Chapter 1: Recombinant Protein Expression and Purification

Why Produce Proteins Recombinantly?

To be used for research, industrial or pharmaceutical purposes, proteins need to be purified in large quantities. Some proteins, like casein, which makes up 20% of the protein content in milk, can easily be extracted from a readily available source in large quantities. However, most proteins are not naturally produced in a form and in amounts that allow easy purification. The techniques of genetic engineering overcome the limitations of naturally produced proteins by making cells synthesize specific proteins in amounts which can be purified for use in fundamental research or for industrial and therapeutic applications.

Biogen, one of the first companies to develop recombinant proteins, is using genetic engineering to produce human interferon beta-1a in Chinese hamster ovary (CHO) cells and is sold under the tradename Avonex. A similar form of recombinant human interferon, interferon beta-1b, is expressed in *E. coli* and sold by Bayer under the drug name of Betaseron. (An interferon is an immune protein produced in response to a virus, bacteria, parasite, or tumor cell.) Both recombinant human interferon beta-1a and 1b have been developed, tested, and brought to the market to help slow the progression of multiple sclerosis. Without recombinant production of these proteins in CHO cells or in *E. coli*, there would not be an easy way to obtain this protein for therapeutic usage.

Table 1.1 Human proteins produced by genetic engineering. Human proteins produced via genetic engineering and the disease or disorder they are used to treat.

Protein	**Used in the treatment of**
Insulin	Diabetes
Somatostatin	Growth disorders
Somatrotropin	Growth disorders
Factor VIII	Hemophilia
Factor IX	Christmas disease
Interferon-alpha	Leukemia and other cancers, MS
Interferon-beta	Cancer, AIDS, MS
Interferon-gamma	Cancers, rheumatoid arthritis
Interleukins	Cancers, immune disorders
Granulocyte colony stimulating factor	Cancers
Tumor necrosis factor	Cancers
Epidermal growth factor	Ulcers
Fibroblast growth factor	Ulcers
Erythropoietin	Anemia
Tissue plasminogen activator	Heart attack
Superoxide dismutase	Free radical damage in kidney transplants
Lung surfactant protein	Respiratory distress
alpha 1-antitrypsin	Emphysema
Serum albumin	Used as a plasma supplement
Relaxin	Used to aid childbirth

BIO-RAD

Choice of Cell Type

The biotechnology industry uses several cell types, both prokaryotic (bacteria) and eukaryotic (animal, fungi, plant), to synthesize recombinant proteins. The choice of the host cell depends on the protein expressed.

Bacteria can express large amounts of recombinant protein, but the expressed proteins sometimes do not fold properly. In addition, bacterial cells cannot carry out the post-translation modifications that are characteristic of some of the proteins made by eukaryotic cells. The most important post-translation modification is glycosylation, the covalent addition of sugar residues to the amino acid residues making up the protein. Glycosylation can change the structure and thus affect the activity of a protein. Many mammalian blood proteins are glycosylated, and the addition of these sugars often changes the rate of turnover (half-life) of the protein in the blood, because proteins that are misfolded will be quickly degraded. If glycosylation is important for the function of the protein, mammalian cells are the cell type of choice, but these cells produce less protein and are more expensive to grow.

In the early days of the biotechnology industry *Escherichia coli* (*E. coli*) was the bacterial host of choice. This species had been used as the primary experimental system to study bacterial genetics for decades. More was known about the molecular biology of *E. coli* than any other species, and many genetic variants were available. In addition *E. coli* grows quickly, can reach high cell concentrations, and can produce large quantities of a single protein. It is also relatively inexpensive to grow. Today *E. coli* remains the bacterial system of choice, and many companies produce recombinant proteins using this bacterial species. Insulin, the first protein produced by genetic engineering, was produced in *E. coli.* Blockbuster products like human growth hormone and granulocyte colony stimulating factor (which increases white cell production in cancer chemotherapy patients) are also produced using this bacterial species. In general, if a protein's properties allow it to be produced in bacteria, then *E. coli* is the system of choice.

For recombinant protein expression in lower eukaryotic cells, two yeast species are commonly used: *Saccharomyces cerevisiae (S. cerevisiae)* and *Pichia pastoris (P. pastoris)*. *S. cerevisiae* is the yeast species used to make bread, wine and beer. Baker's yeast is used in research laboratories as a model system to study the genetics of eukaryotic cells. *P. pastoris* is a yeast species initially discovered by the petroleum industry. It divides rapidly, grows to a very high cell density, and can produce large quantities of a single protein. In addition, it can be genetically engineered to secrete the protein into the surrounding medium to allow easier recovery. Both species can glycosylate proteins, although the glycosylation patterns may differ from mammalian patterns. The sugars that are added to the protein and their position on the amino acid chain may differ between yeast and mammalian cells. Despite these advantages, relatively few biotech companies use yeast as a production system. The exceptions are the vaccine that immunizes against hepatitis B virus and the vaccine Guardasil that immunizes against HPV, the human papillomavirus.

If a protein has a very large and complex structure, or if that protein requires glycosylation to be active, then the protein must be produced in a mammalian cell line. Chinese hamster ovary cells (CHO) is the cell line that is almost always used. CHO cells bear relatively little resemblance to the cells of the hamster from which they were derived in the 1950s; they have adapted to growth in cell culture medium. Cell lines are established when cells from a multicellular organism are separated from one another by a protein-digesting enzyme and grown as if they are really a unicellular organism. The cells require a rich medium that provides them with all of the amino acids, vitamins, and growth factors that they need to grow. This complexity means that mammalian growth medium is many times more expensive than the media used to grow either bacterial or yeast cells. The CHO cell lines can be adapted for growth in suspension culture. The CHO cell lines are most often engineered to synthesize the protein of interest on the ribosomes attached to the rough endoplasmic reticulum, to package and glycosylate the protein in the Golgi apparatus, and to eventually secrete the protein into the extracellular medium where it is easier to purify.

CHO cells can glycosylate proteins with a mammalian glycosylation pattern. If glycosylation is important

to the function of the protein, CHO cells should be used. Because CHO cells divide slowly, the production runs are much longer than with *E. coli* (on the order of weeks rather than days). All equipment and all growth media must be scrupulously sterilized. A single contaminating bacterial cell will overgrow the culture and will lead to that batch being discarded. Although the growth of CHO cells takes longer, uses expensive media, and presents a greater risk of contamination, the isolation of the proteins that these cells produce is usually easier than in bacterial or yeast cells.

Interferon beta provides a good example of how the end product influences the choice of expression system for recombinant proteins. Avonex, the human interferon beta-1a form produced in CHO cells, is glycosylated; while Betaseron, the human interferon beta-1b form produced in *E. coli*, is not glycosylated. Since glycosylation is important for interferon beta-1a function, it is produced in CHO cells.

Table 1.2. Advantages and disadvantages of using bacteria, yeast and mammalian cells to produce recombinant proteins.

Parameter	Bacteria	Yeast	Mammalian
Contamination risk	Low	Low	High
Cost of growth medium	Low	Low	High
Product titer (concentration)	High	High	Low
Folding	Sometimes	Probably	Yes
Glycosylation	No	Yes, but different pattern	Full
Relative ease to grow	Easy	Easy	Difficult
Relative ease of recovery	Difficult	Easy	Easy
Deposition of product	Intracellular	Intracellular or extracellular	Extracellular
Product	Intracellular	Often secreted into media	Secreted

Table 1.3. Examples of pharmaceutical products and the cell line used to produced them.

Product	Cell Line
Insulin	*Escherichia coli*
Human growth hormone	*Escherichia coli*
Granulocyte colony stimulating factor	*Escherichia coli*
Tissue plaminogen activator	CHO cells
Pulmozyme (DNase) cystic fibrosis	CHO cells
Erythropoietin induces red blood cell production	CHO cells
Hepatitus B virus vaccine	Yeast
Human papillomavirus vaccine	Yeast
Rituxan rheumatoid arthritis, non-hodgkins lymphoma, leukemia	CHO cells
Herceptin breast cancer	CHO cells

Choice of Plasmid

Once the cell type has been chosen, the plasmid or vector to express the protein needs to be selected. Different plasmids are used for expression of proteins in bacteria, yeast and higher eukaryotic cells. Some features that need to be considered in the plasmid include: selection system (such as antibiotic resistance), the promoter, the copy number of the plasmid, presence of signal peptide sequence to excrete the expressed protein out of the cell, presence of sequence coding for a protein purification tag or DNA coding for fusion protein partners.

Antibiotic resistance is a common component of both prokaryotic and eukaryotic vector systems. The presence of a gene for antibiotic resistance allows for selective retention of the plasmid and suppression of growth of any cells that do not contain the plasmid. However, to ensure safety for expression systems being used for vaccine and therapeutic protein production, other selection systems can be used such as

BIO-RAD

the expression of a required metabolic enzyme that has otherwise been deleted from the host organism.

The promoter controls the level of gene expression. It can be either constitutive, meaning that it is always active and there is no control over when the protein of interest is expressed, or inducible, meaning that its activity can be triggered by external factors. Examples of inducible promoters are the heat shock promoters, which are activated by a change in temperature. These promoters are derived from naturally occurring sequences in organisms that need to express a different protein when they are in a warm environment versus a cold one. Other inducible promoters are activated by the addition of a chemical such as lactose or its analog IPTG in the case for the LacZ promoter in *E. coli*. The T7 promoter is an example of a chemically induced promoter system. For industrial applications, genes for the protein of interest tend to be under inducible control.

The copy number of a plasmid depends on the origin of replication present in the plasmid. The origin of replication determines the level of control of replication of the plasmid, and if the plasmid is under a relaxed control more copies can be made. The size of the plasmid and size of the insert also affect the number of copies. If there are more copies of the plasmid in cells it is possible for them to produce more protein than cells that have fewer copies of the plasmid. Plasmids used for cloning such as pUC tend to be higher copy number while plasmids used for protein expression tend to be larger and have lower copy number.

To facilitate the purification of the expressed protein, DNA sequences coding for a signal (amino acid sequence) that targets the protein of interest to be secreted into the periplasmic region of *E. coli* or into the extracellular medium for eukaryotic cells, can be fused to the gene of interest. Other tags that are commonly added are fusion proteins to increase the solubility of the expressed protein (such as glutathione-S-transferase, GST) or affinity tags (polyhistidine or GST tags) that can be used to selectively purify the recombinant protein.

