RECOMBINANT PROTEINS IN RESEARCH
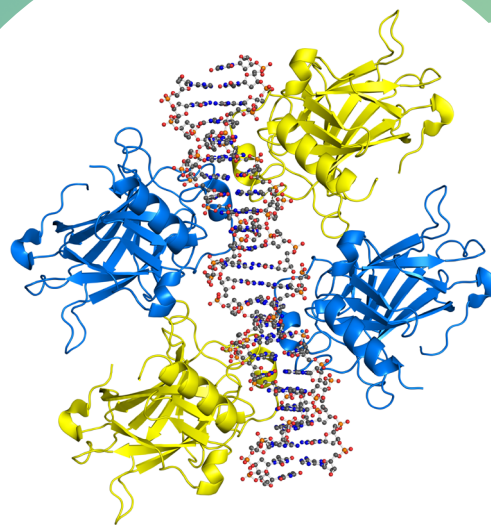

Figure 1. *p53 tetramer binding to DNA.*

Introduction

Proteins play an integral role in human health. As such, basic, applied, and clinical research aim to understand how protein expression, localization, post translational modifications (PTMs), and interactions regulate homeostasis and disease (see yellow box). Our current understanding of how proteins affect cellular phenotype is based on numerous applications, such as affinity purification, western blotting, and in vitro functional assays. Many of these techniques employ recombinant proteins, or proteins expressed in exogenous host systems. In this blog, we will discuss 1) how recombinant proteins are made, 2) the advantages of using recombinant proteins in proteomics research, and 3) how recombinant proteins are used in various proteomics applications.

RayBiotech
Empowering your proteomics

Mutated protein in disease

Alterations to native protein expression or function may contribute to disease initiation, disease progression, or treatment resistance. One well-studied protein to illustrate this point is the tumor suppressor, p53, which is the most frequently mutated protein in cancer.[1] Upon DNA damage, p53 translocates into the nucleus where it forms a homo-tetramer that binds DNA as a transcription factor to mediate cell cycle arrest or apoptosis (Figure 1). p53 acetylation and phosphorylation modulate its activation and DNA binding.[2] Some p53 mutations act in a dominant-negative manner, with only one molecule abolishing the transcriptional activity of the entire p53 tetramer.[3] Other mutations – both loss-of-function and gain-of-function mutations – alter which DNA promoter regions the p53 binds.[1] By inhibiting the activation of pro-apoptotic genes, p53 mutations result in long-lasting cancer cells and contribute to chemotherapy resistance.

How recombinant proteins are made

The first step of making a recombinant protein is synthesizing the gene-of-interest. The nucleic acid sequence is important to consider since one amino acid is encoded by different 3-nucleic acid sequences called codons. For example, four codons (CTT, CTC, CTA, CTG) will be translated to the amino acid residue, leucine. The efficiency of protein translation is improved by optimizing the nucleic acid sequences that reflect the codon bias (due to the varying tRNA pools) of the chosen expression system.[4]
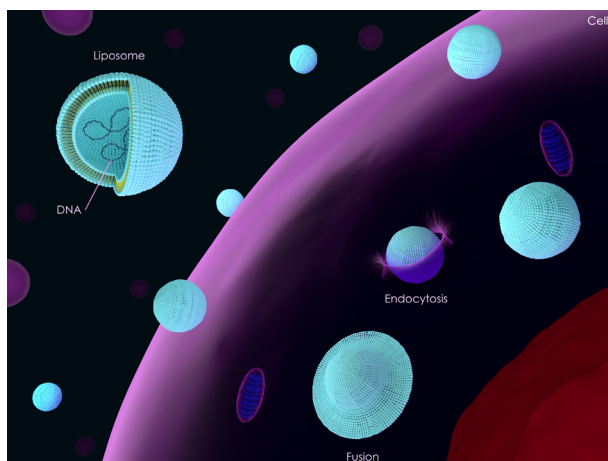

Figure 2. *Delivery of a cDNA plasmid into an expression system using liposomes.*

The synthesized gene is then cloned into an expression or viral vector.[5,6] An expression vector is a cDNA plasmid that includes a promoter sequence and an antibiotic resistant gene. It also often has a sequence at the N- or C-terminus encoding a fusion tag for downstream protein purification or identification. The antibiotic resistant gene enables the selection of cells carrying the plasmid in antibiotic-based media. The viral vector flanks the gene-of-interest by viral sequences, which eventually result in the gene being inserted into the host chromosome permanently.

