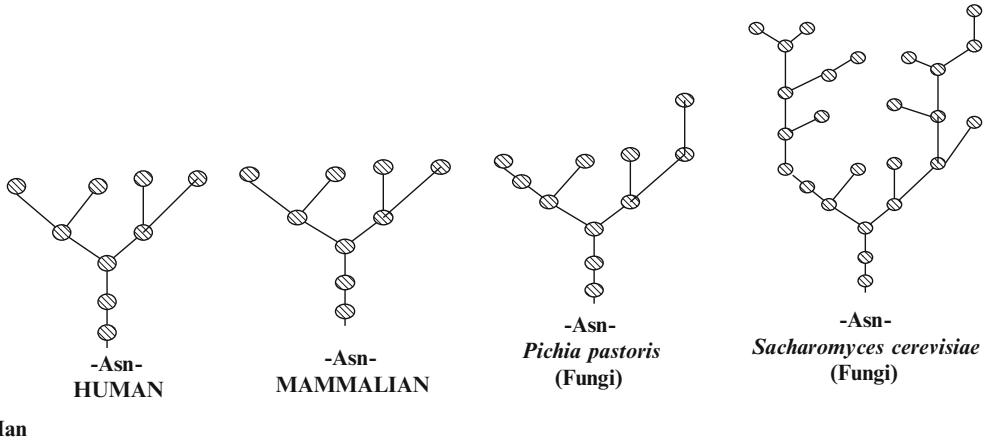
Glycosylation is one of the important posttranslational modifications, which occurs inside the lumen of the endoplasmic reticulum (ER) and in Golgi complex. The ER and Golgi complex are important in protein targeting and transport. N-linked glycosylation starts in the endoplasmic reticulum and continues in the Golgi complex after the polypeptide is synthesized on ribosomes. However, O-linked glycosylation exclusively occurs in the Golgi complex.

In the process, oligosaccharide to be attached to the protein associates with a specialized lipid present in ER, dolichol phosphate, which consists of about 20 isoprene (C5) units. Through phosphate of dolichol phosphate, oligosaccharide is transferred to specific asparagine residue of polypeptide chain on ribosomes. The enzyme responsible for glycosylating protein and activated oligosaccharide are located on the lumen side of ER.

Then these are transported to the Golgi complex, where the carbohydrate units are altered and finalized. Golgi complex is responsible for O-linked sugar attachment and modification of N-linked sugar. Then the proteins are targeted and transported to their destination.

(continued)





**Fig. 4.3** Many of the human proteins are glycosylated (O-linked or N-linked glycosylation). Glycosylation is one of the important posttranslational modifications. The *E. coli* is unable to perform glycosylation. Fungi are simplest eukaryotic systems which can perform glycosylation. The use of fungal host (*Pichia pastoris* and *Saccharomyces*

*cerevisiae*) for production of recombinant proteins results in hyper-glycosylation. The figure shows glycosylation pattern in human, mammalian, *Pichia*, and *Saccharomyces* systems. The hyper-glycosylations or abnormal glycosylations can make the protein highly immunogenic making it unsuitable for therapeutic purposes

#### 4.4 Production of Recombinant Protein in Fungal Hosts

Due to the problems encountered in *E. coli* for production of larger proteins or modified proteins, the next cost-effective, fast, high-density, and easy to handle system is of fungi. *Saccharomyces cerevisiae* (yeast) was the system of choice when it was difficult to obtain therapeutic protein in soluble form and with appropriate posttranslational modifications in bacterial host. In yeast, mutants are available which can give high yield. The approved products obtained from yeast are hormones, vaccines, recombinant granulocyte macrophage colony-stimulating factors (GM-CSF), albumin, hirudin, and platelet-derived growth factor (PDGF).

The advantages of *S. cerevisiae* are the following: (1) it secretes recombinant protein in the culture, (2) protein is properly folded, and (3) it performs most posttranslational modifications. With the yeast system, high amounts of recombinant protein are obtained, and yeast is also capable of performing posttranslational modifications as O-linked glycosylation, phosphorylation, acetylation, and acylation but differs drastically in patterns of N-linked glycosylation (Fig. 4.3).

##### Protein Folding and Molecular Chaperons

Chaperones are family of highly conserved different proteins. The important functions of chaperones are:

- Prevention of aggregation and misfolding of newly synthesized polypeptide chain.
- They prevent irreversible aggregation of nonnative conformation and maintain the protein on the productive folding pathway.
- They prevent nonproductive interactions with other components of the cell.
- They help and guide the direct assembly of multisubunit protein complexes and larger proteins.
- The chaperones involved in folding recognize nonnative substrate proteins mainly via their exposed hydrophobic residues.

The major classes of molecular chaperones are:

Heat shock proteins are present in a variety of systems and prevent damage to the proteins under high heat.

(continued)



**HSP60**

- It is tetradecameric mitochondrial chaperonin.
- It is implicated in protein import and macromolecular assembly.
- Required for folding of precursor polypeptides in ATP-dependent manner.
- Prevents aggregation and mediates refolding of protein after heat shock.

**HSP70**

- They are central components of the cellular network of folding catalysts and molecular chaperones.
- They assist in different types of processes of protein folding in the cell by transient association of their substrate binding domain with short hydrophobic peptide.
- They bind and release their substrate by switching to low-affinity ATP-bound state and the high-affinity ADP-bound state.
- They form complex network of folding machines [15].

**HSP90**

- It is highly abundant chaperone.
- It plays an important role in many cellular processes, for example, cell cycle control, cell survival, and hormone and other signaling pathways.
- It is a key player in maintaining cellular homeostasis during stress.
- Has ATPase activity, whose binding and hydrolysis affects conformational dynamics of the protein.
- It has become a major therapeutic target for cancer, and its role is being explored in neurodegenerative disorders and infectious diseases [10].

**CCT: Chaperonin Containing TCP1**

- This is eukaryotic chaperonin tailless complex polypeptide 1 (TCP1) ring complex (TRiC).
- It facilitates the proper folding of many cellular proteins.

**GroEL**

- It is a bacterial chaperone.
- It binds with partially folded and misfolded proteins.
- For its functionality GroEL requires its cofactor GroES.

Differences in N-glycosylation in yeast are with high or hypermannose which is highly immunogenic. Unmodified proteins are suitable for production in yeast.

Other members of fungi are *Pichia pastoris*, *Pichia methanolica*, *Candida boidinii*, and *Pichia angusta*, which are facultative methylotrophic yeast having great potential. *Pichia pastoris* is favored as high cell densities can be obtained; protein is secreted in high concentration (1 g/l), less hypermannosylation as compared to yeast and thus less immunogenic.

However, the disadvantage is that it requires methanol to induce gene expression as transgene is under the control of the promoter of alcohol oxidase 1 (AOX1) gene. Methanol may be flammable and is toxic to cells and humans if not thoroughly removed. Because of hypermannose-type glycosylation, the fungi are also unsuitable for production of many recombinant proteins.

---

## 4.5 Production of Recombinant Protein in Insect Cell

**Insect cells:** Insect cell can be infected with baculoviruses which are double-stranded circular DNA viruses with arthropods as host. Baculovirus-mediated gene expression in insects is a method of choice and is cost-effective, giving the much higher yield of recombinant protein compared to other systems. It is possible to produce large protein resulting in production of correctly processed and biologically active protein.

A baculovirus *Autographa californica nuclear polyhedrosis virus* (AcMNPV) is used as a cloning vector for insect cell lines. In this viral polyhedron protein is used, which is required in its normal habitat and exhibits high rate of transcrip-

(continued)

tion, but is not needed in cell culture. Thus, the coding sequence of the gene is replaced with foreign DNA. The gene is transcribed under the control of powerful polyhedron promoter with high yields (~30% of total cell protein). The observed yield may be variable due to the course of virus infection and viral titer.