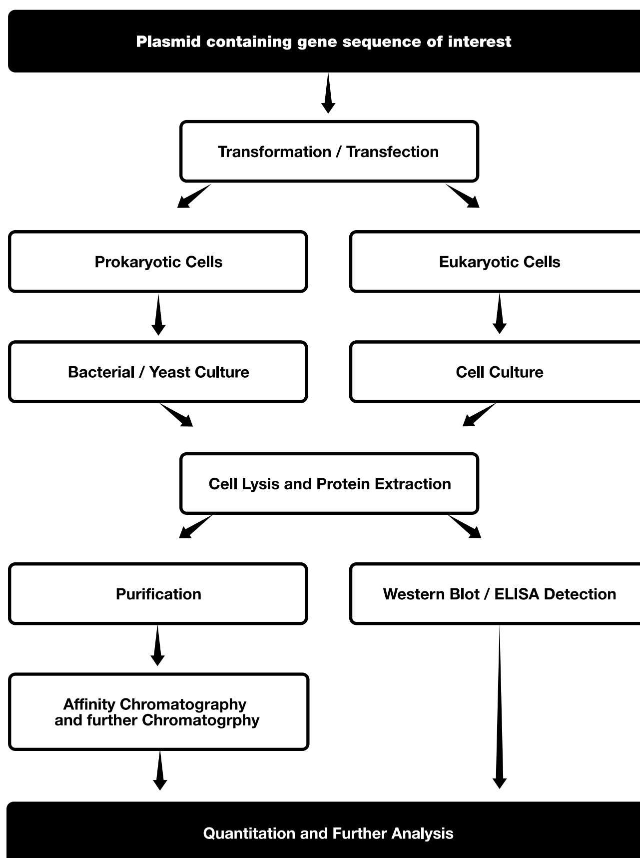


Figure 1.1. Gene Design For Recombinant Protein Production.

Choice of Cell Expression System

Once the cell expression system has been determined for recombinant protein production, the gene construct to express the recombinant protein needs to be designed. This could be as simple as taking the coding sequence of the gene of interest as it exists in the parent organism and inserting this sequence into an expression plasmid for the cell system being used. However, this usually does not produce optimal levels of recombinant protein when the gene is expressed in a heterologous system (a cell or organism different from the one where the gene is naturally found) because the preferred codon usage for various species differs. For example, the codon GGA for glycine can be found at a frequency of 16.4 times per 1000 codons in human genes while it is only used 9.5 times per 1000 codons in *E. coli* genes. Therefore, there is a chance that leaving this codon in the recombinant gene might lead to lower levels of recombinant expression due to a scarcity of tRNA molecules for GGA.

A second consideration for recombinant gene design is whether or not the protein of interest is expressed in a soluble or in an insoluble form. If a protein is expressed in an insoluble form, it can be easier to initially separate it from components, such as nucleic acids, phospholipids and soluble proteins, by centrifugation. An insoluble protein is also relatively protected from the action of proteolytic proteins that are present in the host cell that can be released upon lysis of the cell. However, if a fully functional recombinant protein is desired, it is necessary to refold the insoluble protein to its native conformation, which many times proves extremely problematic especially if the fully refolded protein has disulfide bonds and multiple subunits.

The ability to express a recombinant protein in the soluble form is partially dependent on the protein being expressed as well as the rate of expression of the recombinant protein. If the protein is expressed at an extremely high rate, it could overwhelm the native proteins involved in folding proteins (such as chaperonins) in the cell host. The rate of expression can be controlled by the promoter system involved such as the T7 polymerase system used commonly in *E. coli*.

A final consideration for recombinant protein gene design is how the recombinant protein will be purified from the other host cell components. Some recombinant proteins, such as antibodies, have a specific antigen against which they were raised and hence can be purified by binding to that molecule (affinity chromatography).* Other recombinant proteins have a very large positive or negative charge associated with them and can be purified by binding to charged resins (ion exchange chromatography). Some proteins do not have any strong distinguishing property, and an affinity tag, such as GST or a histidine tag can be added as a DNA sequence to the gene of interest at either the 5' or 3' end of the recombinant gene. The tag attached to the protein enables the specific purification of the recombinant protein using affinity chromatography methods.

*Antibodies can also be purified by binding to Protein A, a surface protein of *Staphylococcus aureus* which has high affinity for immunoglobulins. Protein A is commonly used for the first step in the purification of antibodies in industrial applications.

DHFR—Our Protein of Interest

The Protein Expression and Purification Series focuses on the protein dihydrofolate reductase (DHFR), which is essential for proper cell function and illustrates the importance of basic oxidation–reduction enzymatic reactions.

What is DHFR?

Dihydrofolate reductase, DHFR, is an enzyme that converts dihydrofolate, a folic acid derivative, into tetrahydrofolate (THF) by the addition of a hydride from NADPH. Tetrahydrofolate is a methyl group shuttle required for the synthesis of purines, thymidylic acid, and amino acids, all essential for nucleic acids. DHFR is ubiquitous in prokaryotic and eukaryotic cells, and is found on chromosome 5 in humans. Deficiency in DHFR has been linked to megaloblastic anemia, an anemia disorder with larger-than-normal red blood cells, as well as cerebral folate metabolism disorders. Both are treatable with folic acid and/or Vitamin B_{12}, depending on symptoms. Refer to Bio-Rad's Size Exclusion Chromatography kit (catalog# 166-0008EDU) for an introductory exercise separating Vitamin B_{12} from hemoglobin.

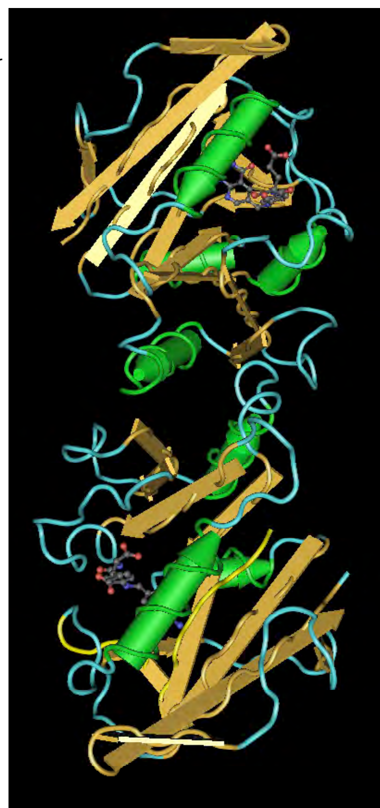


Figure 1.2. DHFR protein structure.

Being able to control DHFR makes it a powerful tool not only for research and gene manipulation but also for medical treatments for cancer and malaria. When DHFR is inhibited or reduced, it leads

BIO-RAD

to a shortage of thymidylates, interfering with nucleic acid synthesis. A lack of nucleic acid synthesis thus interferes with cell growth, proliferation, and ultimately causes cell death.

DHFR Cancer Connection

Cancer occurs when a particular cell loses the ability to control its division. These dividing cells spread, displace normal cells, disrupt the architecture of tissues, and use up the nutrients required by normal cells. Surgery can remove most of the cells in a solid tumor, but malignant cancers send out colonizing cells called metastases that use the blood and lymph systems to spread far from the tumor and non-solid tumors, such as leukemia or lymphoma, are not confined to one specific area. Clinicians use radiation or chemotherapy to kill these cells.

Chemotherapy drugs target cancer cells by disrupting the functions of actively dividing cells. This strategy exploits the fact that most of the cells in an adult are not dividing. Therefore, chemotherapy damages the rapidly dividing cancer cells by disrupting structures required for mitosis, like spindle fibers, or by disrupting the production of nucleotides required for DNA replication. One of the first chemotherapeutic agents was methotrexate, a folic acid analog that interferes with folic acid metabolism. Treatment with methotrexate limits the ability of dividing cells to make nucleotides by competitively inhibiting the enzyme dihydrofolate reductase (DHFR). When the enzyme DHFR is inhibited, cancer cells cannot divide and spread.

Occasionally after repeated treatments with methotrexate, a patient's cancer will develop a methotrexate resistance and will stop responding to the drug. Some of these resistant cells when examined show that the resistance was due to an increased copy number of DHFR genes. This gene amplification leads to increased levels of DHFR protein in the cell and therefore an increased ability to catalyze its reaction and produce nucleotides, even in the presence of methotrexate.

DHFR and Malaria

DHFR is also integral to parasite cell metabolism. Malaria is caused by the parasite Plasmodium falciparum, which is transmitted to humans by mosquitoes. Once in the human bloodstream, the parasite multiplies, eventually causing headaches, fever, coma, and ultimately, death if untreated.

Like in humans, interrupting the DHFR pathway in Plasmodium leads to reduced DNA synthesis: blocking DHFR successfully blocks Plasmodium falciparum multiplication. There are many drugs to treat malaria, but drug-resistant strains are becoming more and more common. Strains resistant to the once effective DHFR inhibitors pyrimethamine, sulphadoxine, and methotrexate are appearing.

Having a system to produce recombinant DHFR to study its enzymatic activity and develop inhibitors for chemotherapy or antimalarial drugs could be a powerful tool in developing new therapeutics. Scientists continue to search for effective drugs to stop the spread of malaria and cure those infected with malaria. (See Appendix J for a link to the World Health Organization's site on malaria.)

Use of DHFR in Biomanufacturing

The regulation of the DHFR gene amplification phenomenon described above is used to produce genetically engineered CHO cells to biomanufacture particular therapeutic proteins. Cells containing the gene of interest and DHFR are treated with methotrexate leading to the amplification of the DHFR gene; since the gene of interest lies next to the DHFR gene, the transgene is amplified too. This increases the amount of protein produced by the cells. Individual clones are separated and independently tested for their ability to produce protein and to divide. A particular clone that produces a large amount of protein and that retains its ability to divide quickly, will become the master cell bank from which all subsequent cells for production runs will be pulled.