After cloning, genes encoding for the gene-of-interest are introduced into the host system, either transiently by a cDNA plasmid (i.e., transformation, transfection) or permanently by viral infection (i.e., transduction).[5] In either case, the plasmids encoding the gene-of-interest passively enter the cell through its membrane, which has been temporarily compromised (i.e., "holy") via calcium phosphate or electroporation. Liposomes are also employed to deliver DNA to cells (Figure 2).[7] For transformation (into bacteria) or transfection (into mammalian cells), the cell will express the recombinant protein. Transduction, on the other hand, is a multi-step process. First, a cell is transfected with numerous plasmids encoding the gene-of-interest along with various viral particles. Second, the gene-of-interest is then packaged into an infectious virion via the expressed viral particles. Third, the viral package is then collected and used to infect the expression host system.

Various exogenous host systems are used to express recombinant proteins, such as *Escherichia coli*, insect cells, yeast, or human cells. *Escherichia coli (E. coli)* is an attractive system because the cost is low and the doubling rate of the bacteria is fast (i.e., 30 min), which means that a lot of recombinant protein can be produced with minimal expense (Table 1).[6] Notably, *E. coli* do not have the same chaperone proteins as a human cell, which means that a recombinant human protein may not be folded properly or have native function. Improperly-folded proteins may result in the expressed protein falling "out of solution" as insoluble aggregates, which require denaturation using denaturant such as > 6 M urea to go back into solution. To remain soluble, either the protein must remain in its denatured state or sometimes the protein can be refolded in buffer (e.g., phosphate-buffered saline). Importantly, "refolding" occurs without the aid of folding chaperones, thus increasing the risk that the protein is not folded into its native structure. *E. coli* also have difficulties making large proteins (> 150 kDa) and do not have PTM capability. However, not all applications require native conformation or activity, which is discussed later in this blog.

Table 1. *Comparison of* E. coli *and mammalian expression systems. "Mammalian" refers to human HEK293 cells.*

|	CHARACTERISTICS	BACTERIA	MAMMALIAN CELLS (TruExp™)
GENERAL FEATURES	**Expression Level**	Low-Medium-High	Low-Medium-High
	Cell Doubling Time	Rapid (30min)	Slow (24hr)
	Cost	Low	High
	Time Line	3 Weeks	4-5 Weeks
POST-TRANSLATIONAL MODIFICATIONS	**Protein Folding**	Not Reliable	Very Reliable
	Glycosylation	None	Complex, authentic
	Phosphorylation	None	Yes
	Acetylation	No	Yes
	Acylation	No	Yes
	gamma-carboxylation	No	Yes
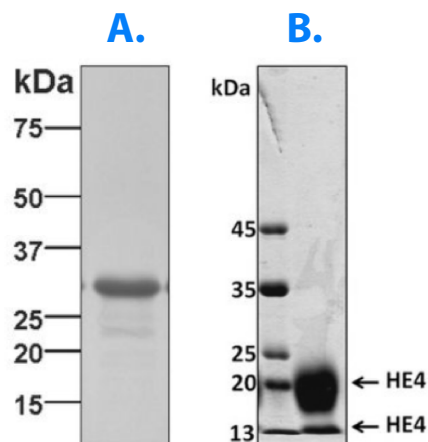
Figure 3 (left) *SDS-PAGE analyses of purified HE4 expressed in E. coli or human HEK293 cells. A) HE4 expressed in E. coli with N-terminal 6x histidine and thioredoxin tags. The protein migrates at ~30 kDa, the estimated molecular weight of the tagged HE4 protein. B) HE4 expressed in human HEK293 cells with a C-terminal 6x histidine tag. The protein migrates at the expected molecular weight of ~14 kDa. The band at ~20 kDa represents the glycosylated form of HE4.*