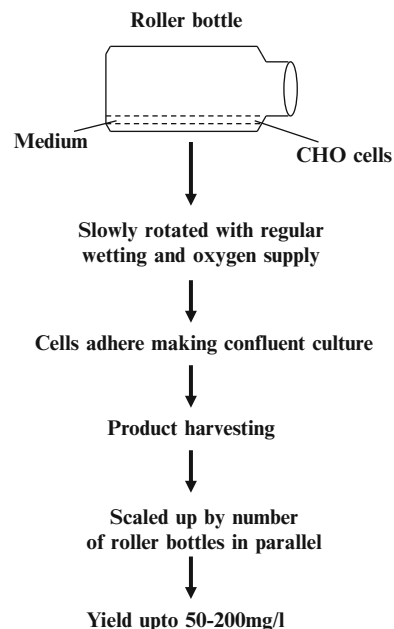
The production of recombinant protein in insect cell is time consuming (as compared to bacterial system) as cell growth is slow and the cost of medium is high. Every time fresh cells are required, viral infection is lethal for cells. It also has limitations in performing posttranslational modifications as it performs non-syalated N-linked glycosylation. All the other optimizations need to be perfect as yield depends upon the virus titer and time taken from infection to expression. Insect cells are preferred when active protein is difficult to obtain in *E. coli* system.

Genetic engineering has been used to select MIMIC™ (Invitrogen) and SfSWT-3, which are transgenic cell lines expressing all necessary enzymes to obtain humanized, complex N-linked glycosylation pattern. The system has been extensively used for structural studies as correctly folded eukaryotic proteins may be obtained in secreted form simplifying purification protocols. Some of the approved biopharmaceuticals from infected insect cell line Hi Five are Cervarix (recombinant papillomavirus C-terminal truncated major capsid protein L1 types 16 and 18, used as cancer vaccine).

Glycosylation is a problem which is encountered when insect cells are used for production of recombinant human glycosylated proteins. Lots of genetic engineering is required to produce human-like glycosylation in insect cell. Thus, the preferred system for therapeutic human protein production is mammalian system (Chinese hamster ovary cell line). Due to time and difficulty in maintaining insect cells, the mammalian cells were explored for production of recombinant protein. Mammalian cells, because of their properties of protein folding, assembly, and posttranslational modifications, have become the preferred system for protein production and are now accounting for major recombinant protein production.

## 4.6 Production of Recombinant Protein in Mammalian Cell

**Mammalian Cells:** Production of complex proteins requires extensive processing and posttranslational modifications. Mammalian cells have the advantage of performing PTMs (Fig. 4.3) correctly; they secrete recombinant protein into the medium in their natural form, thus skipping the critical steps of renaturation and refolding which sometimes leads to inactive proteins. Therefore, major therapeutic proteins (60–70%) are produced in mammalian cells primarily Chinese hamster ovary (CHO) cells and baby hamster kidney cells (BHK). CHO cells are relatively easy to manipulate and their properties favor large-scale production in them [24]. The proteins produced are safe to use in humans with no adverse reactions because of similar glycosylation pattern. Chinese hamster ovary (CHO) and baby hamster kidney (BHK) cells are the prominent producers of recombinant proteins (Fig. 4.4).



**Fig. 4.4** The figure shows the culturing of mammalian cells using roller bottles. These cells are maintained in number of roller bottles. For adherent cells, the microcarrier beads are used. The cells adhere to beads and the beads are maintained in suspension culture

In mammalian cells, genes can be expressed either transiently or stably. Obtaining stably transformed cell lines requires the usage of some selectable marker. Another major advantage of these cells is they can be grown in suspension, in serum-free (SF), protein-free, and chemically defined media. The product is safe without the risk of prions of bovine spongiform encephalopathy (BSE) from bovine serum albumin and infections of variant Creutzfeldt–Jakob disease (vCJD) from the human serum.

Presently due to virus and prions in the donor plasma or blood samples, the manufacturers are opting for plasma-/serum-free growth medium for culturing the cell lines. Cells require some of these components (albumin, transferring) for their growth. Nowadays recombinant human albumin is available like human insulin (produced in *E. coli* and yeast) and yeast-derived animal-free recombinant human transferrin is available. These support plasma-free mammalian cell culture. Else, CHO cells may be engineered to produce its own transferring or insulin-like growth factor. The therapeutic protein obtained by mammalian cell is treated and tested for the presence of any pathogenic agents or viruses as contaminant. The therapeutic products obtained during processing are treated with various agents to remove inactive virus, but the presence of asymptomatic virus poses a serious risk. Common virus inactivation technique is using solvent/detergent, which is effective against lipid-enveloped viruses such as *human immunodeficiency virus* (HIV), *hepatitis B and C virus* (HBV and HCV), *human T-cell lymphotropic virus* (HTLV), and *West Nile virus* (WNV). Non-lipid-enveloped viruses such as parvoviruses, enteroviruses, and circoviruses are resistant to inactivation via solvent detergent treatment. Human herpesvirus 8, responsible for Kaposi's sarcoma, is shown to be transmitted through blood and blood product. Following outbreaks, strict requirements were imposed on manufacturers of biologics and medical devices. Such concerns prompted manufacturers to switch to sugar-based final formulations and develop recombinant plasma-free albumin produced in yeast for usage in biopharmaceutical manufacturing.

Plasma-free manufacturing involves the elimination of plasma derivatives in every step (like cell line development, upstream processing, downstream processing, and final formulations) of the process with appropriate postproduction checks. The manufacturers shifted from the use of serum to serum-free cell culture media with animal product-free media and ultimately to protein-free, completely synthetic chemically defined media. The media consists of protein hydrolysates derived from yeast, soy, and wheat with amino acids, peptides, carbohydrates, vitamins, and essential elements, which are ultrafiltered to remove any unwanted contaminants [8].

### Case Study

The recombinant therapeutic product used for clinical applications was produced from mammalian cells. Mammalian cells were maintained in serum-/blood-/plasma-based medium; therefore, the presence of any infectious agent in the product might be deleterious if not properly removed. Infections may range from HIV, *coronavirus* (severe acute respiratory syndrome (SARS)), non-lipid-enveloped (NLE) viruses as circoviruses (Torque tenovirus (TTV) and Torque tenominivirus (TTMV)), HBV, HCV, HTLV (human T-cell lymphotropic virus), to West Nile virus. Prions that are self-replicating infectious proteins may also be present which may lead to variant Creutzfeldt–Jakob disease (vCJD).

Pathogen transmission was a major concern in the manufacture of blood-derived coagulation factor. In the early 1980s, the factor replacement products derived from plasma, which were used to treat hemophilia, were found to be contaminated with HIV, HBV, and HCV viruses. In the year 1984, up to 78% of US-based hemophilic patients were infected with HIV and 74–90% were infected with HCV. Parvoviruses, B19 (B19V) and PARV4, were present as con-

(continued)

taminant in plasma-derived factor VIII. Therefore, the production urgently required regulatory measures.

Then came first recombinant factor VIII, Advate (Baxter), in the USA in 2003. Advate was produced in CHO cells grown in serum-free and protein-free medium with ultrafiltered soybean peptides with subsequent purification by immunoaffinity chromatography. The usage of Advate helped in eliminating the risk of transmitting emerging blood pathogens.

Prions pose a serious risk, as they are highly resistant to physical/chemical inactivation.

Early stage of prion infection is almost impossible to detect in plasma donor. Iatrogenic transmission of prions has occurred in patients who received human-derived pituitary hormones as human growth hormone (hGH) and gonadotropins. CJD was transmitted to over 160 recipients of cadaveric pituitary hGH before its withdrawal. Cadaveric pituitary-derived gonadotropins for infertility were associated with iatrogenic transmission of CJD. Later on cadaveric pituitary hGH and gonadotropins have been replaced with recombinant GH (produced in microbial system) and recombinant gonadotropins (produced in CHO cell lines).