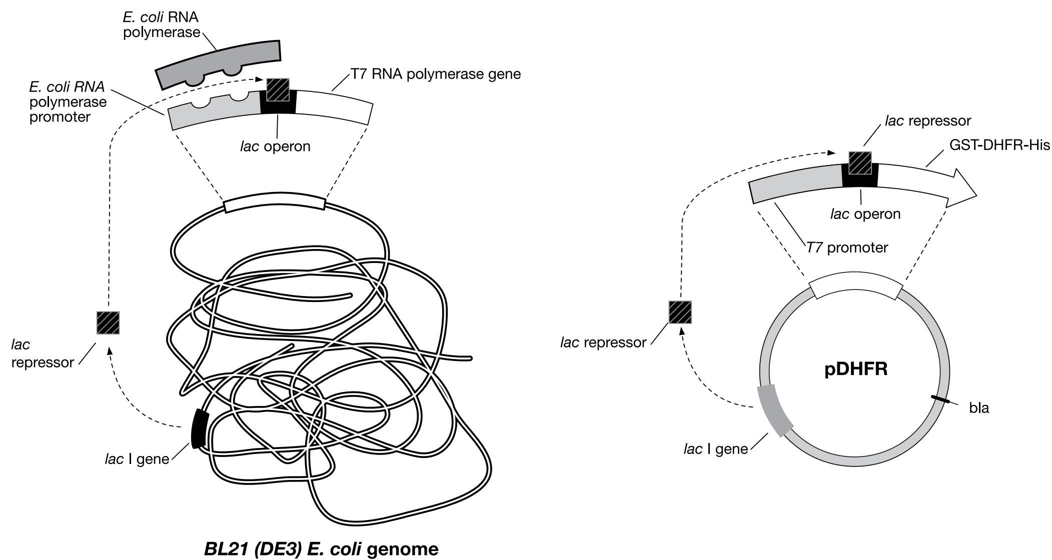
(For more information on biomanufacturing see Appendix G.)

Recombinant Production of DHFR Cell Selection

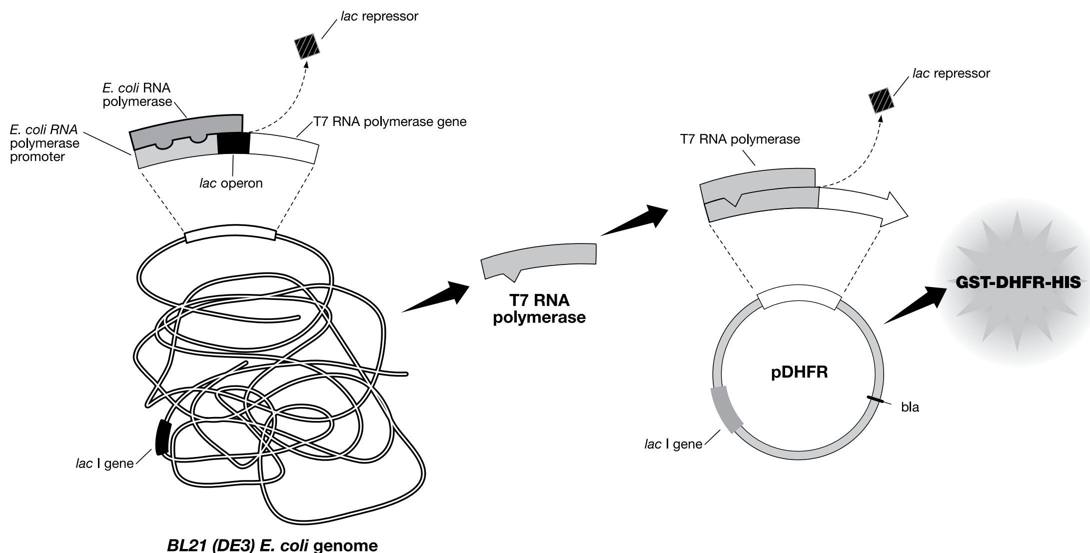
In this series of laboratory exercises, you will express the human DHFR gene in *E. coli*, a recombinant protein host system that is used extensively for research and industrial purposes. The bacteria have a quick doubling time (20 minutes) and are easy and inexpensive to culture, induce, and lyse to release cell contents. Also, posttranslational modifications such as phosphorylation and glycosylation are not required for the human DHFR to function properly, also making *E. coli* a good selection.

Plasmid Selection

A high level of expression with tight regulation is desired so the pET21a plasmid system will be used. This plasmid contains the constitutively expressed β-lactamase gene that confers resistance to ampicillin. The pET21a vector used contains the T7 promoter. This vector is used with a specific type of *E. coli*, the BL21(DE3) strain. This strain has been engineered to contain the T7 RNA polymerase gene (a gene that is not endogenous to bacteria) placed under the control of the inducible lac promoter. Addition of lactose or its analog IPTG to the growth medium, will induce the expression of the T7 RNA polymerase and, thus, the expression of any gene placed downstream of the T7 promoter. In the absence of the inducer, there is no expression of the gene.



Figures 1.3.a (above) and 1.3.b (below). Protein expression control using the pDHFR plasmid in a BL21(DE3) *E. coli*.



BIO-RAD

Figure 1.3a shows the use of the pDHFR plasmid in BL21(DE3) *E. coli* having tightly controlled expression of GST-DHFR-His. Both the BL21(DE3) *E. coli* genomic DNA and the pDHFR plasmid contain the lac I gene which codes for the lac repressor protein. Both the BL21(DE3) *E. coli* genomic DNA and the pDHFR plasmid also have recombinant genes that are inserted after a lac operon site. In the BL21(DE3) *E. coli* genomic DNA, the inserted gene codes for bacteriophage T7 RNA polymerase. In the pDHFR plasmid, the inserted gene codes for the fusion protein GST-DHFR-His. Before the lac operon, the BL21(DE3) *E. coli* genomic DNA has a native *E. coli* RNA polymerase promoter site. This is different than on the pDHFR plasmid where before the lac operon, there is a bacteriophage T7 RNA polymerase promoter site. The lac repressor protein is constitutively expressed, and when there is no lactose or its analog IPTG present, the lac repressor protein binds to the lac operon on both the genomic and plasmid DNA, preventing binding of the appropriate RNA polymerase and transcription of genes downstream from the lac operons, T7 RNA polymerase and GST-DHFR-His, respectively.

Figure 1.3b shows that once lactose or its analog IPTG has been added, the lac repressor protein detaches from the lac operon on both the genomic DNA and the pDHFR plasmid. This allows binding of the constitutively expressed *E. coli* RNA polymerase to the *E. coli* RNA polymerase promoter site on the genomic DNA and starts the transcription of the bacteriophage T7 RNA polymerase gene. Once the bacteriophage T7 RNA polymerase gene has been transcribed and translated into protein, the bacteriophage T7 RNA polymerase binds to the promoter region on the pDHFR gene. Once bound, the bacteriophage T7 RNA polymerase transcribes the recombinant gene GST-DHFR-His coded for on the pDHFR plasmid.

Gene Design

When expressed in *E. coli* using the pET21a system, the human DHFR is insoluble. To increase its solubility, a GST tag was added to its N-terminus. A six histidine sequence (polyhistidine tag) was added to the C-terminus of DHFR to allow for easy purification by metal affinity chromatography. The resulting plasmid codes for a GST-DHFR-His fusion protein that can be expressed in *E. coli* BL21(DE3).

Culturing *E. coli* for Protein Expression

The starting point in the Protein Expression and Purification Series is lyophylized BL21(DE3) *E. coli* containing pDHFR which will be rehydrated and plated to generate individual colonies. An initial culture is grown to saturation from a single bacterial colony picked from a Petri dish. This culture is used to initiate a larger culture that is grown to mid-log phase at which stage expression of the recombinant protein is induced by addition of IPTG to the medium. Determining when cells have reached their mid-log phase of growth is accomplished by measuring the absorbance of the culture at 600 nm. An OD600 of 0.6–1.0 is the typical target for induction. At this point the cells are dividing rapidly, and production of the protein will be optimal.

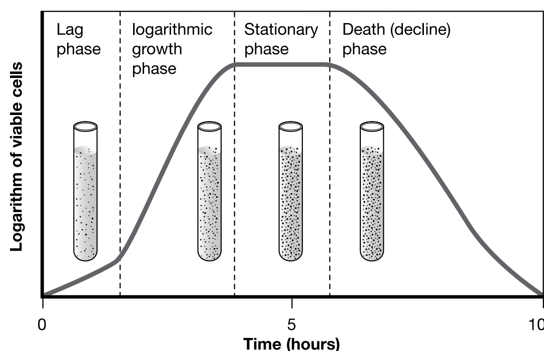


Figure 1.4. Depiction of cell growth phase in relation to overall culture viability.

For the GST-DHFR-His construct in the BL21(DE3) *E. coli* an overnight culture containing 1% glucose is grown from a single colony on an LB/amp plate. The 1% glucose is added to ensure that the lac operon remains repressed and no T7 RNA polymerase or GST-DHFR-His is expressed. Leaky expression (not actively induced) of recombinant proteins is generally undesirable because expression of the recombinant protein may be toxic and prevent bacterial growth. The overnight culture (late log/stationary phase) is diluted in fresh LB medium and allowed to grow until mid-late log phase.

At mid-log phase, expression of GST-DHFR-HIS is induced by addition of IPTG. Cells will use their internal

biochemistry to produce the recombinant protein and will not divide as quickly or at all. After induction has been completed, cells are recovered by centrifugation, and protein is extracted from the cell pellet.

Introduction to Protein Purification

Protein purification is an important step in biotechnology workflows. It is the isolation of the protein of interest so that it may be used in subsequent research, for diagnostic tests, or for pharmaceutical production. The purity needed depends on its end use. For proteins used in research 90–95% purity may be sufficient but for proteins used for pharmaceutical applications, much higher purity levels (up to 99.99%) must be reached. How purification is done will depend on the type of protein engineered, the volume of protein to be purified, the degree of purity required, and the availability of special laboratory equipment.

The first step in the purification workflow is to extract the protein from the cells by lysing, or breaking them open. Techniques used for lysing cells depend on what type of cell—bacterial, plant, or mammalian—was used to produce the protein as well as where the protein is produced in the cell. Freeze-thaw cycles, enzymatic digestion, chemical breakdown, and mechanical disruption methods such as sonication or grinding (mortar and pestle) are some of the cell lysis techniques available. In this laboratory the *E. coli* cells are lysed in a buffer containing lysozyme, an enzyme which digests the cell wall, along with freeze-thaw cycles. Once the cells are lysed, the soluble and insoluble components are separated by centrifugation. The insoluble components form a pellet at the bottom of the tube while soluble components will remain in the aqueous phase. (Not only proteins are found in the cell lysate, but also nucleic acids, sugars, phospholipids and other cell components.) When present in high concentration, nucleic acids (mostly genomic DNA) render the aqueous phase viscous making it difficult to isolate the recombinant protein. To reduce viscosity the genomic DNA can be broken down using sonication (ultrasound), a French Press that shears the DNA by pushing the cell lysate through a narrow opening at very high pressure or shearing the lysate using a narrow gauge syringe needle. These treatments break the genomic DNA down into small fragments and reduce viscosity. The DNA can also be degraded enzymatically by addition of a DNase. A too viscous soluble fraction of protein can be problematic since it may clog the resin and the frit (bed support) of the chromatography column during the subsequent protein purification step.

