Department of BSBE
Indian Institute Of Technology Guwahati

Recombinant production of Insulin

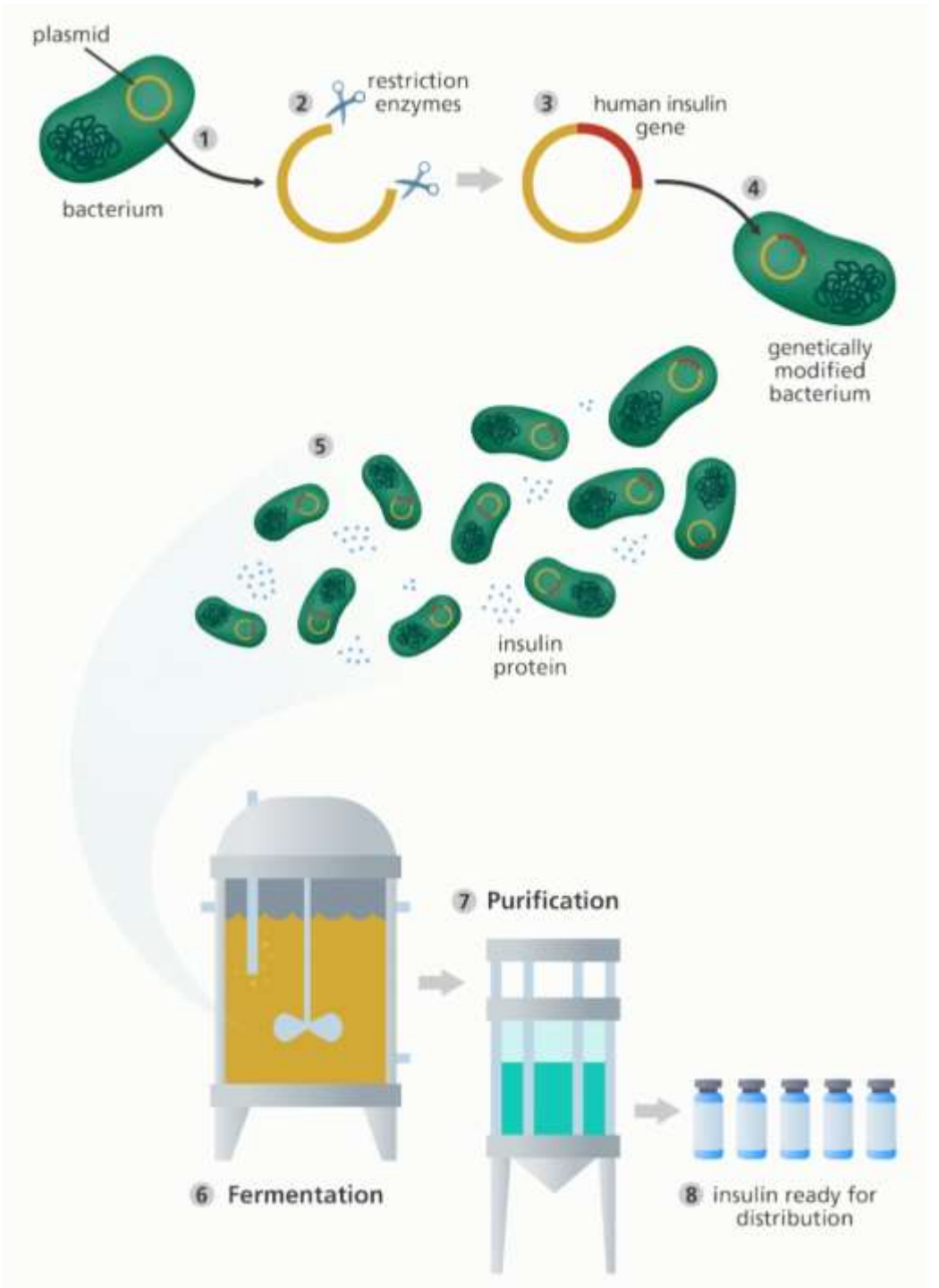
A case study

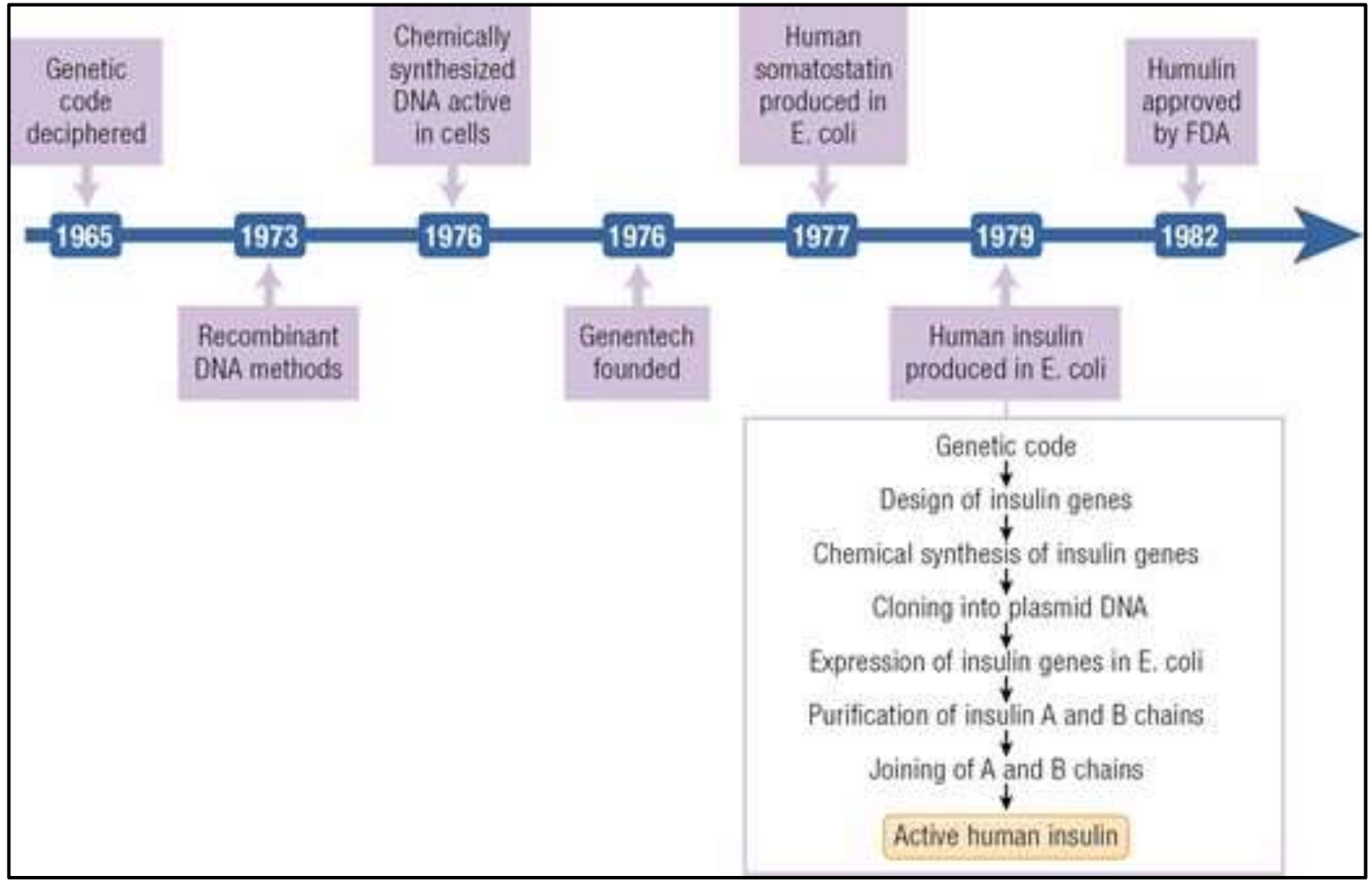
BT 207

Dr. Sanjukta Patra

Genetic engineering for Insulin production

- Diabetes is well managed by taking insulin. Earlier insulin was extracted from the pancreas of killed cattle and pigs. It had shortcomings.
- Another challenge was to cater to the ever increasing demand and large scale production.
- To overcome this, the production of insulin by recombinant DNA technology was done and it has proved to be very beneficial.
- In fact, it was the first recombinant medicine to be used in the USA.
- Biosynthetic insulin produced by rDNA technology is purer than animal insulin. It reduces the formation of antibodies against it.