## 4.7 Using Human Cells for Protein Production

Human cell lines: Human cell line-derived Dynepoerithropoietin (erythropoietin with increased shelf life), Elaprase-irudonate-2-sulfatase (lysosomal enzyme), and Replagal-alfa-galactosidase A (lysosomal hydrolase) have been approved by the European Union (EU) or Food and Drug Administration (FDA, USA). As these products are fully glycosylated when expressed in human cell lines and used as therapeutics in human beings.

## 4.8 Transgenics for Protein Production

### 4.8.1 Transgenic Animals

The transgenic animals are successfully used for production of recombinant proteins (for details refer to Chap. 5). Protein production poses great risk in terms of safety as transmission of infectious agents, allergic responses, immune reactivity, and autoimmune responses might occur.

ATryn was the only approved (approved in 2006 by European medical agency and in 2009 by FDA) recombinant biopharmaceutical using transgenic animals. It contains human antithrombin (432AA) with 15% glycosylated moieties and is secreted into milk of transgenic goats.

Rhucin intended for acute attacks of angioedema in patients with congenital C1 inhibitor activity deficiency, obtained from transgenic rabbit, was denied approval. Transgenic plants are being explored as recombinant protein producers for research and diagnostic uses.

### 4.8.2 Transgenic Plants

Obtaining therapeutic proteins from their natural source poses threat for spread of diseases. Therefore, alternative systems for the production of therapeutic agents have their own benefits. Molecular farming in plants has been widely explored for production of recombinant pharmaceutical proteins. Their advantages are low cost, high mass production, scale-up, lack of human pathogens, and addition of eukaryotic PTMs. The first recombinant protein obtained in 1986 from tobacco plants was human growth hormone. However, sometimes plant-specific PTMs might result in adverse immune reactivity.

Production of recombinant heterologous proteins in plants is simple and is used for production of non-naturally occurring proteins as single chain Fv fragments (ScFvs). High yield of recombinant protein is the main goal of production system in transgenic plants. Therefore appropriate expression vectors and constructs are designed to achieve high yields of the engineered gene products [7, 12].

### Protein Production in Plant System

The transgene in expression construct is chimeric structure as it is surrounded by various active regulatory elements. Polyadenylation sites play an important role for the high level of expression of transgene. *Cauliflower mosaic virus* (CaMV) 35S promoter works well with dicots. It is a strong constitutive promoter that is made more active by duplicating the enhancer region. However, in monocots maize ubiquitin-1 promoter is the preferred promoter. The presence of an intron in the 5'-untranslated region (5'-UTR) enhances transcription in monocots.

For obtaining high yield of the protein, several factors may be appropriately considered:

- The incorporation of polyadenylation sites may be from CaMV 35S transcripts or the *Agrobacterium tumefaciens* nos gene or the pea ssu gene.
- The yield can be controlled by placing the gene under the control of the promoter which is active in a particular tissue or developmental stage or particular environment (e.g., rice glutelin, pea legumin).
- Usage of inducible promoter (e.g., tomato hydroxyl-3-methylglutaryl CoA reductase-2 (HMGR2)) which has mechanical gene activation system developed by Cramer (Crop Tech Corp., Virginia, USA). Transcription starts when harvested tobacco leaves are sheared during processing.
- Codon bias in the host plant may be overcome by engineering of transgene at positions, which might lead to truncation and/or misincorporation, or slowing the process.

(continued)

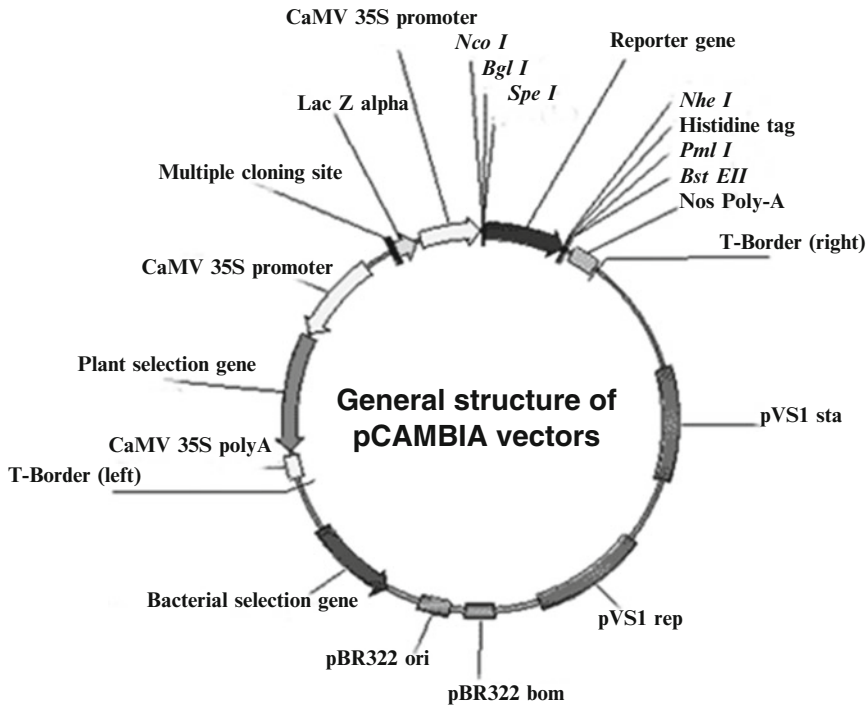
- Subcellular targeting is also very important factor which affects the process of folding, assembly, and posttranslational modification and can be efficiently achieved by inclusion of an N-terminal signal peptide.
- Position of transgene integration.
- Structure of transgene locus.
- Gene copy number.
- Presence of truncated or rearranged transgene copies.
- Affinity tags as His or the FLAG epitopes can be used to ease the process of purification; however, these modifications not only affect the primary structure but also the properties of the protein.

Figure 4.5 shows the general vector pCAMBIA (it is small in size (7–12 kb) maintained in high copy number with pVS1 replicon that imparts chloramphenicol or kanamycin resistance and high stability in *Agrobacterium*) that is used for transgene expression in the plants. The modified vector has shown success in insulin production.

Nowadays plant system is efficiently engineered to produce human growth hormone, human serum albumin, erythropoietin,  $\alpha$ -interferon, antibodies and ScFvs, toxins, subunit vaccines, and insulin [20].

## 4.9 Challenges of Production of Therapeutic Proteins

With increasing demand from the consumer, the companies are trying to increase the productivity [17]. Few challenges with large-scale production of proteins are:



**Fig. 4.5** The figure shows the structure of pCambia vector (cambia.org) used for plant transformation. The vector has CaMV35S promoter, multiple cloning site, and

reporter gene (GUS or GFP may be used). The vector can be modified to express genes for insulin (tomato) or Hep-B surface antigen (HBsAg) for recombinant therapeutics

1. Loss of expression: For high output, it is very important that gene of interest should give adequate protein production. However expression may be lost due to many factors, like, if there are structural changes in the recombinant gene or inactivation or disappearance of the gene from host cell. Other factors influencing yield may be increase in the copy number of insert, maintenance of optimum temperature, and toxicity of the expressed protein to the host. Chromosomal integration of the foreign gene might overcome the problem of expression stability, but in plasmid-based system, high copy number leads to increased yield.
2. Posttranslational processing: Protein folding requires foldases (accelerates protein folding) and chaperones (prevents protein formation of nonnative insoluble folding intermediates). Glycosylation is complex PTM requiring consecutive steps and enzymes. Glycosylation is important as it determines protein stability, solubility, antigenicity, folding, localization,

biological activity, and circulation half-life. Getting correctly glycosylated and folded protein is required for therapeutic usage. In prokaryotes, with the discovery of N-glycosylation system in *Campylobacter jejuni*, several other systems of O-glycosylation were unraveled in both pathogenic and symbiotic bacteria. The production of recombinant proteins is commonly done by using *E. coli*, yeast, or cell lines derived from insects (SF9), mice (SP2/0), or CHO, but obtaining fully human PTMs is a challenging task. The commonly used mammalian cell lines of rodent origin (such as SP2/0, CHO, or BHK) are able to perform humanlike glycosylation. However, some human components are missing (such as the 2,6-linked sialylation), and a number of non-human components have been found to significantly increase (terminal galactose linked to another galactose or terminal sialic acids) which increases the high risk of immunogenic reactions. For this reason, human cell lines



providing a human glycosylation pattern have increased attention, and the efforts are being made for the development of novel glyco-engineered cell lines for production of fully glycosylated protein therapeutic.