Isolating the Recombinant Protein of Interest

Proteins are usually purified by chromatography. There are a variety of chromatographic methods to choose from; the method used will depend on the protein's physicochemical properties. In the following sections some of the common chromatographic techniques available are described.

Introduction to Protein Chromatography

Chromatography is commonly used in biotechnology for separating biological molecules, such as proteins, from complex mixtures. Chromatography consists of a mobile phase (solvent and the molecules to be separated) and a stationary phase, such as paper (in paper chromatography) or glass, resin, or ceramic beads (in column chromatography), through which the mobile phase travels. The stationary phase is typically packed in a cylinder known as a column. Molecules travel through the stationary phase at different rates or bind to the solid phase based on their physicochemical properties.

The liquid used to dissolve the biomolecules to make the mobile phase is usually called a buffer. In column chromatography the mixture of biomolecules dissolved in the buffer is called the sample. The sample is allowed to flow through the column bed, and the biomolecules within the buffer enter the top of the column bed, filter through and around the beads, and ultimately pass through a small opening at the bottom of the column. For this process to be completed additional buffer is placed on the column bed after the sample has entered the bed. The mobile phase liquid is collected, as drops, into collection tubes that are sequentially ordered. A set number of drops, known as a fraction, is collected into each tube. Fractions are collected so that they may later be analyzed to see which one or ones contain the protein or proteins of interest.

Chromatography Techniques

There are many ways to perform liquid column chromatography. The choice of chromatography media and buffers depends on the properties of the protein of interest to be purified.

Hydrophobic Interaction Chromatography (HIC) separates molecules based on their hydrophobicity. Hydrophobic (water-fearing) substances do not mix well with water. Exposing a hydrophobic protein to a high salt buffer causes the three-dimensional structure of the protein to change so that the hydrophobic regions of the protein are more exposed on the surface of the protein and the hydrophilic (water-loving) regions are more shielded. The sample in high salt buffer is then loaded onto a chromatography column packed with hydrophobic interaction beads. The hydrophobic proteins in the sample will stick to the beads in the column. The more hydrophobic the proteins are, the more tightly they will stick. When the salt is removed by flowing in a low salt buffer through the column, the three-dimensional structure of the protein changes again so that the hydrophobic regions of the protein now move to the interior of the protein, and the hydrophilic regions move to the exterior. The result is that the hydrophobic proteins no longer stick to the beads and elute (wash out) from the column, separated from the other proteins.

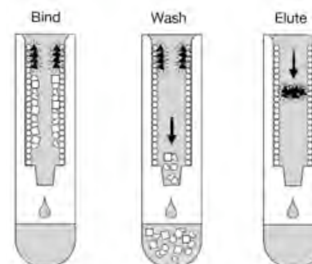


Figure 1.5. Depiction of HIC separation of molecules based on hydrophobicity.

Size Exclusion Chromatography (SEC), also known as gel filtration or desalting chromatography, separates molecules based on their shape and size. The solid phase is made of gel beads that have pores of varying size (think of them like wiffle balls). Larger molecules cannot enter the pores and are excluded, so they merely flow between the beads and are eluted first. Smaller molecules can enter the pores and therefore will take longer to flow down the column. Typically, size exclusion columns are tall, narrow columns so that there is a long path for the molecules to flow through or to be retained by the pores and better separated from each other. Size exclusion chromatography can also be used to exchange the buffer that the molecule of interest is currently in for another buffer.

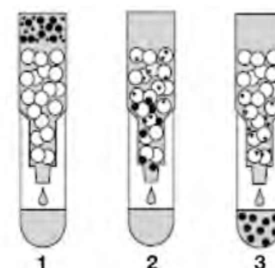


Figure 1.6. SEC separation of molecules based on size. 1) A mixture of large and small proteins is applied to a column of porous beads. 2) As the buffer flows down the column, the small proteins penetrate into the beads and are slowed. 3) The larger protein molecules emerge from the column first.

Ion Exchange Chromatography (IEX) beads have either a positive (cation) or negative (anion) charge.
During ion exchange chromatography, the protein of interest binds to the oppositely charged beads. If the charge of the beads is positive, it will bind negatively charged, or anionic, molecules. This technique is called anion exchange (AEX) chromatography. If the beads are negatively charged, they bind positively charged, or cationic, molecules, called cation exchange (CEX). A scientist picks the resin based on the charge of the protein of interest. After contaminants are separated from the protein of interest, a high salt buffer is used to disrupt the interaction and to elute the protein of interest from the column.

Resin Type	Anion Exchanger	Cation Exchanger
Net Charge of Molecule of Interest	-	+
Charge of Resin	+	-

For example, in anion exchange chromatography, the resin beads have a molecule with a positive charge covalently attached to the resin. The molecules to be separated flow across the resin beads and any positively charged molecules are repelled and do not stick to the column, exiting the column with the flow of the buffer. Negatively charged molecules will bind, or adsorb, to the column. The column is then washed with buffer of increasing salt concentration, and those molecules that are more tightly bound will elute. Cation exchange chromatography works in a comparable way except that the resin beads have molecules with negative charges covalently bound to them.

In ion exchange chromatography knowing the isoelectric point (pI) of the protein allows researchers to manipulate the charge of the protein. The pI of the protein is where the protein has equal positive and negative charges. A buffer with pH higher than the pI will give the protein a negative charge; a buffer with pH below the pI of the protein will have a more positive charge. This change of charge depending on buffer pH can be used to elute protein from ion exchange columns by changing buffer pH.

Mixed Mode, or Multimodal, Chromatography resins combine more than one type of chromatography technique, such as having both anion and cation exchange properties on the same bead. Based upon the properties of the molecule of interest and the buffers used, this can be a very selective chromatography method.

In **Affinity Chromatography (AC)**, a ligand with specific affinity for the molecule to be isolated is covalently attached to the beads. A mixture of proteins is added to the column and everything passes through except the protein of interest which binds to the ligand and is retained on the solid support. The desired molecule is primarily eluted by adding a molecule that competes for the ligand.
The affinity chromatography methodology depends on the presence of a specific tag on the recombinant protein such as a polyhistidine (His) affinity tag; a fusion protein partner, such as maltose binding protein or glutathione-s-transferase (GST-tag), or an antibody-antigen interaction such as protein A or protein G and different classes of antibodies.

Polyhistidine-tagged proteins bind to nickel groups that have been attached to the resin, known as immobilized metal affinity chromatography (IMAC). The histidine groups of the polyhistidine-tag bind to the Ni^{++} groups on the resin. The protein of interest can be eluted by the addition of imidazole which competes for nickel binding sites. A second example is using a fusion partner such as GST. This protein will bind to a resin bead coated with glutathione. In order to elute the recombinant protein fused with GST, glutathione is added to the mobile phase and competes with the binding site on the GST and the fusion protein elutes.

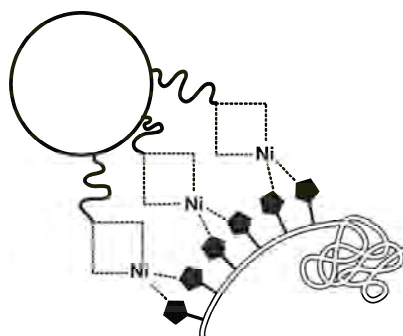


Figure 1.7. Depiction of the interaction of the Ni^{++} groups on the resin and the polyhistidine tag.

In this laboratory purification affinity polyhistidine tag chromatography (more specifically, immobilized metal affinity chromatography (IMAC)), will be used to purify DHFR. The recombinant DHFR in this laboratory also has a GST-tag, but in this instance the tag was added to make the protein more soluble and increase the overall molecular weight of the fusion protein. It would be possible to use the GST-tag in a second round of purification. In research and industrial purification of proteins more than one chromatographic method is needed to reach the level of purity desired.

Chromatography Methods

Once the chromatographic purification strategy is chosen for the target protein, a decision needs to be made as to how the chromatographic separation will be performed.

Batch Purification

The simplest way to perform chromatography is in a beaker. To do this, the chromatography resin of choice is resuspended in buffer in a beaker. The sample is added. The beaker is either gently swirled to mix the sample and resin, or a stir bar and stir plate may be used, being careful not to damage the resin. Next, the resin is allowed to settle, the buffer is decanted and wash buffer is added to the beaker. Again, the resin is mixed, and the buffer is decanted. Finally, an elution buffer is added and the decanted buffer is saved since that is where the protein of interest should be. This method may work well for purifying crude extracts where quantity, and perhaps quality, of protein is not a concern.

BIO-RAD

Gravity Chromatography

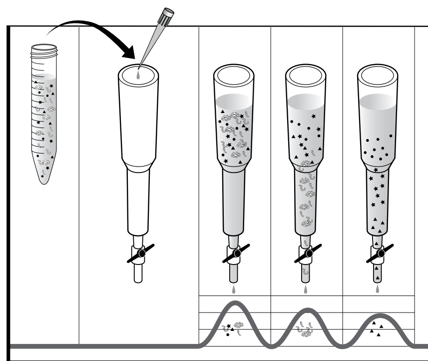


Figure 1.8. Gravity column set-up

The most common way to purify protein is by column chromatography using gravity flow. In gravity chromatography the resin is resuspended in buffer and poured into a column. The column is placed upright in a stand and the resin is allowed to settle into what is known as the resin bed. Once the resin has settled, the sample is loaded onto the column. As the sample flows through the column, buffer is added to the column so that the top of the resin bed stays wet and there is a pressure head to continuously push the sample and buffers through the resin. The buffer and sample rely on gravity to move through the column. During the chromatography process fractions are collected. Samples from these fractions are tested for purity and for the presence of the protein of interest.

The main advantage of gravity chromatography is that it is an inexpensive method of purification. A glass or plastic column, buffers, sample, a column holder, and test tubes to collect the fractions are all that is needed. Columns, and sometimes resin, can be cleaned and reused multiple times. The expense of this method of chromatography is time. A person must monitor the column so that it does not go dry, ensure the correct buffers are used when they should be used, and collect fractions. Gravity chromatography is typically used with larger diameter chromatography beads and softer resins. There needs to be minimal resistance to flow due to gravity or this will not work. (Think of trying to flow water through a small diameter of sand versus around pebbles or rocks.)

Spin Column Chromatography

Figure 1.9. Spin column.