The HEK293 cell line, derived from human embryonic kidney cells, can also be used as an exogenous host system to express recombinant proteins, although it is more expensive expression system than *E. coli* (Table 1). Notably, HEK293 cells have native chaperones for folding human proteins and PTM ability.[8] For example, human epididymis protein 4 (HE4) is a 13 kDa polypeptide with a well-known glycosylation site at asparagine residue 44, which may aid in the protein's secretion from cells.[9] HE4 expressed in *E. coli* (Figure 3A) migrates at the expected molecular weight, whereas HE4 expressed in human HEK293 cells (Figure 3B) has a 6 kDa migration shift, which has been attributed previously to the glycosylation at asparagine residue 44.

Regarding expression efficiency, neither *E. coli* nor human HEK293 cells is the winner. Protein expression in both systems can range from low to high, and it is impossible to predict the expression efficiency since it varies from protein-to-protein. It is also difficult to know beforehand whether a protein expressed in E. coli will form insoluble aggregates. Some proteins cannot be expressed in either system because they are toxic to the cell.[6]

The recombinant protein is separated from the other proteins in the cell lysate by immunoaffinity or affinity purification. In immunoaffinity purification, an antibody is used to bind to and pull out the specific protein-of-interest.[10] However, not all proteins-of-interest have a corresponding antibody that are compatible with this application. This approach is also inconvenient for high throughput experiments in which many proteins must be purified since a reliable antibody specific to each protein-of-interest must be employed. Therefore, most recombinant proteins are purified through their fusion tags.[11] Immunoaffinity purification targeting the fusion tag can be employed, although affinity purification based on the unique characteristics of the tag is more often performed. For example, histidine-tagged proteins are purified using a nickel column. Proteins with a maltose binding protein (MBP) tag are purified using amylose. Flag-tagged proteins are purified using an anti-Flag antibody. Histidine, MBP, and Flag tags are just a few of the options available. Another advantage to using a fusion tag is that protein expression can be ascertained via western blotting using an easily obtainable antibody to the fusion tag. Both immunoaffinity and affinity purification often result in highly pure proteins (i.e., > 95%), although it is important to mention that the purity depends on the specificity of the antibody and type of affinity purification that is used, respectively.

Advantages of recombinant proteins

Native proteins, or proteins expressed endogenously, can and are used in many biochemical applications. They have native structure, PTMs, and activity. Why then are recombinant proteins used so much in proteomics? The advantages of using recombinant proteins include:
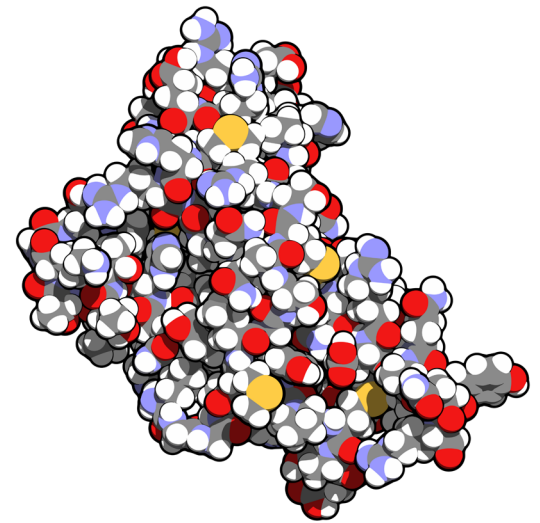
- **The amount of recombinant protein can be controlled.** Some downstream assays require milligrams or even grams of purified protein. Getting this amount of native protein will be difficult and expensive.

- **Proteins can be purified by way of a universal tag.** This enables proteins without a specific antibody to be purified and studied. Purification via a fusion tag also simplifies high throughput experiments in which many proteins must be purified.

- **Protein expression can be ascertained easily by way of a universal tag.** This enables proteins without a specific antibody to be analyzed with western blotting.

- **Approval by the Institutional Review Board (IRB) is NOT required.** Since obtaining native human proteins requires the participation of human volunteers, IRB approval is necessary. This results in more paperwork and waiting time for the paperwork to be approved.