3. Overexpression of therapeutic protein might result in the formation of inclusion bodies in prokaryotic system. Rapid intracellular protein accumulation and expression of large proteins increases the probability of aggregation. Aggregation protects proteins from proteolysis and can facilitate protein recovery. When the expressed protein is toxic to the host, the presence of protein in the inclusion body tends to protect it.

**Precautions** Recombinant protein production requires some precaution resulting in a loss of yield and/or product:

1. Contamination: At any stage, contamination might occur from any source. It poses big challenge and may be adverse when contaminated material is put for human usage.
2. Immune response: For the therapeutic agent, the body can mount the immune reactions which lead to deposition of immune complexes in various tissues, and condition of anaphylactic shock might occur (e.g., when some essential agent is lost since birth like factor VIII, then the patient might raise antibody response against the treatment). This also occurs when antibodies are used as therapeutic agents in the treatment of variety of cancers.
3. Protein aggregation: Any nonfunctional condition might result in aggregation or loss of activity of recombinant protein.
4. Posttranslational modifications and folding: After purification, the proper modifications and refolding are required for therapeutics.
5. Disulfide bond formation: It stabilizes protein structure; thus, strategies for specific extracellular excretion pathway or overexpression of chaperones is required for optimum production.
6. Degradation: Sometimes proteins, which are active in host cells like proteases or protein-modifying chemicals, might degrade the recombinant protein (e.g., PEG interferon is modified

to have polyethylene glycol with prolonged presence and reduction in enzymatic degradation and renal clearance, thus extending its presence with lesser immunogenicity) [17].

---

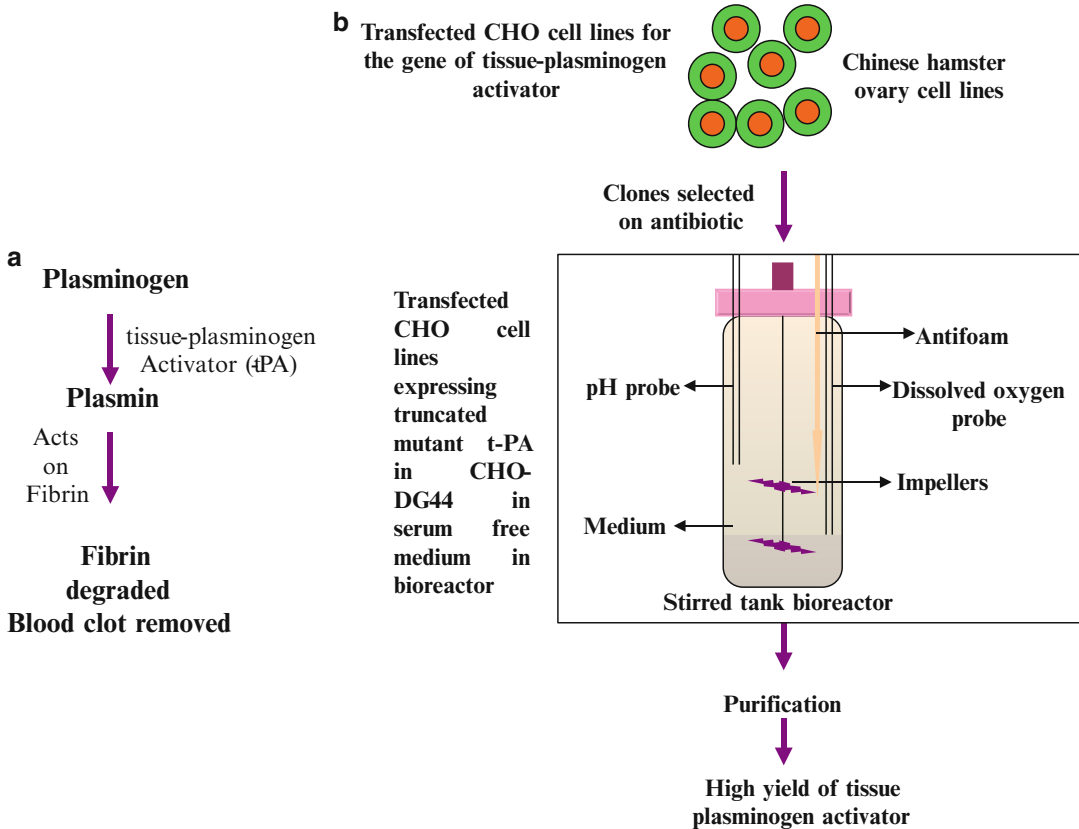
## 4.10 Some Important Biopharmaceuticals

### 4.10.1 Tissue Plasminogen Activator (tPA)

Ischemic stroke and myocardial infarction are one of the leading causes of cardiovascular morbidity and mortality in the world. Important thrombolytic agents are urokinase (obtained from urine), tissue plasminogen activator (tPA), and streptokinase (obtained from bacteria). Among these, t-PA is largely used commercially. Plasminogen activators are serine proteases, which are responsible for conversion of inactive proenzyme plasminogen (Plg-a single chain glycoprotein) to serine protease called plasmin (Plm). The plasmin degrades the network of fibrin of the blood clots (Fig. 4.6a). There are two immunologically unrelated groups of plasminogen activators, the 55 kDa urokinase-type-PA (u-PA) and 72 kDa tissue-type PA (t-PA) (EC 3.4.21.68). The t-PA is the physiological vascular activator consisting of single polypeptide chain of 72 kDa consisting of 527 amino acids. It shows strong activity in the presence of fibrin [23].

The potential inhibitors of the thrombolytic cascade are type I plasminogen activator inhibitor (PAI-1 or serpin E1) and PAI-2 (secreted by placenta and present in significant amount during pregnancy). They act by competing with t-PA for binding sites on fibrin thus preventing the fibrinolytic cascade. PAI-1 complexes with t-PA for binding to fibrin. Thus, truncated t-PA in which the residues responsible for interacting with PAI-1 (296–299) are replaced with four alanine amino acids and three domains (finger, epidermal growth factor (EGF), and kringle 1 domains) are deleted, and chimeric tetrapeptide Gly-His-Arg-Pro (GHRP) with high affinity to fibrin was added. Reteplase is the deletion mutant with a





**Fig. 4.6** The figure shows the function and production of tissue plasminogen activator. (a) This figure shows the function of t-PA that it helps in degradation of blood clot by degrading fibrin by activation of plasminogen into

plasmin. (b) This figure shows the production of truncated t-PA in CHO DG44 cell line. The mutated form of t-PA is resistant to inhibition by plasminogen activator inhibitor and has better effectiveness in clinical usage

prolonged half-life, in which the finger, EGF, and kringle 1 domains of the full-length molecule are all deleted; thus, it is not inhibited.

Deficiency of PAI leads to over fibrinolysis and hemorrhagic diathesis (like deficiency of clotting factors). Tiplaxtinin is the inhibitor and is used for remodeling of blood vessels [2].

#### 4.10.1.1 Production

Plasminogen activator has great clinical relevance for the management of stroke and myocardial infarctions. For production of t-PA, *E. coli* and yeast system did not work properly due to lack of posttranslational modification and over glycosylation, respectively.