A variation on gravity chromatography is the use of a spin column. Spin columns are typically small (3–5 cm in length) plastic chromatography columns that fit in a standard microcentrifuge. They come empty or in many cases prepacked with resin. The sample volume that can be applied to the column is limited by the small size of the column and by how much protein can bind to the amount of resin in the tube; the advantage, however, is that this is a quick way to perform chromatography. Simply load the resin or add buffer to a prepacked column, spin in the centrifuge, add sample in buffer, spin again to bind the protein, add elution buffer, and spin again. The gravitational force from the spin "pulls" the buffer and sample through the column. The spin column fits in a centrifuge tube, and after each spin the column is moved to a fresh tube so at the final spin the protein of interest will be in the elution tube. Each resin or prepacked column has a spin protocol to follow on how to use it in a spin column and how fast to spin it. (See Appendix C for more information on centrifuge spin speed.)

The advantage of the spin column is that it is quick and relatively inexpensive in terms of columns, resin, and time. Many spin columns come prepacked with popular resins. Most spin columns are single use. They require a centrifuge and sometimes require a certain rotor (the inside of the centrifuge, where the tubes are placed), that will depend on the length of the spin column and type of chromatography resin used.

Prepacked Chromatography Cartridges

Figure 1.10. Prepacked chromatography cartridge.

Prepacked cartridges are a good choice for chromatographic needs when sample size is too large for a spin column, but a short purification time is desired. Cartridges are convenient as they are prepacked with resin, consistent in quality of resin bed from column to column, and disposable. Most cartridges have Luer-Lok or similar screw-like connections that allow them to be used with a syringe with a

Luer-Lok end, a mechanical pump or chromatography system.
The simplest way to use a prepacked cartridge is with a Luer-Lok syringe. The syringe is used to deliver the buffers and sample in sequence and collect the eluate or fractions in test tubes. For more complex purifications or for a hands-off approach, a pump can be used to deliver the sample and buffers through the columns. Some monitoring of the pump may be needed depending on the programming capabilities. The cartridge can also be connected to a fraction collection system or apparatus.

Chromatography Systems

Most large-scale purifications, when the sample volume is milliliters to liters and columns range in size from smaller than a tube of toothpaste to 100 L and larger, are performed using a chromatography system. The system will allow for multiple buffers and columns to be used. Systems can range from low pressure to high pressure, and basic to complex. As the complexity of the chromatography equipment rises, so does the price. The decision of which system to use depends on the purification scheme.

Figure 1.11. Bio-Rad BioLogic System.

Protein Expression and Purification Series Laboratory Background

Chromatographic purification of GST-DHFR-His

The purification of the GST-DHFR-His is accomplished in two steps: First the protein will be purified using IMAC. As previously described, the IMAC resin selectively binds polyhistidine-tagged proteins. All other biomolecules will flow through the column. The column is first equilibrated in 20 mM sodium phosphate, 300 mM NaCl and 5 mM imidazole. The 300 mM NaCl prevents the non-specific binding of charged molecules in the *E. coli* lysate soluble fraction from binding to the column. Since imidazole has a similar structure to histidine (see figure 1.12), the 5 mM imidazole prevents non-specific binding of any *E. coli* proteins which contain multiple histidine residues, but the imidazole is not at a high enough level to prevent the binding of the polyhistidine tag on the GST-DHFR-His to the Ni-IMAC beads. After the lysate is added and the GST-DHFR-His binds to the resin and the majority of *E. coli* proteins flow through without binding, the column is washed. The wash buffer contains the same 20 mM sodium phosphate for buffering and 300 mM NaCl to prevent non-specific ionic binding, but also contains a higher level of imidazole. This 10 mM imidazole is slightly more stringent and will wash away many of the *E. coli* proteins that were able to bind under the 5 mM imidazole condition. However, 10 mM imidazole is not stringent enough to effectively compete with the binding of the polyhistidine tag so the GST-DHFR-His remains bound to the column beads during the wash step. GST-DHFR-His is eluted from the column with a buffer containing 250 mM imidazole. At 250 mM, the imidazole is at a much higher concentration than the GST-DHFR-His effectively competing with the polyhistidine tag for binding to the Ni^{++} groups of the resin. There are a lot more molecules of imidazole present than GST-DHFR-His so the GST-DHFR-His is knocked off the Ni^{++} binding sites and elutes from the column.

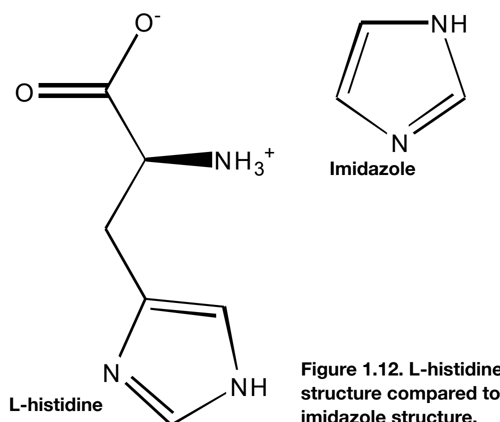


Figure 1.12. L-histidine structure compared to imidazole structure.

The second step in the purification of GST-DHFR-His is removing excess salt and imidazole from the sample. If the sample is not desalted, during subsequent polyacrylamide gel analysis, sample bands may be fuzzy, broad, skewed, or otherwise distorted. The desalting also removes the imidazole, which interferes with measuring the absorbance at 280 nm, used to estimate protein quantity.

BIO-RAD

Desalting the sample is performed using a size exclusion gel prepacked in a spin column. The gel, Bio-Gel P-6, has a fractionation range of 1,000 to 6,000 Da. This means that the pores of the gel are large enough to allow molecules in the 1 to 6 kD range, like salts and small proteins, to enter, but larger molecules, such as GST-DHFR-His, with a molecular weight of 52 kD, will be excluded and flow through the column. In essence, the salts (imidazole, NaCl and phosphate buffer) are trapped in the gel while the protein of interest comes out “clean” in a Tris buffer.

Methods To Quantify Protein Concentration and Check Protein Purity

Once the protein of interest (GST-DHFR-His) is purified it is necessary to check its purity and determine the quantity of protein purified. There are multiple ways to perform these tasks.

Absorbance at 280 nm

The aromatic amino acids (tryptophan, and to a lesser degree tyrosine and phenylalanine) in proteins absorb at 280 nm. If the extinction coefficient (a parameter that helps define how well a substance absorbs light at a specific wavelength at a particular concentration) is known, the amount of protein present can be calculated using Beer’s Law (absorbance = εcL) where ε is the extinction coefficient, c is concentration and L is the pathlength of light. If the extinction coefficient is not known, there are computer programs, such as that from Expasy (see Appendix J), that can calculate an approximate extinction coefficient from empirical relationships.

It should be noted that other molecules such as imidazole absorb at 280 nm and can interfere with calculations. The buffer in which the sample is measured is critical for A280 measurements. This is one of the reasons why the second purification step in this series is performed. It is a desalting step to remove the imidazole from the GST-DHFR-His so that the A280 measurement can be taken.

Colorimetric Protein Assays

There are multiple colorimetric protein assays that have been developed to determine protein concentrations. The first—the Bradford protein assay—is based on a shift in the maximum absorbance of a colored dye, Coomassie Brilliant Blue G-250. The dye interacts mainly with basic amino acid groups (arginine and lysine) as well as aromatic amino acids (phenylalanine, tyrosine and tryptophan). A second method—the Lowry method—is based on reaction of the protein with an alkaline copper tartrate solution and Folin reagent. In the case of the Lowry method, color is mainly due to the presence of tyrosine and tryptophan as well as cysteine and histidine. Each of these assays has its advantages and disadvantages, namely the compatibility of the assay with reagents in the buffer as well as sensitivity.

SDS-PAGE Analysis

Another method of determining which fractions contain the purified protein as well as the progress of the protein purification procedure, is to analyze samples by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). This method is not quantitative unless a protein standard of known concentration is run in one of the lanes and quantitative software is used to examine the gel image. It will, however, allow the determination of which fractions contain the protein of interest if the molecular mass is known and at what relative concentration. It will also allow the assessment of the purity of the protein fraction. If the protein is pure, a single band should be visible on the gel, even when large quantities of the samples are loaded.

Protein Structures and Basic Properties

In contrast to DNA, which is quantified in terms of its length (i.e., the number of base pairs), proteins are quantified in terms of their molecular weights relative to a hydrogen atom, in Daltons. This is because DNA is composed of only 4 nucleotides, which are in roughly equal proportions and about the same molecular

weight. Proteins on the other hand are composed of 20 amino acids with molecular weights from 89 to 204 Daltons (the average is 110). They vary considerably in amino acid composition. One Dalton equals the mass of a hydrogen atom, which is 1.66×10^{-24} grams. Most proteins have masses on the order of thousands of Daltons, so the term kilodalton (kD) is used for protein molecular masses. Proteins range in size from several kilodaltons to thousands of kilodaltons but most fall between the range of 10 kD and 220 kD. DHFR-GST-His has a primary structure of 410 amino acids, a total molecular weight of 52,000 daltons, or 52 kD.

Using Gel Electrophoresis to Separate and Identify Proteins

A protein's electrical charge and its mass affect its mobility through a gel during electrophoresis. The ratio of charge to mass is called charge density. Since every protein is made of a unique combination of amino acids, the net charge of each protein may be different. The electric charge of proteins must be removed as a factor affecting migration in order for polyacrylamide electrophoresis to be effective as a method of protein molecular weight determination. The intrinsic charges of proteins are masked by placing a strongly anionic (negatively charged) detergent, sodium dodecyl sulfate (SDS), in both the sample buffer and the gel running buffer. SDS binds to and coats the proteins and also keeps them denatured as relatively linear chains. In this form, proteins migrate in a polyacrylamide gel as if they have equivalent negative charge densities, and mass becomes the main variable affecting the migration rate of each protein. (**Note:** Posttranslational modifications such as glycosylation can also affect protein migration).

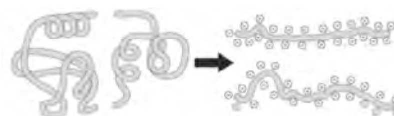


Fig. 1.13. The combination of heat and the detergent SDS denatures proteins for SDS-PAGE analysis.

Aside from obscuring protein charge with SDS, to effectively determine the molecular weight, the secondary (2°), tertiary (3°), and quaternary (4°) structures of the protein complexes within a protein extract are also disrupted prior to electrophoresis. This process of structural disruption is called **denaturation**. A reducing agent, such as β-mercaptoethanol (BME) or dithiothreitol (DTT), is sometimes added to samples to ensure complete breakage of disulfide bonds. (In the case of GST-DHFR-His, no reducing agent is needed as there are no disulfide bonds in the protein structure.) Three factors—heat, ionic detergent, and reducing agent—completely disrupt the 2°, 3°, and 4° structures of proteins and protein complexes, resulting in linear chains of amino acids. The denatured amino acid chains move through the gel at rates proportional to their molecular masses.