- The genetically engineered insulin does not contain the C peptide.

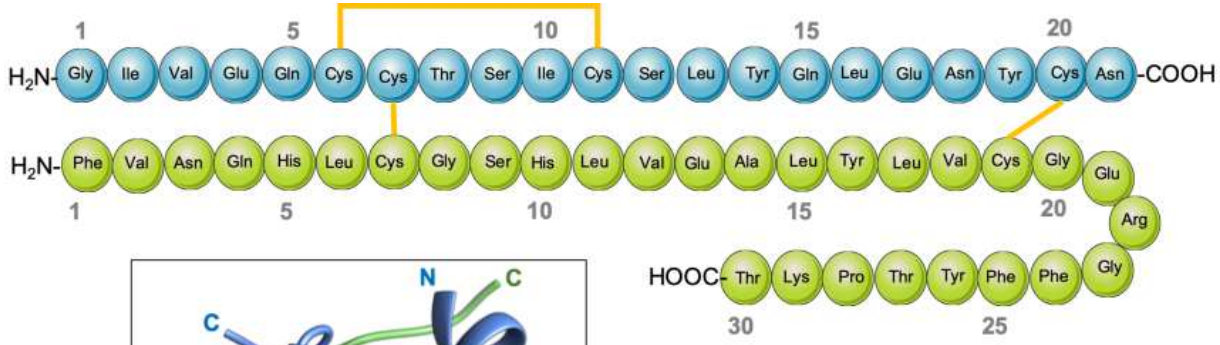


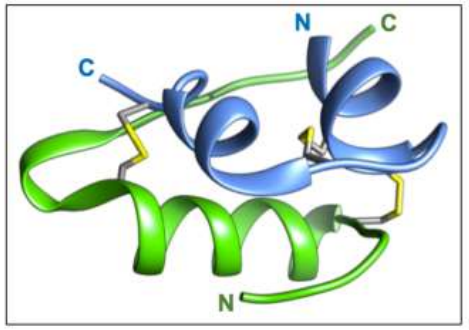


https://doi.org/10.1210/endrev/bnaa029

INSULIN

- Peptide hormone
- Secreted by beta cells of the pancreatic islets
- Decreased or absent insulin activity results in diabetes mellitus
- Gene length: 333
- Chromosome 11: single genetic locus, precursor
- Gene contains 3 exons; exon 2 encodes the signal peptide, the B chain, and part of the C-peptide, while exon 3 encodes the remainder of the C-peptide and the A chain
- Proinsulin is converted to insulin by the enzymatic removal of C segment (for 'connecting') peptide that connects the amino end of the A chain to the carboxyl end of the B chain.