A novel truncated form of t-PA with an improved fibrin affinity and an increased resis-

tance to PAI-1 was expressed in a CHO DG44 expression system. Therapeutic protein was produced in stably transfected CHO DG44 cell lines. These cell lines were maintained in serum-free medium, with glutamine, hypoxanthine, and thymine in stirred tank bioreactor. The cells were grown at 37 °C, 140 rpm with 5% CO<sub>2</sub> and 85% humidity. The protein was then purified, with higher yield (Fig. 4.6b).

#### 4.10.2 Factor VIII

Factor VIII is one of the important factors of all blood clotting factors. Deficiency of factor VIII causes bleeding disorder called hemophilia A. The hemophilia may be mild to severe depend-

ing upon factor VIII concentration in the body. In moderate and severe factor VIII deficiency, there can be spontaneous bleeding episodes in the joints. Hemophilia A affects 1 in 5,000–10,000 males. Replacement therapy is the treatment option for hemophiliacs either with human plasma-derived factor VIII (pdFVIII) or recombinant FVIII (rFVIII) [21].

Transfection of HEK 293 cell cultures in serum-free suspension is being tried for optimal yield. Recombinant factor VIII (rFVIII) is produced by culturing mammalian cells as baby hamster kidney (BHK) or Chinese hamster ovary cells (CHO), using large-scale bioreactors. Standardizations are done to maximize yields.

### 4.10.3 Insulin

Insulin is a peptide hormone consisting of 51 amino acids. It is secreted by  $\beta$  cells of islets of Langerhans of pancreas. The hormone is responsible for maintaining normal blood glucose level in blood. Insulin is stored in the form of proinsulin which contains two polypeptide chains, A and B, and is connected with a third peptide C-chain), which before secretion is cleaved with production of insulin and C-chain. The cleavage results in the removal of C-chain, and the A (21 amino acids) and B chain (30 amino acids) are linked by disulfide linkage to form mature insulin.

In the beginning, efforts were made to isolate mRNA for pre- and proinsulin from rat islets of Langerhans of pancreas and to synthesize cDNA. Thereafter, it was inserted into a plasmid. The recombinant plasmids were transferred into the *E. coli* cells, which secreted proinsulin [4, 5, 6, 19].

Scientists have chemically synthesized DNA sequences for two chains, A and B, of insulin and separately inserted into two pBR322 plasmids by the side of  $\beta$ -galactosidase gene. The recombinant plasmids were separately transferred into *E. coli* cells which secreted fused  $\beta$ -galactosidase-A chain and  $\beta$ -galactosidase-B chain separately. These chains were isolated in pure form by detaching from  $\beta$ -galactosidase with yields of

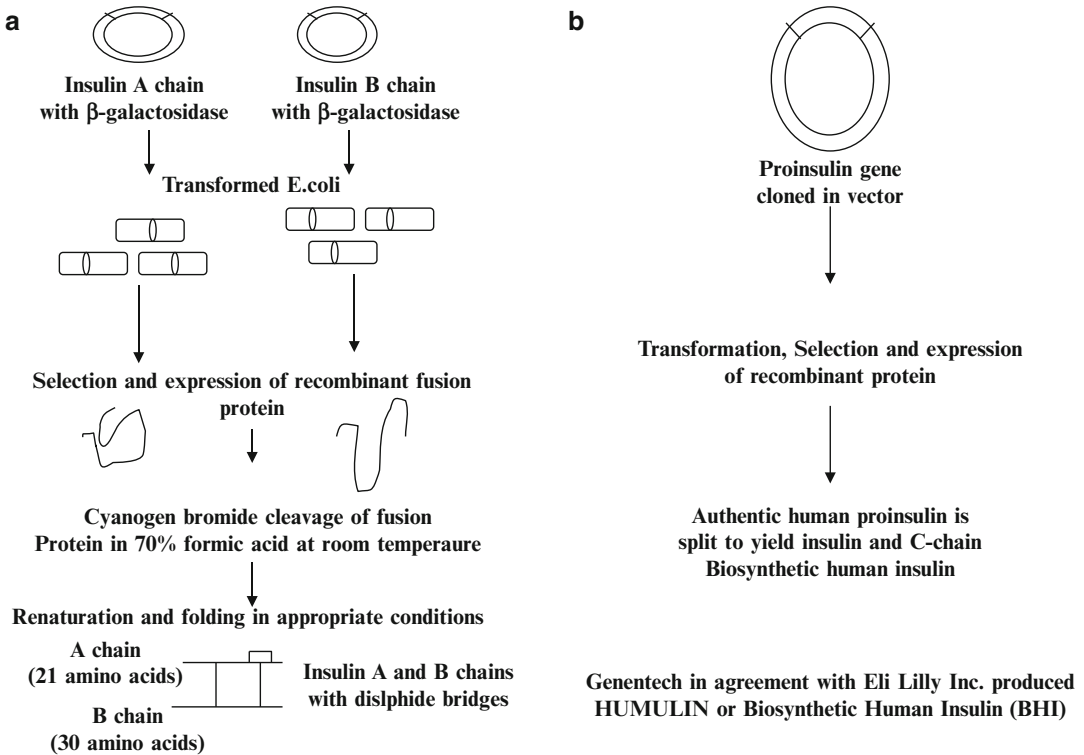
about 10 mg/24 g of healthy and transformed cells. Production of recombinant insulin is shown in (Fig. 4.7a, b).

### 4.10.4 Human Growth Hormone (hGH)

Growth hormone is produced by the anterior lobe of the pituitary gland and is released in multiple pulses. The hGH is encoded by GHN gene cluster (an array of five closely related genes), which is localized on chromosome 17. It belongs to diverse gene family that has evolved by gene duplication events and has lots of structural similarity and some common functions. Of the various forms, the predominant form of hGH is 22 kDa protein of 191 amino acids with two disulfide bonds.

GH does not control the functions directly, but acts on certain hormones or somatomedins for its activity, for example, insulin-like growth factor 1 (IGF-1). It has a wide spectrum of roles to play as promotion of long bone growth, promotion of normal sex organ development through puberty, regulation of metabolism, stimulation of tissue growth and repair, anabolic/anti-catabolic effect via improved nitrogen retention, modulation of bone mineral density and metabolism throughout life, proliferation of some cell types of the immune system, appetite stimulation, and breakdown of fat (lipolysis). In clinical conditions, the therapy of growth hormone is given in the treatment of dwarfism, bone fractures, skin burns, bleeding ulcers, and AIDS.

Recombinant human growth hormone (rhGH) is 22 kDa consisting of long chain of amino acids. It is used in deficiency disorders of growth hormone. In children, it is used for growth abnormality as short stature and is used in chronic renal insufficiency. The therapy of growth hormone is also approved for adult growth hormone deficiency. GH is one of the most widely used hormones with the estimated market of more than 1.7 billion USD. Long experience in its administration has proven the therapy as safe and effective in various conditions of growth abnormality.



**Fig. 4.7** The figure shows the production of insulin. (a) The A and B chains of insulin are cloned separately in the vector. The transformation is done and after selection, the recombinant protein is obtained for both A- and B-chain genes. Cleavage by using cyanogen bromide removes the

bacterial protein. The protein is renatured and provided suitable conditions for folding. (b) This figure shows the production of insulin from proinsulin gene which yields insulin and C-peptide. This is preferred method for production of biosynthetic human insulin

Earlier pituitary-derived hGH was used but later on it was prohibited when found associated with Creutzfeldt–Jakob disease. Because of recombinant DNA technology, safe and abundant recombinant hGH was produced in various heterologous systems. As the non-glycosylated human growth hormone was biologically active, thus, the preferred system for its production is *E. coli*, which allows its rapid and economical production in large amounts [14].