Performing Electrophoresis

In this lab, the induction of expression of GST-DHFR-His, solubility of the expressed GST-DHFR-His, and success of purification will be analyzed by SDS-PAGE. To do this, a portion of the fractions will be run on an SDS-PAGE gel against a protein standard of known molecular weight. The samples of the fractions have Laemmli sample buffer added to them and then will be loaded into the wells of a polyacrylamide gel. Once the gel is placed in the electrophoresis cell, the lower and upper chambers are filled with running buffer. The lower part of the cell contains the anode, or positive pole, and the top one contains the cathode, or negative pole. Once assembled, the electrophoresis cell is connected to a power supply. The proteins, now negatively charged due to the presence of SDS, will flow with the electric field toward the anode. When the dye front (from the sample loading dye) reaches the end of the gel, electrophoresis is stopped. The gel will be stained with the Coomassie blue dye. The stained gel will allow students to check the purity of their fractions and quality and relative quantity of the recombinant protein produced.

DHFR Enzyme Activity

Dihydrofolate reductase is a critical enzyme necessary for the conversion of dihydrofolate (DHF) to tetrahydrofolate (THF). This reaction also requires the presence of the cofactor NADPH (nicotinamide adenine dinucleotide phosphate). A **cofactor** is a molecule that is required to be bound to the enzyme in

BIO-RAD

order for the reaction to occur. NADPH is a cofactor that carries electrons from one reaction to another for use in metabolic enzymatic reactions.

The conversion of DHF to THF by dihydrofolate reductase in the presence of NADPH is an example of a classical oxidation-reduction reaction. Oxidation is defined as the loss of at least one electron, while reduction is the gain of that electron. In the case of organic compounds, oxidation is the loss of a hydride ion (H^-), or a negatively charged hydrogen ion that consists of two electrons and one proton. The NADPH donates a hydride ion to DHF, and when this hydride ion reacts with a specific bond of DHF, it removes the existing double bond and is covalently bound to the carbon group producing THF. Dihydrofolate reductase catalyzes the reactions shown in Figure 1.14. As shown, the NADPH loses a hydride ion that is transferred to DHF converting it to THF.

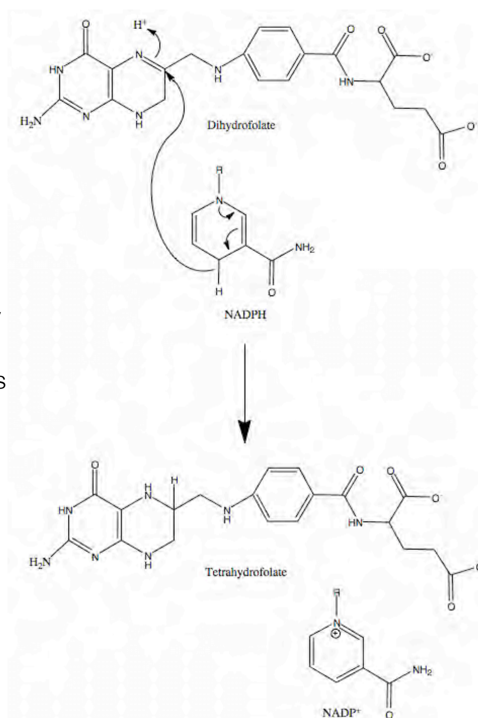


Figure 1.14. The conversion of DHF to THF. NADPH transfers a hydride ion to DHF, converting it to THF.

To measure the enzymatic activity of dihydrofolate reductase, a methodology needs to be developed to track the disappearance of NADPH or DHF or conversely to measure the appearance of $NADP^+$ or THF. For every molecule of NADPH that is converted to $NADP^+$ by the enzyme dihydrofolate reductase, one molecule of THF is produced from the substrate DHF. Unfortunately, none of these four compounds are colored so there is no way to directly visualize the reaction. However, it is possible to measure the rate of this reaction using a spectrophotometer than can detect light in the UV range. NADPH absorbs light in the UV range at 340 nm. However, when NADPH is converted to $NADP^+$, it no longer absorbs at 340 nm. Therefore, by measuring the decrease of absorbance at 340 nm of a reaction containing a known quantity of DHF, NADPH and dihydrofolate reductase, the reaction rate that NADPH is converted to $NADP^+$ can be calculated.

In this laboratory series, students will recombinantly make and purify human dihydrofolate reductase as a fusion construct with GST and a histidine tag (GST-DHFR-His). The concentration of the purified GST-DHFR-His will be calculated using the proteins' intrinsic absorbance of UV light at 280 nm. This known concentration of GST-DHFR-His will then be first combined with a known quantity of NADPH. Since no DHF substrate is present, the NADPH should not be reduced and the absorbance at 340 nm over time should be constant. Once the substrate DHF is added to the solution containing the purified GST-DHFR-His and NADPH, the absorbance at 340 nm should decrease over time as the reaction occurs converting NADPH to $NADP^+$ and DHF to THF.

Chapter 2: Focus Questions

Focus Questions for DHFR Cancer Connection

1. What is DHFR? Why is it important?

2. What role does DHFR play in cancer?

3. How does methotrexate interfere with DHFR?

4. What reaction does DHFR catalyze? What is the cofactor that is necessary for this reaction to occur?

5. Name the primary inhibitor of DHFR. How does this inhibitor affect its action?

Focus Questions for Protein Expression

1. What is a recombinant protein, and why would it be used instead of a native protein?

2. Name at least one pro and one con of using eukaryotic cells to produce a recombinant protein. When would you choose to use eukaryotic cells to produce a recombinant protein?

3. Name at least one pro and one con of using prokaryotic cells to produce a recombinant protein. When would you choose to use prokaryotic cells to produce a recombinant protein?

4. What are three considerations for recombinant protein gene design?

5. What is the log phase of a cell culture, and why is it important to recombinant protein production?

6. Why are subcultures prepared when trying to produce recombinant proteins?

7. What is the purpose of adding glucose to culture media for lac operon systems?

Focus Questions for Protein Purification Introduction

1. What is protein purification?

2. Describe lysis and why it is a necessary part of the protein purification process. Name a lysis technique.

3. Why is it important to remove DNA from the sample?

4. What enzyme is used to break down the DNA in a lysate but leaves the protein intact?

5. What is a common method of capture, or isolation, of protein?

Focus Questions for Introduction to Chromatography

1. What is the purpose of chromatography?

2. What is the mobile phase in chromatography?

3. What is the stationary phase in chromatography? Give three examples of stationary phase.

4. Name four types of liquid chromatography techniques.

5. What is an anion and what is a cation? When would you use anion exchange chromatography versus cation exchange chromatography?

6. What is an affinity chromatography tag? Name two tags.

7. Does adding an affinity tag to the protein change the protein? Explain your answer.

Focus Questions for Centrifugation and Chromatography Instrumentation Purification

1. What is gravity chromatography? Name an advantage and disadvantage of this system.

2. Is centrifugation/spin chromatography the same as gravity chromatography? Explain your answer.

3. What are the advantages of spin chromatography? When would you choose to use centrifugation/spin chromatography and why?

4. What are the advantages of using prepacked cartridges? When would you choose to use prepacked cartridges and why?

5. What is a fraction in chromatography?

Focus Questions for SDS-PAGE Electrophoresis

1. What are four ways to quantify the amount of protein in a sample?

2. What does primary (1°), secondary (2°), tertiary (3°), and quaternary (4°) protein structure refer to?

3. Describe SDS-PAGE analysis. Why would you use SDS-PAGE to analyze your samples?

4. What are BME and DTT and what do they do to a protein?

5. Describe the components and function of each component in Laemmli sample buffer.

Focus Questions for DHFR Enzyme Activity

1. Which cofactor is required in the conversion of dihydrofolate to tetrahydrofolate?

2. What is the purpose of the cofactor?

3. Define oxidation. Define reduction.

4. In the conversion of DHF to THF, which compound donates the hydride ion to DHF?

5. What information does the DHFR enzyme activity provide in the context of recombinant protein expression and purification?

Appendix G: Biomanufacturing—Protein Pharmaceutical Production Using Genetically Engineered Organisms

Introduction

Industrial biotechnology usually refers to the production of protein pharmaceuticals by living organisms whose genes have been altered in the laboratory. The Research and Development teams of a biotechnology company construct recombinant DNA molecules in their laboratories, introduce this DNA into an appropriate cell type, induce that cell to produce the valuable protein product, and then develop a process where the protein of interest can be purified by separating it from other cellular components and other proteins. The manufacturing team then takes this cell and uses it to produce that pharmaceutical protein on a large scale.

The genetic engineering revolution started in academic laboratories in the mid-1970s, produced major business enterprises in the 1980s, and matured into a multibillion dollar industry in the 1990s. In the 21st century this mode of manufacturing likely will emerge as a dominant industry as a way to manufacture many more pharmaceuticals, as well as to produce biofuels, textiles, and a variety of consumer products.

This process bears some similarities to ancient alchemy. Alchemists tried to turn base metals like lead into gold. Modern biotechnology companies convert the components of the cell growth media into pharmaceutically active proteins that are more valuable than gold. The standard laboratory bacterium, *Escherichia coli*, can take the relatively inexpensive components of a simple growth medium, like simple sugars and salts, and convert them into a product worth many times its weight in gold.

Glucose + Air + Salts ----> Pharmaceutical Proteins

Recombinant Protein	Price per gram
Bovine Growth Hormone (BST)	$ 14
Gold	44
Insulin	60
Growth Hormone	227,000
Granulocyte Colony Stimulating Factor	1,357,000
Note: Prices in 2011 dollars	

These proteins are worth many times their weight in gold, and they often treat diseases for which there was no treatment before the large-scale production of the protein had been developed. The first protein produced was insulin to treat diabetes. This bacterial production of insulin replaced pig insulin that had been isolated from pancreas' collected from slaughterhouses. Next, recombinant human growth hormone, to treat Turner syndrome and other diseases that produced short stature, replaced growth hormone isolated from the pituitary glands of cadavers. Tissue plasminogen activator, an enzyme that triggers the digestion of blood clots, became the standard treatment to remove the blood clots in coronary arteries that caused heart attacks and the blood clots in brain arteries that caused strokes. The hormone erythropoietin raised the level of red blood cells to treat the anemia that accompanies cancer chemotherapy; this protein is also given to patients on dialysis to treat the anemia associated with kidney failure. Biotechnology companies have developed antibodies that specifically recognize proteins on the surface of cancer cells. These antibodies have been combined with traditional chemotherapy to dramatically increase the treatment success rates of some of the most aggressive forms of cancer.