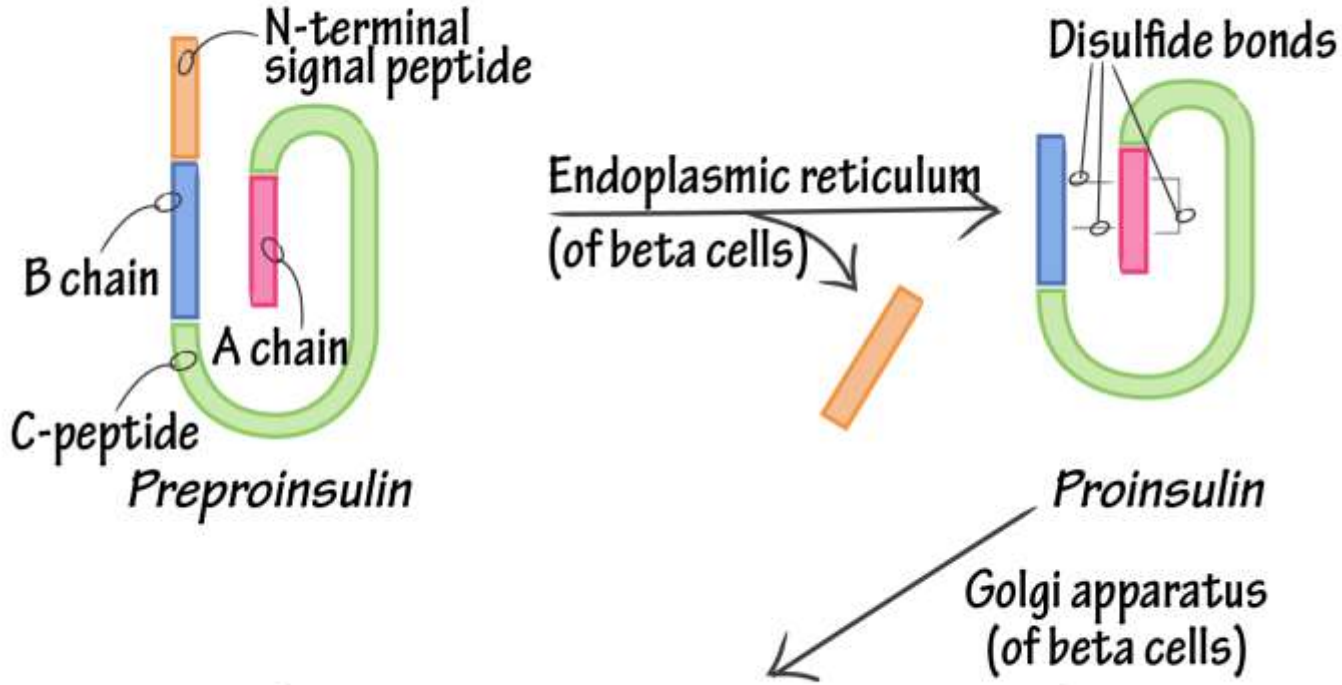




>ENA|CAA23828|CAA23828.1 Homo sapiens (human) preproinsulin
ATGGCCCTGTGGATGCGCCTCCTGCCCCTGCTGGCGCTGCTGGCCCTCTGGGGACCTG
ACCCAGCCGCAGCCTTTGTGAACCAACACCTGTGCGGCTCACACCTGGTGGAAGCTCT
CTACCTAGTGTGCGGGGAACGAGGCTTCTTCTACACACCCAAGACCCGCCGGGAGGC
AGAGGACCTGCAGGTGGGGCAGGTGGAGCTGGGCGGGGGCCCTGGTGCAGGCAGC
CTGCAGCCCTTGGCCCTGGAGGGGTCCCTGCAGAAGCGTGGCATTGTGGAACAATGC
TGTACCAGCATCTGCTCCCTCTACCAGCTGGAGAACTACTGCAACTAG

>sp|P01308|INS_HUMAN Insulin OS=Homo sapiens OX=9606 GN=INS PE=1 SV=1
MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED
LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN

Insulin Structure



Insulin

- $T_{1/2}$ ~ 6 minutes

C-peptide

- $T_{1/2}$ > insulin
- Marker of insulin synthesis & secretion

INSULIN IS THE FIRST EVER PHARMACEUTICAL PRODUCT OF rDNA TECHNOLOGY ADMINISTERED TO THE HUMANS.

BASIS OF RECOMBINENT INSULIN PREPARATION

- The basic technique consisted of inserting human insulin gene and the promoter gene of Lac operon on to the plasmids of *Escherichia coli* (Strain no.812). -The two chains separately synthesized are finally combined to form active insulin.