Recombinant hGH (rhGH) is now used to treat:

- GH-deficient (GHD) short-stature children.
- Acceleration of wound healing.
- Increase in insulin-like growth factor (IGF)-1 levels.

- rhGH increases IGF-1, osteocalcin, type I procollagen pro-peptide (PICP), and bone density, when administered to children with GHD.

For optimal productivity, strong inducible promoters are preferred as IPL, IPR, trc, and T7 in *E. coli*. They are advantageous as they drive overproduction of recombinant proteins. Apart from *E. coli*, human somatotropin (hST) expression was tried in a biologically active, disulfide-bonded form in tobacco chloroplasts. The hormone is used for the treatment of hypopituitary dwarfism in children; additional indications are in treatment of Turner syndrome, chronic renal failure, HIV wasting syndrome, and possibly treatment of the elderly. Growth hormone deficiency in human occurs both in children and adults [18, 22].

### 4.10.5 Interferons

Interferons are a group of proteins which are secreted in response to viral infections. Resistance imparted by INFs is short lived and does not last forever. They are family of naturally occurring proteins that are made and secreted by all the cells (INF- $\alpha$  and INF- $\beta$ ) and lymphocytes (INF- $\gamma$ ). All these modulate the response of cells and immune system to viruses, bacteria, and cancer.

Interferons are produced by either an established cell line (lymphoblastoid) or fresh cells isolated from blood. The production involves induction with virus and priming (incubation with some interferon) with interferon, resulting in better yield. The affinity chromatography with monoclonal antibodies packed in the column has helped in purification of interferons. But before the clinical usage, the removal or inactivation of virus is very important. The interferon therapy is used for cancers and viral infections (INF- $\alpha$ ), multiple sclerosis (INF- $\beta$ ), and chronic granulomatous disease (INF- $\gamma$ ). Multiferon is natural interferon- $\alpha$ , which consists of several subtypes. In some of the cancers like Merkel cell carcinoma, type I interferons (multiferon, which is a mix of various INF- $\alpha$  subtypes and INF- $\beta$ ) are highly effective. In Chap. 11, interferons are mentioned in protein therapeutics. Interferon is marketed as Roferon-A, Infergen, Intron A, and so on.

### 4.10.6 Erythropoietin

Hematopoietic growth factors consist of cytokines and protein hormones produced by the body which govern the production and maturation of the various cells produced during the process of hematopoiesis in the bone marrow from hematopoietic stem cell. The precursor cells in the presence of a particular growth factor differentiate and become a specialized kind of cell as monocyte, macrophage, lymphocytes, or red blood cell.

One of the important growth factor is erythropoietin which is a protein hormone produced by a specific type of cells in the kidney. In the presence

of erythropoietin, progenitor cells are stimulated in the bone marrow to form mature erythrocytes (red blood cells). Thus the patients with chronic kidney disease are unable to maintain adequate amount of erythropoietin for normal development of erythrocytes in blood, resulting in low numbers of red blood cells and subsequent anemia. These patients either require blood transfusion or erythropoietin from outside. As supply is limited from the natural source that is kidney cells, thus a recombinant human erythropoietin EPOGEN® which is Amgen's trade name for epoetin alfa is marketed for anemic condition involving erythropoietin. The human gene encoding erythropoietin was cloned into the Chinese hamster ovary cell line for production of the human protein. This cell line continues to be used today for the production of EPOGEN®.

The half-life of erythropoietin can be increased by incorporating the glycosylation of the protein growth factor. Thus, Darbepoetin- $\alpha$  is an analog which is engineered for two extra amino acids which are substrates for glycosylation. Thus, production is done in CHO cell lines; the product has five N-linked sugar chains and has almost three times longer life than erythropoietin.

### 4.10.7 Platelet-Derived Growth Factor (PDGF)

PDGF regulates cell growth and division and plays a significant role in blood vessel formation (angiogenesis), the growth of blood vessels from already existing blood vessel in tissue, and may act in autocrine and paracrine stimulation of cell growth in vivo. PDGF plays the role in development, cell proliferation, cell migration, and angiogenesis and has been linked to atherosclerosis, fibrosis, and malignant diseases.

PDGF has five different isoforms: PDGFA, PDGFB, PDGFC, PDGFD, and AB heterodimer. PDGF-A and PDGF-B have 60% similarity in amino acid sequence, but experiments suggest different biological functions for the two chains and different locations of these under different transcriptional controls. PGDF-AA is released in the medium, and PDGF-BB are insufficiently

secreted and remain attached to the plasma membrane, and of these PDGF-BB and PDGF-AB are strong mitogens and are probably responsible for biological roles of PDGF. PDGF receptor (PDGFR) is receptor tyrosine kinase (RTK) (alpha and beta type). Upon activation by PDGF, these receptors dimerize leading to autophosphorylation of several sites on their cytosolic domain. PDGF being a mitogen promotes the proliferation of fibroblasts and smooth muscle cells *in vitro*.

PDGF shows considerable heterogeneity with sizes of 27–31 kDa; however, purified PDGF is cationic protein of 30 kDa. Recombinant human platelet-derived growth factor (rh-PDGF) was the first recombinant protein to be approved by the US Food and Drug Administration for treatment of chronic foot ulcers in diabetic patients (Regranex, Ethicon Inc. Somerville, NJ). It has the potential for use in bone regeneration and increasing bone density in long bones and spine. PDGF is commercially produced by using *E. coli* and mammalian cells.

#### 4.10.8 Epidermal Growth Factor (EGF)

Human EGF protein has 53AA and three intracellular disulfide bonds and plays an important role in the regulation of cell growth and proliferation. It shows strong sequential and functional homology with human type- alpha transforming growth factor (hTGF alpha), which is a competitor for EGF receptor site. EGF acts by binding with high affinity to EGFR on cell surface and stimulates the intrinsic protein tyrosine kinase activity. EGF has many biological activities. Initial observations were centered around their proliferative effects on fibroblasts, keratinocytes, and epithelial cells. EGF modulates luteinizing hormone and thyroid hormone. EGF is produced commercially by engineered *E. coli*. The other systems are also being explored for optimum EGF production [9].

#### 4.10.9 Fibroblast Growth Factor (FGF)

FGFs are commonly mitogens with multifunctional proteins with a wide variety of regulatory, morphological, and endocrine effects. There are 18 mammalian FGFs (1–10 and 16–23) which affect growth and functions of a wide variety of mesenchymal, endocrine, and nerve cells. The functions of FGFs in developmental processes include mesoderm induction, anteroposterior patterning, limb formation, neural tube induction, and brain development and in mature tissue/systems angiogenesis, keratinocyte organization, and wound healing processes. FGF is very important during normal development of both vertebrates and invertebrates, and any irregularities in their function lead to a range of developmental defects [1].

FGF1 and FGF2 show strong angiogenic properties with the promotion of endothelial cell proliferation and the physical organization of endothelial cells into tubelike structures and the growth of new blood vessels from the preexisting vasculature. FGF7 and FGF10 (also known as keratinocyte growth factors (KGF) and KGF2, respectively) stimulate the repair of injured skin and mucosal tissues by stimulating the proliferation, migration, and differentiation of epithelial cells, and they have direct chemotactic effects on tissue remodeling. Most FGFs are secreted proteins that bind heparan sulfates and can therefore be caught up in the extracellular matrix of tissues that contain heparan sulfate proteoglycans. This allows them to act locally in a paracrine fashion. However, the FGF19 subfamily (including FGF19, FGF21, and FGF23) which binds less tightly to heparan sulfates can act in an endocrine fashion on far away tissues, such as the intestine, liver, kidney, adipose, and bone. For example, FGF19 is produced by intestinal cells but acts on FGFR4-expressing liver cells to downregulate key genes in the bile acid synthase pathway; FGF23 is produced by the bone but acts on FGFR1-expressing kidney cells to regulate the synthesis of vitamin D and in turn affect calcium homeostasis. FGF may be synthesized using *E. coli* as host system.