Overview—The Biomanufacturing Process

Biomanufacturing is used to describe the method of producing large quantities of pharmaceutical products from scalable, validated processes. The manufacturing facilities that grow these cells and then purify

BIO-RAD

the proteins that they produce cost hundreds of millions of dollars to build. These plants manufacture pharmaceuticals under strict government regulations. Every procedure must be validated; that is, proven to be mistake-proof, and every piece of equipment must be proven to perform its function. All employees are specifically trained and must gown-up before entering manufacturing suites. The manufacturing plants are designed for easy cleaning, for a logical personnel flow, and for the airflow to be filtered ensuring it is free of any microbes that might contaminate the product. Each room is continually monitored to make sure that there is the lowest possible level of microbial contamination in the air, on the countertops, and on the floors. Every process follows a standard operating procedure (SOP), and personnel fill out paperwork called a production batch record in which they record every action. Periodically the government inspects the plant and checks these records to make sure that quality standards are strictly followed.

Genetically engineered cells are grown under strictly controlled conditions in very large tanks called bioreactors. After the cells reach their maximum density in the bioreactor, they are separated from the growth medium, then lysed to release the protein that they have produced. Recovery techniques separate cellular proteins away from other cellular components, like DNA and membranes, and then the protein of interest is purified away from other cellular proteins. The Food and Drug Administration (FDA) requires that the final protein injected into a patient be 99.99% pure.

Since the function of a protein depends upon its three-dimensional shape, all production, purification, and delivery techniques must be designed to have the protein maintain this structure. Therefore, proteins are kept away from temperature extremes, pH extremes, organic chemicals, and extreme agitation that might denature (unfold) the protein and render it useless/ineffective. At each stage in the production process the protein must be surrounded by a solution of buffer molecules that moderate pH changes. It is also this consideration of preventing denaturation that dictates whether protein pharmaceuticals will be delivered directly into the bloodstream (parenteral) rather than delivered in a pill form. (If a protein were taken as a pill, the protein would denature the moment that it reached the acidic environment of the stomach.) Therefore, the final preparation of the protein must be placed in a solution that readies it for injection directly into the bloodstream.

Upstream Process—The Growth of Genetically Engineered Cells

The Research and Development team generates a large quantity of the genetically engineered cells and freezes these cells in liquid nitrogen to form a master cell bank; this cell bank will provide the product-producing cells for the entire patent life of the product—20 years. A sample of cells is taken from the master cell bank and used to produce a working cell bank that will provide the cells for a production run, which is called a campaign). These cells are initially grown in a small flask containing a growth medium matching the requirements of the cell. Their cell division causes the culture to increase in concentration, but before the cells reach stationary phase (get too crowded and stop dividing), they will be transferred to a bioreactor and diluted with new media. This bioreactor will stir and aerate (add sterile air to) the cells to make sure that the culture remains aerobic at all times. The goals for the cells are to always generate their energy by cellular respiration and to prevent them from encountering anaerobic conditions that would cause them to supplement ATP production with fermentation. Probes inside the bioreactor monitor the temperature, pH, and dissolved oxygen concentration of the solution. If any of these parameters change, a controller initiates an action to push the altered parameter back to its set point. If the temperature of the medium goes up, cooling water removes heat as it floods through the jacket surrounding the bioreactor. If the pH becomes acid, then a pump turns on to inject base into the medium until the medium is neutralized. If the dissolved oxygen drops, the bioreactor increases the sparge rate of air being pumped in or can supplement the air supply with pure oxygen. This monitoring and reacting to changes maintains the cells in their optimum growth conditions and at their maximum cell division rates.

As the cells divide they are systematically transferred to larger and larger volume bioreactors. The largest bioreactors in a major biotechnology company may be three stories tall. The gene coding for the recombinant protein of interest is always cloned behind an inducible promoter that allows the gene to be turned off for the entire growth phase of production. During this phase the cells devote their energy to

making more cells. It isn't until after the maximum cell concentration is reached in the final bioreactor that the gene is induced, and the cells begin to produce the pharmaceutical protein product. The cells then produce a large concentration (also known as titer) of protein. An optimized process will produce tens of kilograms of protein during a single production run. The cells and their surrounding medium then go to the recovery and purification process.

Downstream Process—The Recovery and Purification of the Protein

Once the cells have produced the protein, the protein must be purified to the 99.99% purity standard required by the FDA. In recovery, the cells are separated from their growth medium by centrifugation or filtration. Typically in *E. coli* the protein is produced as intracellular, therefore the cells are retained and the medium is discarded. If the production cells have been genetically engineered to secrete the proteins into the medium (typical for mammalian cells), then the cell culture fluid is retained and the cells are discarded.

In *E. coli*, when pressure is used to break the cells open to release cellular proteins, the pharmaceutical protein might be 20% of the protein quantity released. In mammalian cell culture (most typically CHO cells), the cells typically are engineered to secrete the protein, and the harvested cell culture fluid comes directly from the centrifuge or filter that removed the cells to be subjected to multiple rounds of chromatography.

The bioseparation technique, chromatography, exploits the unique properties of the pharmaceutical protein to separate it from other proteins in the solution. Different types of chromatography separate proteins according to different properties:

- Size Exclusion (gel filtration or desalting) Chromatography separates proteins according to their size
- Ion Exchange Chromatography separates proteins according to their charge
- Hydrophobic Interaction Chromatography separates proteins according to their hydrophobicity
- Affinity Chromatography separates proteins according to some specific property that they possess
- Mixed Modal Chromatography can be manipulated to separate proteins according to two properties, such as charge and metal affinity

During chromatography purification, the protein containing solution is passed through a column that contains a resin that binds protein of a particular characteristic; for example, an anion exchange chromatography column employs positively charged resin to bind negatively charged proteins and allows neutrally charged and positively charged proteins to pass through. If the protein is an antibody, purification uses a resin coupled with a protein that specifically binds antibodies, such as protein A. The protein will typically be subjected to three, sometimes up to five, different chromatography steps. Contaminating proteins are discarded at each step until the protein is pure enough to place into the final vial.

All of these chromatography steps are carried out at a very large scale. The chromatography columns can be five meters in diameter and can contain several million dollars of chromatography resin. At this scale, it may take thousands of liters of chromatography buffers to wash the proteins during the chromatography processes

After chromatography, the purified protein is often subjected to ultrafiltration. In ultrafiltration the protein-containing solution is passed through a filter of a defined pore size. This step assures sterility and is used to place the protein into its final solution in a process called formulation.

Formulation, Fill, and Finish

Once the protein has been isolated to 99.99% purity, it must be packaged in such a way that prepares it for injection into the patient. Again, all protein pharmaceuticals currently are injected; the ability of stomach acid to denature proteins renders them useless in pill form. To be injected, the proteins must be placed in a solution with the right components; the design of this solution that delivers the drug is called formulation.

Proteins must be surrounded by the same salt concentration as the blood. (If the protein is delivered in pure water, the water would cause the patient's red blood cells to explode.) In addition, the protein is placed in a solution that contains added molecules that buffer the pH, prevents aggregation (proteins sticking together), and helps protect the protein from denaturation.

Once formulation is completed, an automated system dispenses the protein into sterile vials. This process occurs in a specialized facility and is carried out by gowned personnel with specialized training. The sterilized filling room has air that has been filtered to a high degree to remove contaminating microbes. The personnel gowns-up in a manner leaving no skin exposed to avoid shedding bacteria from their bodies. They learn behaviors that help prevent bacteria from inadvertently being placed in the vials. The protein solution within the vials must be absolutely sterile, since it will be injected directly into the bloodstream.

A precisely measured amount of the protein solution is aliquoted into the sterile vial. The filling machine then covers the vial with a sterilized cap and a closure around that cap.

Many proteins are lyophilized, or freeze-dried, into a powder. Lyophilized proteins are more stable than liquid formulations. Lyophilized proteins have the disadvantage that the physician must carry out the extra step of reconstituting the drug by injecting sterile water into the vial and mixing before withdrawing it to inject into the patient.

Labeling and Packaging

An automated process applies labels to the drug-containing vials. Because a mislabeled drug would represent a tremendous danger, the labeling process is carefully controlled. (Mislabeled drugs are one of the top reasons for the recall of a drug.) Each vial is hand-inspected before being placed in a box. The paper insert with the detailed description of the drug's characteristics must be packaged with the drug. Manufacturers' carefully track the destination of each vial in case of a drug recall.

Regulation

In the United States, the Food and Drug Administration, an agency of the Department of Health and Human Services, which is part of the executive branch of the Federal government, regulates the manufacturing of pharmaceuticals. Some would argue that pharmaceutical manufacturing is so carefully monitored that the only industry with greater oversight might be the production of nuclear energy. A pharmaceutical manufacturing company must prove the purity, safety, potency, identity, dosage, and efficacy (effectiveness) of each batch of drug. Each drug production process is examined closely for its ability to produce a consistent product. The stability of each drug must be determined to give it an expiration date.

The law requires that a Quality Assurance unit exist independent of the manufacturing department. This unit tests raw materials, in-process samples, environmental monitoring, utilities, and the final product.

Federal law requires that each pharmaceutical be manufactured under current Good Manufacturing Practices (GMPs). These defined sets of standards and procedures assure that the manufacturing plant follows each production process in exactly the same way for every batch.

To market a drug, a pharmaceutical company must take that drug through clinical trials. The drug is first tested in several model animals. If the pre-clinical animal studies indicate that the drug appears safe, then the drug goes to human clinical trials. These tests involve hundreds of physicians treating thousands of patients. Clinical trials take seven to ten years and cost the company hundreds of millions of dollars. In the end, the FDA must be convinced that the drug is both safe and effective before they issue a formal approval that allows the company to sell the drug in the United States.

Current Good Manufacturing Practice (cGMP) dictates every element of the manufacturing process. The regulations that communicate the provisions of cGMP are found in the Code of Federal Regulations, which containd the following parts:

General Provisions
Organization and Personnel
Buildings and Facilities
Equipment
Control of Containers and Closures
Process and Production Controls
Packaging and Labeling Controls
Holding and Distribution
Laboratory Controls
Records and Reports
Returned and Salvaged Drug Products

The manufacturing buildings must be built in a specific way; their design separates manufacturing from other parts of the biotech company. The regulations, and thus building design, ensure that all personnel gown up and these personnel flow through an airlock before entering the manufacturing areas. The air must be filtered to reduce bacteria, and all surfaces are cleaned regularly. The FDA inspects and approves the facilities before any manufacturing can begin, a process that can take up to four years.

