# Quality control bioinformatics for recombinant protein production

***How might we confirm the purity of biomanufactured proteins to ensure that only the intended recombinant protein is being produced in a given batch?***

For decades, people with diabetes relied on insulin extracted from cows and pigs to regulate their blood sugar. However, this method was inefficient, costly, and sometimes triggered immune reactions. The advent of **recombinant DNA technology** in the 1980s revolutionized insulin production by enabling genetically modified bacteria to produce human insulin. This breakthrough transformed **biomanufacturing,** allowing for safer, purer, and more scalable production of life-saving proteins.

Recombinant DNA technology involves isolating a human gene of interest—such as insulin, erythropoietin (EPO), or tissue plasminogen activator (tPA)—and inserting it into a small circular piece of bacterial DNA called a **plasmid.** This recombinant plasmid is then introduced into fast-growing bacteria, typically *Escherichia coli* (E. coli), which serve as protein factories. Once inside the bacteria, the gene is expressed, leading to the mass production of the desired protein. The process is further refined through purification steps to ensure that only the correct protein is collected for medical use.

However, successful gene insertion does not always guarantee proper protein production. Mistakes such as mutations, contamination with unintended sequences, or loss of the recombinant plasmid could compromise the quality and efficacy of the final product. To maintain rigorous quality control, scientists use bioinformatics tools like NCBI BLAST to verify the accuracy of bacterial DNA, ensuring that the inserted gene is correct and free from harmful alterations.

## Scenario

A biotech company has developed a new strain of genetically modified bacteria, claiming it produces insulin more efficiently than current strains. To scale up production of recombinant proteins, these bacteria are grown in massive 6,000-gallon fermentation tanks. However, before mass production can proceed, it is critical to confirm the purity of each tank and ensure that only the intended recombinant protein is being produced in a given batch.

In addition to insulin, the company also produces two other recombinant proteins:

- **Erythropoietin (EPO):** A hormone produced by the kidneys that regulates red blood cell production. Recombinant EPO is used to treat anemia in patients with chronic kidney disease.
- **Tissue Plasminogen Activator (tPA):** A life-saving enzyme used to dissolve blood clots in patients suffering from acute ischemic strokes. Administering tPA shortly after a stroke can restore blood flow and reduce brain damage.

While all three recombinant proteins are essential biopharmaceuticals, maintaining purity in production is crucial to avoid cross-contamination and ensure product safety. As part of routine quality control, plasmid DNA has been purified and sequenced from five active fermentation tanks.

## Your task

As part of the quality control team, your job is to analyze the DNA sequencing results to:

1. Identify the most common recombinant DNA present within each fermentation tank.
2. If your sequence is insulin, check for any mutations that could result in a defective, ineffective, or contaminated batch of protein.

By using the bioinformatics tools NCBI BLAST, you will determine whether these genetically modified bacteria are correctly producing human insulin or if further investigation is needed before mass production can continue.

## Materials

- A computer with internet access
- Tank sequence

## Procedure

1. Visit the NCBI website: ***ncbi.nlm.nih.gov***
2. On the homepage, look for the “BLAST” option in the right-hand menu and click it.
3. Select “Nucleotide BLAST” from the list of available tools.
4. Copy and paste the sequence you’ve been assigned into the large text box labeled “Enter Query Sequence.”
5. Under “Database,” ensure the option is set to Standard databases (nr etc.).
6. Scroll down and click the blue “BLAST” button.
7. Wait for the BLAST program to complete the search.
8. On the results page, there will be a section called “Descriptions’ with a table listed as ‘Sequences producing significant alignments”. These are the matches (called “hits”) of your sequence to known nucleotide sequences in the database.
9. Click on the description to see the alignment between your sequence and the database sequence.
10. Record the name of the top hit sequence and other relevant details provided by the alignment shown.
11. Click on the unique “Sequence ID” at the top of the alignment to learn more about the database sequence.

## Reflection

1. What human protein does your sequence correspond to?
2. If the tank has insulin production, did you notice any differences with the aligned sequence within the NCBI database? If so, please describe the difference.
3. If your tank was not producing insulin, describe what protein product was found.
4. Why is it important to verify that the human insulin gene has been correctly inserted into the bacterial DNA before mass production?
5. How does bioinformatics, specifically tools like NCBI BLAST, help scientists ensure the accuracy and safety of recombinant protein production?
6. What are some advantages of using genetically modified bacteria for protein production compared to traditional methods, such as extracting insulin from animal sources?