HOW THE RECOMBINANT HUMAN INSULIN IS MADE?

- The genes for Insulin Chain A and Chain B are separately inserted to the plasmids of two different *Escherichia coli* cultures. **The Lac operon system** consisting of inducer gene, promoter gene, operator gene and structural gene (β-Galactosidase) is used for expression of both the genes.

- The presence of Lactose in the culture medium induces the synthesis of INSULIN A and INSULIN B chains in separate cultures.
- The so formed Insulin chains can be isolated, purified and joined together to give exact human Insulin. The technique was improvised for improving yield and addition of signal peptide was done.

NEED OF SECOND GENERATION RECOMBINANT INSULIN

After injecting Insulin, the plasma concentration of Insulin rises slowly so patient must be injected at least 15 minutes before meal. -Decrease of Insulin levels is also slow , exposing the patients to a danger of Hyperinsulinemia. -All this is due to the existence of therapeutic Insulin as a hexamer (six molecule associated),which dissociates slowly to the biologically active dimer and monomer.

*ATTEMPTS HAVE BEEN MADE IN RECENT YEARS TO PRODUCE SECOND GENERATION INSULINS BY SITE DIRECTED MUTAGENESIS AND PROTEIN ENGINEERING.

SECOND GENERATION RECOMBINANT INSULIN

The second generation recombinant proteins are termed as MUTEINS. -A large number of Insulin Muteins have been constructed with an objective of faster dissociation of hexamers to biologically active forms. -Insulin Lispro is among these with modified amino acid residues at position 29 and 30 of the chain b of Insulin. Other 2nd generation Insulins are Glargine and Lente. *INSULIN LISPRO CAN BE INJECTED IMMEDIATELY BEFORE A MEAL AS IT ATTAINS THE PHARMACOLOGICALLY EFFICIENT LEVELS VERY FAST.

CHEMICALLY ALTERED PORCIN INSULIN

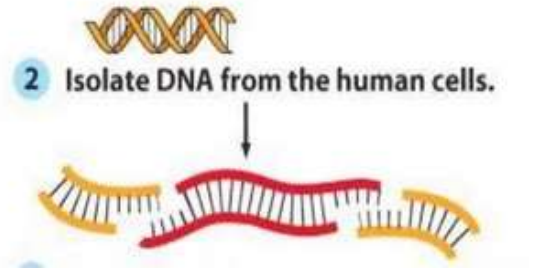
Porcin Insulin differs from human Insulin just by one amino acid (Alanine in place of Threonine) at the C- terminal of the B Chain of Insulin. -Biotechnologists have developed methods to alter the chemical structure of Porcin Insulin to make it identical to human Insulin. -This chemically modified Porcin Insulin can also be employed for treating Diabetes Mellitus.

Knowing more about Insulin from databases..

- **InterPro (Classification of protein families)**: https://www.ebi.ac.uk/interpro/entry/InterPro/IPR004825/
- **PROSITE**: https://prosite.expasy.org/PS00262
- **Pfam (protein family)**: http://pfam.xfam.org/protein/P01308
- **UniProt**: https://www.uniprot.org/uniref/UniRef100_P01308
- **Ensembl (genome annotation)**: https://asia.ensembl.org/Homo_sapiens/Transcript/Summary?g=ENSG00000254647;r=11:2159779-2161221;t=ENST00000250971
- **KEGG database**: https://www.genome.jp/dbget-bin/www_bget?hsa:3630

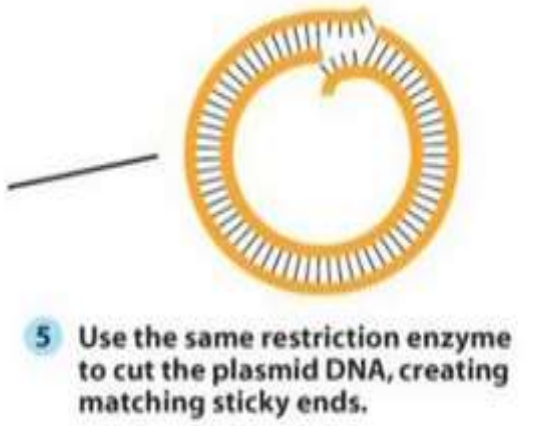
First Step (Preparing)

• The human gene is isolated. The mRNA is taken from the cell of islet of Langerhans. Messenger RNA is a molecule of RNA that encodes a chemical "blueprint" for a protein product. The isolated gene contains the code of the human DNA for the production of insulin. • The plasmid DNA of the bacterial cell is taken out of the cell. NOTE: *Escherichia coli* (E. Coli) bacteria is widely used in producing insulin but yeast may also be used.



Second Step (Cutting)

• The plasmid DNA of the bacteria is cut out producing plasmid ring which is an empty segment of the DNA. A Restriction Enzyme is an enzyme that cuts DNA at specific recognition nucleotide sequences known as restriction sites. A segment of DNA known as sticky ends.