- **Amino acid sequence can be altered easily with recombinant proteins.** Truncated, mutated, or elongated proteins-of-interest can be produced. Chimeric proteins or proteins with altered or improved function can also be made with little difficulty.

- **Unnatural amino acids can be incorporated into recombinant proteins.** This is made possible by using modified aminoacyl-tRNA synthases.[12] Unnatural amino acids can introduce unique properties to the protein, such as enhanced enzymatic activity or stability.

- **Recombinant proteins are cheaper than native proteins.** Once the expression vector has been synthesized and tested, the process of expressing and purifying recombinant proteins is cheaper per milligram than purifying native proteins.

- **Some applications do not require a properly-folded or active protein.** If a protein does not need to be folded, properly folded, or active, it is more economical to produce recombinant proteins than to purify native proteins. *E. coli*, in particular, is a cost-effective expression system.

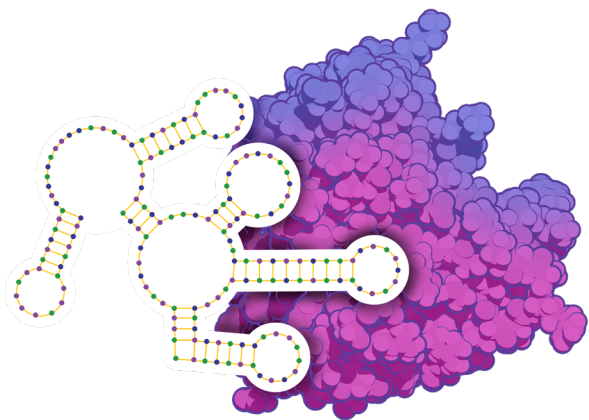# Applications of recombinant protein in research

Antibody production

Figure 4. *Aptamer binding to a protein. An aptamer adopts a three-dimensional structure to specifically bind to a protein or molecule.*

For antibody production, animal hosts are inoculated multiple times with purified protein to obtain primary and secondary immune responses.[13] Milligrams of protein are typically required for antibody production, although the amount is proportional to the size of the animal. For example, mice require a total of ~0.5 mg of protein whereas large animals like sheep require ~4 mg. The animal's blood is then collected when the levels of the immunoglobulin isotype, IgG, reach maximum serological levels. However, sample types other than blood can also be used to obtain antibodies (e.g., chicken eggs). The advantage of using a recombinant protein rather than a short amino acid sequence representing a fragment of the protein (i.e., peptide) is that the protein has more potential immunogenic sites. Thus, proteins are considered to be more immunogenic than peptides.

Aptamer development

Aptamers, or nucleic acid-based antibodies produced in vitro, also employ recombinant proteins (Figure 4). The recombinant protein is first incubated with trillions of random RNA or DNA oligos. Sequences that bind to the protein are then collected and amplified multiple times. It is also possible to perform negative selection during this step to ensure that the aptamer does not bind to proteins similar to the target-of-interest (e.g., wild-type versus mutant, isoform 1 versus isoform 2). Finally, aptamers that bind to the target protein are sequenced and the binding affinity of each individual aptamer is ascertained.

Want to learn more about aptamers? Read our blog comparing antibodies with aptamers here: *https://raybiotech.com/learning-center/aptamers-applications-and-strategies-for-development/*

Western blot controls

Western blots are often used to determine whether and how much a protein-of-interest is expressed in a sample. Recombinant proteins are used as a positive control for western blotting to ensure that the western blotting procedure (including the primary antibody) was successful. Recombinant proteins are also used to verify the migration pattern of the protein. The p53 protein, for example, is 44 kDa, but migrates at 53 kDa (hence, the name "p53").[14] It is not necessary to use a *purified* recombinant protein as a positive control. Indeed, unpurified recombinant proteins are a fraction of the cost of a purified protein.

ELISA standards

Enzyme-linked immunosorbent assay (ELISA) is a technique that measures the protein-of-interest's concentration. A purified recombinant protein is first quantified using a method like the bicinchoninic acid (BCA) assay. The recombinant protein is then diluted to create a standard curve to determine the exact concentration of the protein in the sample.