When technicians grow cells and purify proteins, each part of the process follows a Standard Operating Procedure (SOP). Each procedure is carefully designed and validated, that is, proven to perform the function that it claims to perform. During the carrying out of each procedure, a technician will keep strict records by filling out a production batch record. The paperwork may be checked by an FDA official during a formal inspection of the plant, which happens at least every two years.

The Quality Assurance (QA) department checks the product at every stage of production. According to law, it must operate independently of manufacturing and serve as an independent entity to assure that the manufacturing department is following cGMP regulations to the letter. QA operates several Quality Control laboratories to check raw materials coming into the plant, sample and test the product at every stage of the process, and extensively test the final product for purity, identity, potency, consistency, sterility, and stability. The QA department monitors the quality of the air in every room and checks for the presence of bacteria on every surface. Any violation of the provisions of cGMP will render the drug “adulterated” and will prevent the product from being released for consumption. These very strict provisions assure a safe and effective product for the consumer.

Appendix I: Glossary

Absorbance – The amount of ultraviolet light that is not transmitted (absorbed) through a sample. 280 nm is the ultraviolet wavelength absorbed by the tryptophan, tyrosine, and phenylalanine groups in a protein in solution. The relationship between absorbance of the ultraviolet light and protein concentration is linear.

Affinity chromatography – A chromatography method of separating molecules based on a highly specific biologic interaction such as that between an antigen and antibody, enzyme and substrate, or receptor and ligand.

Anion – A negatively charged ion or biomolecule.

Anode – Positive electrode; attracts negative ions.

Anion exchange chromatography – A chromatography method where a positively charged chromatography resin binds negatively charged molecules, or anions.

Aromatic amino acid groups – Amino acid groups that contain an aromatic ring in the R-group are nonpolar and absorb untraviolet light at 280 nm. Aromatic amino acids are tyrosine (Y, Tyr), tryptophan (W, Trp), and Phenylalanine (F, Phe).

β-mercaptoethanol – (BME) is a reducing agent used to break the diusulfide bonds of proteins, thus disrupting the tertiary and quaternary structure of the protein. It helps to linearize the protein in prepartion for electrophoresis.

Bradford Protein Assay – A test used to measure protein concentration in a sample. The assay relies on the shift in absorbance of Coomassie Brilliant Blue G-250 dye. The dye reacts with mainly basic amino acid and aromatic amino acid groups.

Buffer – The liquid that is used to dissolve the biomolecules that will be applied to the chromatography column.

Cathode – Negative electrode; attracts positive ions.

Cation – A positively charged ion or biomolecule.

Cation exchange chromatography – A chromatography method where negatively charged chromatography resin binds positively charged molecules, or cations.

Cell lysate – All the components, soluble and insoluble, of a cell that have been broken open.

Centrifugation – Spinning a mixture at very high speed to separate heavy and light particles. In protein expression and purification, centrifugation results in a “pellet” found at the bottom of the tube, and a liquid “supernatant” that resides above the pellet.

Charge density – The protein’s ratio of charge to mass.

Chromatogram – A visual output of the chromatographic separation. Peaks on the chromatogram indicate when samples are eluting from the column.

BIO-RAD

Chromatography – A process for separating complex mixtures of proteins or other molecules. In the case of column liquid chromatography, separation is accomplished by passing a liquid mixture over a column containing a solid matrix. The properties of the matrix can be tailored to allow the selective separation of one kind of molecule from another. Properties include hydrophobicity, molecular size, and charge.

Column – A plastic or glass cylinder that is densely filled ("packed") with small, porous spheres or beads.

Column bed – The volume of beads packed within the chromatography column.

Column volume – (CV), the volume of buffer equal to the volume of resin packed in a chromatography column.

Dalton – One Dalton equals the mass of a hydrogen atom, which is 1.66 x 10^{-24} grams. A DNA kilobase pair has a mass of approximately 660 kD.

Decant – Gently removing liquid or buffer from a column or other vessel so as not to disturb the resin or sediment bed.

ΔOD/min – Change in optical density, or measured absorbance, per minute. For this series, it is used in calculating the activity of an enzyme.

Denaturation – Process of disrupting a protein's structure.

DHF – Dihydrofolate or dihydrofolic acid.

DHFR – Dihydrofolate reductase.

Disulfide bond – S—S (sulfide—sulfide) bond between amino acids in a polypeptide chain; contributes to tertiary and quaternary structure of proteins.

Dithiothreitol – (DTT) is a reducing agent used to break the disulfide bonds of proteins, thus disrupting the tertiary and quaternary structure of the protein. It linearizes and prepares the protein for electrophoresis.

DNase – Enzyme that specifically breaks down DNA.

Electrophoresis – Means "to carry with electricity." It is the migration of charged molecules in an electric field toward the electrode with the opposite charge.

Eluate – The solution of buffer and biomolecules from elution.

Elute – The removal of a bound molecule from a chromatography resin.

Elution buffer – The chromatography buffer containing chemicals used for the removal of a bound molecule from a chromatography resin.

Exclusion limit – The upper size limit for molecules that cannot penetrate the pores of the porous beads. See *Fractionation range*.

Fraction – A tube that contains material that has flowed through the chromatography column. Multiple tubes or fractions are collected during each chromatography run.

Fractionation range – For size exclusion chromatography gels, the fractionation range is the range of molecular weights that will enter the gel. A fractionation range of 1,000–6,000 kD will have pores large enough for molecules in that size range to pass through. Molecules larger than 6,000 kD will be excluded from entering the gel. The fractionation range is sometimes referred to as the “exclusion limit.”

Frit – Bed support of the chromatography column.

Gel electrophoresis – Technique used to separate, or sieve, molecules that carry electric charges. The molecules separate from each other according to the different rates at which they migrate through an electric field set up in a gel that is soaked in a chemical solution.

Glycoslyation – An enzymatic process that adds glycans, or sugars, to a protein or other organic molecule. Glycosylation is known to aid in protein folding.

GST-tag – Glutathione-s-transferase, an enzyme that binds to the substrate glutathione, is a small amino acid sequence (27 kD mw) that is added to the sequence of a recombinant protein. Glutathione is bound to chromatography resin and thus used to purify proteins with the GST-tag. GST is also added to recombinant proteins to aid in solubility.

His-tag – A series of histidine residues (usually 6) fused to a protein that aids protein purification because of its strong binding to nickel (IMAC) columns. Also known as a “polyhistidine tag.”

Hydrophilic – A molecule that has a strong affinity for water, “water loving.”

Hydrophobic – A molecule that has a strong dislike for or is insoluble in water, “water fearing.”

Hydrophobic interaction chromatography – A chromatography method that separates molecules based on their level of hydrophobicity.

IMAC – Immobilized Metal Affinity Chromatography; a chromatography method where the affinity of histidines to metals, such as nickel, is used to purify proteins tagged with polyhistidine sequences.

Inclusion body – Aggregated and precipitated expressed proteins found inside bacteria induced to make high levels of recombinant protein.

Insoluble – The parts of the cell that are not dissolved in water or buffer.

Ion exchange chromatography – A chromatography method where the charge of the molecule is exploited to bind to oppositely charged chromatography media.

Isoelectric point – (pI) The pH at which a molecule has a net charge of 0.

Laemmli sample buffer – The first, and most common, sample buffer used for protein electrophoresis. First described in 1970, this buffer consists of 62.5 mM Tris buffer to maintain pH conducive to electrophoresis; 10% glycerol to increase density of the protein so that it stays sunk in the gel well, 2% SDS to equalize the protein charge; 5% DTT (or BME) can be added to reduce disulfide bonds in the protein; and 0.01% bromophenol blue, which gives the sample color.

Ligand – A molecule, such as an antibody, enzyme, or protein tag, with specific affinity for another molecule.

Loading buffer – (Equilibration buffer) The buffer used to add sample to a chromatography column. The loading buffer is formulated to exploit properties of the biomolecule of interest for the particular chromatography resin and allows the biomolecule to bind to the resin.

Lowry Protein Assay – An absorbance test used to measure the protein concentration in a sample. The assay relies on the reaction of protein with alkaline copper tartrate and Folin and the change in color of the sample.

Luer-Lok – A standardized system of low-pressure fluid fittings used for making leak-free connections between a male-taper fitting and its mating female part on medical and laboratory instruments.

Lyse – To break open a cell.

Mixed-mode chromatography – A method of chromatography in which the resin used utilizes multiple chromatography techniques, such as a combination of anion and cation exchange properties, to resolve a mixture of biomolecules.

Mobile phase – The liquid, solvent, buffer, or sample, that moves through the stationary phase or chromatography resin.

Pellet – The insoluble components of a lysed cell that settle in the bottom of the centrifuge tube during centrifugation of the cell lysate.

Polyhistidine tag – A series of histidine residues (usually 6) fused to a protein that aids protein purification because of its strong binding to nickel (IMAC) columns. Also known as a "His-tag."

Prepacked cartridge – A chromatography column that is prepacked with chromatography resin. The cartridge is capped, keeping the resin contained. The cartridge has a quick-connect fitting on its bottom and top allowing for easy connection to a syringe or chromatography pump or system.

Protein assay – A test using the shift in absorbance of colorimetric dye to determine protein concentration. A spectrophotometer is required to perform the assay. Two popular protein assay methods are the Bradford Protein Assay and the Lowry Method. Which method to use is based on compatibility of the method with reagents in the sample buffer, as well as sensitivity.

Resin bed – The settled, packed chromatography resin in a column.

Sample – A mixture of biomolecules that is dissolved in a buffer and which is applied to a chromatography column.

SDS-PAGE – Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis is a technique of separating a mixture of proteins based solely on its size with the use of a gel medium and an electric field.

Size exclusion chromatography – Also known as gel filtration chromatography, desalting chromatography, and buffer exchange chromatography, size exclusion chromatography uses beads containing tiny holes, or pores, to separate a mixture of molecules based on its size. The larger molecules, which cannot fit through the holes, pass quickly around the beads whereas smaller molecules enter the holes and pass through the column more slowly.

Sodium Dodecyl Sulfate – Also known as SDS, it is a strongly anionic detergent used to coat proteins and give them an overall negative charge.

Soluble – The parts of the cell that are easily dissolved in water or buffer.

Spin column – A small chromatography column that fits into a standard bench-top centrifuge and allows for quick purification of biomolecules. The column may be prepacked or be empty, allowing the user to choose the resin.

Stationary phase – In chromatography, the stationary phase is the chromatography resin that is used to bind the molecule of interest or separate the sample mixture.

Tertiary structure – 3-dimensional shape of a folded polypeptide, maintained by disulfide bonds, electrostatic interactions, and hydrophobic effects.

BIO-RAD