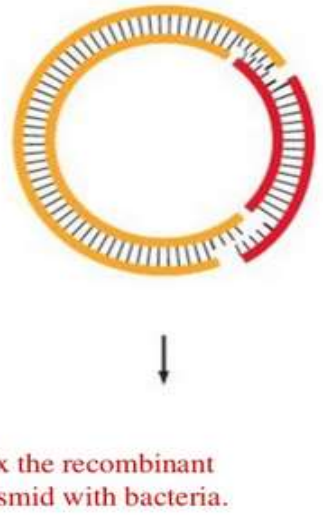


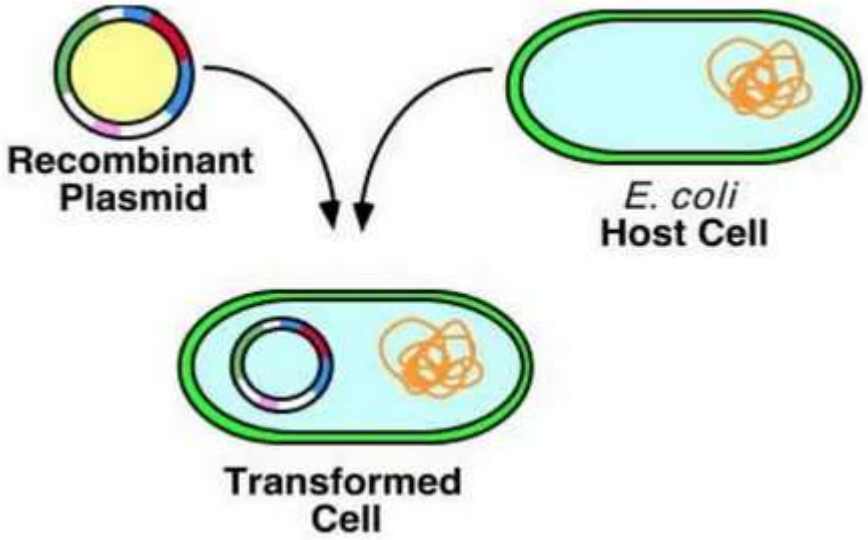
Third Step (Combining)

• With the plasmid ring open, the gene obtained from human cell that contains the code of protein responsible for the production of insulin is inserted into the plasmid ring and the ring is closed. • The human insulin gene is now combined with the bacterial DNA plasmid. Mix the recombinant plasmid with bacteria.

Fourth Step (Inserting)

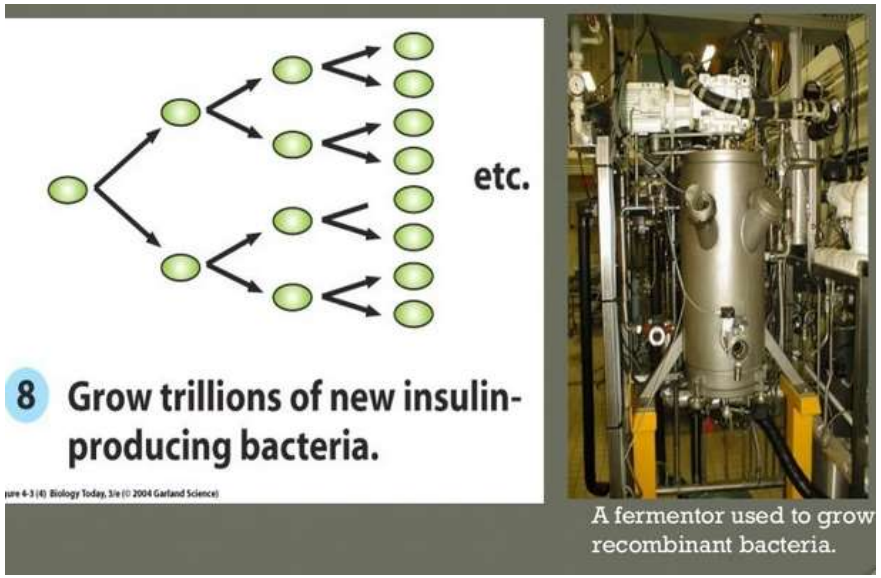
• Transformed: resulting DNA is inserted back to the bacteria.





Fifth Step (Production)

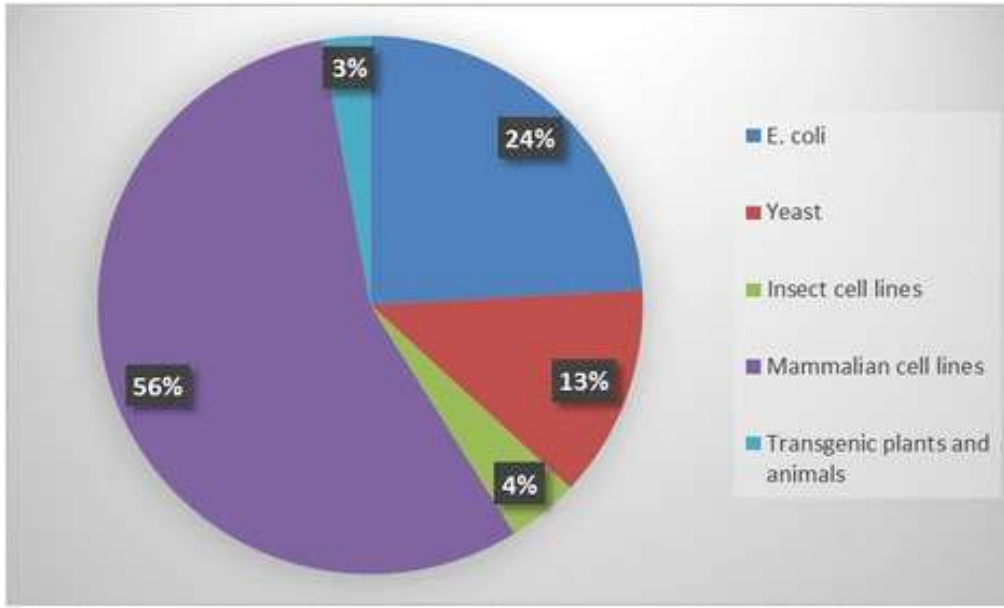
• The cells need nutrients in order to grow, divide, and live. While they live, the bacterial cell processes turn on the gene for human insulin and the insulin is produced in the cell. When the bacterial cells reproduce by dividing, the human insulin gene is also reproduced in the newly created cells.