Protein interactions

Protein interactions with other proteins, DNA, or small molecules are characterized with recombinant proteins, either in solution or on a solid substrate.[15] These data may include information regarding binding partners, binding affinity, or kinetics using techniques like immunoprecipitation, isothermal calorimetry, and surface plasmon resonance (SPR). Antibody specificity and protein modifications during enzymatic reactions can also be characterized. It is important to consider the expression system and fusion tags of the recombinant proteins for these types of functional assays. For example, glutathione S-transferase (GST) may not be the best tag. In addition to its large size (26 kDa) that could block potential binding sites, its dimerization will affect the molarity and downstream binding kinetics and affinity.

Protein structure

X-ray crystallography, nuclear magnetic resonance (NMR), and mass spectrometry are used to determine protein structure. X-ray crystallography creates an electron map using a minimum of 5 – 10 milligrams of purified, crystallized recombinant protein.[16] NMR applies a magnetic field and radiofrequencies to 2 – 30 mg of purified protein to measure the distance between atomic nuclei.[17] Although less common than X-ray crystallography and NMR, mass spectrometry can also be used to determine protein structure.[18] Briefly, mass spectrometry measures the amount of deuterium-hydrogen exchange following the protein's incubation in a deuterium-based solution. The amount of incorporated deuterium reflects the location of the amino acid relative to the buffer, such that peptides on the surface of the protein will have more deuterium than those in the central part of the protein.

Cell culture experiments

Recombinant proteins are also employed in cell culture experiments. They help determine whether a protein can recover signaling after that pathway has been inhibited. This can provide insights into why a patient may be resistant to specific targeted therapies (e.g., small molecule inhibitor).[19] In the same way, recombinant proteins can help map the relative location (i.e., upstream, downstream) of the protein in a signaling pathway. Comparisons of wild-type and mutant recombinant proteins on cellular phenotype can help delineate their native function in homeostasis and disease.

Antibody profiling

Antibodies reflect the immune response. As such, antibody biomarkers have been identified for various diseases, including cancer, multiple sclerosis, and arthritis.[20-22] Antibody profiling is possible with protein arrays where recombinant proteins are printed onto a solid substrate in an arrayed and addressable format. Protein arrays are usually probed with serum or plasma, during which serological antibodies bind to their specific antigen. The array is washed to remove unbound antibodies, and a secondary fluorophore-conjugated anti-human antibody is applied for detection. Antibody biomarkers are then discerned by comparing the patterns across sample groups (e.g., healthy versus diseased).

Improved or altered functions

Gene synthesis, the first step of producing a recombinant protein, allows the protein sequence to be altered easily. This has enabled scientists to improve or alter protein function, most notably with enzymes. For example, a wild-type Xanthobacter dehalogenase binds and releases chloroalkanes. A His272Phe mutation results in a covalent interaction between the mutated dehalogenase and a chloroalkane ligand.[23] The mutated dehalogenase called HaloTag® is now used as a fusion tag, which is growing in popularity because it binds covalently with its ligand. Autolysis of the protease, trypsin, is reduced by reductive methylation.[24] The mutated trypsin is the most common protease used in bottom-up (i.e., peptide-to-protein identification) mass spectrometry.

Summary

Recombinant proteins are invaluable in proteomics research. In comparison to native proteins, recombinant have two primary advantages: 1) more flexibility regarding which and how much recombinant proteins are produced, and 2) purified recombinant proteins are cheaper than purifying native proteins. The selection of the appropriate recombinant protein for an application is not a menial task. The researcher should ask themselves: Does the protein need to be purified? Does the protein need to be active? Does the protein need to have specific PTMs? Will the tag or the location of the tag interfere with the downstream assay? Functional assays that need an active protein should either obtain proteins that have already been tested for activity or produced in a similar expression milieu. Proteins that require PTMs should be expressed in systems capable of PTMs. Large tags may block binding epitopes. The usefulness and versatility of recombinant proteins in research is reflected by the large (i.e., tens of thousands) and ever-growing offering of commercially-available recombinant proteins. Furthermore, recombinant proteins that are not on the market can be produced using custom protein services.