#### 4.10.9.1 Therapeutic Potential of FGFRs

FGF is involved in stimulating collateral vascularization and recovery from ischemia as well as enhancing wound healing, nerve regeneration, and repair of cartilage and has been alternately referred to as “pluripotent” (capable of developing into more than one cell type or tissue) growth factors and as “promiscuous” (biochemistry and pharmacology concept of how a variety of molecules can bind to and elicit a response from single receptor) growth factors due to its multiple actions on multiple cell types. The FGFs and small-molecule FGF receptor kinase inhibitor are used in the treatment of cancer and cardiovascular disease and have potential in the treatment of metabolic syndrome and hypophosphatemic diseases:

- Receptor tyrosine kinase inhibitor (Sunitinib) is approved for indications in renal cell carcinoma and gastrointestinal stromal tumors.
- Small FGFR inhibitors, SU5402, PD173074, and nordihydroguaiaretic acid, are effective in multiple myeloma cell lines.
- PD173074 can induce cell cycle arrest in endometrial cancer cells with mutated FGFR2.
- Antibody against FGFR3 has been shown to effectively cause apoptosis in mouse models of multiple myeloma and bladder cancer.

Thus FGFR inhibition can be very effective in the treatment of cancer.

#### 4.10.10 Nerve Growth Factor (NGF)

The nerve growth factor (NGF) is an important member of the family of neurotrophins. Five protein nerve growth factors of the neurotrophin family are important. They regulate the development of the nervous system and play an important role in maintaining the structure, plasticity, and repair of the adult nervous system. All the neurotrophins are basic proteins of about 120AA, share 50% sequence homology, and are highly conserved in mammalian species.

Nerve growth factor (NGF) is a small secreted protein which induces the differentiation and survival of particular target neurons (nerve cells). Little is known of the biological action of neurotrophin apart from NGF. The nucleotide sequence of cDNA predicts that NGF is synthesized as pre-pro-NGF. Upon removal of hydrophobic signal, either 34 kDa or 27 kDa pro-NGF is generated depending on the size of the transcript. However, processing of the precursors in different tissues is not well understood. They are essential for normal development, growth, and differentiation of the sympathetic and sensory neurons and are also essential to maintain the normal function of these cells in adults. Thus it is important for maturation and survival of neurons and prevents degeneration of adult neurons. Apart from its important role in the nervous system, it has been shown to possess protective action of human pressure ulcer, corneal ulcer, and glaucoma. Reduced sensation may be observed in leprosy, wound healing, nerve injury, and diabetes. NGF may help to regulate the sensory fiber sensitivity and function directly or indirectly by stimulating other effectors. Administration of recombinant NGF may improve sensation and pain [13].

Cholinergic neurons of the basal forebrain show receptors for NGF; specific mRNAs for various NGFs have been identified in different areas of the brain. Cholinergic neuron loss is a cardinal feature of Alzheimer's disease. Nerve growth factor (NGF) stimulates cholinergic function, improves memory, and prevents cholinergic degeneration in animal models of injury, amyloid overexpression, and aging. NGF acts on intracellular calcium through tyrosine kinase receptor mechanism. Nerve growth factor enhances early regeneration of severed axons and is also important in maintaining the biochemical and morphological phenotype of mature basal forebrain cholinergic neurons (BFCNLs) after lesions or injury of the central nervous system (CNS). Thus, NGF may provide therapeutic option for preventing death of cholinergic neurons and other clinical conditions and is produced using *E. coli*.

#### 4.10.11 Transforming Growth Factor Alpha (TGF- $\alpha$ )

TGF- $\alpha$  exerts its function in an autocrine and endocrine fashion for various cell types of ectodermal origin, including most epithelial cells. It is normally a transmembrane protein and functions in cell communication through its ability to activate a receptor tyrosine kinase. Ectodomain of TGF- $\alpha$  is cleaved in a highly regulated manner, releasing soluble TGF- $\alpha$  which activates paracrine signaling. The receptor is EGF/TGF alpha receptor; therefore, the focus is on understanding the important roles of TGF- $\alpha$  and EGF receptor signaling in carcinoma development.

#### 4.10.12 Transforming Growth Factor Beta (TGF- $\beta$ )

TGF- $\beta$  is a large group of related proteins. Some of its family members include bone morphogenetic protein (BMPs), growth, and differentiation factors (GDF). It affects tissue remodeling, wound repair, hematopoiesis, morphogenesis, embryonic development, adult stem cell differentiation, immune regulation (is switch factor for IgA), and inflammation. It exists as multiple forms as TGF- $\beta$ 1, TGF- $\beta$ 2, and TGF- $\beta$ 3. It acts through transmembrane serine/threonine receptor kinase leading to the activation of Smads. It acts as tumor suppressor during cancer initiation but promoter during tumor progression. It has a role in the control of embryonic development, cellular differentiation, hormone secretion, and immune function. Its role as mesenchymal differentiation factor, with focus on the muscle, fat, and bone cell, might provide insights into its deregulation in skeletal and developmental diseases and is the area of active research. It is produced using CHO cell lines.

### 4.11 Future Prospects

There is huge potential for future therapies using proteins as therapeutic agents. The recombinant proteins are not only beneficial, but the researchers can further engineer them to improve their

activity and prolonged stay in the body, for example, engineering of monoclonal antibodies to have toxin or radioisotope or generation of bi-specific antibody. Still the technology is struggling hard to make the diseases completely a text of books and having a society free of diseases and the pathogens.

### 4.12 Chapter End Summary

- The production of proteins for therapeutic purpose is very important not only because of their specific action with minimum side effects but also due to their unique form and functions.
- Nowadays therapeutics (pancreatic enzymes from hog and pig pancreas or a-1-proteinase inhibitor from pooled human plasma) are not extracted from their native sources (from humans, animals) due to risk of transmission of pathogens. Other problems faced for extraction from native sources were the scarce availability of animal tissue for production, high cost of purification with less yield, and immunological reactions in the recipients.
- Majority of the biological therapeutic agents are produced by using various advance tools and technologies as cloning, selection, purification, and stability monitoring. It is also very important to monitor risks and side effects of therapeutics (safety analysis) obtained from a wide range of cells.
- The various biological systems are available for production of recombinant protein as bacteria, fungal system (yeast), insect cell, mammalian cell, human cell, and transgenic plants and animals.
- The system of choice is dependent upon the total cost of production and obtaining fully functional and appropriately modified (PTMs) (glycosylation, phosphorylation, or properly folded) protein. All these are essential for the biological activity of the protein.
- The bacterial system is the simplest with short cell division times, easy maintenance, easy to modify, and cost-effective production. The bacteria have limitation in production of



large-sized protein and are unable to perform posttranslational modification as glycosylation. The glycosylation is very important as it affects the activity, half-life, and immunogenicity of the recombinant protein.

- Due to limitations, the other host systems were used for production which could give better product with minimal immune reactions and are cost-effective. Fungal and insect systems are utilized for protein production but sometimes result in inappropriate glycosylation. Improper glycosylation may result in nonfunctional and highly immunogenic protein. Mammalian cells are the most preferred system for production of these proteins. CHO system is the system of choice for therapeutic protein production. It accounts for nearly 70% production of all products available in the market.
- The protein products are having important therapeutic role. They are approved for marketing and use by various government agencies like the Food and Drug Administration (FDA) and European Union (EU). The approved recombinant biotechnology medicines include replacement products, monoclonal antibodies, interferons, vaccines, hormones, modified natural products, and many others. Many of them are in usage and helping to alleviate the symptoms or cause of the diseases.