There are more than 300 biopharmaceutical products including therapeutic proteins and antibodies in the market with sales exceeding USD100 billion.

Percentage of biopharmaceuticals produced in different expression systems

At present, insulin is being produced predominantly in *E. coli* and *Saccharomyces cerevisiae* for treatment of diabetic patients.



Systems employed as a host for large-scale production of recombinant insulin

1. *E.coli*
2. and yeast,
3. mammalian cells,
4. transgenic animals
5. plant expression systems

A.E. Coli expression system

Using *E. coli* expression system:

1. the insulin precursors (IP) are produced as inclusion bodies
2. and fully functional polypeptides are obtained finally by solubilization and refolding procedures.

B.Yeast based expression system yield soluble IP which is secreted into the culture supernatant.

Saccharomyces cerevisiae is the most preferred and predominant yeast for large scale commercial production of insulin, however several other alternate yeast strains have been explored for insulin production.

Since the early 1920s, diabetic patients were treated with insulin, which was purified from bovine or porcine pancreas.
Nowadays, recombinant human insulin is mainly produced either in *E. coli* or *Saccharomyces cerevisiae*.

Using *E. coli* expression system, the insulin precursors (IP) are produced as inclusion bodies and fully functional polypeptides are obtained finally by solubilization and refolding procedures.

Yeast based expression system yield soluble IP which is secreted into the culture supernatant.

Saccharomyces cerevisiae is the most preferred and predominant yeast for large scale commercial production of insulin, however several other alternate yeast strains have been explored for insulin production.

Besides, *E.coli* and yeast, mammalian cells, transgenic animals and plant expression systems are also employed as a host for large-scale production of recombinant insulin.

Advantages of *E.coli*:

1. high growth rate,
2. simple media requirement,
3. easy to handle,
4. high yield and very cost effective.

disadvantages using *E. coli* expression system

1. intracellular accumulation of heterologous proteins as inclusion bodies,
2. improper protein refolding,
3. lack of post- translational modifications (including unable to form disulphide bonds),
4. protein-mediated metabolic burden and stress,
5. poor secretion,
6. proteolytic digestion
7. and complexity in downstream process
8. such as loss of plasmid and antibiotic property,
9. unsolicited inducers for gene expression,
10. endotoxin contamination,

Structure, production and function of insulin

- The human insulin is comprised of 51 amino acids and has a molecular weight of 5808 Da.
- It is produced by beta cells of the pancreas and plays a key role in regulating carbohydrate and fat metabolism in the body.
- Insulin is synthesized as a single polypeptide known as preproinsulin in pancreatic beta cells.
- Preproinsulin harbours a 24-residue signal peptide, which directs the nascent polypeptide to the endoplasmic reticulum.
- The signal peptide is cleaved as the polypeptide is translocated into the endoplasmic reticulum resulting in the formation of proinsulin.
- In the Endoplasmic reticulum, the proinsulin is folded in proper confirmation with the formation of 3 disulphide bonds.
- Folded proinsulin is then transported to the trans-Golgi network, where it is converted into active insulin by cellular endopeptidases called as prohormone convertases (PC1 and PC2) and exoprotease carboxypeptidase E.

- The endopeptidases cleaves at two positions, resulting in the release of a fragment termed as C-peptide.

- The mature insulin, thus formed consists of an A-chain with 21 aminoacids and a B-chain containing 30 aminoacids and both polypeptides linked together by two disulphide bonds.

- Besides, the A-chain has an intrachain disulphide bond.

E. coli expression system for production of insulin

E. coli is a preferred microorganism for large-scale production of recombinant proteins.

However, several disadvantages limit its use for production of recombinant biopharmaceuticals.

Various post-translational modifications (PTMs) such as glycosylation, phosphorylation, proteolytic processing and formations of disulfide bonds which are very crucial for biological activity, do not occur in *E. coli*.

N-linked glycosylation is the most common posttranslational modification of proteins in eukaryotes. It has been discovered that the bacterium *Campylobacter jejuni* possess the capability to glycosylate the proteins and it was also shown that a functionally active N-glycosylation pathway could be transferred to *E. coli*.

Although the structure of bacterial N-glycan is different from that observed in eukaryotes, engineering of *Campylobacter* N-linked glycosylation pathway into *E. coli,* provides an opportunity to express heterologous proteins in glycosylated form in *E. coli.*

Expression of Pglb oligosaccharyl transferase or (OTase) from *C. jejuni* in *E. coli* showed a significant increase in glycopepetide yield. Recently efforts has been made to produce glycosylated proteins with substrates other than native and non-native to *E. coli* and *C.jejuni* .