RayBiotech has a diverse catalog of native and recombinant proteins as well as a comprehensive custom protein service to help meet all of your research needs.

References

1. Kastenhuber ER, Lowe SW. Putting p53 in Context. Cell. 2017;170(6):1062-1078.

2. Gu B, Zhu WG. Surf the post-translational modification network of p53 regulation. Int J Biol Sci. 2012;8(5):672-684.

3. Chan WM, Siu WY, Lau A, Poon RY. How many mutant p53 molecules are needed to inactivate a tetramer? Mol Cell Biol. 2004;24(8):3536-3551.

4. Quax TE, Claassens NJ, Soll D, van der Oost J. Codon Bias as a Means to Fine-Tune Gene Expression. Mol Cell. 2015;59(2):149-161.

5. Kim TK, Eberwine JH. Mammalian cell transfection: the present and the future. Anal Bioanal Chem. 2010;397(8):3173-3178.

6. Rosano GL, Ceccarelli EA. Recombinant protein expression in Escherichia coli: advances and challenges. Front Microbiol. 2014;5:172.

7. Dean DA, Gasiorowski JZ. Liposome-mediated transfection. Cold Spring Harb Protoc. 2011;2011(3):prot5583.

8. Subedi GP, Johnson RW, Moniz HA, Moremen KW, Barb A. High Yield Expression of Recombinant Human Proteins with the Transient Transfection of HEK293 Cells in Suspension. J Vis Exp. 2015(106):e53568.

9. James NE, Chichester C, Ribeiro JR. Beyond the Biomarker: Understanding the Diverse Roles of Human Epididymis Protein 4 in the Pathogenesis of Epithelial Ovarian Cancer. Front Oncol. 2018;8:124.

10. Mayes EL. Immunoaffinity purification of protein antigens. Methods Mol Biol. 1984;1:13-20.

11. Kimple ME, Brill AL, Pasker RL. Overview of affinity tags for protein purification. Curr Protoc Protein Sci. 2013;73:Unit 9 9.

12. Wals K, Ovaa H. Unnatural amino acid incorporation in E. coli: current and future applications in the design of therapeutic proteins. Front Chem. 2014;2:15.

13. Murphy K, Travers P, Walport M, Janeway C. Janeway's immunobiology. 8th ed. New York: Garland Science; 2012.

14. Linzer DI, Levine AJ. Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. Cell. 1979;17(1):43-52.

15. Meyerkord CL, Fu H. Protein-protein interactions : methods and applications. Second edition. ed. New York: Humana Press; 2015.

16. Dessau MA, Modis Y. Protein crystallization for X-ray crystallography. J Vis Exp. 2011(47).

17. Marion D. An introduction to biological NMR spectroscopy. Mol Cell Proteomics. 2013;12(11):3006-3025.

18. Tsutsui Y, Wintrode PL. Hydrogen/deuterium exchange-mass spectrometry: a powerful tool for probing protein structure, dynamics and interactions. Curr Med Chem. 2007;14(22):2344-2358.

19. Straussman R, Morikawa T, Shee K, et al. Tumour micro-environment elicits innate resistance to RAF inhibitors through HGF secretion. Nature. 2012;487(7408):500-504.

20. Gavrila BI, Ciofu C, Stoica V. Biomarkers in Rheumatoid Arthritis, what is new? J Med Life. 2016;9(2):144-148.

21. Harris VK, Tuddenham JF, Sadiq SA. Biomarkers of multiple sclerosis: current findings. Degener Neurol Neuromuscul Dis. 2017;7:19-29.

22. Jaras K, Anderson K. Autoantibodies in cancer: prognostic biomarkers and immune activation. Expert Rev Proteomics. 2011;8(5):577-589.

23. England CG, Luo H, Cai W. HaloTag technology: a versatile platform for biomedical applications. Bioconjug Chem. 2015;26(6):975-986.

24. Rice RH, Means GE, Brown WD. Stabilization of bovine trypsin by reductive methylation. Biochim Biophys Acta. 1977;492(2):316-321.