### Multiple Choice Questions

1. The function of proteins in the body is:
  - (a) Initiation of transcription
  - (b) Mediators of metabolic processes
  - (c) Acts as enzymes
  - (d) All of the above
2. The purpose of recombinant DNA technology is:
  - (a) Genes can be cloned in *E. coli*
  - (b) The product is cost-effective
  - (c) Mammalian cells may be used for production
  - (d) None of the above
3. The advantages of using *E. coli* as host for production of recombinant protein are:
  - (a) Easy scale-up and simple media requirement
  - (b) Can easily perform splicing of foreign DNA
  - (c) Produces processed and properly modified protein
  - (d) All of the above
4. The limitation of *E. coli* which poses a problem for recombinant protein production is:
  - (a) Codon biasing
  - (b) Splicing
  - (c) Posttranslational modification
  - (d) None of these
5. The fungal being eukaryote offers tremendous advantages in recombinant protein production but it does:
  - (a) Proper protein folding
  - (b) Posttranslational modification
  - (c) Hypermannosylation
  - (d) Easy scale-up
6. Insect cells may be infected with which viral vector for production of recombinant protein?
  - (a) Retrovirus
  - (b) *Coronavirus*
  - (c) Baculovirus
  - (d) Parvovirus
7. Most commonly used mammalian cell line is:
  - (a) BHK
  - (b) CHO
  - (c) HeLa
  - (d) All of these
8. Why is there a need of serum- or plasma-free medium in the production of therapeutic protein?
  - (a) As their cost is high, thus the cost of product is high
  - (b) Risk of transmission of unknown pathogens leading to diseases
  - (c) Variation in each lot results in different yield
  - (d) None of these
9. The first recombinant protein released in the market was:
  - (a) Activase
  - (b) Humulin
  - (c) Trastuzumab
  - (d) All of these
10. Growth factor inhibitors are finding place in therapeutics because:
  - (a) Growth factors are harmful
  - (b) Growth factor deactivation leads to healthy body

- (c) Growth factors are required for progression of tumors  
 (d) All of the above
11. Darbepoetin is used for the treatment of:
- (a) Short stature  
 (b) Anemia  
 (c) Blood clotting  
 (d) None of these

### Answers

1. (d); 2. (b); 3. (a); 4. (c); 5. (c); 6. (c); 7. (b);  
 8. (b); 9. (b); 10. (c); 11. (b)

### Review Questions

- Q1. What are the advantages and disadvantages of using *E. coli* as host for production of recombinant proteins?
- Q2. What is the role of codon biasing in expression of foreign gene?
- Q3. What are the posttranslational modifications?
- Q4. Write a note on fungal host for production of proteins.
- Q5. When are insect cells preferred for the recombinant protein production?
- Q6. Write a note on (a) insulin, (b) factor VIII, (c) GH, and (d) tissue plasminogen activator.
- Q7. Write the trade names of insulin, tPA, and GH.

### References

1. Beenken A, Mohammadi M (2009) The FGF family: biology, pathophysiology and therapy Nature reviews. Drug Discov 8:235–253
2. Fatemeh D, Barkhordari F, Alebouyeh M, Adeli A, Mahboudi F (2011) Combined TGE-SGE expression of novel PAI-1-resistant t-PA in CHO DG44 cells using orbitally shaking disposable bioreactors. J Microbiol Biotechnol 21:1299–1305
3. Ferrer MN, Domingo EJ, Corchero JL, Vazquez E, Villaverde A (2009) Microbial factories for recombinant pharmaceuticals. Microb Cell Fact 8:17
4. Frank BH, Chance RE (1985) The preparation and characterization of human insulin of recombinant DNA origin. In: Joyeaux A, Leygue G, Morre M, Roncucci R, Schmelck PH (eds) Therapeutic agents produced by genetic engineering “Quo Vadis”. Toulouse-Labege, Symposium Sanofi Group, Paris MEDSI, pp 137–46
5. Galloway JA (1988) Chemistry and clinical use of insulin. In: Galloway JA, Potwin JH, Shuman CR (eds) Diabetes Mellitus, 9th edn. Lilly, Indianapolis, pp 106–137
6. Galloway JA, Hooper SA, Spradlin CT, Howey DC, Frank BH, Bowsher RR, Anderson JH (1992) Biosynthetic human proinsulin: review of chemistry, in vitro and in vivo receptor binding, animal and human pharmacology studies, and clinical trial experience. Diabetes Care 15:666–692
7. Goldstein DA, Thomas JA (2004) Biopharmaceuticals derived from genetically modified plants. Q J Med 97:705–716
8. Grillberger L, Kreil TR, Nasr S, Reiter M (2009) Emerging trends in plasma-free manufacturing of recombinant protein therapeutics expressed in mammalian cells. Biotechnol J 4:186–201
9. Guglietta A, Sullivan PB (1995) Clinical applications of epidermal growth factor. Eur J Gastroenterol Hepatol 7:945–950
10. Jackson SE (2013) Hsp90: structure and function. Top Curr Chem 328:155–240
11. Kamionka M (2011) Engineering of therapeutic proteins production in *Escherichia coli*. Curr Pharm Biotechnol 12:268–274
12. Ma JK, Drake PMW, Christou P (2003) The production of recombinant pharmaceutical proteins in plants. Nat Rev Genet 4:794–805
13. Manni L, Rocco ML, Bianchi P, Soligo M, Guaragna M, Barbaro SP, Aloe L (2013) Nerve growth factor: basic studies and possible therapeutic applications. Growth Factors 31:115–122
14. Martínez JAA, Saldaña HAB (2012) Genetic engineering and biotechnology of growth hormones, genetic engineering – basics, new applications and responsibilities, Prof. Hugo A. Barrera-Saldaña (Ed.), ISBN: 978-953-307-790-1, InTech, Available from: <http://www.intechopen.com/books/genetic-engineering-basics-new-applications-and-responsibilities/geneticengineering-and-biotechnology-of-growth-hormones>
15. Mayer MP, Bukau B (2005) Hsp70 chaperones: cellular functions and molecular mechanism. Cell Mol Life Sci 62:670–684
16. Miralles (2009) Microbial factories for recombinant pharmaceuticals. Microb Cell Fact 8:17
17. Palomares LA, Mondaca SE, Ramirez OT (2004) Production of recombinant proteins. In: Balbas P, Lorence A (eds) From methods in molecular biology. Humana Press, Totowa 267:15–51.
18. Rezaei M, Esfahani SHZ (2012) Optimization of production of recombinant human growth hormone in *Escherichia coli*. J Res Med Sci 17:681–685

19. Riggs AD (1981) Bacterial production of human insulin. *Diabetes Care* 4:65–68
20. Soltanmohammadi B et al (2014) Cloning, transformation and expression of proinsulin gene in tomato (*Lycopersicon esculentum* Mill). *Jun J Nat Pharma Prod* 9:9–15
21. Swiech K, Kamen A et al (2011) Transient transfection of serum-free suspension HEK 293 cell culture for efficient production of human rFVIII. *BMC Biotechnol* 11:11–114
22. Tabandeh F, Shojaosadati SA, Zomorodipour A, Khodabandeh M, Sanati MH, Yakhchali B (2004) Heat-induced production of human growth hormone by high cell density cultivation of recombinant *Escherichia coli*. *Biotechnol Lett* 26:245–250
23. Tor NY, Fredrik E, Lund B (1984) The structure of the human tissue-type plasminogen activator gene: correlation of intron and exon structures to functional and structural domains. *Proc Natl Acad Sci U S A* 81:5355–5359
24. Wurm FM (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotechnol* 22:1393–1398

### Some Selected Resources

- [www.ibclifesciences.com](http://www.ibclifesciences.com)
- [www.link.springer.com](http://www.link.springer.com)
- [www.ncbi.nlm.nih.gov/pubmed](http://www.ncbi.nlm.nih.gov/pubmed)
- [www.qjmed.oxfordjournals.org](http://www.qjmed.oxfordjournals.org)
- [www.sciencedaily.com](http://www.sciencedaily.com)