The codon usage of the heterologous protein also plays a major role in determining the expression level of recombinant protein.

If the codon usage of heterologous protein differs significantly from the average codon usage of the *E. coli* host, it could result in very low expression.

Usually, the frequency of the codon usage reflects the abundance of their corresponding tRNA. Therefore, significant differences in codon usage could result in premature termination of translation, misincorporation of aminoacids and inhibition of protein synthesis.

Expression of heterologous proteins in *E. coli* can be improved by replacing codons that are rarely found in highly expressed *E. coli* genes with more favorable major codons. Similarly, co-expression of the genes encoding for a number of the tRNA for rare codon, may enhance the expression of heterologous proteins in *E. coli.*

There are some commercial *E. coli* strains available that encodes for tRNA for rare codons such as BL21 (DE3) CodonPlus-RIL, BL21 (DE3) CodonPlus-RP (Stratagene, USA) and Rosetta (DE3). BL21 (DE3) CodonPlus-RIL harbors tRNA genes for rare codons like AGG, AGA (arginine), AUA (isoleucine) and CUA (leucine).

Similarly, Rosetta (DE3) strain harbors tRNA genes for rare codons like AGG, AGA (arginine), CGG (arginine), AUA (isoleucine), CUA (leucine), CCC (proline) and GGA (glycine). These rare codons have been associated with low expression of proteins in *E. coli,* hence application of these genetically engineered *E. coli* host strains may improve the expression level of heterologous proteins and thus might result in higher yield of desired protein.

The use of protease-deficient *E. coli* strains, which carry mutations that eliminate the production of proteases may also improve the yield of recombinant protein by reducing proteolytic degradation. *E. coli* strain BL-21, is deficient in two proteases encoded by the *lon* (cytoplasmic) and *ompT* (periplasmic) genes. Rather than the external parameters, targeted methods such as modifications in protease or secretion pathways can provide the insight into biology of recombinant proteins.

In *E. coli,* complex and large therapeutic proteins can be secreted in periplasm as it provides an oxidizing environment and help in forming disulphide bonds, which facilitate the proper folding of recombinant proteins and likely to yield reliable N- terminus of expressed protein.

Periplasm has advantages over cytoplasm in less protein concentration and proteolytic activity, improve the production titer, and enhance the solubility of recombinant protein. Altogether, with these advanced modifications and developments ease the process of target protein production thus accelerating the drug development.

Heterologous proteins generally accumulate in *E. coli* as inclusion bodies, which comprise of insoluble misfolded aggregates of proteins. Use of molecular chaperones may increase the protein solubility and assist in proper folding of recombinant protein. Some of the chaperones prevent aggregation of protein and some assist in refolding and solubilization of misfolded proteins. The most important chaperones in *E. coli* are GroEL, GroES, DnaK, DnaJ, GrpE and Trigger factor. These chaperones may be used singly, or in combination to enhance the protein solubility in *E. coli* .

1. Recombinant human insulin was first produced in *E. coli* by Genentech in 1978, using a approach that required the expression of chemically synthesized cDNA encoding for the insulin A and B chains separately in *E. coli* .

After expressing independently, the two chains are purified and co-incubated under optimum reaction conditions that promoted the generation of intact and bioactive insulin by disulphide bond formation.

The first commercial recombinant insulin was developed for therapeutic use in human by this two-chain combination procedure.

2.Another approach involves the expression of a single chemically synthesized cDNA encoding for human proinsulin in *E. coli* followed by purification and subsequent excision of C-peptide by proteolytic digestion.

This approach was more efficient and convenient for large scale production of therapeutic insulin as compared to the two chain combination approach and has been used commercially since 1986.

Eli Lilly followed this technology to produce Humulin, the first recombinant insulin approved in 1982, for the treatment of diabetic patients. These first generation recombinant insulins have an amino acid sequence identical to native human insulin and are preferred over animal derived insulin products

However, advancement in the field of genetic engineering and development of technology to chemically synthesize genes with altered nucleotide sequence, facilitated the development of insulin analogues with altered amino acid sequence.

It had been observed that native insulin in commercial preparations usually exist in oligomeric form, as zinc-containing hexamer due to very high concentration, but in blood, biologically active insulin is in monomeric form.

Hence, this oligomeric complex should dissociate so that insulin can be absorbed from the site of injection into the blood. Due to this, subcutaneously injected recombinant insulin usually have a slow onset with peak plasma concentration after 2 hours of injection and longer duration of action that last for 6–8 hours.

Hence, in order to develop a fast- acting insulin analogue, it was required to modify the amino acids residues whose side chains are involved in dimer or oligomer formation. It has been shown that amino acids residues in insulin B-chain particularly B8, 9,12, 13, 16 and 23-28 play critical role in oligomerization.

Lispro, developed by Eli Lilly, was the first fast acting insulin analogue to obtain regulatory approval in 1996, for therapeutic use. Insulin Lispro is engineered in such a way that it has similar amino acid sequence as the native insulin but has an inversion of proline-lysine sequence at position 28 and 29 of the B-chain, which resulted in reduced hydrophobic interactions and thus prevented dimer formations.

For commercial production of insulin Lispro, a synthetic cDNA encoding for Lys ^{B28}- Pro ^{B29} human proinsulin was expressed in *E. coli* and insulin Lispro was excised proteolytically from the proinsulin by treating with trypsin and carboxypeptidase. Another rapid-acting insulin analogue, produced in *E. coli* is Glulisine (Apidra) which was developed by Aventis Pharmaceuticals and approved by US regulatory authorities in 2004. Insulin Glulisine have been generated by replacing B3 asparagine by a lysine and B29 lysine replaced by glutamic acid.

To avoid multiple injection, long-acting insulin analogues with prolonged duration of actions have also generated.

Insulin Glargine is one of such long-acting insulin analogues, which was developed by Aventis Pharmaceuticals and approved by regulatory authorities of USA and EU in 2000. Insulin Glargine was generated by replacing the C-terminal asparagine of the A-chain with a glycine residue and the C-terminal of the B- chain was modified by adding two arginine residues.

These modifications resulted in increase of the isoelectric point (pI) from 5.4 to neutral values. Glargine was produced as proinsulin and expressed in *E. coli* and was finally formulated at pH 4 in soluble form.

However, after subcutaneous administration, it precipitated due to neutral pH in the subcutaneous tissue. Resolubilization of insulin occur slowly, resulting in longer duration for its release in the blood.

Yeast expression system for the production of insulin

Saccharomyces cerevisiae has been extensively used to produce recombinant human insulin since early 1980s and a large proportion of recombinant commercial insulins are produced by this yeast expression system.

For efficient expression and secretion of recombinant proinsulin in yeast, insulin construct was engineered to contain the native A-chain and a B-chain lacking the C-terminal B30 threonine, linked via a short synthetic C peptide (like AAK).

The cDNA sequence encoding for this construct was fused with α-factor signal sequence of *Saccharomyces cerevisiae* for secreted expression of proinsulin which gave yield upto 80 mg/ml of insulin. The single chain proinsulin was purified and converted to active insulin by a trypsin-mediated transpeptidation reaction in presence of threonine ester.

Besides native recombinant insulin, various insulin analogues are also being produced in *S. cerevisiae.*

Insulin Aspart is another fast-acting insulin analogue, which was produced in *S. cerevisiae*, developed by Novo Nordisk and approved by US FDA in 2001 for therapeutic use in human. Insulin Aspart was generated by replacing proline residue at position 28 with aspartic acid in the B-chain. This genetic modification resulted in an increase in inter-chain charge repulsion, decrease in self-association and thus causing rapid entry into the blood from the site of subcutaneous injection.

Transgenic plants as host for insulin production

Recombinant human insulin has been successfully expressed and produced in oilseeds of plant *Arabidopsis thaliana* .
This technology involved the targeted expression of insulin in subcellular organelles known as oilbodies that allowed very high level of expression with easy recovery of recombinant insulin.

Oilbodies are storage organelles inside the oilseeds, which comprises of hydrophobic triacylglycerol core encapsulated by phospholipid membrane and an outer wall of proteins known as oleosins. Genetically engineered oil seeds have been generated with recombinant protein specifically targeted to oilbodies as oleosin fusion.

Then the oilbodies are easily separated from other seed components by liquid-liquid phase separation, which reduced the number of chromatography steps required to obtain purified insulin. It has been observed that insulin accumulated to high level in transgenic seed (0.13% of total seed protein).

Recombinant insulin was cleaved from the oleosin fusion partner and matured with trypsin digestion following oil body purification to yield a biologically active insulin. This study clearly demonstrated that expression of insulin as oleosin fusion protein in plant allow accumulation of large amount of recombinant insulin within the seed and also provide simple downstream purification by centrifugation i.e. oilbody purification.

Subsequent maturation to obtain biologically active insulin can be accomplished using standard enzymatic methods currently used for commercial production of insulin from *E. coli* and yeast. Oilseeds also act as a natural cellular warehouse, where recombinant insulin can be stockpiled until required.

In another approach, transgenic plants have been generated, in which, tobacco and lettuce chloroplasts were transformed with human proinsulin comprised of A, B and C-chains fused with the cholera toxin B subunit.

It has been observed that, old tobacco leaves accumulated proinsulin upto 47% of total leaf protein and similarly, old lettuce leaves amassed proinsulin up to 53% of total leaf protein. Proinsulin stored in leaves of lettuce was found to be very stable as up to 40% of proinsulin was detected even in senescent and dried leaves .

Proinsulin from tobacco leaves was extracted with 98% purity and cleaved by Furin protease to release insulin peptides. Oral delivery of unprocessed proinsulin encapsulated in plant cell or by injection into mice revealed lowering of blood glucose levels similar to commercially available insulins. Based on the yield (3 mg of proinsulin/gm of leaves), it was estimated that one acre of tobacco plantation could yield upto 20 million daily doses of insulin per year.

C-peptide of proinsulin, which is not present in current commercially available insulin and insulin analogues derived from *E. coli* and *S. cereviciae*, would be a great advantage in long-term treatment of diabetic complications such as stimulation of nerve and renal functions.

Very high level of expression of biologically active proinsulin in tobacco and lettuce leaves and long-term stability in dried leaves offers a reliable low-cost technology for both injectable as well as oral delivery of proinsulin.

Thank You